Universität Karlsruhe (TH)
Fakultät für Informatik
Institut für Telematik
Zirkel 2,76128 Karlsruhe

Mobilkommunikation

Seminar SS 2001

Herausgeber:
Dipl.-Ing. Kilian Weniger
MSc. Jidong Wu
Prof. Dr. Martina Zitterbart

Interner Bericht 2001-13
ISSN 1432–7864

Zusammenfassung

Der vorliegende Interne Bericht enthält die Beiträge des Seminars Mobilkommunikation vom Sommersemester 2001.

Die Themenauswahl kann grob in folgende vier Blöcke gegliedert werden:

1. **Protokolle und Konzepte zur Mikromobilitätsunterstützung**, die als Erweiterung des Makromobilitätsprotokolls Mobile IP schnelle Handover innerhalb einer Domäne als Ziel haben und den Skalierbarkeitsaspekt von Mobile IP ansprechen. Insbesondere werden die Protokolle *Hierarchical Mobile IPv6* und *Cellular IP* behandelt.

2. **IP Routingalgorithmen in mobilen ad-hoc Netzen**. In einer Doppelarbeit werden diverse *Unicast- und Multicast Routingalgorithmen*, in einer weiteren Arbeit diverse *Clusteringalgorithmen* behandelt.

3. **Sicherheits- und Privacy Konzepte**. Eine Arbeit widmet sich der *Location Privacy* in der Mobilkommunikation, während sich eine weitere mit dem Aufbau einer *AAA-Infrastruktur in Mobile IP-Netzen* befasst.

4. **IP in Mobilfunknetzen**. In dieser Arbeit geht es um eine *IP-basierte UMTS Systemarchitektur*.

Vorwort

Das Seminar "Mobilkommunikation" wurde in Form eines Teleseminars das erste Mal am Institut für Telematik abgehalten und erfuhr sofort starken Andrang. Anstatt der geplanten sechs Teilnehmer wurden letztendlich neun Teilnehmer zugelassen, deren Vortäge mit Hilfe einer Audio- und Videoschaltung per Multicast an das Institut für allgemeine Nachrichtentechnik der Universität Hannover und an das Institut für Betriebssysteme und Netze der Technischen Universität Braunschweig übetragen wurden. Durch die Konferenzschaltung fand ein reger Austausch der verschiedenen Institute und Fachrichtungen statt.

In diesem Seminarband sollen nun die Beiträge der Karlsruher Studenten in Form eines Internen Berichts zusammengefasst werden. Folgende Themengebiete werden hier behandelt:

Protokolle und Konzepte zur Mikromobilitätsunterstützung

Protokolle zur Mobilitätsunterstützung, deren bekanntester Vertreter wohl Mobile IP ist, bieten die Möglichkeit sich an jedem beliebigen Ort ans Internet anzuschliessen (Makromobilität) und dabei weiterhin über die IP-Adresse seines Heimatorts erreichbar zu sein. Ausserdem kann der Verbindungspunkt zum Internet gewechselt werden (Handover), ohne bestehende Verbindungen neu aufbauen zu müssen. Ein Hauptproblem von Mobile IP ist allerdings, dass bei weit vom Heimatort entfernten Handovern mitunter sehr lange Verzögerungen auftreten können und dass bei internetweitem Einsatz des Protokolls die Skalierbarkeit vermutlich nicht gegeben ist. Um diesen und weitere Nachteile zu beseitigen wurden verschiedene Erweiterungen zur Mikromobilitätsunterstützung vorgeschlagen.

Algorithmen für mobile ad-hoc Netze

Um in drahtlosen Netzen nicht nur Endgeräte erreichen zu können, die sich in direkter Reichweite des Senders befinden, können andere mobile Endgeräte die Rolle eines Routers übernehmen und Pakete zu einem entfernten Endgerät weiterleiten. An Algorithmen, die für die Wegewahl zuständig sind, müssen auf Grund der hohen Dynamik hohe Anforderungen gestellt werden. Einige dieser Algorithmen, sowohl für Unicast- als auch für Multicastverkehr, sollen in diesem Abschnitt vorgestellt werden.

Sicherheits- und Privacy Konzepte in der Mobilkommunikation

Um bei einem flächendeckenden Einsatz von Mobile IP Rechnungsstellungen von Netzbetreibern zu ermöglichen, müssen die Dienstnehmer authentifiziert und Daten über deren Nutzungsdauer aufgezeichnet werden. Gerade in dem Bereich der Rechnungsstellung müssen erhöhte Sicherheitsanforderungen an das System gestellt werden, um Mißbrauch vorzubeugen. Auf der anderen Seite möchten die meisten Nutzer eines mobilen Endgerätes nicht, dass Fremde den Aufenthaltsort des mobilen Gerätes und damit den des Nutzers erfahren. In diesem Abschnitt werden Konzepte aus diesen beiden Gebieten vorgestellt.

IP in Mobilfunknetzen

Um die Flexibilität und Effizienz der Core-Netze zu erhöhen, gerade im Hinblick auf die zunehmend paketbasierten Datenströme, denken die Mobilfunknetzbetreiber über eine Umstellung ihre Netze auf IP nach. Diesbezügliche Konzepte und Architekturen werden am Beispiel UMTS in diesem Beitrag vorgestellt.

Inhaltsverzeichnis

Zusammenfassung i

Vorwort ii

Peter Weik:
Hierarchisches Mobilitätsmanagement für Mobile IP 3

Tingchao Yan:
Mikromobilität mit Cellular IP 17

Tilmann Rothhammer:
AAA-Infrastruktur in Mobile IP Netzen 31

Markus Klein:
Location Privacy in der Mobilkommunikation 47

Rami Nassar und Philipp Geipel:
Unicast- und Multicast Algorithmen in Ad-Hoc Netzen 59

Steffen Kamuf:
IP-basierende UMTS System Architektur 81

David Divisek:
Clustering für Ad-hoc Netze 95

Hierarchisches Mobilitätsmanagement für Mobile IP

Peter Weik

Kurzfassung

In dieser Arbeit soll das hierarchische MOBILE IPv6 (HMIPv6) als eine Erweiterung von MOBILE IP vorgestellt werden, das sich neue IPv6 Funktionalitäten wie den vergrößerten Adressraum und Neighbor discovery zunutze macht, um ein bezüglich der Signalisierungslast optimiertes Mobilitätsmanagement vorzuschlagen. HMIPv6 unterscheidet dabei zwischen Bewegungen eines mobilen Hosts innerhalb und zwischen einzelnen Domains und erreicht somit eine Reduzierung der für die Signalisierung verwandten Bandbreite von 69 bis zu 100 %. Lokal werden Bewegungen verborgen und gleichzeitig optimales Routing und eine schnelle Dienstleistung beim Übergang von einer Domain zur nächsten geboten. Als Ergebnis werden die globale Belastung des Internets sowie Verluste von Binding updates und damit auch Verbindungsverluste eines mobilen Hosts reduziert. Zusätzlich könnte dieser Ansatz das Vertrauen in Mobile Computing erhöhen, da die genaue Lokation eines mobilen Hosts für seine externen Kommunikationspartner verborgen bleibt.

1 MOBILE IP

1.1 Einleitung

MOBILE IP ist die Erweiterung des klassischen Netzwerkschichtprotokolls IP, das in den 60er Jahren für die Kommunikation zwischen immobilen Computern entwickelt wurde. Mit der zunehmenden Zahl von mobilen Endgeräten und einer in den nächsten Jahren zu erwartenden großen Wachstumsquote wird mit Sicherheit auch immer mehr die Notwendigkeit für ein zuverlässiges und robustes Netzwerkprotokoll für mobile Rechner aufkommen. Diesen Anforderungen möchte MOBILE IP, wenn auch bisher noch mit Sicherheitsproblemen behaftet, Rechnung tragen. Der Schritt, der sich in den letzten 20 Jahren von der ortsgebundenen zur orts*un*gebundenen Kommunikation vollzogen hat, ist momentan im Begriff, sich auch immer mehr in der Nutzung von Computern zu vollziehen. Die bisher zur Computerkommunikation verwandten Netzwerkprotokolle der Schicht 3, sei es nun IP, IPX oder auch Appletalk, gehen alle von einem an einem bestimmten und immer gleichbleibenden Punkt des Netzes angeschlossenen Computer aus, der sich über seine IP- Adresse als Mitglied „seines“ Netzwerks identifiziert. Der Grund für die Erweiterung von IP bestand in dem Problem, daß ein mobiler Computer, im folgenden nur noch als mobile host = MH bezeichnet, der sein Heimatnetz verlassen hatte, um an einem anderen Netzzugangspunkt angeschlossen zu werden, vor der Wahl stand entweder seine IP- Adresse um den Preis zu behalten, daß an ihn gesendete IP-Pakete ihn nicht mehr erreichen und auch von ihn gesendete Pakete wegen einer topologisch unkorrekten Absenderadresse ihr Ziel nicht mehr erreichen oder seine IP-Adresse zu ändern um den Preis, daß bestehende Verbindungen der Schicht 4 aufwärts (vor allem TCP/UDP-Verbindungen) abbrechen würden. Für sie wäre nach jedem Netzwechsel in diesem Fall ein Neuaufbau aller Verbindungen nötig. Das Ziel für MOBILE IP ist es also die Mobilität des Rechners für höhere Schichten transparent zu gestalten.

1.2 Die Archtitektur von MOBILE IP

MOBILE IP stellt eine Erweiterung zu IP dar, die es einem MH ermöglicht, sich unter Beibehaltung seiner topologisch „korrekten“ IP- Adresse frei zu bewegen. Die MOBILE IP Architektur definiert zu diesem Zweck spezielle Objekte: einen Home Agent (HA), einen Foreign Agent (FA) sowie eine Care of Adresse (CoA). Die Agenten sind Computer (aber nicht zwingend Router), die in ihren jeweiligen Netzen Pakete abfangen und weiterleiten. Das Prinzip von MOBILE IP ist dabei wie folgt :

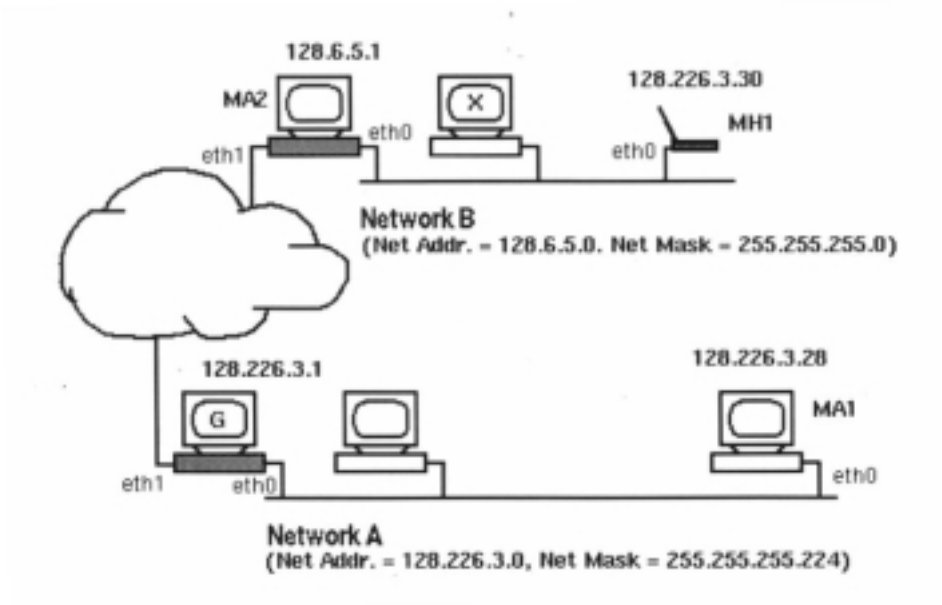

Abbildung 1: Wechsel eines MH in ein fremdes Netz.

- Der MH bewegt sich von seinem Heimatnetz A in das Fremdnetz B.
- Er registriert sich beim FA, der ihm eine CoA zuteilt und ein mobility binding (Tabelleneintrag) macht.
- Der MH informiert seinen HA über seine neue Erreichbarkeit unter der CoA.
- Ein correspondend host (CH) will ein IP- Paket an den MH senden, das vom HA anhand der Zieladresse des Paketes ausgefiltert wird und dem er einen neuer IP- Header voranstellt, der als Zieladresse die CoA des FA hat.
- Dieses Paket wird zum FA geschickt („tunnelling“), der den Endpunkt des Tunnels darstellt, den äußeren IP- Kopf abstreift und das Paket an die Adresse des MH schickt. (siehe Abbildung 2)

MOBILE IP funktioniert also nach dem Prinzip, daß ein MH zwecks Transparenz für höhere Schichten seine topologisch korrekte IP- Adresse auch nach Verlassen des Heimatnetzes behält. Die CoA, die ein MH erhält, kann entweder beim FA liegen, der die CoA über *Agent Advertisements* den MH mitteilt (*Foreign Agent CoA*), oder direkt vom MH über DHCP (oder auch PPP) als *co-located CoA* übernommen werden. Im zweiten Fall liegt der Tunnelendpunkt beim MH selbst, der dann allerdings auch die Entkapselung der Pakete zu bewerkstelligen hätte, was eine Modifizierung des Kernels des MH notwendig machen würde.

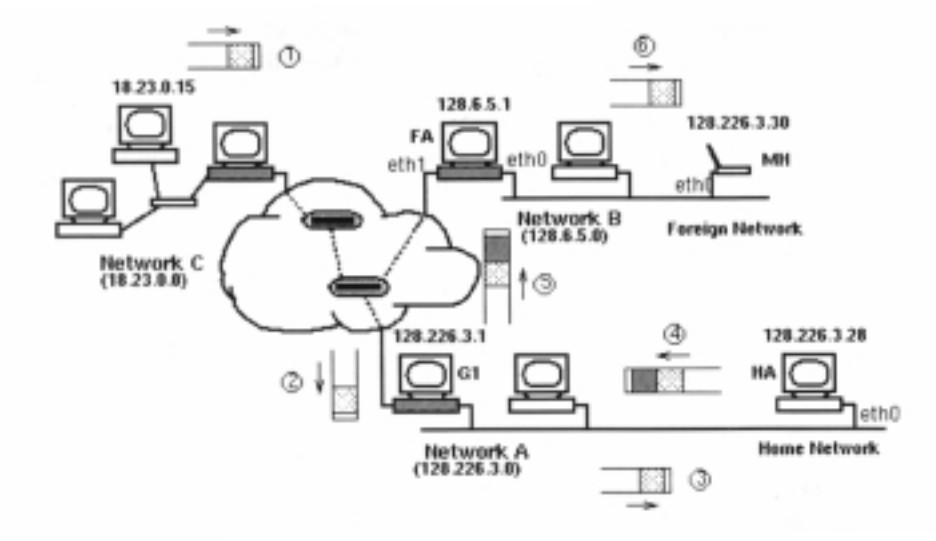

Abbildung 2: Tunnelling von Pakete zu einer CoA in MOBILE IP.

2 MOBILE IPv6

In MOBILE IPv4 gibt es in jedem Fall immer *genau einen* Tunnelendpunkt. Die Architektur von MOBILE IPv6 wird durch einige Neuerungen von IPv6 vereinfacht. „Einer der Vorteile der Foreign Agents in IPv4 war, daß sie eine CoA bereitstellte, die sich sehr viele mobile Knoten teilen konnten. Im Besonderen eliminieren FA's die Notwendigkeit, jedem einzelnen mobilen Knoten eine eindeutige, allgemeingültige co-located CoA zuzuweisen. In IPv6 ist jedoch die Verfügbarkeit von Adressen kein Problem. IPv6 erlaubt bis zu $2^{128} = 3.4028237 \times 10^{38}$ adressierbare Knoten (man beachte, daß 10^{12} schon eine Trillion ist !) " [Solo98]
Außer dem größeren Adressenraum bietet IPv6 zum anderen auch die Möglichkeit, benachbarte Rechner über Neighbor discovery aufzuspüren und vor allem die sich abzeichnende Adressenknappheit lässt einen Wechsel von IPv4 zu IPv6 in absehbarer Zeit möglich erscheinen. „Mehrere Mechanismen, die in MOBILE IP speziell für die Unterstüzung der Mobilität integriert werden mussten, sind bereits ein fester Bestandteil in IPv6. Ein Mechanismus betrifft die Sicherheit von Registrierungsnachrichten, was nun ein verpflichtender Mechanismus für alle IPv6 Knoten ist. Hierzu sind also keine besonderen Maßnahmen während der Registrierung mehr zu treffen. Jeder IPv6 Knoten beherrscht eine automatische Konfiguration, daher sind die Mechanismen zur Erlangung einer CoA bereits in IPv6 integriert worden. Die Entdeckung von topologischen Nachbarn (neighbor discovery) ist weiterhin ein verpflichtender Bestandteil aller Knotenimplementierungen, daher werden keine speziellen Fremdagenten mehr benötigt, die ihre Dienste in gesonderten Nachrichten verkünden. Die Verbindung der integrierten Mechanismen, der automatischen Konfiguration und Entdeckung von Nachbarknoten erlaubt es, jedem mobilen Knoten automatisch, eine topologisch korrekte Adresse zu erzeugen oder zu erlangen, die zum aktuellen Zugangspunkt im Netz passt." [Schi00] Durch das eigene Versenden von Aktualisierungsnachrichten über seine aktuelle CoA an die CH's, den sogenannten *binding updates* (im folgenden nur noch BU genannt) wird ein mögliches Problem von MOBILE IP ausgeschaltet, nämlich, daß bei einem häufigen Wechsel der CoA und einer großen Anzahl von CH's der HA, der in MIPv4 mit dem Versenden der BU's betraut ist, zu einem Bremsklotz in der Kommunikation werden konnte. Vor allem wird aber das Problem des triangle rounting verhindert, da ein CH nun nicht mehr alle seine Pakete über den HA senden muss.
In MOBILE IPv6 erhält ein MH jedesmal wenn er sich von einem Subnetz zu einem anderen

Subnetz bewegt eine neue CoA, die immer co-located ist und die er bei seinem HA registriert. Der HA arbeitet quasi als Proxy für den MH, bis dieser Binding Eintrag im Router ausläuft. Der HA fängt Pakete, die an die Heimatadresse des MH gerichtet sind, ab und leitet diese unter Benutzung der IPv6 Kapselung an die MH- CoA weiter. Ebenso informiert der MH aber auch seine CH's per BU über seine neue CoA. Wenn diese in der Lage sind, Änderungen in ihren lokalen Wegewahltabellen durchzuführen, schicken sie die Pakete dann direkt an den MH (Routenoptimierung), ansonsten muß der Weg über den HA gewählt werden. Bedenkt man, daß die Zahl der mobilen Hosts in Zukunft immer mehr ansteigen wird, führt dies dazu, daß die Zahl der BU's proportional dazu mitansteigen und damit auch zu einer deutlichen Mehrbelastung von Netzressourcen führen wird.
Das MOBILE IPv6 Protokoll wird derzeit von einer working group der IETF (Internet Engineering Task Force) spezifiziert. Die Ziele dieser Spezifikation bleiben (wie in MOBILE IPv4 auch):

1. den Ortswechel von einem Subnetz zu einem beliebigen anderen vor den User vollkommen zu verbergen. Die bestehenden Verbindungen sollen auch bei einem Wechsel des Anschlußpunktes erhalten bleiben. (*Mobilitäts- und Performancetransparenz*)

2. das Protokoll sollte auch mit rasch steigender Benutzerzahl *skalieren*.

Die Skalierbarkeit des MOBILE IPv6 Protokolls ist dabei mit Sicherheit ein nicht zu unterschätzender Punkt, da das Internet zurzeit immer noch anwächst und da die Zahl der zu erwartenden mobilen Hosts sehr hoch werden kann, wenn man bedenkt, daß ein MH nicht unbedingt ein Laptop sein muss, sondern daß ein MH durchaus auch in Lastwagen, Autos, Flugzeugen, Schiffen usw. zum Einsatz kommen kann. Um den dazu notwendigen IP- Adressraum bereitzustellen, eignet sich IPv6 optimal. Castellucia et al. [eal.98] schreiben in ihrem HMIP- Vorschlag, daß der IETF MOBILE IPv6 Vorschlag zwar ein Mobilitätsmanagementschema für das Internet bietet, doch nicht vollständig diesen Designzielen gerecht wird. Obwohl er Performancetransparenz bietet, bezweifeln sie allerdings, daß MOBILE IPv6 skaliert. Es wird dabei argumentiert, daß die periodisch von MH ausgesandten BU's mit steigender Zahl von Benutzern eine zu große Signalisierungslast auf das Netz legen. Um diesem Problem zu begegnen, wurde von ihnen ein Vorschlag zur Erweiterung von MOBILE IP bei der IETF eingereicht, der ein hierarchisches Mobilitätsmanagement ermöglicht.

3 Hierarchisches Mobilitätsmanagement für MOBILE IP

In MOBILE IP (sowohl in v4 also auch in v6) wird zwischen Mobilität auf globaler und auf lokaler Ebene nicht unterschieden. Dies bedeutet, daß es bei der momentanen Architektur von MOBILE IP egal ist, ob sich ein MH z.B. von dem Netz einer Firma in ein Netz bewegt, das unter der Administration einer vollständig anderen Firma befindet oder ob er von Subnetz A zu Subnetz B innerhalb einer Domain wechselt. Das MOBILE IP Protokoll , sowohl in seiner Version 4 als auch in seiner Version 6 kennt keinerlei Hierarchiestufen.

Eine Studie von G. Kirby [Kirb95] untersuchte die Bewegungsmuster von Berufstätigen, unabhängig davon, ob sie nun mit tragbaren Geräten ausgestattet waren oder nicht. Kirby kam dabei zu dem Ergebnis, daß 69 Prozent der Benutzerbewegung lokal ist (sich also innerhalb einer Firma, eines Campus ... abspielt). Wenn also die Mehrheit der Bewegungen der potentiellen Nutzer von MOBILE IP lokal geschehen, wäre die Signallast für das Internet durch binding updates viel zu hoch da die BU's, selbst bei minimaler Mobilität innerhalb einer Domain, periodisch durch das Internet gesendet werden.

3.1 Der Ansatz

Ungeachtet der bisher noch ungeklärten Sicherheitslücken von MOBILE IPv6 soll das MIPv6 Protokoll noch um die Unterscheidung zwischen Mikro- und Makromobilität durch die Einführung einiger zusätzlicher Features erweitert werden. Natürlich stellt der hier vorgestellte Ansatz der Einführung einer Hierarchie nur einen von vielen denkbaren Ansätzen zur Erlangung von Mikromobilität dar. Die Verwendung eines solchen hierarchischen Ansatzes bietet zumindest zwei Vorteile. Erstens verbessert er die handoff performance da lokale Handoffs nun auch wirklich lokal vollzogen werden, was die Handoff- Geschwindigkeit erhöht und er minimiert die Paketverluste, die während der Übergabe vorkommen können. Zweitens reduziert er signifikant die Signalisierungsbelastung für das Internet durch das Mobilitätsmanagement da die Signalisierungsnachrichten, die zu lokalen Bewegungen gehören, nicht das ganze Internet durchqueren, sondern auf die Domain begrenzt bleiben. Die Mikromobilität eines MH bleibt also einem der Domain außenstehenden Host komplett verborgen. Dies macht dann eine genaue Lokalisierung und Wegaufzeichnung des MH für den HA und externe Kommunikationspartner unmöglich.

3.2 Protokollüberblick

(Zu den einzelnen Begriffsdefinitonen siehe Tabelle 1)
Ziel des Aufbaus einer hierarchischen Mobilitätsmanagementstruktur sollte es sein, die Signalisierungslast durch BU's zu reduzieren. Gleichzeitig sollte sie es ermöglichen, zwischen der Mobilität *zwischen* einzelnen Domains (inter-site Mobilität) und *innerhalb* einer Domain (intra-site Mobilität) differenzieren zu können. Zu diesem Zweck wird eine zweite CoA eingführt, die Virtuelle CoA, welche nur innerhalb eines Mobilitätsnetzes Gültigkeit hat.

3.2.1 Inter-site Mobilität

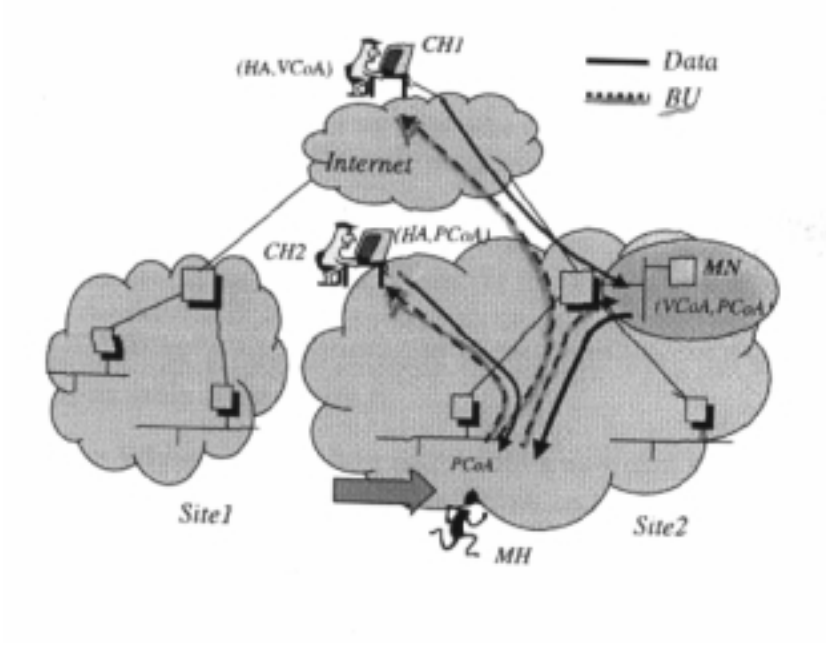

Abbildung 3: Inter-site Mobilität

„Wenn ein MH eine neue Domain betritt, bekommt er zwei CoA's: eine private (oder physikalische) CoA (PCoA), die eine CoA zu dem Verbindungspunkt ist, an dem er angeschlossen

Domain	Eine Domain (oder auch site) kann definiert werden als der höchste Level einer hierarchischen Architektur und muss keiner bestimmten Struktur folgen. Es kann sich dabei um ein ISP-, ein Campus- oder ein Firmennetzwerk handeln, um eine Menge von LAN's oder auch nur um ein einzelnes LAN.
Borderrouter (BR)	Eine Domain hat über einen oder mehrere miteinander verbundene Borderrouter Verbindung zum Internet.
Mobilitätsnetz (MN)	Ein Mobilitätsnetz einer Domain ist ein LAN, das einen Adressraum für MH's definiert, die sich innerhalb einer Domain bewegen. Das MN kann jedes Subnetz einer Domain sein und muss sich nicht ausschließlich um MH's kümmern, sondern kann durchaus auch (gewöhnliche) festangeschlossene Hosts versorgen.
Mobilitätsagent (MA)	Ein Mobilitätsagent ist ein Router des Mobilitätsnetzes, der pro MH, der die Domain besucht, einen Verbindungseintrag hält. An welcher räumlichen Stelle des MN der MA platziert wird, spielt keine Rolle. (Das Konzept des Mobilitätsagenten ist dem des Home Agent sehr ähnlich.)
Physikalische Care-of Adresse (PCoA)	Der Verbindungspunkt eines Netzes am dem ein MH tatsächlich (physikalisch) angeschlossen ist. Der Anschluß muss dabei nicht unbedingt über eine Kabelverbindung, sondern kann v.a. auch „wireless“ realisiert sein.
Virtuelle Care-of Adresse (VCoA)	Die CoA, die einem MH innerhalb eines MN zugewiesen wird.

Tabelle 1: Begriffsdefinitionen für HMIPv6

ist, und eine virtuelle CoA (VCoA), die eine CoA im Mobilitätsnetz (MN) der Domain ist. Der MH versendet dann einige BU's. Er sendet:

- ein BU bei dem das Acknowledge-Bit (A-bit) gesetzt ist, das die Verbindung seiner VCoA und seiner PCoA spezifiziert, zum MA der Domain. Nach dem Empfang dieses BU führt der MA einige Zugangskontrollen durch wie Authentifizierung und Rechnungsstellung. Wenn die Anfrage angenommen wird, wird ein Acknowledgement zurück zum MH gesendet. Die Thematik der Authentifizierung und der Rechnungsstellung sollen außerhalb der Betrachtung dieses Reports bleiben.
- ein BU, das die Verbindung zwischen seiner Heimatadresse und seiner VCoA spezifiziert an seinen HA und an jeden seiner externen CH's (CH's außerhalb der Domain)
- ein BU, das die Verbindung zwischen seiner Heimatadresse und seiner PCoA spezifiziert an jeden seiner lokalen CH's (CH's innerhalb der Domain)

Als Ergebnis erhält man :

- Ein externer Host, der Pakete zum MH sendet, benutzt dessen VCoA. Die Pakete werden dann zum Mobilitätsnetzwerk der besuchten Domain geroutet, vom MA abgefangen und zur aktuellen PCoA getunnelt.

- Ein lokaler Host, der Pakete zum MH sendet, benutzt dessen PCoA. Die Pakete werden dann direkt zum MH geliefert. (siehe Abbildung 3)

3.2.2 Intra-site Mobilität

Wenn sich ein MH innerhalb einer Domain bewegt, bekommt er an jedem neuen Anschlußpunkt eine neue PCoA. Die virtuelle CoA bleibt dieselbe solange sich der MH lokal fortbewegt. Der MH sendet dann die folgenden BU's:

- ein BU, das die Verbindung zwischen seiner Heimatadresse und seiner neuen PCoA spezifiziert an jeden seiner lokalen CH's

- ein BU, das die Verbindung zwischen seiner VCoA und PCoA spezifiziert an den MA der Domain.“ [eal.98] (siehe Abbildung 4)

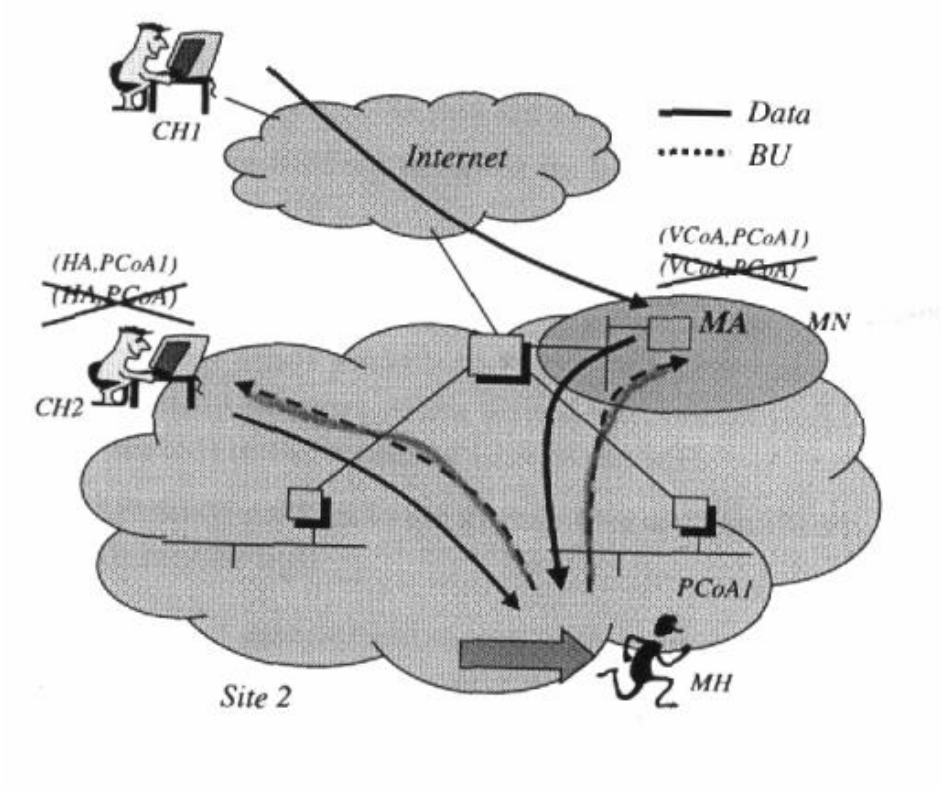

Abbildung 4: Intra-site Mobilität

Durch die Einführung einer VCoA wird erreicht, daß während lokalen Bewegungen des MH innerhalb einer Domain *kein* BU ins Internet gesendet werden muss da Übergaben lokal durchgeführt werden können. Während des lokalen Roamens ist die Transparenz für alle höheren Schichten gewährleistet, da die Heimatadresse nicht geändert werden muss. Dieser Ansatz skaliert also bei großer Anzahl von MH's innerhalb einer Domain besser als das „reine“ MOBILE IPv6.

3.3 Verwendung von Hierarchien

Die Einführung einer VCoA zur Reduktion der nach außerhalb einer Domain gesendeten BU's bildet die Basisüberlegung für ein hierarchisches Mobilitätsmanagement. Dessen Grundgedanke besteht darin, eine Domain in verschiedene Hierarchieebenen zu unterteilen, die ihrerseits auch wieder unterteilt sein können, und ein Mobilitätsnetz in jeder Hierarchieebene zu installieren. Auf diese Art kann die räumliche Ausdehnung einer Domain doch beträchtlich groß werden. Die Strukturierung der Mobilitätsnetze kann dabei eleganterweise der bestehenden Hierarchie einer Domain folgen. Castellucia et. al. schlagen eine Strukturierung der MN in Baumform vor mit einem MN an der Wurzel des Baumes (als höchste Hierarchiestufe) auf das MN's eine Hierarchieebene tiefer folgen usw. Diese Strukturierung kann sich u.U. bis hinunter zu einzelnen Subnetzen fortsetzen. Damit bekommt ein MH eine VCoA in jedem MN vom hierarchisch „höchsten" MN, das hier mit M_1 bezeichnet werden soll, wobei die 1 repräsentativ für die höchste Hierarchiestufe steht, bis hin zu seinem Anschlußpunkt. (siehe Abbildung 5)

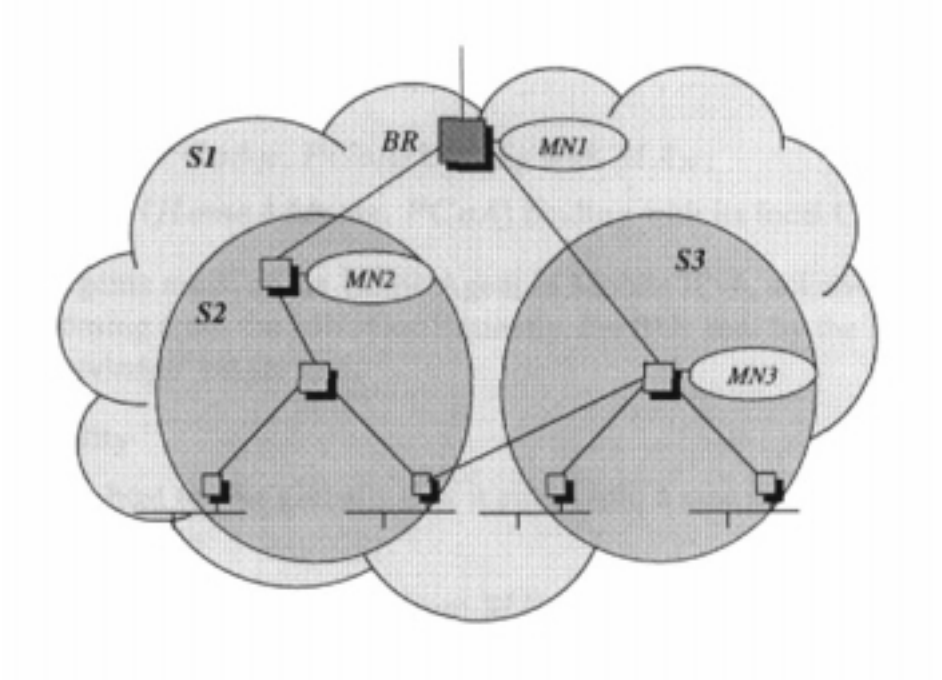

Abbildung 5: Hierarchie einer Domain

3.3.1 Inter-site

Beim Eintritt in eine neue Domain mit n Hierarchiestufen, werden folgende Aktionen ausgeführt:

- Der MH bekommt in jeder Hierarchiestufe (von 1 bis n) eine neue VCoA.
- Er bekommt eine neue PCoA.
- Er registriert die ($VCoA_{i-1}$,$VCoA_i$) Verbindung bei den MA_{i-1} von 1 bis n (quasi von der Wurzel des Baumes bis zu den Blättern).
- Er registriert die ($VCoA_n$,PCoA) Verbindung beim MA_n.
- Er registriert die (Heimatadresse,$VCoA_1$) Verbindungen bei allen externen CH's und seinem HA.

3.3.2 Intra-site

Bewegt sich der MH nun innerhalb einer Domain mit z.B. 5 Hierarchiestufen von einem Subnetz zum nächsten, so wird über die Mobility Agent Information Option des Router Advertisements bestimmt, bis wiehoch in der Baumstruktur neue VCoA's vergeben werden müssen. (In diesen Beispiel sei die Veränderung bis hoch zu Ebene 3.)

- Der MH bekommt in jedem MN von M_3 an bis hinunter zur Hierarchiestufe 5 (M_5) eine neue VCoA.
- Er bekommt eine neue PCoA an seinem neuen Anschlußpunkt.
- Er registriert die ($VCoA_{i-1}$,$VCoA_i$) Verbindungen bei den MA_{i-1} für alle i von 3 bis 5 (VCoA sei dabei die VCoA des MH im MN_i).
- Er registriert die ($VCoA_5$,PCoA) Verbindung bei dem MA_5.
- Er registriert die (Heimatadresse,PCoA) Verbindung bei seinen lokalen CH's.

Die jeweiligen BU's müssen alle jeweils von den Mobilitätsagenten bestätigt werden (A-bit muss auf 1 gesetzt sein). Durch den Aufbau einer Hierarchie ist es für den MH nur für den Fall des Wechsels von einer Domain zur nächsten notwendig, BU's nach außerhalb der Domain zu schicken. Damit diese Registrierungsoperationen alle funktionieren, muß der MH einige Informationen über die Domain und die MN's bekommen. Diese Informationen bezieht er aus einer neuen Option der Router Advertisement Nachrichten des Neighbor discovery von IPv6, der „Mobility Information Option". Er erfährt dabei:

- das Präfix der Domain, um deren „Grenzen" festlegen zu können
- die Hierarchiestufe des MN auf der er sich momentan befindet
- für jedes MN des aktuellen Zweiges des Baums bis hinauf zu dessen Wurzel, deren jeweiliges Netzwerkpräfix sowie die IP- Adresse des zugehörigen MA.

Pakete, die den MH von außerhalb der aktuellen Domain erreichen sollen, werden somit direkt zur $VCoA_1$ geschickt, wo der MA_1 das Paket entkapselt und neu eingekapselt zum MA_2 schickt usw. (Anstatt ihrerseits Pakete wieder zu kapseln können die MA ab MA_2 auch einfach nur die Quell- und Zieladressen des Pakets anhand ihrer Tabelleneinträge verändern.) Das Paket wandert dann in der Hierarchie nach unten bis es an die PCoA geschickt wird. Die lokalen CH's senden ihre Pakete gleich direkt an die PCoA des MH. Bei eigenen zu sendenden Paketen setzt der MH immer in das Quelladressfeld seine PCoA ein (egal nach wo gesendet werden soll) und liefert seine eigentlich Heimatadresse durch Setzen der „Home Adress Option" in IP- Paket. Sicherheitstechnische Aspekte bei diesem Verfahren sollen bei dieser Arbeit nicht in Betracht gezogen werden.

3.3.3 Mehrere Mobilitätsagenten pro Mobilitätsnetz einsetzen

Ein Problem, das sich bei dem hierarchischen Aufbau in einer Baumstruktur mit einem MA pro MN ergeben kann, ist, daß die hierarchisch hoch angesiedelten MA's bei einer großen Anzahl von MH's zum Flaschenhals in Sachen Performance werden können. Ein MA hat pro von ihn betreuten MH schließlich einen Tabelleneintrag zu verwalten. Deshalb wird in HMIPv6 vorgeschlagen, mehrere MA's pro MN in den höheren Hierarchiestufen zu implementieren. Für die CH's geschieht dies dabei transparent da ein die VCoA des MH adressiertes Paket wenn

es zum MN gelangt von dessen MA abgefangen wird. Die tatsächliche MA Identität bleibt der Quelle des Pakets aber verborgen. Die Mobilitätsagenten können dabei dynamisch oder von einem Administrator dupliziert werden. Bei mehreren MA pro MN wäre jeder MA für einen Teil der niedrigeren Hierarchieebenen verantwortlich, was z.B. auch eine geographische Partitionierung einer Domain ermöglichen würde. Die „neuen“ MA's würden dabei durch die neue “Mobility Information Option“ den niedrigeren Netzen angekündigt werden. Mit dieser Technik lässt sich die Last, die durch BU- Verarbeitung, Paketweiterleitung und Verbindungsspeicherung entsteht, auf mehrere MA's verteilen. Als weiterer Vorteil wird die Robustheit der Architektur gegen eventuelle Ausfälle eines MA in den höheren Hierarchieebenen erhöht.

3.3.4 Optionales Vorgehen bei einer Domain mit mehreren Borderroutern

Castellucia et.al. [eal.98] schlagen vor, ein MN pro Domain einzurichten und dieses falls möglich direkt mit dem Borderrouter zu verbinden. Verfügt eine Domain über mehere Borderrouter, die sie mit dem Internet verbinden und gibt es nur ein MN, so kommen von außerhalb an den MH gerichtete Pakete eventuell über eine suboptimale Route zum MH, da die Pakete erst zum MA geleitet werden müssen, bevor sie zum MH geschickt werden können. Es wird ein Algorithmus vorgeschlagen, der ein MN pro Borderrouter einer Domain einsetzt und jedes MN mit zwei Präfixen versieht, einem dem MN eigenen Präfix (P_y) und einem Präfix für alle MN's (P). Die Informationen über die jeweiligen Präfixe und die IP- Adressen des jeweiligen MA werden über eine erweiterte Option der “ Mobility Information Option“ mit den Router Advertisements an die MH's verschickt. Ein MH konfiguriert seine VCoA mit dem Präfix P und registriert seine Verbindung (PCoA,VCoA) mit jeden einzelnen MN. Wenn zusätzlich noch jeder Borderrouter so konfiguriert wird, daß er Pakete, die Zieladressen mit dem Präfix P haben an das direkt an ihn angeschlossene MN weiterleitet, wird das Problem des suboptimalem Routings gelöst. Pakete an VCoA eines MH werden dann einfach an das nächstgelegene MN weitergeleitet, vom MA aufgenommen, entkapselt und weitergeleitet. (siehe Abbildung 6)

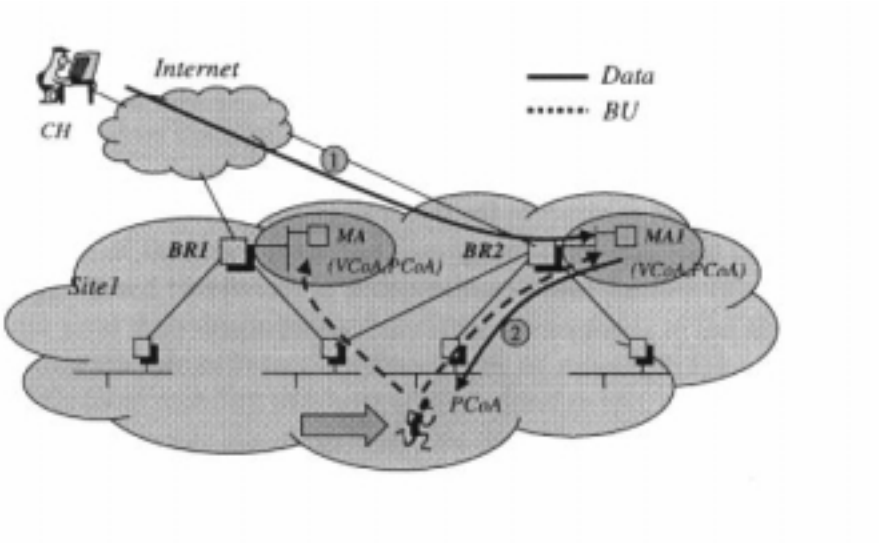

Abbildung 6: Lösung für das Multiborderrouterproblem.

Für eine Domain mit mehreren Hierarchiestufen und mehreren Borderroutern können sich zwei Probleme ergeben.

- die Router Advertisement Nachrichtenpakete können sehr schnell sehr groß werden, da für jedes MN Informationen mitgeschickt werden müssen

- die Signalisierungsbelastung durch lokale BU's kann sehr hoch werden, da jedes an einen Borderrouter angeschlossene MN von den MH's periodisch benachrichtigt werden muss

Deshalb wurde vorgeschlagen, die Multi- Borderrouter Registrationserweiterung für größere Domains optional zu machen.

3.4 Vergleich von HMIPv6 mit MIPv6

Castellucia et. al. [eal.98] stellen einen Vergleich zwischen MOBILE IPv6 und der von ihnen vorgeschlagenen Erweiterung des hierarchischen MOBILE IPv6 an in den Punkten:

- Routing performance, d.h. welche zusätzlichen Routingverzögerungen werden erzeugt ?
- Handover performance, d.h. wie schnell können die Übergabephasen durchgeführt werden ?
- Skalierungseigenschaften, d.h. wie verhält sich das Schema wenn die Netzwerke anwachsen und die Anzahl der mobilen Hosts steigt ?

In den ersten beiden Punkten kamen sie zu dem Ergebnis, daß die Routing- und Übergabe performance der beiden Schemata sich ziemlich ähnlich sind. Bei MOBILE IP ist das Routing optimal, d.h. die Pakete nehmen den kürzesten Weg von den CH's zum MH außer den ersten Paketen, die über den HA des MH laufen müssen. Beim hierarchischen MOBILE IP ist eine extra Umleitung über den Mobilitätsagenten notwendig. Allerdings sind die Kosten dieser Umleitung gering, vor allem wenn der MA wie von vorgeschlagen nahe am Borderrouter ist. Im HMIPv6 Vorschlag werden lokale Handoffs innerhalb der Domain abgewickelt. In MOBILE IPv6, wo location updates das ganze Internet überqueren müssen, um die CH's des MH zu erreichen, wird ein Mechanismus bereitgestellt, um die Übergaben abzuglätten. Ein mobiler Knoten kann nachdem er zu seinem neuen Defaultrouter gewechselt hat, ein BU zu seinem bisherigen Defaultrouter schicken, das ihn bittet, alle ankommenden Pakete zu seiner neuen CoA zu schicken.

Da Skalierbarkeit das Ziel von HMIPv6 ist, ist es nicht weiter verwunderlich, daß sich die Hauptunterschiede zwischen den beiden Ansätzen in puncto Skalierungseigenschaften ergaben. Als wichtigstes Kriterium, um Skalierungseigenschaften eines Mobilitätsmanagementschemas zu charakterisieren, wurde die Signalisierungsbelastung auf den Internetbackbone gesehen, d.h. die Bandbreite, die durch mobilitätsunterstützende Kontrollnachrichten, wie z.B. binding updates, verursacht werden. Die lokale Signalisierungsbelastung wurde außer Acht gelassen, da, so die Argumentation, lokale Ressourcenknappheit nicht zu einen kritischen Punkt bei der Gestaltung werden dürfte. Für beide Schemata wurde die aggregierte Bandbreite abgeschätzt, die durch Signalisierungsbelastung im Internet eingenommen wurde. Diese aggregierte Bandbreite war unabhängig von der Anzahl von Knoten, die die BU's bis zu ihrem Ziel zu überqueren hatten, sondern entspricht eher der Signalisierungsbandbreite für einen Link.

In der Schätzung wurde zwischen den zwei Arten von Intra-site Mobilität, innerhalb der eigenen (G_{heim}) und innerhalb einer fremdem (G_{fremd}) Domain), und der Inter-site Mobilität ($G_{transit}$) eines mobilen Hosts unterschieden und die durchschnittliche Signalisierungsbelasung dieser drei Bewegungsarten abgeschätzt. Dabei wurde die Frequenz, mit der die verschidenen BU's ausgesandt werden, der erzeugte Overhead in Paketen, die Anzahl der CH's sowie, abgeleitet aus der Untersuchung von G. Kirby [Kirb95] , die Tatsache, daß sich ein User zu 69 % in seinem Heimatnetz befindet in die Abschätzung mit einbezogen. Als Ergebnis erhielten sie eine Reduzierung der Signalisierungslast für einen Backbone durch emittierte BU's von

69 % bis hin zu 100 % je nach Bewegungsverhalten des Users. Eine Reduzierung um 100 % wird dann erreicht, wenn sich der MH nur innerhalb seiner Heimatdomain bewegt.

Errechnet wurde dies über

$$G_{durchschnitt} = 0.69 \times G_{heim} + 0.31 \times (\alpha \times G_{fremd} + \beta \times G_{transit})$$

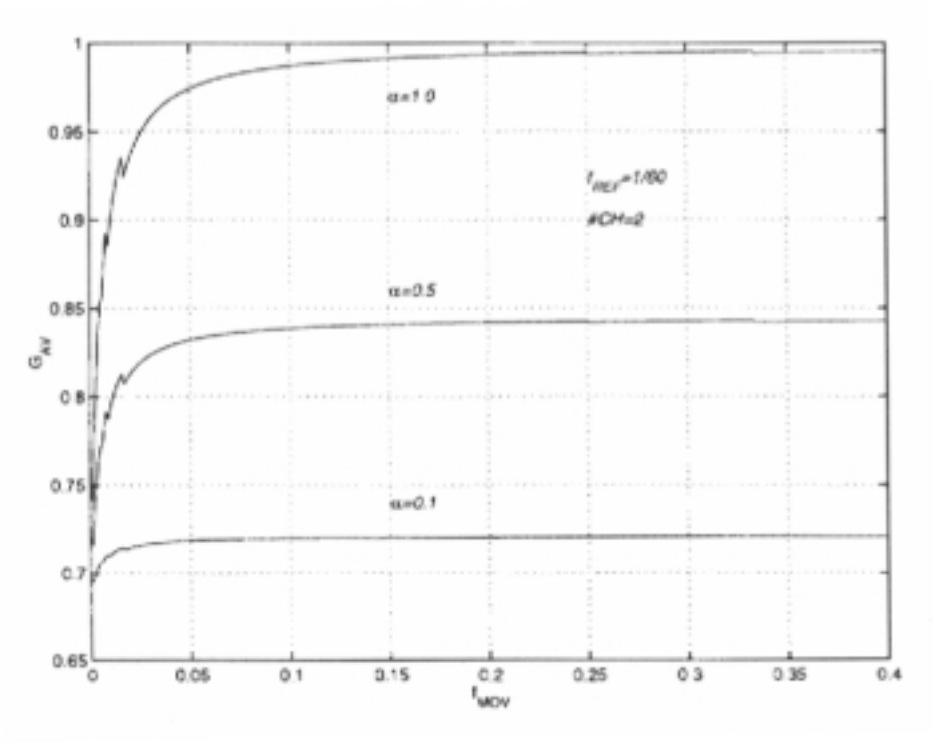

Abbildung 7: durschnittlicher Gewinn $G_{durchschnitt}$ von HMIPv6 über MIPv6

wobei

- $\alpha = \frac{N-1}{N}$
  mit $N = \emptyset$ - Anzahl verschiedener Anschlußpunkte, die ein MH innerhalb einer Domain bekommt, bevor er zu einer anderen Domain wechselt
- $\beta = \frac{1}{N}$
- $\alpha + \beta = 1$
- $G_{heim} = 1.0$
  da 100 % der BU's, die übers Internet geschickt werden bei HMIPv6 eingespart werden
- $G_{transit} = 0$
  da sich bei Inter-site Bewegungen MIPv6 und HMIPv6 exakt gleich verhalten
- G_{fremd} ist eine Funktion der Anzahl der CH's, der BU- Frequenz für Kommunikationspartner und bewegt sich zwischen 0 für keine Bewegung des MH und einer Zahl nahe 1.0 für häufige Bewegung des MH

4 Fazit

Obwohl der Aufbau einer Hierarchie zu Zwecken des Mobilitätsmanagements auch schon für MOBILE IPv4 von Caceres et. al. vorgeschlagen wurde [CaPa96] und es auch einen dem HMIPv6 ähnlichen Vorschlag von Charles Perkins gibt [Perk96] , so ist der hier vorgestellte Vorschlag diesen in puncto Skalierbarkeit, Flexibilität und Robustheit voraus, da er Gebrauch von neuen IPv6 Funktionalitäten macht. Die Herangehensweisen von Caceres und Perkins benutzen das Agent Advertisement auf dem niedrigsten Level einer Hierarchie, um einem MH eine Foreign Agent Hierarchie anzuzeigen. Dies verlangt allerdings auch, daß ein FA in jedem Subnetz verwandt werden muss, das MH's verherbergen soll, was aber eine starke Entwurfseinschränkung darstellt, die durch die Verwendung des Neighbor Discovery Mechanismus in HMIPv6 wegfällt. Somit wird in HMIPv6 keine Beschränkung bezüglich des Standortes des MA verlangt. Der Vorschlag des HMIPv6 ist im Vergleich zu den bisherigen hierarchischen Mobilitätsmanagementansätzen :

- *leichter einzusetzen* : Die vorgeschlagene Lösung kann in das aktuelle MOBILE IPv6 Protokoll ohne weitere Modifikationen eingebunden werden. Sie erfordert lediglich die Definition einer neuen Router Advertisement Option und einige kleinere Modifikationen auf seiten der mobilen Hosts.

- *flexibler* : Ein MH kann sich entscheiden einige Hierarchiestufen zu überspringen, wenn es nötig ist. So kann z.B. ein MH, der sich nicht allzu regelmäßig fortbewegt und/oder der Bandbreite auf dem letzten Hop (der auch wireless sein kann) durch Bergenzen der ausgesandten BU's sparen möchte, sich nur beim „höchsten" MA registrieren, um Kosten, die durch Umleitung und MA Verarbeitung entstehen können, zu umgehen.

- *skaliert besser* : Caceres und Perkins Vorschläge schreiben vor, daß die FA's in einer Baumstruktur arrangiert werden müssen, wobei der FA an der Wurzel einen Eintrag für jeden MH in der Domain verwalten muss. Dies kann bei einer großen Anzahl von MH's zu einem Problem werden. Im HMIPv6 können dagegen mehrere MA's zum Zweck der Lastverteilung in jeder Hierarchiebene eingebunden werden.

Zusammenfassend kann festgestellt werden, daß durch den Vorschlag des HMIPv6 die Signalisierungsbelastung für das Internet durch BU's im Vergleich zu MIPv6 klar reduziert wird und durch die Verwendung neuer IPv6 Funktionalitäten wie dem großen Adressraum und dem Neighbor Discovery Mechanismus ein Mobilitätsmanagementschema vorgestellt wird, das skaliert und dabei robust und flexibel ist.

Literatur

[CaPa96] Ramon Caceres und Venkata N. Padmanabhan. Fast and Scalable Handoffs for Wireless Internetworks. *Proc. 2st Annual International Conference on Mobile Computing and Networking*, 1996.

[eal.98] Claude Castelluccia et. al. (Hrsg.). A Hierarchical Mobile IPv6 Proposal. White Paper, Institut National de Recherche en Informatiquee et en Automatique, November 1998.

[Kirb95] G. Kirby. Locatin the User. *Communication International*, 1995.

[Perk96] Charles Perkins. Mobile-IP Local Registration with Hierarchical Foreign Agents. *Internet draft*, 1996.

[Schi00] Jochen Schiller. *Mobilkommunikation - Techniken für das allgegenwärtige Internet*. Addison Wesley. 2000.

[Solo98] James D. Solomon. *MOBILE IP - The Internet unplugged*. Prentice Hall. 1998.

Abbildungsverzeichnis

1 Wechsel eines MH in ein fremdes Netz. 4

2 Tunnelling von Pakete zu einer CoA in MOBILE IP. 5

3 Inter-site Mobilität 7

4 Intra-site Mobilität 9

5 Hierarchie einer Domain 10

6 Lösung für das Multiborderrouterproblem. 12

7 durschnittlicher Gewinn $G_{durchschnitt}$ von HMIPv6 über MIPv6 14

Tabellenverzeichnis

1 Begriffsdefinitionen für HMIPv6 8

Mikromobilität mit Cellular IP

Tingchao Yan

Kurzfassung

Das Papier stellt einen neuen Ansatz zur Unterstützung des Mobilehosts im Internet dar. Während existierende mobile host Protokolle (z.B.Mobile IP) gute Dienstleistungen im Globalmobilitätsbereich bieten können, funktionieren sie nicht effizient im Mikromobilitätsbereich, weil das Mobile IP Protokoll nicht für häufige Ortwechsel des mobile hosts geeignet ist. Cellular IP, das auf Mobile IP aufbaut, unterstützt den häufigen Ortwechsel der Mobilehosts dagegen viel besser, und ist deshalb eine gute Lösung für Mikromobilität. Es wird hier ausführlich diskutiert, wie Cellular IP funktioniert. Das Besondere des Protokolls ist, dass mittels der paging- und route- Methode das Cellular IP Cheap passive connectivity, um mehr Mobilehosts in einer Zelle aufnehmen zu können, und seamless Handoffs realisieren kann. Cellular IP setzt voraus, dass der Mobilehost zwei Zuständen (active und idle) hat. Damit unterhält jede base station in einem wireless access network zwei Caches (Paging Cache und Routing Cache). Durch die von den Mobilehosts hop-by-hop geschickten Kontrollpakete (oder Datenpakete) bekommt das wireless access network die umgekehrten Wege vom Gateway Router zu den Mobilehosts für paging und routing. Die Kontrollpakete laufen nur im wireless access network. D.h., sie sind vom Internet isoliert. Deshalb ist Cellular IP "transparent" zum Internet und ohne Änderung der IP Paketformats oder anderer Infrastrukturen im Internet zu implementieren

1 Einleitung

Schon seit längerem gibt es weltweit verschiedene Systeme zur mobilen und drahtlosen Kommunikation, gerade aber in den letzten zwei Jahren wachsen die Teilnehmerzahlen an Mobilfunksystemen explosionsartig an. Die mobile Nutzung von Computer ist ganz klar die Anwendung der Zukunft. Deswegen ist es sinnvoll, dass wir in diesem Papier die Hostmobilität in einer Umgebung annehmen, in der eine drahtlose Verbindung zum Internet typisch ist, im Gegensatz zur heutigen Zeit, in der es meist eine Ausnahme ist.

Wie wir schon wissen, bietet Mobile IP eine einfache und skalierbare Lösung für Globalmobilität. Aber es ist nicht günstig für häufigen Ortwechsel des Mobile Hosts (MH/MHs) und nahtlose Handoffs, weil nach jedem Ortswechsel des MHs eine lokale Adresse beachtet und mit dem Home Agent (HA) oder möglicherweise Distant Location Directory kommuniziert werden muss. Dagegen bieten third generation cellular systems (z.B. GSM) nahtlose Mobilität an. Aber es wird auf eine komplexe und aufwendig verbindungsorientierte Netzinfrastruktur aufgebaut, der Flexibilität, Robustheit und Skalierbarkeit fehlt.

Deshalb wurde Cellular IP vorgeschlagen, ein neues leichtes und robustes Protokoll, das Lokalmobilität unterstützt und durch Zusammenarbeit mit Mobile IP auch Globalmobilität unterstützen kann. Es ist ein Ansatz, der versucht möglichst viele Vorteile beider Systeme zu kombinieren.

Abbildung 1 zeigt die zwei Ebenen in dieser Struktur. Ein wireless access network besteht aus mehreren Zellen. Es greift auf das Internet durch einen gateway router zu. Die Globalmobilität wird durch den Handoff zwischen zwei wireless access networks unterstützt. Local

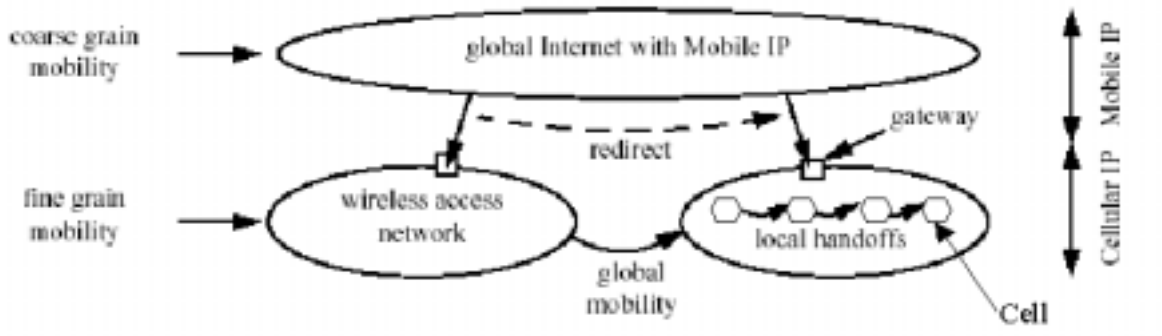

Abbildung 1: Mobile IP und Cellular IP

handoffs, auch Mikromobilität genannt, bedeutet der Zellenwechsel des MHs in einem wireless access network selbst. Die Globalmobilität funktioniert durch die Unterstützung von Mobile IP während die Mikromobilität durch die Unterstützung von Cellular IP funktioniert.

Dieses Papier ist folgendermaßen organisiert. In Teil 2 wird das Szenario von Cellular Internet erklärt. Danach wird Cellular IP ausführlich diskutiert, insbesondere die Idee von Cheap passive connectvity und seamless Handoffs mit der Hilfe der Paging- und Route- Methode. In Teil 4 gibt es eine kleine Zusammenfassung.

2 Das Szenario von Cellular Internet

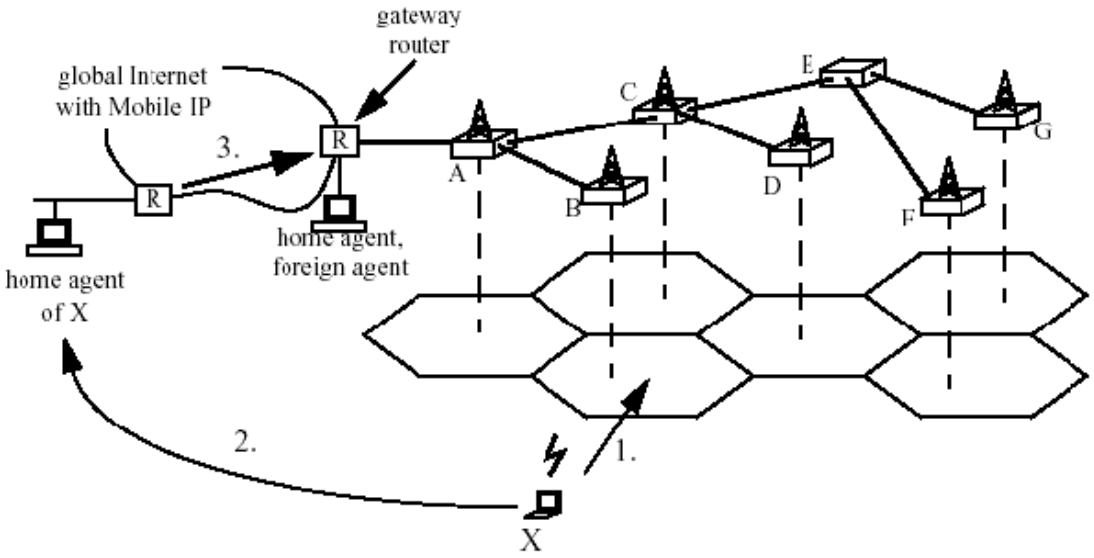

Abbildung 2: Wireless Access Network Modell

Wie Abbildung 2 zeigt, besteht ein wireless access network aus Knoten (base stations) A, B, C, D, F, G , die mit Leitungen verbunden sind. Außer diesen base stations, kann das Netz auch den Knoten E enthalten, der keine Funkanbindung bietet. Seine einzige Funktion ist die Verbindung von base stations.

Das wireless access networks wird über den Router mit dem Internet verbunden. Der Router, auch Gateway Router genannt, ist die beste Position für den HA oder FA. Es wird vorausgesetzt, dass das Mobile IP im ganzen Internet Globalmobilität zwischen verschiedenen

Wireless access networks (HA oder FA) unterstützt. Wenn ein Mobile Host (MH) (hier X) auf das Wireless access network zugreift (Schritt 1), meldet er sich bei seinem HA (Schritt 2) an, der später die Pakete, die mit X's IP Adresse ankommen, weiterleiten wird (Schritt 3). Solange der MH (X) noch in dem selbem Wireless access network (FA) ist, braucht sein HA keine neue Information über seine Position. D.h. die Lokalmobilität vom MH zwischen Zellen ist "transparent" zum HA. Nur wenn der MH über die Grenze zwischen zwei Wireless access networks läuft (dies gehört zu Globalmobilität), muss er sich an seinem HA neu anmelden. Normalerweise passiert der Wechsel zwischen Wireless access networks nicht so häufig. Ein MH wandert oft nur in einem wireless access network. Das braucht keine neue Anmeldung an den HA vom MH. Deshalb ermöglicht Cellular IP geringere Netzlast im Vergleich zu Mobile IP.

Nun werden einige wichtige Anforderungen für wireless access networks idenfiziert.

Base stations emittieren periodisch "beacon signals", damit die MHs eine verfügbare base station identifizieren können. Besuchende MHs werden behandelt, als ob sie zu diesem wireless access network gehören würden . Um den globalen Positionswechsel zu erleichtern, ist eine einfache und schnelle Lösung für die Anmeldung des Zugriffs auf wireless access networks zu finden.

Wie in GSM werden hier auch zwei Zuständen vom MH difiniert: active und idle. Ein MH ist active, falls er Datenpakete mit seinem Internetpartner austauscht. Andernfalls ist ein MH idle. Abbildung 3 stellt das Zustanddiagramm dar.

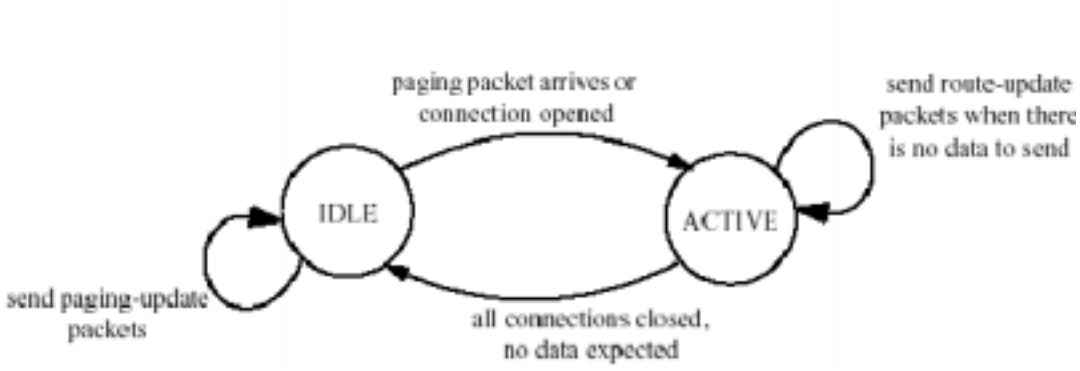

Abbildung 3: Mobile host state machine

Eine andere wichtige Anforderung ist, so viele MHs wie möglich in einem gegebenen wireless access network zu erlauben. Während active MHs hohen Kommunikationsaufwand (z.B. Bandbreite) brauchen, müssen idle MHs nicht so häufig mit base stations kommunizieren, um ihre Standorte genau mitzuteilen. Diese Anforderung wird mit cheap passive connectivity bezeichnet. Es geht um Page-update-Packets und Route-update-Packets, die in Teil 3 weiter diskutiert werden.

Manchmal überlappen sich Funkzellen. Diese Eigenschaft kann nahtlose Handoffs unterstützen. D.h., das wireless access network hat die Fähigkeit, dass der active MH ohne Unterbrechung der Erreichbarkeit durch die Zellen laufen kann. Jedoch soll das Protokoll auch ohne Überlappung der Funkzellen effizient funktionieren.

Da die Zelle einen sehr kleinen Bereich abdecken dürfte, ist es notwendig, dass das wireless access network gut funktioniert, wenn der Zellenwechsel des MHs häufig passieren. Es wäre nicht sinnvoll, dass nach jedem Zellenwechsel eine Kontrollnachricht vom MH an den Router geschickt wird. Jedoch ist es auch nicht nützlich, der MH im ganzen wireless access network keine Kontrollnachricht schicken zu lassen und im ganzen wireless access network zu suchen, wenn ein zu diesem MH geschicktes Packet ankommt. Hier gibt es einen "Trade

off"(Kompromiß). Die genaue Entscheidung vom Schicken der Kontrollnachricht ist ja nach der praktischen Situation. Deswegen muss das location management flexibel sein.

Ein wirksames Verfahren wird vom Location management gefordert, so dass die Standortinformation vom MH unterstützt werden kann, ohne das Netzwerk zu überlasten. Ferner muss Location management auch einen schnellen und effizienten Suchalgorithmus enthalten.

Ist der MH in einer Umgebung mit geringerer Zellenwechselhäufigkeit, ist genauere Standortinformation des MHs zu fordern. (Deshalb ist Paging (um den MH zu finden) einfacher.) Dagegen hängt hohe Zellenwechselhäufigkeit des MHs mehr vom schnellen Suchalgorithmus ab, weil das wireless access network weniger Standortinformationen von MHs bekommt, um die Netzwerklast zu beschränken.

Um billige MHs zu unterstützen, soll das wireless access network eine geringe Komplexität des MHs voraussetzen. Ein einfacher MH hat vielleicht keinen Speicher. Entweder im active Zustand oder im idle Zustand hat er gleiche elementare Aktionen. Beim Handoff ist auch keine besondere Aktion zu fordern.

Zusammenfassend können fünf wichtige Anforderungen vom Entwurf des Cellular IP Protokolls im wireless access network definiert werden:

- einfacher globaler Ortwechsel
- billige passive Verbindung (cheap passive connectivity)
- flexible Unterstürzung für den Handoff
- effizient location management
- wenige Anforderungen für einfachen MH

3 Cellular IP

Zusätzlich zu den in Teil 2 diskutierten Anforderungen ist das primäre Entwurfsziel von Cellular IP maximale Skalierbarkeit und Robustheit mit minimaler Komplexität zu bieten. Ein Cellular IP Netzwerk ist verteilt, d.h.:

- Kein Knoten kennt die gesamte Netzwerktopologie.
- Es existiert keine zentrale Datenbank oder andere "single points of failure".
- Kein Element im Netzwerk muss sich die Komplexität erhöhen, falls der abgedeckte Bereich vergrößert wird(Das bedeutet, dass die Zahl möglicher MHs in einem Bereich steigt.)

3.1 Paging und Routing Mappings

Wegen Einfachheit und Skalierbarkeit weiß kein Knoten in einem Cellular IP Netzwerk den genauen Standort von einem MH. Die an einem MH adressierten Pakete werden zu seinem aktuellen Base Station durch eine hop-by-hop Methode geleitet, indem jeder Knoten nur den Port vom nächsten Knoten weiß, um die Pakete weiterzuleiten. Diese begrenzte Routeinformation ist lokal für dem MH. Es ist nicht notwendig, dass die Knoten die ganze Topologie von dem wireless access network wissen müssen. Das Informationselement wird "mapping" genannt, weil es die ID vom MH zum entsprechenden Knotenport abbildet.

Mappings werden von den Paketen erzeugt und aktualisiert, die vom MH zum gateway router gesendet werden. Diese Pakete laufen im wireless access network zum gateway router durch die hop-by-hop Methode. Die Knoten auf dem Weg überwachen diese Pakete und bilden die Adresse von Sender zu eingehendem Port ab, um die mappings zu erzeugen. Damit erstellt das wireless access network einen umgekehrten Weg vom gateway router zum MH für zukünftig zum MH geschickten Pakete.

Um die Kontrollnachrichten zu minimieren, wird nach dem Handoff des MHs die mappings in den entsprechenden Knoten nicht sofort explizit gelöscht, sondern es wird eine Timeout verwendet. Das bedeutet, um seinen Weg vom gateway router zu seiner base station zu behalten, muss der MH periodisch Scheinpakete(dummy packets) schicken, falls es keine Datenpakete zu schicken braucht. Das heißt, solange ein MH in einen Dienstbereich(einer Funkzelle) ist, existiert der aktuelle Weg vom gateway router zur Base Station des MHs. Deshalb kann die Datenpakete später durch diesen Weg zum MH geschickt werden. Dieses Schema vereinfacht auch die Wanderung zwischen wireless access networks , weil Knoten keine Information vom MH brauchen, um mappings zu erstellen. Knoten müssen auch nicht informiert werden, wenn der MH dessen Reichweite verlässt.

Bezüglich Timer müssen zwei Faktoren berücksichtigt werden. Nach dem Handoff des MHs bleibt sein Weg zur alten base station noch gütig bis die mappings gelöscht werden. Falls Datenpakete zum MH in dieser Zeit geleitet werden, werden sie nicht nur zur aktuellen base station sondern auch zur alten base station geliefert. Diese verschwendeten Ressourcen können durch die Auswahl eines kleines Timerwertes in den base stations vermieden werden. Andererseits muss ein idle MH periodisch Scheinpaketen schicken, um den Weg zwischen seiner base station und seinem gateway router zu aktualisieren. Das wird aufwendig sein, falls der Timerwert zu klein ist.

Um das Problem zu überschwinden werden zuerst die Eigenschaften von diesen zwei Faktoren beobachtet. Um die Netzwerklast wegen nicht gelöschter mappings zu minimieren sollte der Timerwert der Paketübertragung gewählt. Dagegen um die Netzwerklast durch Scheinpakete zu minimieren sollte die Zeit von dem Zellenwechsel der MHs gewählt werden. Der letztere ist sicher länger.

Cellular IP löst das Problem durch die Benutzung zweier paralleler Strukturen von mappings. Zunächst unterhält jeder Knoten ein sogenannten Routing Cache(RC) für active MHs. Diese mappings haben einen kurzen Timerwert. Unabhängig davon unterhalten der Knoten einen anderen mapping Cache, der Paging Caches(PC) heißt. Diese mappings werden für beide idle und active MHs unterhalten. Der Timerwert ist länger. Normalerweise ist er einige Sekunden oder einige Minuten, entsprechend der Lokalmobilität des MHs. Der PC gibt Auskunft über den groben Standort des MHs, der RC dagegen über den genauen Standort des MHs.

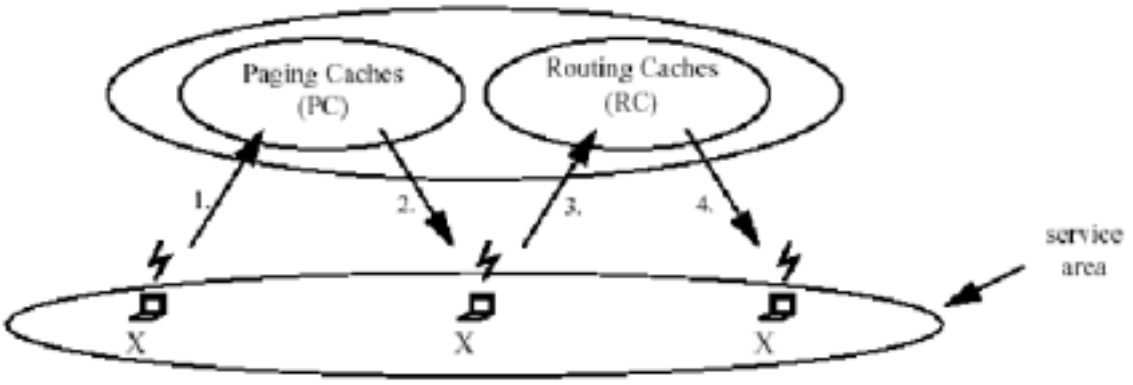

Abbildung 4: Paging und Routing

Abbildung 4 stellt die Beziehung zwischen PCs und RCs dar. Der idle MH behält die aktuelle PCs in einer niedriger Häufigkeit durch das Schicken der Scheinpakete (schritt 1). PCs haben einen relative längeren Timeout. Wenn ein zum MH geleitetes paging packet kommt, werden PC mappings verwendet, um den MH zu finden (schritt 2). Dann beginnt der Datenaustausch. Solange die Datenpakte dauernd ankommen, unterhält das MH RC mappings, indem es entweder Datenpakete oder ein Scheinpaket schickt (schritt 3). Die an den MH adressierte Datenpakete werden durch RCs, die nicht wie PCs einen genauen Weg zum MH bieten können, unmittelbar geleitet.

Diese Trennung von paging und routing hat noch einen Vorteil. Ein wireless access network kann eine große Zahl von MHs haben, die zufällig gleichzeitig auf das wireless access network zugreifen, aber von denen nur ein kleiner Prozent Satz aktiv ist. In diesem Fall enthalten PCs eine große Zahl von MHs, die Datenbank von PCs ist deshalb viel größer als von RCs. Da PCs nur für das Suchen des MHs verwendet werden, und keine hohe Rate von Daten geliefert wird, ist es möglich, dass der Netzwerkoperator einige bestimmte Knoten auswählen kann, die PCs unterhalten. Die andere Knoten, die keine PCs haben, broadcasten einfach die Suchnachrichten (paging packets). Mehr PCs bedeutet größere Datenbanken. Aber Je mehr PCs erstellt werden, desto genauere Standortinformationen werden geboten. Dadurch wird die Größe des Suchbereichs verkleinert. Diese Entwurfeigenschaft bietet dem Netzwerkoperator die Freiheit, das location management gemäß der praktischen Situation zu unterhalten.

3.2 Paging

Idle MHs erzeugen periodisch kurze Kontrollpakete, die paging update packets genannt werden, und schicken sie zu der nächsten erreichbaren base station. Wie Abbildung 5 darstellt, werden die paging update packets durch das wireless access network per hop-by-hop zum gateway router geleitet. Die PCs unterhaltende Knoten überwachen die durchgehende paging update packets und aktualisieren ihre Paging Cache, indem sie den Identifier des MHs zum Port, wodurch die paging update packets ankommen, abbilden. Der gateway router lässt einfach die paging update packets weg. Deshalb wird Cellular IP vom Internet isoliert. Das bedeutet, Cellular IP wird nur im wireless access network implementiert. Es ist “transparent” zum globalen Internet.

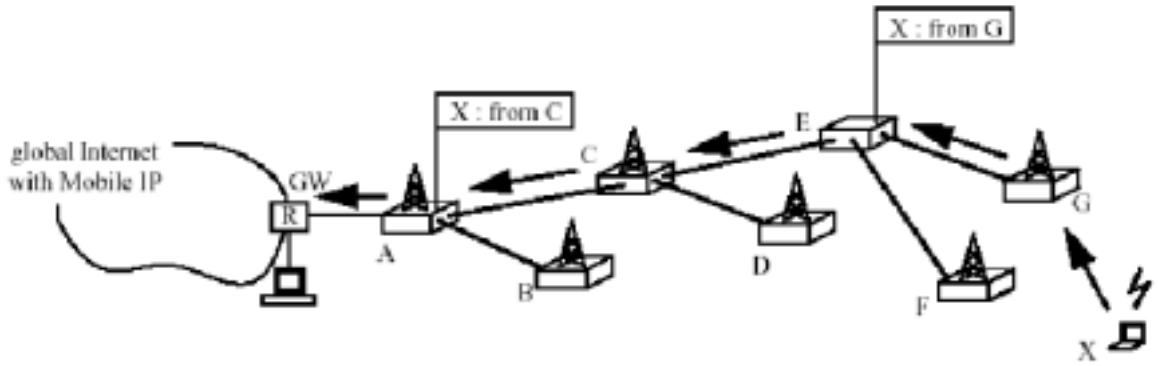

Abbildung 5: Paging update packet erstellt mappings in PCs

Jetzt wird Abbildung 5 erklärt. Der MH X ist gegenwärtig in der Zelle von Knoten G. Die von X erzeugte Paging update packets laufen zum Gateway Router durch Knoten G, E, C und A. In diesem Beispielsnetzwerk enthalten A und E PCs, aber C nicht. C leitet die Paging update packets einfach zum Gateway weiter. Der Knoten A weiß, dass die Paging update packets von X aus dem mit Knoten C verbundenen Port ankommen, während der Knoten E weiß, dass die Paging update packets aus dem mit Knoten G verbundenen Port ankommen.

Hier hat der Knoten E drei Ports. Der mit G (und F) verbundetn Port wird downlink Port von E genannt und der mit C verbundetn Port wird uplink Port von E genannt. Ein uplink Port ist der Weg zum Gateway Router, und ein downlink Port ist der Weg zum MH.

Der idle MH muss dauernd die paging update packets senden, um die PCs in allen Knoten zu aktualisieren. Überalterte mappings in PCs werden nach Timeout gelöscht. Wenn zum Beispiel der MH X zur Zelle F wandert, sind seine paging update packets nun zur Base Station F zu senden (siehe Abbildung 6). Während der Knoten A den Unterschied nicht weiß, erstellt der Knoten E ein neues mapping für X. Nach einiger Zeit wird das alte mapping im Knoten E gelöscht. Es ist möglich, dass es einen Zeitraum gibt, in dem die beide mappings (zu F und zu G) gütig sind. Diese Eigenschaft garantiert, dass beim Zellenwechsel der MH immer erreichbar ist.

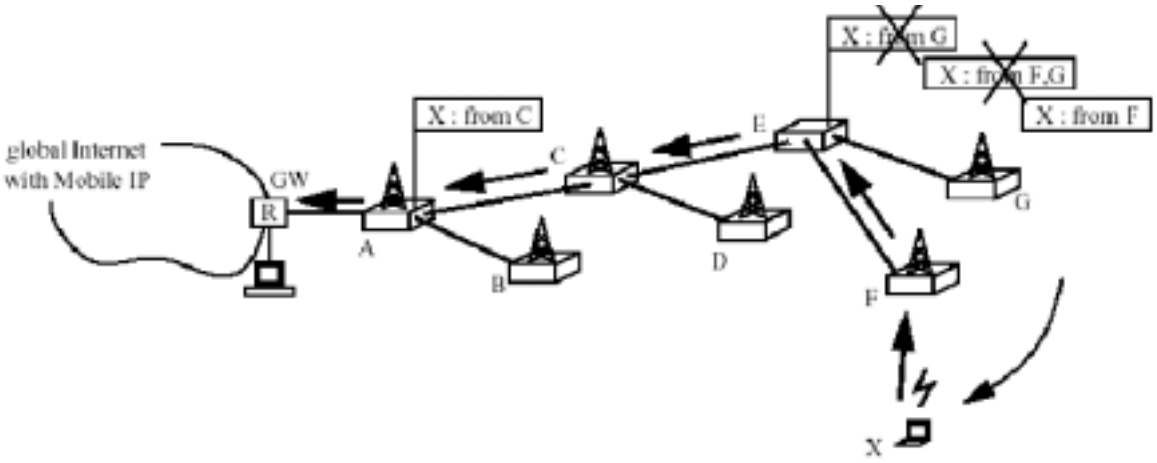

Abbildung 6: die Aktualisieren der PC für ein MH

Es wird angenommen, dass nun die an den MH X adressierte IP Pakete auf den Gateway Router ankommen. Es gibt keine aktuelle Routeinformation für X. (d.h. keine aktuelle RC Informationen, die in Teil 3.3 mehr diskutiert werden.) Jetzt werden PCs verwendet, um X zu finden. Der Gateway Router speichert die ankommende IP Pakete, und erzeugt ein Kontrollpaket, ein sogenanntes paging packet, das die ID von X enthält. Das Paket wird im wireless access network mit Hilfe von PCs per hop-by-hop ausgeliefert. Das ist der Weg, der umgekehrt von dem letzten von X zum Gateway Router geleiteten Paging update Packet besteht. Falls alle Knoten PCs haben, tritt eine volle hop-by-hop Lieferung auf. Anderenfalls liefern die Knoten, die keine PCs haben, einfach das Paket zu allen downlink ports weiter.

Um das paging packet zu routen, prüft A seinen PC und bemerkt, dass das letzte paging update packet durch das mit C verbundete Port angekommen ist. (siehe Abbildung 7) Deshalb leitet A das paging packet zu C, der keine PC Information hat und leitet das Paket in beide Richtungen (D und E) weiter. Das zu D geschickte Paket wird weggelassen, weil D weiß, dass der MH nicht in seiner Zelle ist. Das zu E geschickte Paket wird zu F geleitet. Endlich broadcastet F das Paket in seiner Zelle und findet MH X.

Nachdem der MH das paging packet bekommen hat, erzeugt er ein Kontrollpaket, ein sogenanntes route update packet, und schickt das Paket zu seiner base station(Knoten F). Ähnlich wie paging update packets, läuft das route update packet zum gateway router per hop-by-hop Methode. Gleichzeitig werden alle RCs in seinem Weg aktualisieren. Wenn das route update packet auf dem gateway router ankommt, sind alle RCs auf seinen Weg im aktuellem Zustand. (Jeder Knoten hat RC.) Die im Gateway Router zwischengespeicherten IP Pakete können jetzt zu ihrem MH ausgeliefert werden.

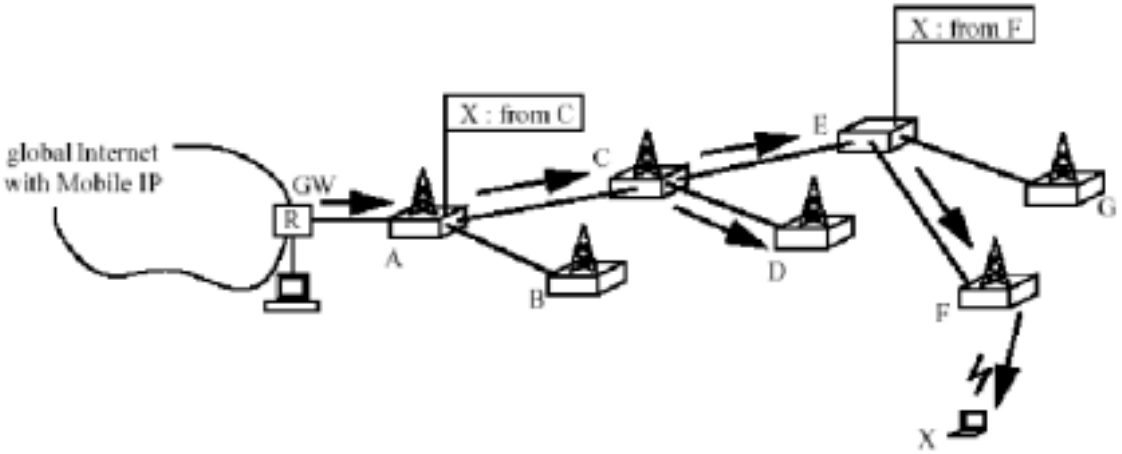

Abbildung 7: Paging Packets werden durch PC zu den MH geleitet

Der Suchprozess verzögert die Lieferung des ersten IP Pakete. Aber sobald der Weg aufgebaut wird, verwenden die folgende IP Pakete diesen Weg ohne einen weiteren Suchprozess. Jedoch wenn das MH zeitweilig unerreichbar ist und dann der Timeout des RCs ausläuft, wird das nächste ankommende IP Paket einen neuen Suchprozess auslösen, wenn der MH wieder erreichbar ist.

Ein weiteres Phänomen ist folgendes. Der idle MH schickt das Paging update Packet, wenn einer der Timer abläuft, oder wenn der MH beobachtet, dass er nun in einer neuen Paging area ist. Der active MH braucht kein Paging update Packet zu schicken, weil die andere vom MH geschickte Pakete auch die PCs in den Knoten aktualisieren können. Die Paging area kann eine Zelle sein, sie kann aber auch mehrere Zellen sein. In diesem Beispiel wird angenommen , dass G zu paging area M gehört und D und F zu paging area N gehören. Nach dem Zellenwechsel des MHs von M zu N sind alle PCs schon aktuell. Nun wird das Paging Paket zur paging area N geleitet. Das heißt, D und F broadcasten das Paket, um den MH X zu finden. (siehe Abbildung 8) Deshalb kann das wireless access network den MH finden, obwohl es nur ungefähre Standortinformation vom MH weiß.

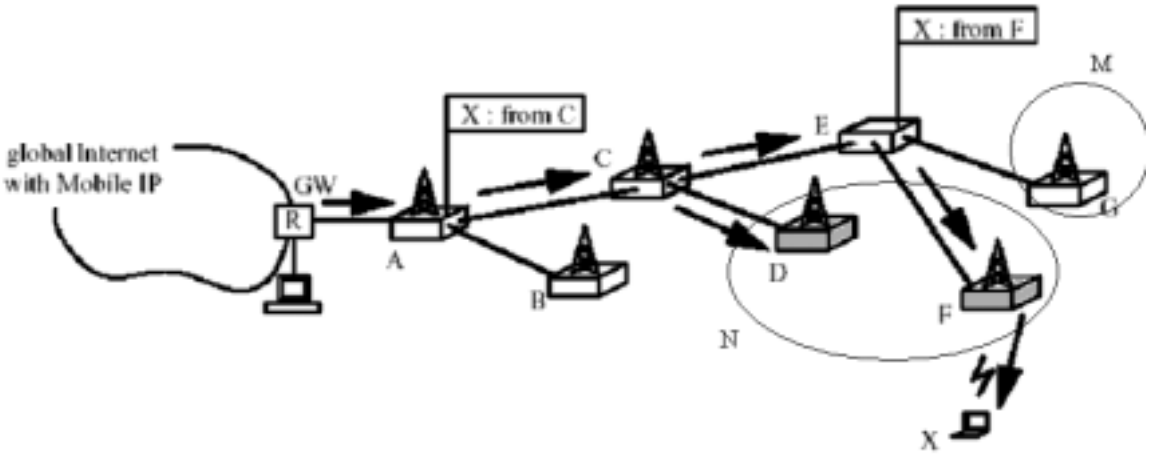

Abbildung 8: Paging area

3.3 Routing

Die vom MH geschickte Datenpakete werden zum Gateway Router per hop-by-hop geleitet. Knoten, die RCs enthalten, überwachen die laufende Datenpakete und verwenden sie, um die mappings vom Identifier des MHs zum entsprechenden downlink port zu erzeugen. Die an den MH adressierte Datenpakete werden über den umgekehrten Weg per hop-by-hop zum MH geleitet.

	Paging Cache (PC)	Paging Cache (PC)
Getrieben von	alle vom MH geschickten Pakete (Datenpakete, route update packet, paging update packet)	Datenpakete, route update packet
MH Zustand	idle MH und active MH	Nur active MH
Ziel	paging packets leiten	Zum MH geschickte Datenpakete leiten
Zeitskala	Mobilität vom MH	Datenpaketerate
Cache Timeout (möglicherweise)	9 Minuten	9 Sekunden
Zeit zwischen vom MH versendeten update packets (möglicherweise)	3 Minuten	3 Sekunden

Tabelle 1: Vergleich zwischen Routing und Paging

Die Struktur und die grundlegende Operation von Routing ist ähnlich wie Paging. In der Tabelle 1 gibt es eine Zusammenfassung für den Vergleich ihrer Eigenschaften. Wegen ihrer Ähnlichkeit ist es nicht notwendig, hier alles über Routing noch einmal darzustellen. Das Besondere ist, dass der RC nur active MHs behandelt. Der Timeout von RC ist viel kleiner als der Timeout von PC.

Manchmal empfängt der MH die Datenpakete, ohne Datenpakete zu schicken (z.B. in UDP). Um die aktuelle RCs in den base stations zu unterhalten (so dass das nochmalige paging vermieden werden kann), muss der MH periodisch route update packets schicken. Genau wie Datenpakete aktualisieren route update packets die RCs und gewährleisten, dass der per hop-by-hop vom Gateway Router zum MH Weg aktuell ist. PCs werden auch aktualisiert, wenn MHs active sind. Aber active MHs brauchen die paging update packets nicht zu schicken, weil die PCs auch von route update packets und von denen von MHs verschickten Datenpaketen aktualisiert werden können.

Wenn ein Handoff passiert, ist es möglich, dass es einen Zeitraum gibt, in dem die beide mappings von RCs (die alte und die neue) gütig sind. D.h, die zum MH geschickte Datenpakete werden zu beiden Wegen geliefert. Nach einiger Zeit werden die alte mappings in den Knoten gelöscht. Diese Eigenschaft unterstützt den nahtlosen Handoff beim Zellenwechsel des MHs, falls die Zellen überlappen . (Man kann auf Abbildung 6 verweisen.)

3.4 Handoff

Der oben dargestellte Mechanismus garantiert, dass der Handoff bei Datenaustausch automatisch behandelt wird. Der Handoff in Cellular IP ist immer vom MH ausgelöst. Wenn der active MH auf eine neue base Station zugreift, leitet er seine Pakete (Datenpakete oder Route

update packets) von der alten zur neuen base station um. Das erste umgeleitete Paket wird automatisch einen Weg, der zwischen dem Gateway Router und der neuen base station ist, durch RCs mappings erzeugen.

In einem Zeitraum vor dem Timeout der RC mappings werden die an den MH adressierte Datenpakete zu beiden der alten und der neuen base stations geliefert. Deshalb darf der Handoff soft sein, falls der MH die beide base stations gleichzeitig hören kann. (d.h. Die Zellen überlappen.) Anderenfalls ist der Handoff hart. Nach einiger Zeit wird der Weg zur alten base station gelöscht werden, während die Datenpakete dauernde zur neuen base station geleitet werden.

Abbildung 9 stellt ein Szenario vom Handoff dar. Beim Datenaustausch wandert der MH X von der Zelle F zur Zelle D. Wie die Fahnen zeigen, ist der alte Routingweg vom gateway router zum MH durch A, C, E, F (schritt 1). Gerade nach dem Zellenwechsel schickt der MH X Pakete (Datenpakete oder Route update packets) zum gateway router, um den neuen Routingweg zu erzeugen(durchgezugener Pfeil). Die neue Routingwege sind dann A, C, D und A, C, E, F (schritt 2). Nach dem Timeout der RCs ist der schließliche neue Routingweg A, C, D. (Schritt 3) Der Handoffprozess ist dann beendet. In dem ganzen Prozess wird der RC von A nicht geändert.

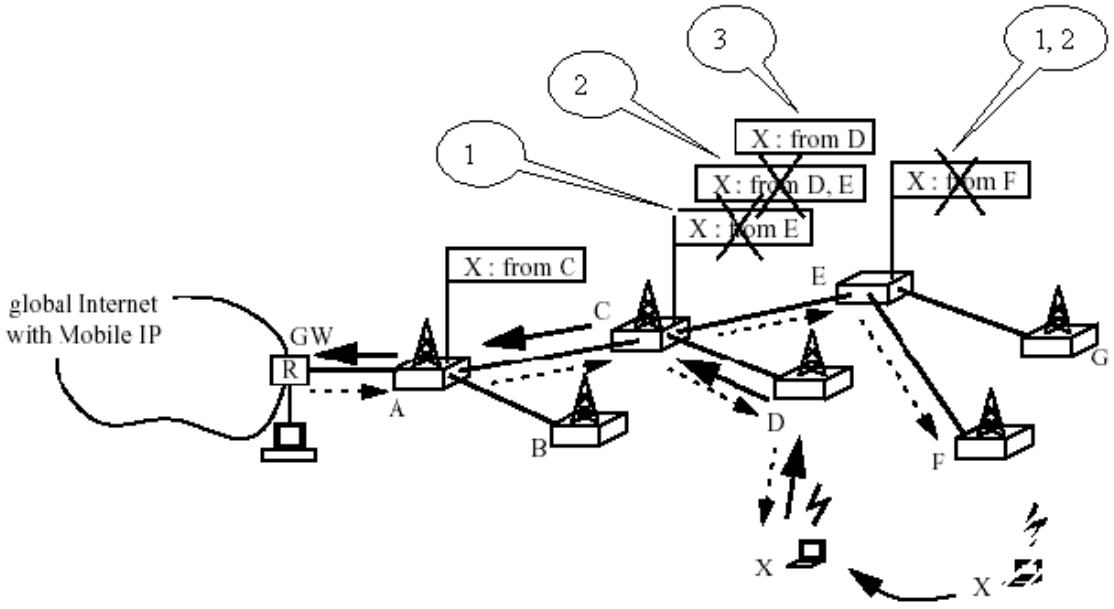

Abbildung 9: Handoff

Normalerweise findet der Datenaustausch zwischen dem MH und dem gateway router in beide Richtungen statt. Die vom MH geschickte Paketen können dann den Routingweg aufrechterhalten. Falls in einiger Zeit (die von timeout abhängt) der MH keine Datenpaketen schickt (z.B., in UDP), muss er stattdessen ein route update packet schicken.

Ob der Handoff hard oder soft ist, ist eine Implementierungssache und abhängig von der Funkanbindung der base stations.

3.5 Sicherheit

Jedes Cellular IP Netzwerk(wireless access network) hat einen geheimen Schlüssel, der arbiträr lang ist. Alle base stations in dem wireless access network kennen den Schlüssel, aber

MHs und andere base stations kennen ihn nicht. Wenn der MH sich bei dem gateway router eines wireless access network erst mal anmeldet, muss er authentifiziert und ermächtigt werden. Die anfängliche Authentifizierung und Ermächtigung kann auf beliebigen symmetrischen oder asymmetrischen Verschlüsselungsmethoden basieren. Nach der Authentifizierung kann der gateway router den Netzwerkschlüssel mit der IP Adresse des MHs mittels MD5 Hash verbinden, um die PID vom MH zu bekommen:

PID := MD5(network key, IP address of MH)

Dann erfasst der gateway router den public key des MHs von der CA (Certificate Association), entschlüsselt die PID und sendet sie zum MH. Nun haben der MH und das Cellular IP Netzwerk denselben Schlüssel. Die PID bleibt unverändert beim Handoff. Sie ist einfach von jeder base station zu rechnen.

Die PID ist anwendbar bei der Authentifizierung der Pakete, die vom MH geschickt werden. Die Authentifizierung wird schnell durchgeführt. Sogar beim Handoff des MHs ist der Vorgang der Authentifizierung auch nicht langsam. Deshalb können die base stations leicht die Gültigkeit des Pakets prüfen.

Um den Angriff durch die Fabrication des MHs (d.h. Der böswillige Host schickt die Pakete, als ob es ein zum wireless access network gehörendes MH wäre.) zu verteidigen, ist es notwendig, dass jedes vom MH geschickte Datenpaket und Kontrollpaket authentifiziert wird, bevor es die Paging Caches und die Routing Caches beeinflusst und zum Gateway Router geleitet wird. Deshalb ist die Geschwindigkeit der Authentifizierung sehr wichtig.

3.6 Rechnungsstellung(Charging)

Die Anbieter des Cellular IP Netzwerks (wireless access network) können die Kosten der MHs gemäß der Verbindungszeit oder der ausgetauschten Datenpakete berechnen. Rechnungsinformationen sind am besten vom gateway router zu sammeln. Der gateway router empfängt alle Kontrollpakete und kann deshalb alle notwendigen Informationen bekommen. Es ist auch möglich, bei der Rechnung beide Faktoren (Zeit und Datenflut) zu bewerten.

4 Zusammenfassung

In diesem Papier sind zuerst die Nachteile von existierenden mobile host protocols (z.B.Mobile IP) diskutiert, die trotz ihrer guten Dienstleistungen im Globalmobilitätsbereich nicht effizient im Mikromobilitätsbereich funktionieren. Danach wird ein neues Protokoll—-Cellular IP erklärt. Das auf Mobile IP aufgebaute Cellular IP kann den häufigen Ortwechsel des Mobilehosts unterstützen. Damit ist es sehr geeignet für Mikromobilität. In Teil 2 wird das Szenario vom Cellular Internet dargestellt. Dadurch werden die wichtigen Anforderungen vom Entwurf des Cellular IP Protokolls im wireless access network difiniert. In Teil 3 sind die Details von Cellular IP erklärt. Das Besondere des Protokoll ist, dass mittels der paging- und route- Methode das Cellular IP Cheap passive connectvity (um mehr Mobilehosts in einer Zelle aufnehmen zu können)und seamless Handoffs realisieren kann. Cellular IP setzt voraus, dass der Mobilehost zwei Zuständen (active und idle) hat. Damit unterhält jede Base station in einem wireless access network zwei Caches (Paging Cache und Routing Cache). Durch die von den Mobilehosts hop-by-hop geschickten Kontrollepakete (oder Datenpakete) bekommt das wireless access network die umgekehrten Wege vom Gateway Router zu den Mobilehosts für paging und routing. Die Kontrollepakete laufen nur im wireless access network. D.h., sie sind vom Internet isoliert. Deshalb ist Cellular IP “transparent” zum Internet und einfach zu verwirklichen im ganzen Internet in der Zukunft ohne Änderung der IP

Paketformats oder anderer Infrastrukturen. Der aktuelle Forschungsstand steht im Internet http://comet.ctr.columbia.edu/cellularip/ zur Verfügung.

5 Danksagung

Ich möchte mich bei meinem Betreuer Dipl.-Ing. Kilian Weniger bedanken, welcher mit vielen Hilfen und Hinweisen zur meiner Ausarbeitung beitrug.

Literatur

[CaGo00] Andrew T. Campbell und Javier Gomez (Hrsg.). *Design, Implementation, and Evaluation of Cellular IP*, IEEE, 2000.

[CaGV99] Andrew T. Campbell, Javier Gomez und Andras G. Valko (Hrsg.). *An Overview of Cellular IP*, IEEE, 1999.

[CGWK+99] A. Campbell, J. Gomez, C-Y. Wan, S. Kim, Z. Turanyi und A. Valko (Hrsg.). *Cellular IP, Internet-draft*, IEEE, 1999.

[Schi00] Jochen Schiller. *Mobilkommunikation*. Addison-Wesley. 2000.

[Valk99] Andras G. Valko (Hrsg.). *Cellular IP:A New Approach to Internet Host Mobility*, 1999.

[VGKC99] Andras G. Valko, Javier Gomez, Sanghyo Kim und Andrew T. Campbell (Hrsg.). *On the Analysis of Cellular IP Access Networks*, 1999.

Abbildungsverzeichnis

1 Mobile IP und Cellular IP 18

2 Wireless Access Network Modell 18

3 Mobile host state machine 19

4 Paging und Routing 21

5 Paging update packet erstellt mappings in PCs 22

6 die Aktualisieren der PC für ein MH 23

7 Paging Packets werden durch PC zu den MH geleitet 24

8 Paging area 24

9 Handoff 26

Tabellenverzeichnis

1 Vergleich zwischen Routing und Paging 25

AAA-Infrastruktur in Mobile IP Netzen

Tilmann Rothhammer

Kurzfassung

Mobile IP- und AAA-Arbeitsgruppen arbeiten gegenwärtig die Anforderungen an eine Infrastruktur zur Authentifizierung, Authorisierung und Rechnungsstellung aus. Dieser Beitrag fasst die bisher aufgestellten Anforderungen aufgrund des entwickelten Basis-Modells, der Besonderheiten für Mobile IP und der notwendigen Skalierbarkeit in der Praxis zusammen. Anschließend werden die benutzten Protokolle RADIUS und Diameter eingeführt, die grundsätzlichen Protokollabläufe erläutert und mit den aufgestellten Anforderungen verglichen. Zum Abschluß erfolgt ein Ausblick auf Veränderungen im Zusammenhang mit IPv6.

1 Einleitung

Die zunehmende Popularität mobiler Geräte stellt die Netzbetreiber gleichzeitig vor neue Herausforderungen und Chancen: Allgegenwärtiger Zugang zum mobilen Internet verbunden mit Zusatzdiensten wie email usw. verspricht ein wichtiger Bestandteil im Angebot von Mobilfunkbetreibern und ein milliardenschwerer Markt zu werden. Hierfür sind die technischen Voraussetzungen zu schaffen, dass die existierenden Sprachdienste neben den paketvermittelten Datendiensten bestehen können [McHi00]. Im Speziellen dürfen die Paketdienste nicht zu unnötigen Kosten in der Netzinfrastruktur führen, d.h. die vorhandene Netzwerkstruktur soll am besten unverändert bleiben.
Normalerweise erhalten Kunden mittels eines Internet Service Providers (ISP) Zugang zu Internet-Diensten in ihrer „Home Domain“. Die Moblilität der Nutzer erfordert aber abhängig vom Aufenthaltsort den Zugriff auf „network access servers“ (NAS) fremder Netzbertreiber [Metz99]. Um ihren Kunden größtmögliche Netzdeckung anbieten zu können, müssen sie partnerschaftlich die Ressourcen zusammen nutzen und dabei aber die Kontrolle über den Zugang, die Nutzung und die Rechnungsinformation behalten.
Mobile IP stellt keine Lösungen für Zugangskontrolle und Rechnungsinformationsübertragung zur Verfügung. Die Lösung dieser Aufgaben in einem Rahmenwerk ist der Kern von Authentication, Authorization und Accounting, kurz AAA. Ziel ist die Koordination der verschiedenen Anforderungen über verschiedene Netzwerktechnologien und Plattformen. Zum Verständnis der Probleme werden im nächsten Abschnitt die Begriffe, die im Zusammenhang mit AAA stehen kurz erläutert und gegenseitige Abhängigkeiten aufgezeigt. Die Darstellung stützt sich dabei auf Metz [Metz99] und Perkins et al. [GHJP00].

1.1 Begriffserklärungen AAA

Authentication beschreibt den Vorgang der Identifikation des Nutzers, bevor ihm Zugriff zum Netzwerk gewährt wird. Dieser Prozess beruht auf der Überlegung, dass der Endnutzer sich mittels eindeutiger Informationen, in der Literatur „credentials“ genannt, identifizieren kann.

Denkbar sind Nutzername und Passwort, geheime Schlüssel oder ähnliches. Diese Daten werden an den AAA-Server weitergeleitet und überprüft. Nur im Falle positiver Übereinstimmung wird der Zugang zum Netzwerk gewährt.

Authorization bestimmt die Rechte des Nutzers, d.h. zu welchen Ressourcen er Zugang erhält. Beispiele sind der Zugang zum Internet, die Bereitstellung einer IP-Adresse oder Zugang zu privaten Ressourcen innerhalb der fremden Domäne. Authentication und Authorization werden im allgemeinen zusammen ausgeführt.

Accounting bezeichnet den Vorgang der Buchführung über die Ressourcennutzung. Diese Information wird unter anderem zur Erstellung der Rechnung, Überwachung oder Kapazitätsplanung verwendet.

Es wird deutlich, dass die AAA Dienste teils aufeinander aufbauen bzw. voneinander abgeleitet sind, auf jeden Fall aber eng miteinander verbunden sind.

2 Anforderungen an AAA

Zur Nutzung von Mobile IP zwischen verschiedenen Domänen, die AAA Dienste benötigen, müssen kompatible Protokolle entwickelt werden. Dies kann nur geschehen, wenn die Anforderungen an die Protokolle klar definiert sind. Ziel dieses Abschnitts ist es, die Anforderungen im Umfeld von Authentifizierung, Autorisierung und Accounting darzustellen.
Dafür wird zuerst das Basis-Modell vorgestellt und die daraus folgenden Anforderungen abgeleitet. Anschließend wird auf die Anforderungen in Zusammenhang mit IP hingewiesen und dann auf die Besonderheiten bei Mobile IP eingegangen. Im letzten Teil wird eine abgewandelte Architektur, das Broker Modell, vorgestellt, die in der Praxis das Problem der Skalierbarkeit löst.

2.1 Das Basis Modell

Wie in der Einleitung beschrieben nutzt ein mobiles Endgerät, im Folgenden client genannt, häufig Ressourcen einer fremden adminstrativen Domäne, der sogenannten „foreign domain“. Ein Agent in der foreign domain, attendant genannt, antwortet auf die Anfrage des client. Bevor er aber den Zugriff auf die angeforderten Ressourcen gewährt, wird er eine Authentifizierung mittels „credentials“ verlangen. Unter Umständen ist die foreign domain in der Lage, diese zu verifizieren, im Allgemeinen aber kann dies nur die home domain.
Der attendant hat also im allgemeinen keinen direkten Zugriff auf die benötigten Daten. Daher kontaktiert er den AAA-Server seiner Domäne, den AAAL. Dieser verfügt über ausreichend Informationen um die home domain des client zu identifizieren und eine Verbindung ohne weitere AAA Agenten mit dem AAAH aufzubauen. Die home domain verfügt über alle notwendigen Informationen zur Authentifizierung der geheimen Informationen. Bestätigt sie die Identität des client, wird der AAAL und über diesen der attendant informiert und die angeforderten Ressourcen freigegeben. Abbildung 1 verdeutlicht die beschriebenen Beziehungen.
Ein Beispiel für das beschriebene Modell ist die Nutzung von RADIUS zum Zugang mobiler clients in das Internet via lokaler ISPs (Internet Service Provider). Der ISP wird den Zugang erst ermöglichen, wenn die Bezahlung seiner Leistung gesichert ist. Der ISP lässt dafür die angebotenen Sicherheitsinformationen von der home authority überprüfen.

Das beschriebene Modell impliziert ein Sicherheitsmodell, dessen einzelnen Beziehungen es zu identifizieren gilt.

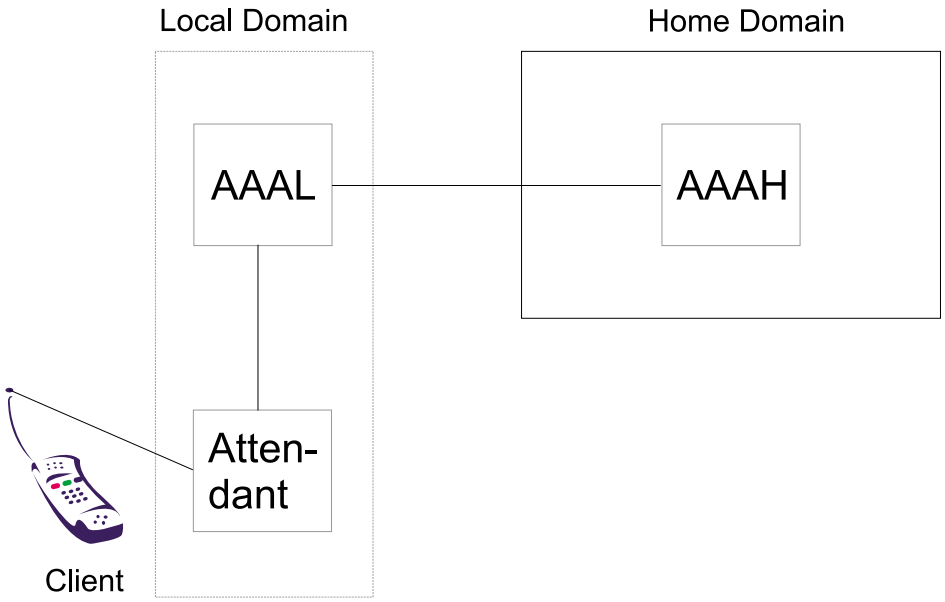

Abbildung 1: Das Basis Modell

1. Der Kunde, z.B. der mobile Laptop und der AAAH haben eine Sicherheitsbeziehung, da auf diese Weise die home domain des Kunden festlegt wird. Durch diese Sicherheitsbeziehung gehört der Kunde zur Domäne des AAA-Servers.

2. Zwischen dem AAAL und dem AAAH muss eine dauerhafte oder dynamisch erstellte Sicherheitsbeziehung bestehen, damit die credentials des Kunden überprüft werden können und auch die Rechnungsinformation übertragen werden kann. In dieser zweiseitigen Sicherheitsbeziehung liegt der Haupthinderungsgrund für das beschriebene Modell, da diese Architektur in Anbetracht der zahlreichen verschiedenen AAAH nicht skaliert. Eine Lösung dieses Problems wird im Abschnitt 2.4 Das Broker Modell gezeigt.

3. Auch der attendant und der AAAL teilen sich eine Sicherheitsbeziehung, damit eine gesicherte Übertragung der Kundendaten möglich ist.

Die obige Aufzählung lässt nach [Perk99] weitere Anforderungen an die Architektur erkennen:

- Während der Überprüfung der credentials muss der attendant die Verbindung mit dem client halten und auf dessen Anfragen reagieren.
- Es sind Fälle denkbar, bei denen der attendant gleichzeitig Anfragen verschiedener clients behandeln muss.
- Die im attendant benötigte Infrastruktur sollte möglichst billig sein, damit sie so oft wie möglich aufgebaut wird, um zahlreiche Kunden bedienen zu können.
- Das mobile Endgerät muss in der Lage sein ohne vorherigen Kontakt mit seiner home domain fälschungssichere credentials zu erzeugen.
- Die credentials sollten derart beschaffen sein, dass die zwischengeschalteten Knoten keine geheimen Information daraus lernen können.
- Der attendant soll Quality of Service Anforderungen des client erfüllen können.
- Der attendant muss sich gegen Wiederholungsangriffe schützen.

Aus dem praktischen Betrieb von RADIUS-Servern haben sich weitere nützliche Anforderungen an die AAA-Architektur gezeigt. Der interessierte Leser findet diese in [GHJP00].
Das gezeigte Modell ist im Großen und Ganzen kompatibel mit den Bedürfnissen von Mobile IP. Die notwendigen Änderungen werden weiter unten im Abschnitt 2.3 erläutert.

2.2 Anforderungen bezogen auf Basis-IP-Verbindungen

Die oben aufgelisteten Anforderungen beziehen sich auf das Zusammenspiel zwischen den funktionalen Einheiten. Sie sind unabhängig von der zugrundeliegenden Netzwerkaddressierung. Viele mobile Endgeräte benötigen aber während der Initialisierungsphase IP-spezifische Dienste:

- Der AAA-Server muss dem client eine geeignete IP-Adresse zuweisen können
- Der AAA-Server kann den client nicht mittels der IP-Adresse identifizieren

Der zweite Punkt ergibt sich aus der Überlegung, dass der client unter Umständen noch keine IP-Adresse besitzt und daher anders erkannt werden muss. AAA-Server benutzen heute den Network Access Identifier (NAI) zur Identifikation der clients. Die Form des NAI (user@realm) ermöglicht dem AAAL die home domain des client zu erkennen und so den Authentifizierungsprozess, wie im Basis-Modell ausgeführt, durchzuführen.

[CaPe00] schlagen jetzt einen Weg vor, mit dem sich der mobile client mittels der Mobile Node NAI extension selbst identifizieren kann, da unter Umständen der NAI nur den client, nicht aber seine home address eindeutig identifiziert. Dadurch wird es möglich, dass der client zur Verbindung mit der foreign domain berechtigt wird, ohne eine home address zu haben.
Wenn die NAI extension im Registration Request verwendet wird, kann das home address-Feld zu 0 gesetzt werden. In diesem Fall muss der foreign agent in seinen Registrierungs-Protokollen den NAI zusammen mit dem Idenfication Feld verwenden.
Im Zusammenhang mit der Mobile-IPv4-Konfigurations-Option mit Internet Protocol Control Protocol, kurz IPCP[1], im Laufe derer der client eine IP-Adresse vom home agent erhalten will, ergeben sich folgende Änderungen: Der PPP peer soll dem mobile Endgerät eine co-located care-of address zuweisen. Falls in der IP Address Configuration Option zu IPCP bereits eine IP-Adresse enthalten ist, soll der PPP peer diese dem mobilen Endgerät als seine co-located care-of address zuweisen.
Falls die IP-Adresse Konfigurations-Option im IPCP Configure-Request ausgelassen wird, muss dem mobilen Endgerät zu einem späteren Zeitpunkt eine IP-Adresse zugewiesen werden.

2.3 Mobile IP

Clients, die Mobile IP verwenden, brauchen spezielle AAA Dienste. Um die Anpassungen des Basis-Modells zu verstehen, soll der mobile node (MN) dem client in Abbildung 1 entsprechen und der attendant dem foreign agent (FA). Der home agent (HA), der für Mobile IP sehr wichtig ist, spielt bei der anfänglichen Registrierung eine im Vergleich zum AAAH untergeordnete Rolle. Die angesprochenen Änderungen sind in Abbildung 2 wiedergegeben. Nach der anfänglichen Registrierung wird der MN berechtigt, Mobile IP in der foreign domain ohne weitere Einflussnahme der AAA-Server zu benutzen.

Der Mehraufwand bei der Registrierung (AAA-Registrierung und Mobile IP-Anmeldung) wird zu einer verlängerten Anmeldezeit führen. Um lange Wartezeiten zu vermeiden, sollte

[1] IPCP ist für Konfiguration, Aufbau und Abbau der IP-Protokoll Module auf beiden Seiten der Punkt-zu-Punkt-Verbindung verantwortlich [Meri92]

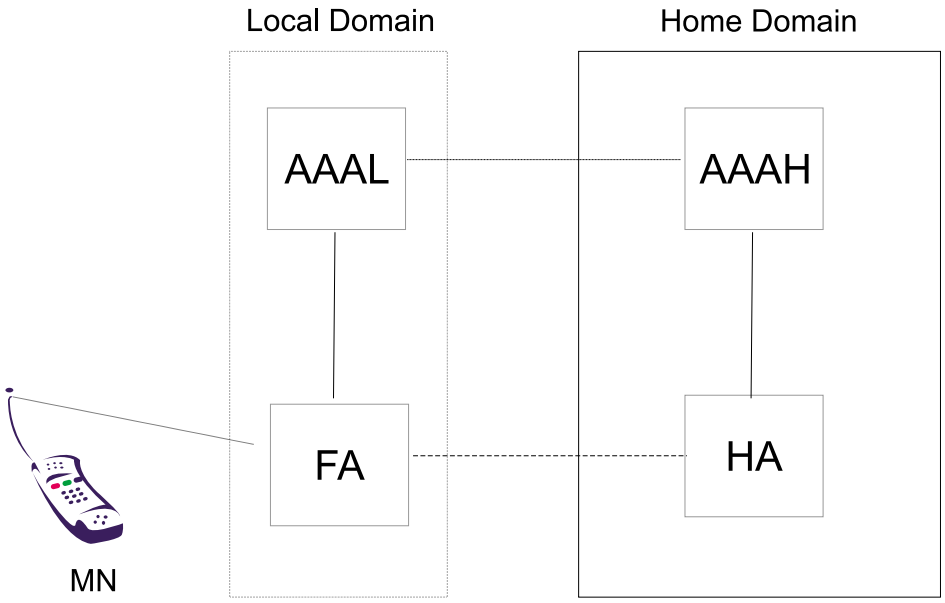

Abbildung 2: AAA Server mit Mobile IP Agenten

der Kommunikationsaufwand zwischen den AAA-Servern soweit wie möglich verringert werden. Der Großteil der Verzögerung kommt durch die Durchquerung des Internets, das AAAL und AAAH trennt, zustande. Daher sollte die anfängliche Registrierung von AAA und Mobile IP zusammen von den beteiligten AAA-Servern und dem FA und HA in nur einem Nachrichtenaustausch ausgeführt werden.
Ein Spezialfall ergibt sich, wenn die AAA home domain und die HA home domain verschieden sind. Dieses Szenario ergibt sich z.B., wenn die home address des MN von einem ISP bereitgestellt wird, während die Authorisierung und die Berechnung von einer Kreditkartenfirma übernommen wird. Dann wird zuerst die Authentifizierungsinformation vom FA zum AAAL gesandt und anschließend die Mobile IP-Registrierung durchgeführt. Diese beiden Schritte können nur nacheinander ausgeführt werden, was die Anmeldezeit verlängert.

Ziel ist also die Durchführung aller Anmeldeprozeduren mit einem Nachrichtenaustausch über das Internet. Dafür müssen die AAA-Server die Mobile IP Agenten identifizieren und ihnen die notwendigen Informationen übertragen, während sie im eigentlichen Mobile IP-Anmeldeprozeß unbeteiligt bleiben. Daraus lassen sich nach [Metz99] und [Perk99] folgende Aufgaben für AAAH und AAAL ableiten : Ermöglichung der Mobile IP Registrierung und der Authorisierung des MN zur Nutzung der angeforderten Dienste, Anstoßen der Rechungsinformationen und Verwendung geeigneter AAA-Protokolle.

Zur Gewährleistung der Sicherheit beim Austausch der Kontrollinformationen sind Verschlüsselungsalgorithmen zu verwenden. Um auch nachfolgende Registrierungen zu ermöglichen, müssen die AAA-Server Schlüsselinformationen innerhalb der ersten Mobile IP-Registrierung austauschen können. Die Verteilung der Schlüsselinformationen liefert folgende Sicherheitsfunktionen:

- Erstellung einer Sicherheitsbeziehung zwischen MN und HA, damit Mobile IP überhaupt arbeiten kann.
- Erstellung einer Sicherheitsbeziehung zwischen MN und FA, damit der FA den Registrierungsprozeß mit dem MN fortsetzen kann.
- Erstellung einer Sicherheitsbeziehung zwischen HA und FA, damit für den FA sichergestellt ist, dass er die weitere Registrierung mit dem selben HA durchführt.

- Teilnahme am Austausch der Sicherheitsverbindung (und des Sicherheitsparameterindexes) zwischen den Mobile IP Einheiten.
- Validierung von Zertifikaten des MN

Dabei sollte die Lebenszeit jeder Sicherheitsbeziehung gewährleisten, dass nicht zu häufig neue Schlüssel ausgetauscht werden müssen, was zu längeren Verzögerungen führen würde.

Bei der Verwendung der dynamischen IP-Zuweisung kann das beschriebene Modell nach [GHJP00] vereinfacht werden. Die Ausführung dessen geht allerdings über den Rahmen dieser Arbeit hinaus. Der interessierte Leser findet dort auch eine Behandlung der Problematik beim Einsatz von Firewalls und AAA und die Beschreibung eines Szenarios, bei dem sich das Problem bei lokaler Bezahlung vereinfachen lässt. Im Folgenden wird hier auf ein Problem der beschriebenen Architektur aus diesem Dokument eingegangen, die sie im praktischen Einsatz scheitern lässt.

2.4 Das Broker Modell

Das beschriebene Modell fordert eine Sicherheitsbeziehung zwischen AAAH und AAAL. Das führt zu exponentiellem Wachstum der erforderlichen Sicherheitsbeziehungen bei linearem Wachstum der AAA-Server. So eine Architektur kann in der Praxis nicht skalieren, wie auch [AbZo98] festgestellt haben. Insbesondere wenn man die Anzahl der ISPs heute bedenkt, ist es nicht vorstellbar, dass alle untereinander Roaming-Abkommen abschließen. Als Lösung bieten sich Broker Modelle an. Diese erlauben den lokalen AAA-Servern im Extremfall ihre Sicherheitsbeziehungen auf eine einzige zum AAA-Broker (AAAB) zu reduzieren. Denkbar sind auch mehrstufige Broker-Architekturen, damit die Skalierbarkeit in großen Netzwerken gewährleistet bleibt.
Kritisch beim Einsatz von Brokern ist die Garantie von Sicherheit bei der Übertragung. Sie kann entweder durch hop-by-hop- oder Ende-zu-Ende-Sicherheitsbeziehungen realisiert werden. In jedem Fall steigt die Zahl der benötigten Bestätigungen bevor mit dem Datenaustausch begonnen werden kann, was zu höherer Latenz führt. Dies ist bei zeitkritischen Anwendungen unbedingt zu beachten.

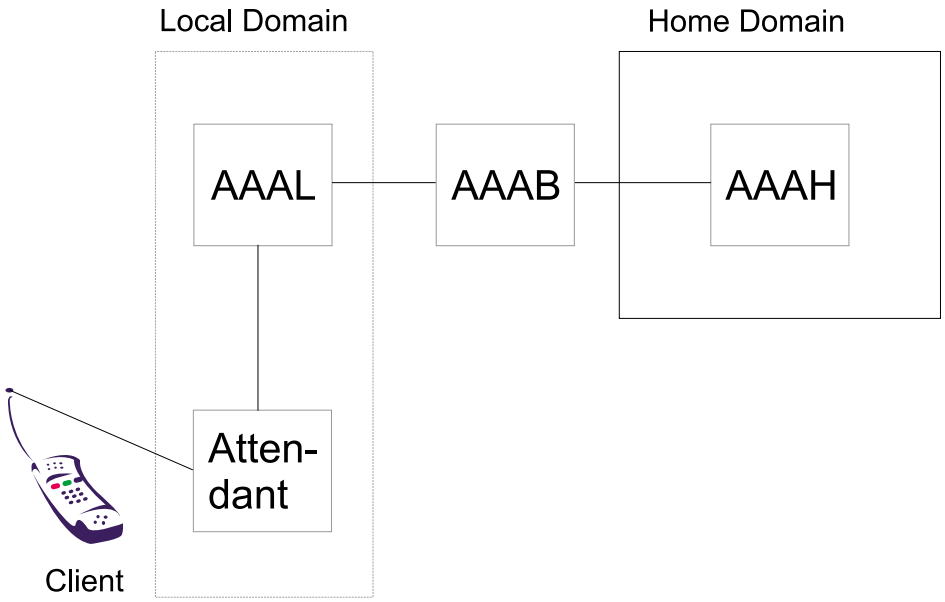

Abbildung 3: AAA Server mit Broker

Abbildung 3 zeigt die schematische Darstellung der Broker-Architektur zwischen zwei AAA-Servern. Der AAAB ermöglicht ein Zusammenarbeiten zwischen den beiden, ohne dass eine direkte Sicherheitsbeziehungen zwischen den beiden bestehen muss. Damit kann die geforderte Skalierbarkeit zwischen unabhängigen Domänen erreicht werden.
Im Spezialfall der Authorisierung für mobile Nutzer stellen [AbVo99] folgende Anforderungen an den Broker:

- AAA-Broker müssen eine Vertrauensbasis zwischen fremden Domänen schaffen.
- Für Rechnungsinformationsdaten, die einen substantiellen Paketverlust erfahren können, müssen Übertragungswiederholungen an Zwischenstationen vorgesehen werden, um Verzögerungen zu vermeiden.
- Obwohl für den Transport zwischen AAAL und AAAH Broker eingeschaltet werden, muss die Ende-zu-Ende-Sicherheit erhalten bleiben.
- Auch der Transport verschlüsselter Daten, wie Schlüssel, zwischen den AAA-Servern muss garantiert werden.

3 Protokolle

Die vorgestellte Architektur und die aufgestellten Anforderungen bestimmen den Rahmen für die einzusetzenden Protokolle. Im Folgenden werden die beiden meistbenutzen AAA-Protokolle, RADIUS und Diameter, vorgstellt und anschließend wird ein Vergleich beider Spezifikationen mit den aufgestellten Anforderungen erfolgen.
Diese beiden Protokolle sind nicht für die Nutzung von AAA-Diensten entwickelt worden, sondern vielmehr handelt es sich um vorhandene Protokollarchitekturen, die aber für den Einsatz auf der beschriebenen Architektur geeignet sind. Damit der Vergleich mit den vorgestellten Anforderungen möglich ist und die notwendigen Anpassungen direkt bei der Beschreibung der Protokolle dargestellt werden können, werden sie erst jetzt nach der Einführung der Architektur vorgestellt.

3.1 RADIUS

RADIUS steht für Remote Access Dial-In User Service und ist das bekannteste und noch am häufigsten eingesetzte Protokoll. Ursprünglich wurde es für kleine Netzwerke mit wenigen Endnutzern entwickelt. Die Grundidee war ein Modem-Pool mit Wahlverbindungen nach außen, die es zu überwachen galt. Das konnte am effizientesten mit einer einzigen Datenbank mit Nutzerdaten zur Authentifizierung gewährleistet werden. Dadurch war keine aufwendige und teure Infrastruktur in den Modems selbst nötig. Aus dieser Idee heraus leiten sich die zentralen Eigenschaften von RADIUS nach [RRSW97] ab.

3.1.1 Zentrale Eigenschaften

Client/Server Model: Der NAS ist der client bei RADIUS. Er ist für das Weiterleiten der Nutzerinformationen zum RADIUS-Server zuständig und muss auf die zurückgelieferten Antworten reagieren. Der RADIUS-Server erhält den connection request des client und ist für dessen Authentifizierung und bei Erfolg für die notwendige Konfiguration verantwortlich. Der RADIUS-Server kann auch als proxy client zu anderen RADIUS- oder Authentifizierungs-Servern dienen.

Netzwerk Sicherheit Die zwischen RADIUS client und Server ausgetauschten Nachrichten werden mittels eines gemeinsamen Schlüssels authentifiziert, der niemals über das Netzwerk ausgetauscht wird. Zusätzlich werden Nutzerpasswörter für die Übertragung verschlüsselt.

Flexibler Authentifizierungsmechanismus: RADIUS Server können eine Vielzahl von Methoden zur Authentifizierung eines Nutzers unterstützen. Zu nennen sind unter anderem PPP PAP oder CHAP, UNIX login.

Erweiterbares Protokoll: Alle Transaktionen sind in Attributen-Tuppeln der Länge 3 eingeschlossen. Neue Attribute können, ohne vorhandene Implementierungen des Protokolls abzuändern, hinzugefügt werden.

3.1.2 Protokollablauf

Ein client wählt sich in ein NAS, der RADIUS benutzt, ein und übermittelt ihm eine Authentifizierungsinformation. Der NAS mit dem RADIUS-client erzeugt daraus ein „Access Request“, das den Nutzername, die ID des client und den Port, an dem sich der Nutzer eingewählt hat, enthält. Falls ein Passwort verwendet wird, wird es verschlüsselt.
Der Access-Request wird über das Netzwerk zum RADIUS-Server übermittelt. Sollte nach einer gewissen Zeitspanne keine Antwort eingetroffen sein, kommt es zur Sendungswiederholung zum selben oder einem anderen RADIUS-Server. Sobald die Nachricht beim Server eingeht, gleicht er den Nutzernamen in einer Datenbank ab. Dort sind die notwendigen Anforderungen, wie z.B. das Passwort, für einen Zugang hinterlegt. Unter Umständen muss der RADIUS-Server dafür weitere Server kontaktieren. Wird ein passender Eintrag gefunden, sendet der Server ein „Access Challenge“ an den NAS zurück. Der NAS wiederholt daraufhin seine Übertragung des Access-Request mit einer neuen request ID. Die Nutzer-Passwort-Attribute ersetzt er durch die Antwort.
Sind alle Anforderungen erfüllt, stellt der Server ein „Access-Accept“ mit allen notwendigen Informationen für den client. Dazu gehören unter anderem die IP Adresse oder die Service-Art.

Für die Übertragung innerhalb des Netzwerkes verwendet RADIUS UDP und nicht das gesicherte TCP als Transprotprotokoll. Dies hat nach [RRSW97] folgende technische Gründe:

1. Im Falle des Ausfalls eines Authentifizierungsservers muss ein weiterer Server befragt werden. Dies kann nur geschehen, wenn eine Kopie des requests über dem transport layer behalten wird.

2. Die timing-Anforderungen von RADIUS unterscheiden sich deutlich von den von TCP bereitgestellten. Denn einerseits braucht RADIUS keine aggressive Erkennung verlorener Pakete, da die Authentifizierung des mobilen Endnutzers durchaus mehrere Sekunden dauern kann. In dieser Zeit würde TCP unnötige Übertragungswiederholungen nach Ablauf der round trip Zeit veranlassen. Andererseits darf die Wartezeit gewisse Grenzen nicht überschreiten. Daher ist die zuverlässige Zustellung eines verlorengegangen Pakets einige Minuten später nicht sinnvoll. Die Verwendung eines anderen Servers führt in diesem Fall schneller zur Authentifizierung.

3. Die RADIUS-Architektur ist dynamisch. Das Hinzukommen oder Verlassen des Netzwerks einzelner Server stellt kein Problem dar. Mit UDP ist keine spezielle Behandlung dieser Ereignisse nötig.

4. Die Server-Architektur muss auf Grund der zahlreichen gleichzeitigen Anfragen von clients multi-threaded sein. Sonst würde eine Authentifizierungsanfrage, die bis zu 30 sec. dauern kann, den gesamten Server blockieren und die Wartezeit für die Kunden

würde inakzeptabel werden. Gleichzeitige Bearbeitung der Anfragen ist also notwendig. Dies konnte mit dem Einsatz von UDP leicht realisiert werden, weil mittels einfacher UDP-Pakete der client über den NAS erreicht werden kann.

Allerdings hat die Wahl von UDP auch einen gravierenden Nachteil. Der retransmission-timer von TCP muss in abgeschwächter Form künstlich nachgebildet werden. Dies wiegt aber die Vorteile von UDP nicht auf.

Zusätzlich zur Authentifizierung und Authorisierung wurde RADIUS erweitert, um Rechnungsinformationsdaten zum Kommunikationsverhalten des Nutzers zu sammeln. Dafür gibt es eigene RADIUS-Accounting-Server. Dieses Verfahren wird in RFC2139 beschrieben.

3.2 Diameter

Mit der Zunahme der Komplexität der geforderten Dienste und der zunehmenden Dichte von NAS erweist sich RADIUS als ungeeignet zur Nutzung in größeren Netzwerken. Insbesondere hat sich gezeigt, dass roaming-Vereinbarungen mit allen ISPs untereinander nicht skalieren und man geht deswegen auf den Einsatz von Brokern über, wie in Abschnitt 3 dargestellt. Dies machte die Entwicklung eines neuen Protokolls notwendig.
Das Diameter Protokoll wurde aber laut [CAAG+01] nicht von Grund auf neu entwickelt, sondern ein Großteil des RADIUS-Protokolls wurde erhalten und dessen Fehler beseitigt. Das Hauptziel von Diameter ist die Schaffung eines erweiterbaren Basisprotokolls, das neue Zugangsmöglichkeiten für AAA Dienste schafft. Bisher wird allerdings nur PPP, das ROAMOPS Modell und Mobile IP berücksichtigt.

3.2.1 Zentrale Eigenschaften

Diameter nutzt wie RADIUS Attribut/Wert Paare (AVP) zur Vermittlung von Daten und UDP als Transportprotokoll. Darüber hinaus ist es durch das Hinzufügen neuer Befehle und AVPs erweiterbar. AVPs werden in Diameter für folgende Aufgaben eingesetzt:

- Transport von Nutzer-Authentifizierungsdaten zum Diameter-Server
- Transport dienstspezifischer Authorisierungsinformation zwischen client und Server zur Entscheidung, ob dem Nutzer die angeforderte Ressource bereitgestellt werden soll.
- Austausch von Ressourcen-Nutzungsinformationen für die Rechnungserstellung oder Kapazitätsplanung usw.
- Routing von Diameter Nachrichten in einer Serverhierarchie.

Diameter stellt ein Basis-Protokoll dar, das die Minimalanforderungen eines AAA Transportprotokolls erfüllt. Es ist nicht dafür gedacht, alleine eingesetzt zu werden, sondern sollte immer mit einer anwendungsspezifischen Erweiterung benutzt werden.

Diameter ist ein peer-to-peer-Protokoll. Der Diameter client initiiert normalerweise eine Authentifizierungs- oder Authorisierungs-Anfrage eines Nutzers. Der Diameter Server nimmt diese Anfrage entgegen und beantwortet sie entweder oder leitet sie an einen Proxy-Server weiter. Der Server kann aber auch von sich aus aktiv werden, z.B. zur Aufforderung eines clients, die Rechnungsinformation zu aktualisieren.

3.2.2 Protokollablauf

Sowohl client als auch Server müssen das Basis-Diameter Protokoll inklusive accounting-Informationsverarbeitung beherrschen. Der Server muss darüber hinaus auf jeden Fall die NASREQ und die Mobile IP-Erweiterung unterstützen.

Der genaue Ablauf des Protokolls wird in der jeweiligen Spezifikation festgelegt. Das Basis Protokoll definiert hauptsächlich die Verteilung der Nachrichten über das Netzwerk und zwischen den peers, sowie das Verlassen des Netzwerks von peers. Hier wird jetzt die Mobile IP-Erweiterung [CaPe01] vorgestellt.
Das Mobile IP-Protokoll basiert auf einer bestehenden Sicherheitsbeziehung zwischen FA und HA. Dies führt dazu, dass Diameter als Schlüsselverteilzentrum fungiert, das dynamisch Schlüssel erzeugt und über das Netzwerk verteilt. Der Nachrichtenfluss ist im Falle einer Einwahl bei einem fremden ISP dabei wie folgt:
Der mobile node fordert den gewünschten Dienst beim FA mittels der AA-Mobile-Node-Request-Nachricht (AMR), die die AVPs enthält, an. Die Home Address, der HA, der Mobile Node NAI und weitere benötigte Informationen werden extrahiert und in Diameter AVPs eingeschlossen. Diese Nachricht wird an den lokalen Diameter-Server, der AAAF bezeichnet wird, weitergeleitet. Der AAAF leitet die Nachricht an den AAA-Home-Server (AAAH) weiter. Kann der AAAH den MN erfolgreich authentifizieren, lokalisiert er einen HA. Der genaue Ablauf dieses Schrittes würde zu weit führen und wird hier deshalb ausgelassen. An den HA sendet der AAAH dann den Home-Agent-MIP-Request (HAR). Dieser verarbeitet nach Erhalt der HAR zuerst die Diameter-Nachricht. Daraufhin erstellt er die HAA mit den benötigten Daten, wie Session-ID usw. und sendet sie an den AAAH. Dieser erstellt die AA-Mobile-Node-Answer (AMA), die unter anderem Informationen für das Tunneln von Nachrichten zwischen FA und HA enthält, und sendet sie an den AAAF, der die Zustellung an den richtigen FA sicherstellen muss. Damit ist die Verbindung aufgebaut.

Die Mobile-IP-Erweiterung definiert darüber hinaus zahlreiche Sonderfälle, wie die Behandlung von Handoffs. Hier muss sichergestellt werden, dass die Home Domain diese Handoffs als eine Session ansieht, um korrekte Rechnungsinformationen zu ermöglichen. Zusätzlich ergibt sich die Schwierigkeit, dass die FA im allgemeinen nicht untereinander kommunizieren und man daher nicht davon ausgehen kann, dass dem neuen FA der vorangegangen Authentifizierungsprozeß bekannt ist.
Die erste Registrierungsanfrage beim neuen FA wird daher zu einer AMR-Nachricht zu seinem AAAF führen. Diese AMR wird eine neuen Session Identifier enthalten und der gewählte AAAF kann sich vom alten AAAF unterscheiden. Ebenso ist es vorstellbar, dass auch ein anderer AAAH die AMR-Nachricht erhält. Dieser wiederum hat keinen Zugang zum alten Session Identifier. Daher ist es notwendig, dass der HA, der die Fortsetzung der Session erkennt, seinen internen Status ändert und den neuen Session Identifier übernimmt. Zusätzlich muss er den alten AAAH mittels eines STR (Session Termination Request) über den Abbau der alten Verbindung informieren.
Genauere Ausführungen können in der Literatur [CaPe01] nachgelesen werden.

3.3 Vergleich RADIUS und Diameter

Nach der Darstellung der Grundkonzeption der beiden AAA-Protokolle sollen die Unterschiede zwischen den beiden in Bezug auf Erweiterbarkeit, Verlässlichkeit, Flusskontrolle und Skalierbarkeit zusammengefasst werden. Grundlage stellt der Artikel von [EkSP00] dar.

Neben der unterschiedlichen Architektur unterscheidet zuerst die Erweiterbarkeit die beiden Protokolle: Während RADIUS nur einen begrenzten Befehl- und Attributplatz von maximal

256 Attributen bietet und deswegen als nicht sehr erweiterbar anzusehen ist, wird dieses Problem bei Diameter mittels eines erweiterbaren Basis Protokolls und 32 Bit Attributen gelöst. Jeder denkbare neue Dienst kann in Diameter durch die Erweiterung des Basis Protokolls implementiert werden, ohne die Basis ändern zu müssen.

Zuverlässigkeit beschäftigt sich mit Fragen zur Zustellung von Nachrichten zwischen den Netzwerkelementen und Flusskontrolle.
RADIUS setzt, wie oben erklärt, UDP zum Datentransport zwischen client und Server ein. Zusätzlich ist eine time-out- und Übertragungswiederholungsstrategie implementiert. Jedoch wurde kein Standardschema definiert, was zu verschiedenen Implementierungen und daraus resultierenden Problemen geführt hat.
Diameter arbeitet zwar ebenfalls mit UDP, hat aber im Basis Protokoll explizit einige Erweiterungen zur Garantie der Zuverlässigkeit über den verbindungslosen Dienst. Dazu gehört eine intelligente time-out- und Übertragungswiederholungsstrategie, die TCP nicht leisten kann.
RADIUS implementiert keine Flusskontrolle über UDP. Diameter hingegen benutzt einen sliding window-Mechanismus mit dynamischer Veränderung der Fenstergröße.

Skalierbarkeit ist im Hinblick auf die zu erwartenden Nutzerzahlen der AAA-Dienste sehr wichtig. Zu beachten sind Szenarien, in denen viele Nutzer gleichzeitig AAA-Funktionalitäten beim gleichen Server nutzen wollen.
Bezüglich implementationsspezifischen Gesichtspunkten ergibt sich folgender Unterschied: RADIUS Nachrichten werden byteweise angeordnet, während sie bei Diameter 32-bit angeordnet sind. Dies ermöglicht eine höhere Zahl von Transaktion pro Sekunde pro Server bei Diameter.
Auf der Transportschicht hängt die Skalierbarkeit nicht von der Anzahl der Nutzer, sondern nur von der benötigten Menge an client/Server-Beziehungen ab. Da RADIUS UDP verwendet, muss auf der client Seite eine Beziehung nur während der request/reply Interaktion bestehen. Wenn sich Diameter auf die erweiterten UDP Prozeduren stützt, sollte eine Beziehung aufgrund des sliding window Mechanismus aufrechterhalten werden. Diese Forderung führt dazu, dass RADIUS Vorteile bei der Skalierbarkeit hat.

Ein großes Thema stellen Sicherheitsmechanismen gegenüber Attacken auf die AAA Infrastruktur dar. Überlegungen zu diesem großem Themenkreis gehen über den Rahmen dieses Dokuments hinaus. Einen Überblick über die verwendeten Mechanismen gibt [EkSP00].

Abschließend läßt sich festhalten, dass Diameter die meisten Protokoll-Möglichkeiten bietet und wohl die Plattform für die weitere AAA Entwicklung sein wird. Allerdings müssen auch insbesondere noch Unterstützung für QoS-Umsetzungen eingefügt werden.

4 AAA bei Einsatz von IPv6

In IPv4 erfolgte bisher eine strikte Trennung zwischen den AAA-Aufgaben und den Funktionen, die von Routern und DHCP-Servern wahrgenommen werden. Mit der Einführung von IPv6 hingegen kann angenommen werden, dass Router und DHCPv6-Server in Zusammenarbeit mit den AAA-Servern die Authentifizierung eines client anhand von credentials übernehmen werden. Auf Protokollebene kommt es zu einer stärkeren Zusammenarbeit.
Dafür ist eine Abänderung des in Abschnitt 2.1 vorgestellten Rahmenwerks nötig. Abbildung 4 aus [AsPE00] stellt sie dar.

Aufbauend auf dieser neuen Modellvorstellung ergibt sich unter Verwendung eines IPv6-Protokolls folgender Ablauf: Der Paketfilter im kontrollierten Teil des Routers nimmt den

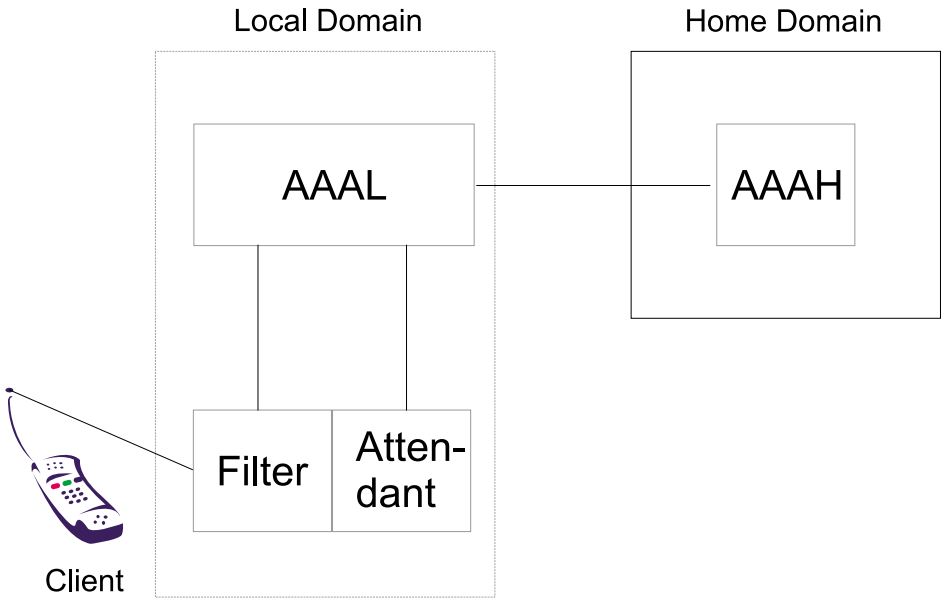

Abbildung 4: Das IPv6 Modell

Verbindungswunsch eines client entgegen. Nur AAA-Authorisierungspakete und Pakete bereits authorisierter Nutzer können den Filter passieren. Die AAA-Pakete werden dem attendant im unkontrollierten Bereich zugänglich gemacht. Dieser extrahiert die notwendigen Informationen und leitet sie seinerseits zum AAAL weiter. Ab dann läuft der bereits im Diameter-Protokoll in Abschnitt 3.2.2 vorgestellt Prozess zwischen AAAL, AAAH und attendant ab. Bei erfolgreicher Authentifizierung schließlich aktualisiert der attendant die Konfiguration des Paketfileters, so dass ab sofort der Verkehr des authentifizierten clients diesen passieren kann und nicht verworfen wird.
In diesem Modell ist der attendant im Router lokalisiert. Diese Vorstellung wird durch die Annahme möglich, dass im Gegensatz zu IPv4 die Router auch AAA-Aufgaben wahrnehmen können.

Zur Identifikation des client kann entweder wie in IPv4 der NAI (Network Access Identifier) oder die IPv6-Adresse herangezogen werden. Beide ermöglichen dem AAAL den passenden AAAH zu identifizieren.

Die vom client bereitgestellten AAA credentials zur Authentifizierung sollten folgende Informationen enthalten:

- Client identifier
- Lokale AAA challenge[2], falls sie vom attendant bereitgestellt wurde
- Abhängig vom Mechanismus der zum Schutz gegen Wiederholungsattacken eingesetzt wird, entweder ein Zeitstempel oder ein Challenge-Paar

Zur Anpassung an Mobile IPv6 ergeben sich zwei Möglichkeiten. Im ersten Fall führt der client zuerst die AAA Prozedur aus sobald er eine care-of-address hat und kontaktiert dann, falls nötig seinen HA und die früheren Kommunikationspartner.
Eine zweite Möglickkeit besteht in der gleichzeitigen Ausführung von AAA Authentifizierung und Bindungsnachricht an den HA, falls AAAH und HA zur selben Domäne gehören.

[2]Enthält die Aufforderung vom attendant an den Client, sich zu identifizieren

Dafür erzeugt der client ein IPv6-Paket mit der Nachricht an den HA, das aber als Nutzlast in ein AAA-Request eingebettet wird. Der AAAH kann dann die Nachricht trennen und das IPv6-Paket zum HA weiterleiten, während er die Authentifizierung durchführt. Die Bindungsbestätigung des HA wird dann an den AAAH geschickt und dieser sendet sie zusammen mit seiner Antwort zurück an den client. Auf diese Weise kann die Verbindungsaufbauphase erheblich verkürzt werden.

5 Schlußbemerkung

Die Revolution des mobilen Internets kann nur gelingen, falls neue interessante Dienste den Kunden angeboten werden können. Nach der Entwicklung derartiger Angebote wird die Nutzung von längeren Sitzungen und intelligenten Quality of Service-Anforderungen gekennzeichnet sein ([Tubi01]). Die daraus auftretenden Anforderungen an die Infrastruktur der Netze sind die Kernkompetenz von AAA. Die vorgestelle Konzeption sollte insbesondere den steigenden Nutzerzahlen und der gleichzeitigen Nutzung angebotener Dienste standhalten und seine Stärken ausspielen können. Denn eines ist deutlich festzuhalten: Ohne eine zusätzliche AAA-Infrastruktur sind die Anforderungen nicht erfüllbar.
Auch die Netzbetreiber können ein gesteigertes Interesse an der Technologie haben, denn sie ermöglicht es ihnen, die Kontrolle über die genutzten Dienste über das Abrechnungsverfahren zu behalten und schützt sie auf diesem Wege davor, zum reinen Verbindungslieferanten zu werden. Sie können sich mit Partnern, die für neue Inhalte sorgen, verbünden ohne die Kontrolle über ihre Kernkompetenzen aufgeben zu müssen. Die AAA-Infrastruktur stellt das Verteilzentrum dar, das aber die Rechnungsstellung allein den Netzwerkbetreibern überlässt.

Literatur

[AbVo99] B. Aboba und J. Vollbrecht. Proxy Chaining and Policy Implementations in Roaming. *RFC 2607*, June 1999.

[AbZo98] B. Aboba und G. Zorn. Criteria for Evaluating Roaming Protocols. Technischer Bericht, RFC 2477, December 1998.

[AsPE00] N. Asokan, C. Perkins und T. Eklund. AAA for IPv6 Network Access. Internet draft, IETF IPng Working Group, March 2000.

[CAAG+01] P. Calhoun, H. Akhtar, J. Arkko, E. Gutman, A. Rubens und G. Zorn. Diameter Base Protocol. Internet draft, IETF AAA Working Group, May 2001.

[CaPe00] P. Calhoun und C. Perkins. Mobile IP Network Access Identifier Extension for IPv4. Standards track, Network Working Group, March 2000.

[CaPe01] P. Calhoun und C. Perkins. Diameter Mobile IPv4 Extension. Internet draft, IETF AAA Working Group, May 2001.

[EkSP00] R. Ekstein, B. Sales und O. Paridaens. AAA Protocols: Comparison between RADIUS, DIAMETER and COPS. Internet draft, NASREQ Working Group, January 2000.

[GHJP00] S. Glass, T. Hiller, S. Jacobs und C. Perkins. Mobile IP Authentication, Authorization, and Accounting Requirements. Internet draft, Network Working Group, October 2000.

[McHi00] P. McCann und T. Hiller. An Internet Infrastructure for Cellular CDMA Networks Using Mobile IP. *IEEE Personal Communications*, August 2000, S. 26–32.

[Meri92] G. Merit. RFC-1332 The PPP Internet Protocol Control Protocol (IPCP). Technischer Bericht, IETF Network Working Group, May 1992.

[Metz99] C. Metz. AAA Protocols: Authentication, Authorization, and Accounting for the Internet. *IEEE Internet Computing*, November, December 1999, S. 75–79.

[Perk99] C. Perkins. Mobile IP and Security Issue: An Overview. Technischer Bericht, Proceedings of 1st IEEE Workshop on Internet Technologies and Services, 1999.

[Perk00] C. Perkins. Mobile IP Joins Forces with AAA. *IEEE Personal Communications*, August 2000, S. 59–61.

[RRSW97] C. Rigney, A. Rubens, S. Simpson und S. Willens. RFC2138 - Remote Authentication Dial In User Service (Radius). Technischer Bericht, IETF Network Working Group, April 1997.

[Tubi01] M Tubinis. Converging IN and AAA Systems to Unlock the Potential of Mobile Data. Intelligent Network Workshop 2001, May 2001.

Abbildungsverzeichnis

1 Das Basis Modell 33
2 AAA Server mit Mobile IP Agenten 35
3 AAA Server mit Broker 36
4 Das IPv6 Modell 42

Location Privacy in der Mobilkommunikation

Markus Klein

Kurzfassung

In mobilen Sprach- und Datennetzen sind die Daten eines Benutzers besonders gefährdet. Zum einen erfolgt die Kommunikation meist über funkgebundene Netze bei denen das Abhören von unverschlüsselten Daten besonders leicht ist. Die Geheimhaltung der übertragenen Daten kann durch geeignete Verschlüsselung gewährleistet werden. Dafür sind die entsprechenden Kommunikationsprogramme zuständig. Eine weitere Gefahr geht nicht von den Daten an sich aus, sondern von den Informationen die normalerweise benötigt werden, um die Daten korrekt zu übertragen, also z.B. der Benutzername, Heimzone,.... Durch Mithören der Kommunikation kann so ein Bewegungsprofil des Benutzers erstellt werden. Dieser Text beschreibt dieses Problem detailierter, untersucht, was davon in bisherigen mobilen Netzen umgesetzt ist und geht dann auf einige Ansätze für „Location Privacy“ in zukünftigen mobilen Netzen ein.

1 Einleitung

Ein Benutzer eines mobilen Gerätes, z.B. eines Handys oder eines PDAs, ist normalerweise bei einem Dienstanbieter angemeldet, der ihm mobile Dienste zur Verfügung stellt und berechnet. Dieser Anbieter hat normalerweise ein eigenes, lokales Netz. Dieses Netz wird von nun an mit Heimzone bezeichnet. Ein Beispiel dafür sind die bestehenden Mobilfunknetze. Diese sind lokal begrenzt, meist innerhalb eines Landes. Der Benutzer kann jedoch auch außerhalb dieser Heimzone Kommunikationsdienste nutzen, und zwar über einen dortigen, lokalen Anbieter solcher Dienste. Dieser bietet die Dienste dem Benutzer an und rechnet evtl. über den Anbieter in der Heimzone ab. Dieser Anbieter wird Fremdzone genannt.

2 Klassifikation von Location Privacy

Was versteht man unter „Location Privacy“? Diese Frage ist nicht so leicht zu beantworten und wird unterschiedlich bewertet. Manche verstehen unter Location Privacy die Geheimhaltung des Ortes eines mobilen Gerätes, der Identität, des Kommunikationsverhaltens, ... Andere bewerten „Location Privacy“ anders. Um nun die verschiedenen Bedeutungen des Begriffs “Location Privacy“ einordnen zu können, kann man eine in [SaMA] vorgestellte Einteilung mittels einer Matrix benutzen. Die Spalten bezeichnen dabei die an dem Vorgang teilnehmenden Parteien und die Zeilen stehen für das, was diese Parteien wissen dürfen. An teilnehmenden Parteien gibt es die Heimzone (Home-Zone H), die Fremdzone (Remote-Zone R), andere legitimisierte Netzkomponenten dazwischen (L) und die möglichen Mithörer (X). An zu schützenden Geheimnissen gibt es die Benutzerkennung (user identity u), die Identität der Heimzone (h) und die Identität der Fremdzone (r). Wenn eine Partei keine Information erhalten darf steht als Eintrag eine 0, wenn die Partei nur einige der Informationen erhalten darf, steht als Eintrag ein s (für some), für alle Informationen eine 1 im entsprechenden

Kästchen. Es wurden die englischen Bezeichnungen gewählt um konsistent mit [SaMA] zu sein. Mit Hilfe dieser Matrix kann man versuchen den Begriff „Location Privacy“ genauer zu beschreiben und in Klassen einzuteilen, die aufeinander aufbauen: $C_1 \subset C_2 \subset C_3 \ldots$. Neben diesen Klassen sind natürlich auch weitere Klassen (bzw. Kombinationen) denkbar.

2.1 Klasse C_1: Benutzeridentität vor Mithörern schützen

Die Klasse C_1 ist die grundlegende Klasse. Die Identität des Benutzers wird hier vor fremden Mithörern geschützt, nicht jedoch vor der Fremdzone und anderen legalisierten Netzkomponenten dazwischen. Auch der Aufenthaltsort des Benutzers und die Heimzone der er angehört sind überhaupt nicht geschützt.

	H	R	L	X
u	1	1	1	0
h	1	1	1	1
r	1	1	1	1

Tabelle 1: Klasse C_1

2.2 Klasse C_2: Benutzeridentität vor anderen Netzkomponenten schützen

Die Klasse C_2 bietet etwas mehr Schutz, sie verbirgt die Benutzeridentität vor allen Beteiligten außer der Heimzone.

	H	R	L	X
u	1	0	0	0
h	1	1	1	1
r	1	1	1	1

Tabelle 2: Klasse C_2

2.3 Klasse C_3: Identität der Heimzone vor Dritten schützen

Die Klasse C_3 verbirgt zusätzlich die Identität der Heimzone vor Mithörern und dazwischenliegenden legalisierten Netzkomponenten.

	H	R	L	X
u	1	0	0	0
h	1	1	0	0
r	1	1	s	s

Tabelle 3: Klasse C_3

2.4 Klasse C_4: Identität der Heimzone vor der Fremdzone schützen

Die Klasse C_4 schützt die Identität der Heimzone nicht nur vor Dritten sondern auch vor der Fremdzone.

	H	R	L	X
u	1	0	0	0
h	1	0	0	0
r	1	1	s	s

Tabelle 4: Klasse C_4

2.5 Klasse C_5: Benutzerverhalten vor der Heimzone schützen

Die Klasse C_5 verbirgt die Identität der Fremdzone (den Aufenthaltsort) vor der Heimzone. Auch eine Migration von einer Fremdzone in eine Andere bleibt der Heimzone verborgen

	H	R	L	X
u	0	0	0	0
h	1	0	0	0
r	s	1	s	s

Tabelle 5: Klasse C_5

3 Existierende Standards der Mobilkommunikation

3.1 GSM

Der Mobilfunkstandard GSM (Global System for Mobile) ist der meistverbreitete Standard in Europa und der übrigen Welt, von Nord-Amerika abgesehen. In GSM ist ein Mechanismus vorgesehen um die Identität des Benutzers vor Mithörern zu schützen, also eine Klasse C_1 Funktionalität. Der Benutzer bekommt einen Alias (TMSI Temporary Mobile System Identifier). Bei jeder Kommunikation mit einer Basisstation in einem Fremdnetz benutzt das Handy diesen Alias. Wenn der Benutzer von einem Fremdnetz in ein anderes migriert, versucht die neue Fremdzone von der Alten die TMSI und die dazugehörigen Daten zu bekommen. Das Verfahren hat jedoch eine große Lücke. Wenn die alte Fremdzone nicht erreichbar ist gibt es einen „fall-back“ Prozess. Dieser kann auch angewandt werden, wenn die Synchronisation zwischen Benutzer und Fremdnetz aus anderen Gründen nicht mehr gewährleistet ist und ein neuer Alias vereinbart werden muß. Auch beim erstmaligen Einschalten eines Handys in einer Fremdzone wird dieses Verfahren benutzt. Dabei sendet das Handy seine IMSI (International Mobile System Identifier) im Klartext an die Basisstation im Fremdnetz. Dieses kontrolliert dann die IMSI beim Heimnetz und vergibt bei erfolgreicher Authentifikation eine neue TMSI. Durch ständiges Mithören der Funkverbindungen kann dadurch die Identität des Benutzers aufgedeckt werden.

Ein ähnlicher Schwachpunkt in GSM ist die Verwundbarkeit gegen „man in the middle attacks“. Mit entsprechender Ausrüstung können Mithörer vorgeben eine Basisstation eines bestimmten Fremdnetzes zu sein. Die gefälschte Basisstation meldet an das Handy dass eine Übertragung des temporären Alias (TMSI) von der vorherigen Basisstation nicht erfolgreich war und fordert die IMSI des Handys an um einen neuen Alias zu vergeben. Die gefälschte Basisstation kommuniziert nun mit der richtigen und gibt sich als das Handy des abgehörten Benutzers aus und leitet die darauffolgende Kommunikation einfach durch und hört alles mit.

Ein weiterer Kritikpunkt an GSM ist, dass alle Kommunikation auf den Backbones, den Verbindungen zwischen den verschiedenen GSM-Netzen im Klartext abläuft. Professionelle

Organisationen können diese Netze abhören und somit ein Bewegungsprofil eines Benutzers in Fremdzonen erstellen.

GSM ist also eigentlich ein Standard der Klasse C_1, kann diesen Sicherheitsstandard jedoch nicht komplett garantieren, da es einige Lücken aufweist.

3.2 CDPD

Der Mobilfunkstandard CDPD (Cellular Digital Packet Data) ist hauptsächlich in Nordamerika verbreitet. Auch er verspricht dem Benutzer eine gewisse „Privacy“, die er nicht halten kann. Wenn sich in CDPD ein Handy in einer Zone anmelden will, wird zunächst nach dem Diffie-Hellman Schlüsselaustauschprotokoll ein geheimer Schlüssel ausgetauscht. Mit diesem Schlüssel verschlüsselt überträgt nun das Handy seine Benutzerkennung an die Basisstation der Zone. Diese beantragt gegebenenfalls bei der Heimzone die Authentifikation. Die Fremdzone kennt also die Identität des Benutzers.

Das Problem ist nun, dass auch CDPD nicht gegen eine „man in the middle attack“ geschützt ist. Eine Mithörer kann sich mit entsprechender Ausrüstung als eine Basisstation ausgeben und mit dem Handy einen Schlüssel austauschen. Damit erfährt er die Identität des Benutzers.

Auch die Backbones zwischen verschiedenen Zonen in CDPD sind nicht immer verschlüsselt, so dass auch hier eine gewisse Unsicherheit besteht.

CDPD ist sicherer als GSM, kann jedoch ebenso wie GSM nichteinmal die Klasse C_1 vollständig garantieren.

3.3 MobileIP

Das IP-Protokoll wurde mit der Annahme implementiert, dass der Benutzer fest mit einem Netzwerk verbunden ist (über einen Zugangspunkt) und eine IP-Adresse innerhalb des eigenen Netzwerks besitzt. Bei der Spezifizierung hatte man noch nicht an mobile Geräte gedacht. Um nun auch in mobilen Geräten IP unterstützen zu können, wurden Erweiterungen für IP spezifiziert. Für den jetzigen Standard (IPv4) gibt es einen Request for Comment (RFC 2002) mit dem Titel IP Mobility Support, der IP für mobile Geräte ermöglicht.

Jedes mobile Gerät hat eine feste Heimatadresse aus seinem Heimatnetzwerk. Will nun irgendein Host im Internet mit dem mobilen Gerät kommunizieren, werden die Pakete von den dazwischenliegenden Routern automatisch an das Heimatnetzwerk geroutet. Wenn der Benutzer nicht in seiner Heimzone ist, sondern in einer Fremdzone, erhält das mobile Gerät des Benutzers eine temporäre IP-Adresse. Diese wird Care-Of-Adresse genannt und wird der Heimzone mitgeteilt. Die IP-Pakete die für das mobile Gerät bestimmt sind und beim Heimnetzwerk landen, werden nun mit einem IP-Header umhüllt und an die Care-Of-Adresse getunnelt. In der Fremdzone nimmt ein spezieller Rechner die Pakete, die an die Care-Of-Adresse adressiert sind, entgegen, entfernt den äusseren IP-Header wieder und leitet das Paket dann ins eigene Netzwerk, in dem das mobile Gerät das Paket mit der Heimatadresse dann empfängt.

Da die gesamte Kommunikation von einem beliebigen Host zum mobilen Gerät über die Heimzone abläuft sind die Wege immer länger als bei direkter Kommunikation. (Dreiecks-Routing). Deswegen ist eine Erweiterung zu MobileIP in der Diskussion, die „Route Optimization“ als Lösung zu diesem Problem vorschlägt. Die Kommunikation zwischen einem beliebigen Host und dem mobilen Gerät verläuft dabei Anfangs nach dem oberen Verfahren. Das mobile Gerät kann nun aber dem sendenden Host die aktuelle Care-Of-Adresse mitteilen. Der Host speichert diese Care-Of-Adresse in einem sogenannten „Binding Cache“, und tunnelt die folgenden

Pakete die er an das mobile Gerät senden will direkt an die Care-Of-Adresse. Wechselt das mobile Gerät während der Verbindung das Netzwerk, bekommt der Sender eine Nachricht mit der aktualisierten Care-Of-Adresse. Die Kommunikation erfolgt damit nicht mit den Umweg über die Heimzone sondern direkt vom Host zum mobilen Gerät.

Für den kommenden IP-Standard (IPv6) ist auch eine MobileIP-Erweiterung in der Diskussion. Das Verfahren ist ziemlich ähnlich. Allerdings ist „Route Optimization“ direkt in MobileIP integriert und in IPv6 selbst wurden einige Vorkehrungen getroffen, die weitere Optimierungen von MobileIP ermöglichen. Im IP-Header gibt es ein eigenes „Home-Adress-Field“ in dem die Heimadresse steht. Dadurch kann das mobile Gerät mit seiner Care-Of-Adresse als Sendeadresse senden und damit einige Probleme mit Firewalls umgehen und das Routing von Multicast-Paketen vereinfachen.

In diesen bisherigen Drafts von MobileIP wurde Location Privacy fast überhaupt nicht beachtet. MobileIP erfüllt in der vorliegenden Form keine der vorgestellten Klassen bezüglich Location Privacy. Da die Kommunikation im Klartext abläuft (auf Vermittlungsebene) wissen eigentlich alle Beteiligen über alles Bescheid. Der Vermittlungsknoten der Fremdzone kennt sowohl die wahre Identität des Benutzers, als auch seine Heimzone und die Kommunikationspartner. Auch die Kommunikationspartner wissen die Heimadresse des Benutzers. Die Heimzone wiederum weiss den momentanen Ort des Benutzers (die care-of-address) und bekommt manche Verbindungen mit, da die Updates der „Binding Caches“ teilweise mit Hilfe der Heimzone geschehen. Insgesamt gesehen ist MobileIP in dieser Form also für Location Privacy untauglich.

4 Sicherheitsmodelle für zukünftige Standards

Da man die mangelnde Location Privacy in den bisherigen mobilen Kommunikationsnetzen mit den dazugehörigen Standards erkannt hat, ist man in den letzten Jahren auf der Suche nach Lösungen. Dazu wurden eine Unmenge verschiedener Verfahren vorgeschlagen, die alle ihre Vor- und Nachteile haben. Die erfolgversprechendsten Verfahren sind gerade auf dem Weg durch die Standardisierungsgremien. Deswegen können hier nur beispielhaft einige Verfahren vorgestellt werden. Zunächst werden einige konkrete Verfahren am Beispiel MobileIP vorgestellt, danach folgen noch ein paar allgemeiner gehaltene Ansätze.

4.1 Erweiterungen von MobileIP

Um die fehlende Location Privacy im MobileIP-Standard zumindest für die nächste Version von IP hinzuzufügen, ist im Moment ein IETF-Draft in der Diskussion: [Cast01]. In diesem Dokument werden 2 Verfahren für grundlegende Location Privacy in MobileIPv6 beschrieben. Danach wird das Konzept der Mixe vorgestellt, auf dem die auch noch beschriebene Non Disclosure Methode aufbaut.

4.1.1 Privacy Extension für MIPv6

Wie schon beschrieben liegt eines der Hauptprobleme von MIPv6 darin, dass bei der Kommunikation mit „Route Optimization“ nicht nur die aktuelle care-of Adresse, sondern auch die Heimatadresse mit im IP-Header übertragen wird und somit jeder beteiligte Host den Ort und die Identität eines Benutzers herausfinden kann, je nach Standort des Hosts auch noch mit wem der Benutzer kommuniziert. Um dieses Problem zu umgehen, führt Castelluccia in [Cast01] eine sogenannte TMI (Temporary Mobile Identifier) ein, die vom Heimnetz und den

mit dem mobilen Gerät kommunizierenden Hosts statt der Heimadresse benutzt wird. Um die TMI zu generieren schlägt Castelluccia vor eine MD5-Funktion über die Heimadresse konkateniert mit der vorhergehenden TMI zu generieren. Mit dieser Funktion ist sichergestellt, dass die Wahrscheinlichkeit, dass 2 mobile Geräte mit der gleichen TMI mit einem gemeinsamen Server kommunizieren verschwindend klein ist, und nur in diesem Fall gäbe es ein Problem mit der Zuordnung der TMI´s. Außerdem wird vorgeschlagen diese TMI´s in einem standardisierten Subnetz anzusiedeln, indem man die ersten 16 Bit festlegt. Damit wäre die Wahrscheinlichkeit einer Kollision zweier TMI´s immer noch sehr gering. Die TMI´s wären dann zwar leicht erkennbar, auf der anderen Seite wäre auch sichergestellt, dass eine TMI nicht mit einer normalen IP kollidiert, und TMI´s könnten in Routern als nicht weiterleitbar deklariert werden.

Mit diesem Konzept lässt sich Location Privacy bei Kommunikation in beide Richtungen verwirklichen.

Wenn das mobile Gerät die Kommunikation mit einem andern Host anstösst, benutzt er die TMI als Heimadresse. Damit weiss weder der Kommunikationspartner, noch evtl. mithörenden dazwischenliegende Netzkomponenten die Identität und das Heimnetz des Benutzers, denn die TMI sagt nichts in dieser Hinsicht aus und ist nicht weiterleitbar. Es muss die aktuelle care-of Adresse benutzt werden.

Wenn ein anderer Host eine Kommunikation mit dem mobilen Gerät wünscht (z.B. wenn man einen Anruf bekommt) dann weiss der andere Host naturgemäss die Identität des Benutzers und damit seine Heimadresse. Die Kommunikationsanfrage geht daher zunächst an die Heimzone des Benutzers. Diese setzt einen Tunnel zu der aktuellen care-of Adresse des Benutzers auf. Damit dazwischenliegende Netzkomponenten bzw. Mithörer die Kombination aus care-of Adresse und Heimadresse, die beim Tunnelaufbau benötigt wird nicht mitbekommen wird vorgeschlagen die Kommunikation verschlüsselt ablaufen zu lassen. Die Kommunikation läuft also vom Kommuniktionspartner über die Heimzone und von dort über einen sicheren IP-Tunnel zum mobilen Gerät. Wenn das mobile Gerät den Kommunikationspartner für vertrauenswürdig hält kann es dem anderen Host ein „Binding Cache Update“ schicken, damit die Kommunikation mittels Route Optimisation direkt von Sender zu Empfänger ohne den Umweg über die Heimzone laufen kann. Als Heimadresse wird dann natürlich wieder die TMI verwendet. Damit dieses „Binding Cache Update“ nicht mitgehört werden kann muss es auch verschlüsselt gesendet werden.

Dieses Verfahren passt nicht so richtig in eine der von [SaMA] vorgeschlagenen Klassen. Bei Kommunikation vom mobilen Gerät aus wird weder die Heimzone noch die Identität preisgegeben. Ob die Fremdzone die Identität des Benutzers weiss hängt vom Authentifizierungsverfahren des mobilen Gerätes in der Fremdzone ab, das aber auf höheren Ebenen stattfindet und in Castelluccias Paper nicht erwähnt wird. Um das Schema sinnvoller verwenden zu können wird es um eine Partei erweitert: K für Kommunikationspartner. Der Kommunikationspartner weiss in diesem Fall nur die aktuelle care-of Adresse, also praktisch die Identität der Fremdzone. Die daraus folgende Matrix ist in Tabelle 6 dargestellt.

	H	R	L	X	K
u	1		0	0	0
h	1	1	0	0	0
r	1	1	1	1	1

Tabelle 6: Kommunikation vom mobilen Gerät zu anderem Host

Bei der Kommunikation von einem anderen Kommunikationspartner zum mobilen Gerät über einen IP-Tunnel sieht das Ganze etwas anders aus. Der anfragende Host und die dazwischenliegenden Netzkomponenten wissen die Heimadresse des Benutzers, aber nicht die momentane

Fremdzone. Das Ganze ist in Tabelle 7 dargestellt. Teilt der Benutzer nun dem Sender die Heimadresse mit, geschieht die restliche Kommunikation nach dem Schema wie in Tabelle 6.

	H	R	L	X	K
u	1		0	0	1
h	1	1	1	1	1
r	1	1	0	0	0

Tabelle 7: Kommunikation von einem Host zum mobilen Gerät

4.1.2 Privacy Extension mit Hierarchical MobileIPv6

In [Cast01] wird auch noch eine erweiterte Version von Location Privacy in MIPv6 angesprochen, und zwar in Zusammenhang mit Hierarchical Mobile IPv6, das in [SCEMB01] diskutiert wird. Der Basismodus von HMIPv6 funktioniert vereinfacht geschildert folgendermassen: Die gesamte Kommunikation mit dem mobilen Gerät erfolgt über einen „Proxy“, einen „Mobile Anchor Point“ (MAP). Jedes mobile Gerät hat nun neben seiner normalen care-of Adresse noch einer „Regional Care of Adress“ (RCoA). Der Benutzer kommuniziert nun mit dieser RCoA statt mit seiner normalen care-of Adresse. Alle Pakete die zum mobilen Gerät (an die RCoA) geschickt werden landen nun beim MAP. Dieser weiss als einziger die Zuordnung von CoA und RCoA eines mobilen Gerätes. Wenn die Authentifizierung im Fremdnetz nun ausserhalb der Gültigkeit der CoA geschieht, also mit der RCoA, und das Authorisierungsverfahren in geeigneter Weise verschlüsselt ist kann man erreichen, dass der MAP die Benutzeridentität nicht erfahren kann. Mit diesem Verfahren kann man nun erreichen, dass kein beteiligter Rechner gleichzeitig die care-of Adresse und die Heimadresse bzw. die Identität des Benutzers weiss. Es wird angenommen, dass R die Authentifizierungsstelle im Fremdnetz ist, die aber auch nur auf der RCoA arbeitet, und daher nicht den genauen Aufenthaltort weiss. Wer was wissen darf ist in Tabelle 8 dargestellt. Das Verfahren ist der Klasse C_5 ähnlich, geht jedoch in einigen Punkten darüber hinaus.

	H	R	L	X	K
u	1		0	0	0
h	1	1	0	0	0
r	0	0	0	0	0

Tabelle 8: Privacy Extension mit HMIPv6

4.1.3 Mixe und Non Disclosure Methode

D. Chaum schlug in den 80er Jahren ein Verfahren vor, mit dem man Sender und Empfänger von E-Mails geheimhalten kann [Chau85], das jedoch auch verallgemeinert werden kann für die Kommunikation mit einzelnen IP-Paketen. Das ganze basiert auf einem Public-Key-Kryptographieverfahren. Der Sender, der dem Empfänger eine Nachricht senden will, verschlüsselt diese mit dem öffentlichen Schlüssel des Empfängers. Vor die Nachricht wird nun ein IP-Header mit der Adresse des Empfängers gesetzt. Dieses komplette IP-Paket wird nun aber nicht direkt verschickt, sondern wird nochmal verschlüsselt mit dem öffentlichen Schlüssel eines MIX-Servers und dann an diesen gesendet. Dieser entschlüsselt das Paket und erhält somit die Adresse des Zielrechners. Das Paket wird dann an die Zieladresse gesendet. Alle Mithörer zwischen dem Sender und dem Mix bekommen nur die Adressen des

Senders und des MIX-Servers mit, die zwischen dem Mix und dem Empfänger die Adressen des Mix und des Empfängers. Nur der Mix selbst könnte die Zuordnung von Empfänger zu Sender machen. Damit Fremde nicht mit Hilfe einer Verkehrsanalyse an beiden „Kommunikationspunkten“ des Mix eine Korrelation zwischen Sender und Empfänger errechnen können, werden die Pakete in einem Mix neu gemischt, d.h. die Eingangsreihenfolge ist nicht dieselbe wie die Ausgangsreihenfolge. Auch wird ein Paket, das schon einmal weitergeleitet wurde, nicht nochmal weitergeleitet. Unter einem gleichem Paket versteht man hier ein Paket gleichen Inhalts mit gleicher Timestamp. Alle verschlüsselten Nachrichten werden mit zufälligen Bitfolgen auf eine einheitliche Länge gebracht, um auch hier keine Angriffspunkte für eine Analyse zu geben. Das ganze Verfahren basiert in dieser Form auf der Annahme, dass der Mix vertrauenswürdig ist.

Um die Sicherheit bei dieser Methode zu erhöhen und nicht nur vom Betreiber eines einzigen Mix abhängig zu sein, kann man die Methode erweitern. Anstatt nur einen Mix zu verwenden, nimmt man eine ganze Kette von Mixen. Man verschlüsselt die Nachricht mit dem öffentlichen Schlüssel des Empfängers, schreibt die Adresse davor, verschlüsselt das Ganze mit dem öffentlichen Schlüssel des ersten Mix in der Kette vom Empfänger aus gesehen, schreibt dessen Adresse davor und verschlüsselt das ganze mit dem öffentlichen Schlüssel des nächsten Mix, und so weiter. Dann sendet man die Nachricht an den ersten Mix von der Senderseite her gesehen. Dieser entpackt die Nachricht mit seinem privaten Schlüssel und leitet sie an den zweiten Mix weiter,.... Jeder Mix in der Kette kennt durch die Verschlüsselung nur die Adresse seines Vorgängers und seines Nachfolgers. Somit reicht ein sicherer bzw. vertrauenswürdiger Mix in der Kette um die Korrelation zwischen Empfänger und Sender geheim zu halten. Wenn man nun genug Mixe von verschiedenen Organisationen nimmt ist die Wahrscheinlichkeit, dass jemand alle diese Mixe unter seiner Kontrolle hat, sehr gering. Wegen der Verzögerung durch die zeitliche Umordnung in den Mixen ist das Konzept nicht für eine Kommunikation mit Echtzeitanforderungen geeignet. Das Verfahren lässt sich nicht in das vorgestellte Klassifizierungsschema pressen, da es die sichere Kommunikation zwischen zwei beliebigen Partnen behandelt, und nicht auf Fragen der Authentifizierung, etc. eingeht, mit der Tatsache, dass die Kommunikation an sich geheim bleibt.

Es gibt einen Ansatz, der das Verfahren jedoch mit MobileIP verwendet [FaKK96]. Dabei wird grundsätzlich von einem Tunnel zwischen der Heimzone und dem mobilen Gerät ausgegangen. Genau diese Verbindung wird nun mit der Methode der Mixe zu seiner sicheren Verbindung, ohne allen Mithörern die Identität, bzw. den Ort des Benutzers mitzuteilen. In die andere Richtung benutzt das mobile Gerät auch mehrere Mixe, der Tunnel muss aber nicht bis zur Heimzone reichen. Dabei ist offen gelassen, ob der Tunnel direkt bis zum mobilen Gerät reicht oder nur bis zur Fremdzone. Da die Fremdzone durch eine Art von Authentifizierung (über die in [FaKK96] übrigens nichts ausgesagt wird) sowieso meistens die Heimzone des Benutzers und seine temporäre Adresse im Fremdnetz kennt, macht das nichts. Wenn die Kommunikation zwischen mobilem Gerät und der Fremdzone verschlüsselt abläuft, und die Kommunikation zwischen mobilem Gerät und Kommunikationspartner, die die Fremdzone über weitere Mixe tunneln soll, auch verschlüsselt ist, ist das Verfahren trotzdem sicher. Es stellt mindestens die Klasse C_1 und je nach Authentifizierungsverfahren die Klasse C_2 sicher, ermöglicht darüber hinaus noch weitere „Geheimnisse“ die in Tabelle 9 dargestellt sind.

	H	R	L	X	K
u	1		0	0	1
h	1	1	0	0	1
r	1	1	0	0	0

Tabelle 9: Privacyverhalten der Non Disclosure Methode

Die Kommunikation läuft bei diesem Verfahren also über die Heimzone ab bzw. über Mixe. An welchem Ort der Benutzer tatsächlich ist bleibt den Kommunikationspartnern verborgen. [FaKK96] zeigten auch, dass sich die Verzögerungen, die bei der Kommunikation durch die zusätzlichen Mixe entstehen, auf einen erträglichen Wert zu bringen sind. Ausserdem ist das Verfahren auf die persönlichen Sicherheitsbedürfnisse des Benutzers flexibel einstellbar. Benötigt er kürzere Antwortzeiten und ist ihm die Verbergung seines Ortes nicht so wichtig, so kann er die Anzahl der Mixe herabsetzen, benötigt er mehr Sicherheit, so kann er weitere Mixe hinzunehmen.

4.2 Basic Authentication Protocol

Das Basic Authentication Protocol von [SaMA] ist ein abstraktes Verfahren, das sich speziell mit der Authentifizierung von mobilen Geräten in einer Fremdzone beschäftigt. Da es sehr komplex ist, wird hier nur kurz und vereinfachend darauf eingegangen. Das ganze funktioniert mit Hilfe eines Public-Key Verschlüsselungsverfahrens. Das mobile Gerät verschlüsselt seine reale Benutzerkennung mit einem Schlüssel der Heimzone, den es gespeichert hat. Das Ergebnis ist dann die temporäre Benutzerkennung des Gerätes im Fremdnetz. Vor dem Verschlüsseln wird die Benutzerkennung noch mit einer Zufallsfolge aufgefüllt, sonst wäre die temporäre Benutzerkennung jedes mal identisch. Diese temporäre Benutzerkennung und die ebenfalls verschlüsselte Identität der Fremdzone wird nun an die Heimzone geschickt. Diese kann beide Informationen entschlüsseln und anhand der realen Benutzerkennung den Benutzer eindeutig identifizieren. Danach wird die temporäre Benutzerkennung und eine digitale Signatur an die Fremdzone gesendet, um dieser mitzuteilen, dass der Benutzer die Dienste (Netzzugang) nutzen darf. Das mobile Gerät kann nun mit dem öffentlichen Schlüssel der Fremdzone, seiner temporären Benutzerkennung und einer Zufallsfolge weitere temporäre Benutzerkennungen errechnen, die die Fremdzone ohne Mithilfe der Heimzone authentifizieren kann. Somit kann die Benutzerkennung regelmässig gewechselt werden und somit die Zuordnung von Nachrichten zu (wenn auch nur temporären) Benutzerkennungen unterbunden werden.

Die wahre Identität des Benutzers bleibt also der Fremdzone und weiteren Beteiligten ausser der Heimzone verborgen. Das Protokoll garantiert Location Privacy der Klassen C_1 und C_2:

	H	R	L	X
u	1	0	0	0
h	1	1	1	1
r	1	1	1	1

Tabelle 10: Basic Authentication Protocol

Mit einem einfachen Trick lässt sich das Verfahren auf eine höhere Klasse bringen. Zu Beginn des Protokolls benötigt das mobile Gerät dazu den öffentlichen Schlüssel der Fremdzone. Diesen bekommt sie von der Fremdzone direkt oder von einem weiteren Schlüsselserver. Damit überträgt das mobile Gerät die Identität der Heimzone verschlüsselt an die Fremdzone. Danach läuft das Protokoll wie oben beschrieben ab. Somit kann man die Identität der Heimzone vor Dritten schützen, das Protokoll steigt also in Klasse C_3 auf:

4.3 Homeless Authentication Protocol

Das Homeless Authentication Protokoll, auch von [SaMA], ist eine Erweiterung des Basic Authentication Protokolls. Auch dieses Protokoll ist recht kompliziert und wird hier nur in

	H	R	L	X
u	1	0	0	0
h	1	1	0	0
r	1	1	s	s

Tabelle 11: Basic Authentication Protocol mit Erweiterung

den Grundzügen dargestellt. Es geht davon aus, dass der Benutzer zwischen verschiedenen Fremdzonen migriert. Die neue Fremdzone fragt bei der ersten Anmeldung des Benutzers in der neuen Fremdzone bei der alten Fremdzone nach. Die alte Fremdzone bestätigt der neuen Fremdzone, dass der Benutzer authentifiziert war. Die Kommunikation zwischen der alten und der neuen Fremdzone läuft verschlüsselt ab. Dabei wird kurz gesagt die temporäre Benutzerkennung an die neue Fremdzone übermittelt, und der Benutzer damit in der neuen Fremdzone authentifiziert. Die Fremdzonen kennen somit die Identität der Heimzone nicht, da die Authentifizierung komplett ohne Nachfragen bei der Heimzone abläuft. Das Protokoll garantiert also Location Privacy nach der Klasse C_4. Nun könnte man einwenden, dass durch die erste Migration (von der Heimzone in die erste Fremdzone) wenigstens die erste Fremdzone Kenntnis von der Heimzone erlangt. Da dabei jedoch das Protokoll nicht unterschiedlich ist, weiss die erste Fremdzone nur, dass der Benutzer in der vorherigen Zone authentifiziert war. Dass das die Heimzone war spielt dabei keine Rolle.

Das Protokoll garantiert sogar teilweise die Klasse C_5. Die Heimzone kennt nur die erste Fremdzone, alle weiteren bleiben ihr verborgen, also steht bei den Kentnissen der Heimzone über die Fremdzonen ein s (für some). Die Fremdzonen wiederum wissen nur teilweise Bescheid über die Bewegungen des Benutzers. Fremdzonen kennen die vorangegangene Fremdzone und die Nachfolgende, da eine Fremdzone mit diesen beiden Kontakt hat wegen der Authentifizierung. Weitere Fremdzonen, in denen sich der Benutzer aufhält sind jedoch vor einer Fremdzone verborgen, deshalb auch hier ein s in der Tabelle:

	H	R	L	X
u	1	0	0	0
h	1	0	0	0
r	s	s	s	s

Tabelle 12: Homeless Authentication Protocol

5 Fazit

In den bisher verbreiteten Mobilfunknetzen hat der Benutzer wenig Location Privacy. Bei der Spezifizierung von MobileIP wurde auf Location Privacy auch noch keine Rücksicht genommen. Es sind jedoch Vorschläge für Erweiterungen des MobileIP-Standards vorhanden, die Mobilität und Privacy verbinden. Hier wurden die Erweiterungen von Castelluccia und die Non Disclosure Methode erwähnt. Diese beiden Verfahren legen ihren Schwerpunkt auf die Kommunikation zwischen mobilem Gerät und anderen Kommunikationspartnern. Die beiden anderen vorgestellten Verfahren, das Basic Authentication Protocol und das Homeless Protocol, behandeln jedoch hauptsächlich die Authentifikation in Fremdnetzen. Alle diese Verfahren zusammen zeigen, dass Location Privacy machbar ist, mit vertretbarem Aufwand. Wieviel und welche dieser Verfahren in der Zukunft tatsächlich implementiert und benutzt werden, wird sich zeigen.

Literatur

[AHKT98] Guiseppe Atinese, Amir Herzberg, Hugo Krawczyk und Gene Tsudik (Hrsg.). *On Traveling Incognito*, IEE Workshop on Mobile Systems and Applications, November 1998.

[Cast01] Claude Castelluccia. A Simple Privacy Extension for Mobile IPv6. DRAFT http://ietf.org/internet-drafts/draft-castelluccia-mobileip-privacy-00.txt, February 2001.

[Chau85] D. Chaum (Hrsg.). *Untraceable Electronic Mail, Return Addresses and Digital Pseudonyms*, Band 24, October 1985.

[CoBi95] David A. Cooper und Kenneth P. Birman (Hrsg.). *Preserving Privacy in a Network of Mobile Computers*, Nr. TR 95-1490 der Proceedings of the 1995 IEEE Symposium on Security and Privacy, May 1995.

[FaKK96] Andreas Fasbender, Dogan Kesdogan und Olaf Kubitz (Hrsg.). *Analysis of Security and Privacy in Mobile IP*, 4th International Conference on Telecommunication Systems, Modelling and Analysis, March 1996.

[LaDG96] Ben Lancki, Abhijit Dixit und Vipul Gupta. Mobile-IP: Supporting Transparent Host Migration on the Internet. *Linux Journal*, August 1996.

[SaMA] Didier Samfat, Refik Molva und N. Asokan. Anonymity and Untracebility in Mobile Networks. http://www.acm.org/pubs/citations/proceedings/comm/215530/p26-samfat/.

[SCEMB01] Hesman Soliman, Claude Castelluccia, Karim El-Malki und Ludovic Bellier. Hierarchical MIPv6 mobility management. DRAFT, February 2001.

Tabellenverzeichnis

1 Klasse C_1 48

2 Klasse C_2 48

3 Klasse C_3 48

4 Klasse C_4 49

5 Klasse C_5 49

6 Kommunikation vom mobilen Gerät zu anderem Host 52

7 Kommunikation von einem Host zum mobilen Gerät 53

8 Privacy Extension mit HMIPv6 53

9 Privacyverhalten der Non Disclosure Methode 54

10 Basic Authentication Protocol 55

11 Basic Authentication Protocol mit Erweiterung 56

12 Homeless Authentication Protocol 56

Unicast- und Multicast Algorithmen in Ad-Hoc Netzen

Rami Nassar und Philipp Geipel

Kurzfassung

Diese Seminararbeit soll einen Überblick über bestehende Algorithmen in der Ad-Hoc Netzwerk Technologie geben. Ausgangssituation ist die rapide Entwicklung drahtloser, mobiler Rechner, die gerade aktuell mit Sende- und Empfangsadapter für die Anbindung an Ad-Hoc Netzwerke angeboten werden. Auch der immer größer werdende Trend, ständig mobil, flexibel und ungebunden zu sein, fordert eine passende, drahtlose und spontane Anbindung an Netzwerke, den sogenannten Ad-Hoc Netzwerken. Dass für die technische Umsetzung solcher Netzwerke, geeignete und effiziente Routing Algorithmen benötigt werden, liegt auf der Hand. In diesem Zusammenhang werden die wichtigsten Protokolle vorgestellt und beschrieben. Im Fordergrund stehen dabei diejenigen Algorithmen, die sich einmal in der Vergangenheit schon bewährt haben und die für zukünftige Anwendungen im Bereich des mobilen Ad-Hoc Netzwerks überhaupt einen Sinn machen. Weiterhin soll der Unterschied zwischen Unicast und Multicast Algorithmen, in bezug auf Anwendungen, Performance in der Praxis sowie Leistungsgrenzen erläutert werden.

1 Was ist ein Ad-Hoc Netzwerk?

Definition : „*Ein Ad-Hoc Netzwerk ist ein drahtloses, sich selbst organisierendes, Netzwerk, dass auf mobilen Knoten basiert und keine feste Infrastruktur benötigt.*"
Dies bedeutet, dass mehrere Computer ein Netzwerk bilden, unabhängig von vorinstallierten Strukturen. Es sind also weder Verkabelungen, noch anderswertige Eingriffe notwendig, die an der Umgebung vorzunehmen sind. Es existiert keine Server/Client-Architektur, sondern gleichberechtigte Einheiten, die jeweils die Knoten des Kommunikationsnetzes bilden. Lediglich die mobilen Geräte bilden untereinander das Netzwerk, das ohne einen hierarchischen Aufbau auskommt. Jedes Gerät muss über eine eingebaute Empfangs- und Sendeeinheit verfügen, um Kontakt zu den anderen Teilnehmern im Netzwerk aufzubauen zu können. Die mobilen Geräte organisieren sich untereinander zu einem Netz-Verbund, in dem Informationen, Daten und Anwendungen ausgetauscht werden können.

1.1 Anwendungen von Ad-Hoc Netzwerken

Aufgrund der großen Fortschritte im Bereich der Hardware, insbesondere der Minimalisierung in Größe und Gewicht der technischen Geräte, erfreuen sich mobile Einsätze sowohl für den kommerziellen als auch militärischen Bereich einer immer größer werdenden Beliebtheit. Vor allem die Leistungsfähigkeit hat sich in den letzten Jahren enorm gesteigert: so kommen heutige mobile Geräte mit einer besonders geringen Leistung aus und können trotzdem eine beachtliche Performance bieten. Das bedeutet, dass viele verschiedene Anwendungen auch auf sehr kleinen mobilen Geräten mit einer vergleichbaren Funktionsfähigkeit wie auf Desktops laufen. Diese Geräte können in verschiedene Plattformen eingebettet werden, so dass sich das Umfeld zu einer hochkontrollierbaren, unterstützenden und kooperativen Umgebung wandelt. Anwendungen dieser Technologie sind leicht für beinahe jeden Aspekt des menschlichen

Lebens vorstellbar: zu Hause, bei der Arbeit, im Supermarkt, im Auto, in der Bahn usw. Ebenso hat sich die Leistungsfähigkeit der Sende- und Empfangsantennen um ein Vielfaches erhöht, so dass sich das äußere Format auf ein praktikables, komfortables Niveau reduziert hat und längere Standbyzeiten möglich geworden sind. Auch die Datendurchsatzraten im Funkbereich stehen heut zu Tage denen von festinstallierten Netzwerken nur im geringen Maße nach, so dass Arbeiten im Funknetzwerk keinerlei Beeinträchtigungen mit sich bringt. So sind zum Beispiel interaktive Vorlesungen und Seminare an Universitäten und Schulen technisch vorstellbar. Weitere Anwendungsgebiete finden sich im Businesssektor, im militärischen Bereich und bei Katastrophen zur Koordination von Rettungseinsätzen. Hier ergeben sich völlig neue Möglichkeiten zum Aufbau von Kommunikationsstrukturen, die keinerlei Infrastruktur benötigen. Besonders in abgelegenen Gegenden, wo eine Einrichtung der technischen Infrastruktur wegen Kostenaspekten oder geographischen Gründen wenig Sinn macht, kommen Anwendungen von Ad-Hoc Netzwerken zum Einsatz. Vorstellbar wäre hier die Vernetzung eines Messestandes.

1.2 Aufbau eines Ad-hoc Netzwerkes

Es existieren zur Zeit zwei Arten von mobilen, drahtlosen Netzwerkstrukturen. Bei der ersten handelt es sich um ein infrastrukturbasiertes Netzwerk mit fixen, verdrahteten Gateways. Die „Bridges" für diese Netzwerke sind unter dem Namen „Base Stations" oder im Deutschen als Basisstationen bekannt. Eine mobile Einheit nimmt Kontakt zur nächstliegenden Basisstation innerhalb ihres Kommunikationsradius auf und kommuniziert mit ihr. Falls sich eine mobile Einheit aus dem Bereich einer Basisstation in den Bereich einer anderen bewegt findet ein sogenannter „handoff" statt. Somit findet ein nahtloser Übergang beim Wechsel der Basisstationen statt, ohne dass die Kommunikation unterbrochen wird. Typisches Anwendungen solcher Netzwerke sind drahtlose Funknetze für zum Beispiel Büroumgebungen, die auch unter dem Namen Wireless Local Area Network (WLAN) zusammengefasst werden.
Die zweite bekannte, drahtlose Netzwerkstruktur, das Ad-Hoc Netzwerk, kommt ohne feststehende Infrastruktur aus. Die Architektur eines Ad-Hoc Netzwerkes ist eine Ansammlung von sich im Raum frei bewegenden Routern, die selbständig und dynamisch andere Router suchen und Verbindung zu ihnen halten. Der Status der Verbindungen zwischen diesen Routern ist zu jeder Zeit abhängig von ihrer Position, Übertragungsstärke ihres Signals, Struktur der Antennenaufteilung im Raum (pattern) und Interferenzen mit anderen Sendern. Die Mobilität der Router und die Unsicherheit der anderen Übertragungsfaktoren resultieren in einer schnell wechselnden und unvorhersehbar schwankenden Topologie. Durch die sicherlich geringere Kapazität schnurloser Funknetzwerke gegenüber Festnetzen, sind Ad-Hoc Netzwerke sehr anfällig, schnell an ihre Kapazitätsgrenzen zu stossen, was eine Überlastung des Netzes zur Folge hat. Praktische Anwendungen solcher Ad-Hoc Netzwerke sind für Personen denkbar, die sich in einer Umgebung ohne feststehende Netzwerkinfrastruktur bewegen und schnell Informationen teilen und Daten aufnehmen wollen.
Grundsätzlich unterscheidet man zwei Arten von Ad-Hoc Netzwerkarchitekturen: das in Abbildung 1 dargestellte „gleichberechtigte", einstufige Modell und das in Abbildung 2 „hierarchische", zum Beispiel zweistufige Netzwerk.

In „gleichberechtigten" Netzwerken sind alle Knoten gleichwertig und das Paketrouting, das hier zum Einsatz kommt, basiert auf peer-to-peer Verbindungen und ist nur durch die Ausbreitungsbedingungen beschränkt. Bei den „hierarchichen" Netzwerken existieren mindestens zwei Ebenen. Auf der unteren Ebene bilden nahe aneinanderliegende Knoten ein peer-to-peer Netzwerk, in dem mindestens ein Knoten als Gateway zur höheren Ebene dient. Die Gatewayknoten aus den peer-to-peer Netzwerken stellen, wie in Abbildung 2 zu sehen, die zweite Hierarchieschicht dar und benötigen üblicherweise stärkere Übertragungs- und Empfangskapazitäten. Das Routing zwischen Knoten des selben Netzwerks in einer unteren Ebene basiert

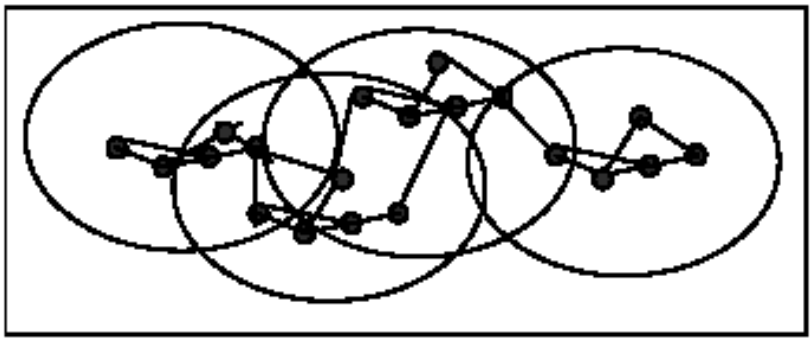

Abbildung 1: Darstellung eines gleichberechtigten, einstufigen Ad-Hoc Netzwerks

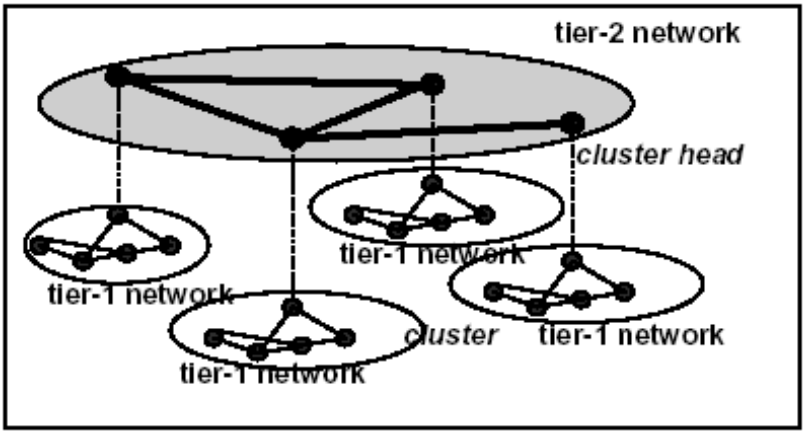

Abbildung 2: Darstellung eines hierarchichen, zweistufigen Ad-Hoc Netzwerks

auf einer peer-to-peer Verbindung, wobei das Routing zwischen verschiedenen Netzwerken unterer Schichten über die Gatewayknoten stattfindet.
Diese Seminararbeit beschäftigt sich ausschließlich mit Routingprotokollen für die eben beschriebenen Ad-Hoc Netzwerke, indem zuerst die wichtigsten in ihrer Funktionsweise vorgestellt werden und im Anschluss die verschiedenen Eigenschaften miteinander verglichen werden.

2 Unicastprotokolle

In den frühen 70er Jahren im 20. Jahrhundert wurden verschiedene Protokolle für Ad-Hoc Netzwerke entwickelt. Unicast stellt die grundlegende Verbindung von einem Sender zu einem Empfänger dar. Wenn ein Datenpaket an mehrere Empfänger gesendet werden soll, dupliziert der Sender dieses Datenpaket und adressiert es an jeden einzelnen Empfänger. Dadurch hat man in den Anfängen der Netzwerktechnologie eine einfache und klare Struktur zwischen Quelle und Ziel geschaffen. Die Quelle kann auf Grundlage verschiedener Protokolle den Weg zum Ziel herausfinden. Abbildung 3 zeigt das Unicastverfahren bei der Versendung eines Datenpaketes an drei verschiedene Empfänger. Es lassen sich wie in Abbildung 4 zu sehen zwei grundlegende Protokollkategorien unterscheiden:

- Table-Driven
- Source-Initiated (demand-driven)

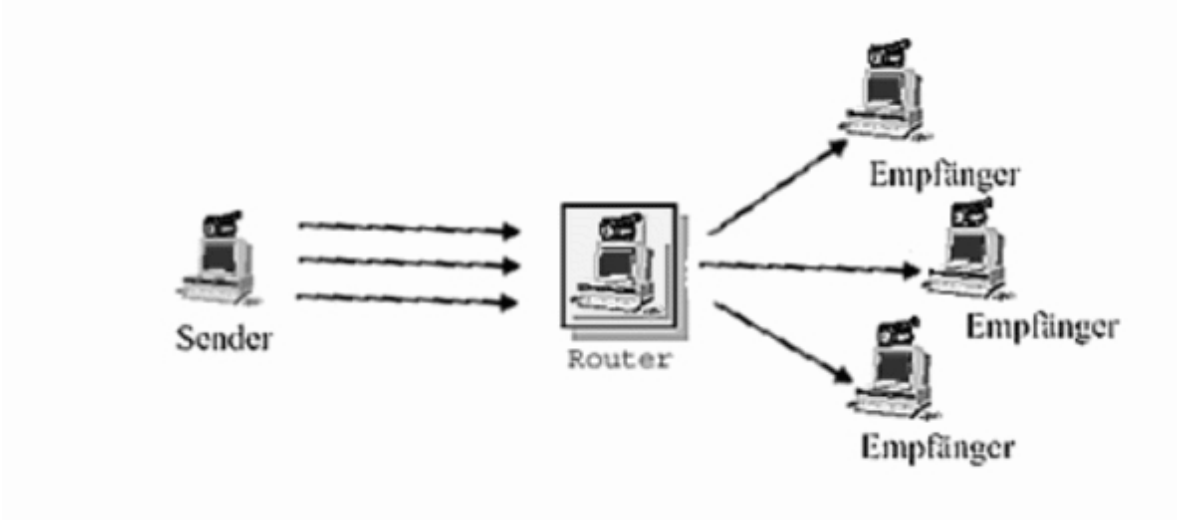

Abbildung 3: Vorgehensweise von Unicast

Die durchgezogenen Linien in Abbildung 4 repräsentieren direkte Weiterentwicklungen, die gestrichelten logische. Obwohl alle Protokolle für den selben Typ von Netzwerken entwickelt

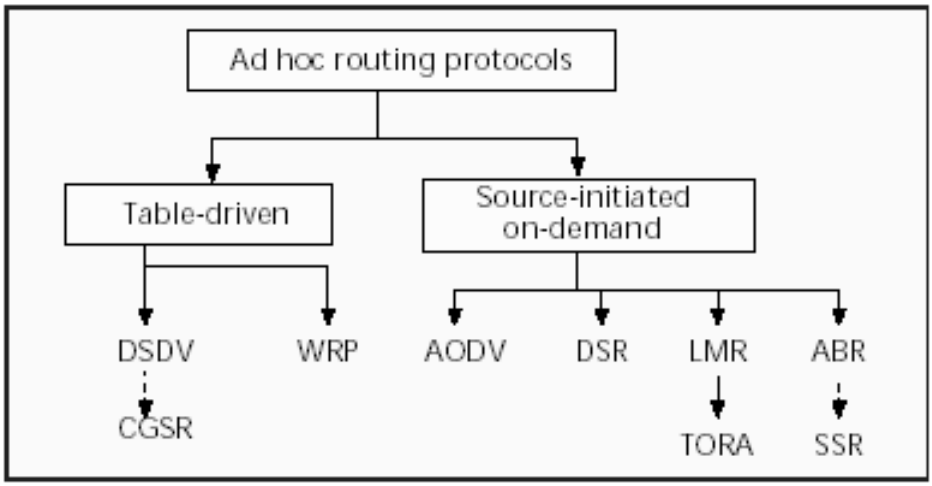

Abbildung 4: Kategorisierung von Unicastprotokollen

wurden, weisen sie unterschiedliche Charakteristiken auf. Diese werden in den folgenden Unterkapiteln untersucht. Diese Protokolle müssen mit den typischen Eigenschaften von hohem Stromverbrauch, niedrigen Bandbreiten und hohen Fehlerraten zurechtkommen.

2.1 Table-Driven Routing Protocols

Die Table-Driven Routing Protokolle (Tabellenbasierte Routingprotokolle) speichern konsistente, stets aktuelle Informationen über bestehende Verbindungen zwischen den Netzknoten. Dafür muss jeder Knoten eine oder mehrere solcher Tabellen anlegen. Bei einer Veränderung der Netzwerktopologie melden die entsprechenden Knoten den aktuellsten Stand an die übrigen Knoten weiter, um die Netzwerkkonsistenz aufrechtzuerhalten. Die Protokolle unterscheiden sich durch die Anzahl der benötigten Tabellen und durch die Übertragungsmethoden von Netzwerkveränderungen.

2.1.1 Destination-Sequenced Distance Vector Routing

Das Destination-Sequenced Distance Vector Routing (DSDV) basiert auf dem klassischen Bellmann-Ford Routing Algorithmus. Es wurden Verbesserungen durchgeführt, so dass auf Schleifen in den Routing Tabellen verzichtet werden konnte. Jeder Knoten unterhält eine Tabelle, in der alle möglichen Zieladressen, sowie die benötigte Anzahl von Sprüngen zum entsprechenden Ziel gespeichert werden. Jeder Eintrag bekommt von dem Zielknoten eine Sequenznummer zugeteilt, die eine Unterscheidung zwischen alten und aktuellen Routen erlaubt, wodurch Schleifenbildungen vermieden werden. Tabellenaktualisierungen werden regelmäßig über das Netzwerk geschickt. Um die dadurch entstehende Netzwerkbelastung zu reduzieren, können zwei unterschiedliche Methoden angewendet werden. Die erste, unter dem Namen *full dump* bekannt, verschickt alle verfügbaren Routinginformationen und kann unter Umständen mehrere Network Protocol Data Units (NPDU) benötigen. Abhängig von der Netzbelastung werden diese Pakete in unregelmäßigen Abständen verschickt. Die andere Methode, *incremental Packets*, verschickt nur diejenigen Informationen, die sich seit dem letzten full dump geändert haben. Jedes dieser Updates sollte die Ressource einer NPDU nicht überschreiten und somit die Netzbelastung reduzieren. Die mobilen Knoten unterhalten eine zusätzliche Tabelle, in der diese Updates gespeichert werden.
Neue Routingübertragungen enthalten sowohl die Zieladresse, die Anzahl der Sprünge zu diesem Ziel, die erhaltene Sequenzenummer die dieses Ziel betrifft als auch einer einmaligen Sequenznummer für die Übertragung. Die Route mit der aktuellesten Sequenznummer wird für die Übertragung benutzt. Im Falle zwei gleicher Sequenznummern, wird die Route mit der günstigeren Eigenschaft verwendet, um den optimalen Weg zu nutzen. Dazu überwachen die Mobilgeräte die durchschnittlich benötigte Zeit, um die Route zu durchlaufen, erst dann wird die günstigste Route ausgewählt. Durch die absichtliche Verzögerung der Übertragung von Updates um die „settling time" kann die Netzwerkbelastung reduziert werden. Der optimalste Weg ergibt sich dadurch, dass die Übertragungen verhindert werden, die nicht die aktuellsten Netzwerkinformationen beinhalten.

2.1.2 Clusterhead Gateway Switch Routing

Das (CGSR) unterscheidet sich von dem vorigen Protokoll in der Adressierung und dem angewendeten Netzwerk Organisationsschema. Im Gegensatz zum einstufigen (flat) Netzwerk ist CGSR ein zu Clustern zusammengefasstes mobiles Netzwerk, das über heuristische Routing Schemata aufgebaut ist. Innerhalb eines Clusters wird mittels eines speziellen Algorithmus ein Knoten als Kopf bestimmt, der für die Kommunikation zwischen den Clustermitgliedern die Rahmenbedingungen vorgibt. Ein Nachteil dieses Protokolls liegt in der ständigen Netzbelastung durch die Suche eines geeigneten Clusterkopfes. Um dieses Problem zu vermeiden, wird der Clusterkopf nur geändert, falls er in die Reichweite eines anderen kommt oder wenn ein Knoten aus der Reichweite aller Clusterköpfe wandert. Dieses Verfahren wird Least Cluster Change (LCC) Clusteralgorithmus genannt.
CGSR benutzt DSDV als zugrunde liegendes Routingschema, mit dem Unterschied, dass ein Datenpaket zunächst an den Clusterkopf geschickt wird und es dieser über Gatewayknoten an den Clusterkopf des Zielknotens weiterleitet. Ein Gatewayknoten steht in Reichweite zu zwei oder mehr Clusterköpfen. Dieses Schema ist in Abbildung 5 dargestellt.

Jeder Knoten speichert in einer Tabelle zu jedem Netzknoten den zuständigen Clusterkopf. Diese Cluster Mitgliedstabellen werden von jedem Knoten mit dem DSDV Algorithmus verschickt. Bei Erhalt einer aktualisierten Clustermitgliedstabelle von einem Nachbarn, wird die eigene Tabelle auf den neuen Stand gebracht. Zusätzlich zu der Clustermitgliedstabelle wird eine Routing Tabelle geführt, anhand welcher der nächste Schritt bei der Weiterleitung eines Pakets an den nächsten Clusterkopf auf dem Weg zum Ziel festgelegt wird.

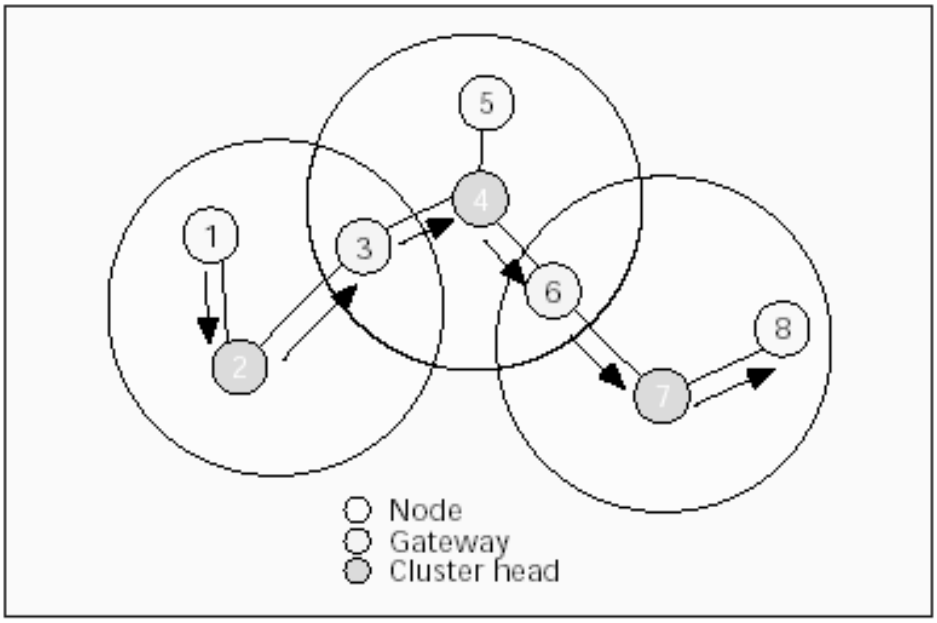

Abbildung 5: CGSR: Routing von Knoten 1 zu Knoten 8

2.1.3 The Wireless Routing Protocol

Das Wireless Routing Protocol (WRP) ist wie die vorherigen Protokolle tabellenbasiert und hat das Ziel, über alle Knoten im Netzwerk Routinginformationen zu speichern. Dafür muss jeder Knoten im Netzwerk vier Tabellen unterhalten:

- Entfernungstabelle
- Routing Tabelle
- Link-cost Tabelle
- Message retransmission list (MRL) Tabelle

Jeder Eintrag in der MRL-Tabelle beinhaltet die Sequenznummer der Aktualisierungsnachricht, einen Wiederübertragungszähler, ein **Bestätigungsvermerk für jeden Nachbarn**, sowie eine Liste von Aktualisierungen, die in der Aktualisierungsnachricht mitversendet wurde. Die MRL speichert welche Aktualisierungen in einer Aktualisierungsnachricht wieder verschickt werden müssen und welche Nachbarn diese bestätigen müssen.
Die mobilen Einheiten informieren sich gegenseitig über Verbindungsveränderungen mit Hilfe der Aktualisierungsnachrichten. Eine Aktualisierungsnachricht wird nur zwischen benachbarten Knoten versendet und beinhaltet sowohl eine Liste von Updates (Ziel, Entfernung zum Ziel, Vorgänger vom Ziel), als auch eine Liste, welche mobilen Einheiten dieses Update bestätigen (ACK) müssen. Die mobilen Einheiten versenden Aktualisierungsnachrichten nachdem sie Updates von ihren Nachbarn verarbeitet haben oder Verbindungsveränderungen zu einem Nachbarn festgestellt haben. Falls zwei Knoten den Kontakt verlieren sollten, informieren sie ihre Nachbarn darüber. Die Nachbarknoten passen dann ihre Entfernungstabelle an und suchen nach einem Weg über andere Knoten. Informationen über neue Wege werden an den ursprünglichen Knoten weitervermittelt. Falls über längere Zeit von einem Knoten keine Nachrichten verschickt wurden, muss eine „Hallo-Nachricht" in bestimmten Zeitperioden versendet werden, damit die Nachbarschaft von seiner Existenz erfährt und der Kontakt erhalten bleibt. Andersfalls wird die Verbindung von den Nachbarn als verloren betrachtet. Wenn ein mobiles

Gerät eine Hallo-Nachricht von einem neuen Knoten empfängt, wird dieser in seine Tabelle aufgenommen und erhält eine Kopie der Routingtabelleninformationen.

Ein großer Bestandteil der Neuerung des WRP ist mit dem Erreichen der Schleifenfreiheit verbunden. WRP gehört zu der Gruppe der pfadsuchenden Algorithmen, vermeidet aber das *count to infinity-* Problem[1], dadurch dass jeder Knoten gezwungen wird, Konsistenzprüfungen durchzuführen. Dies löscht endgültig, wenn auch nicht sofort, Schleifen und sorgt deshalb für schnellere Konvergenz, wenn ein Linkfehler auftritt.

2.1.4 Eigenschaftsvergleich der tabellenbasierten Routingprotokolle

Da nun die drei wichtigsten Protokolle untersucht wurden, soll hier nun ein Vergleich zwischen ihnen durchgeführt werden. WRP hat gegenüber DSDV eine höhere Zeitkomplexität bei auftretenden Verbindungsfehlern, da DSDV nur seine Nachbarknoten über einen verlorenen oder fehlerhaften Kontakt informiert. DSDV muss auch beim Suchen einer neuen Verbindung lediglich die oben beschriebenen Hallo-Nachrichten an seine Nachbarknoten versenden. Auch in dem CGSR Protokoll ist die Zeitkomplexität um einiges höher als beim DSDV, da die Routingperformance von speziellen Knoten abhängt (Clusterkopf- oder Gatewayknoten). Falls also ein Clusterkopf den Kontakt zu seinen übrigen Clustermitgliedern verliert oder Verbindungsfehler auftreten, muss zunächst ein neuer Clusterkopf mit zeitaufwendigen Algorithmen bestimmt werden. Der gleiche zeitliche Aufwand wird beim Hinzukommen eines neuen Clustermitglieds benötigt, da in dieser Situation eine erneute Bestimmung des Clusterkopfes fällig wird. Die Wahl eines Gatewayknotens findet nicht statt, da jeder Knoten selbst in der Lage ist, seinen Nachbarn mitzuteilen, dass er Kontakt zu mindestens zwei Clustern hat und damit Gatewayknoten ist. Bei Verlust eines Gatewayknotens durch Herausbewegung aus dem entsprechenden Cluster, fällt dem Clusterkopf die Aufgabe zu, einen neuen Gatewayknoten zu finden. Auch dieser Vorgang trägt zu einer Erhöhung der Zeitkomplexität bei. Da sich aber alle drei Protokolle für den kürzesten Weg zum Ziel entscheiden, besitzen alle den gleichen zeitlichen Komplexitätsgrad, sowohl beim Zusammenbruch einer Verbindung als auch bei der Neusuche von Wegen. Ein zusammenfassender Überblick findet man in Tabelle 1, wo die Zeit- und Kommunikatonskomplexitäten in Groß O-Notationen gegenübergestellt werden und andere Charakteristiken wie Schleifenfreiheit, Multicastfähigkeiten usw. verglichen werden.

2.2 Source Initiated On-Demand Routing Protocols

Im Gegensatz zu den tabellenbasierten Protokollen, berechnen quelleninitierte Protokolle Wege zum Ziel erst bei Bedarf. Falls ein Knoten den Weg zu einem Ziel benötigt, regt er einen Suchprozess an, der dann beendet wird, wenn ein möglicher Weg gefunden wurde oder aber alle Wegpermutationen untersucht wurden. Ist ein Weg gefunden, wird dieser solange aufrecht erhalten, bis das Ziel auf diesem Pfad nicht mehr erreichbar ist oder der Weg nicht mehr gebraucht wird.

2.2.1 Ad Hoc On-Demand Distance Vector Routing

Das Ad-Hoc On-Demand Distance Vector Routing Protokoll (AODV) baut auf dem DSDV Protokoll auf, verbessert es aber, da es die Anzahl der benötigten Übertragungen minimiert, in dem nur bei Bedarf der Weg zu einem Zielknoten berechnet wird, anstatt eine komplette Liste an Wegen aufzuführen. Knoten, die nicht auf dem ausgesuchten Pfad liegen, benötigen

[1] vgl.[Tann96]

Parameter	DSDV	CGSR	WRP
Zeitkomplexität (Verbindungsaufbau/-fehler)	O(d)	O(d)	O(h)
Kommunikationskomplexität (Verbindungsaufbau/-fehler)	O(x=N)	O(x=N)	O(x=N)
Routing Topologie	flach	hierarchisch	flach
Schleifenfreiheit	ja	ja	ja, aber nicht sofort
Multicastfähigkeit	nein	nein	nein
Anzahl benötigter Tabellen	zwei	zwei	vier
Häufigkeit von Update Übertragungen	periodisch und falls benötigt	periodisch	periodisch und falls benötigt
Updates werden übertragen an	Nachbarn	Nachbarn und Clusterkopf	Nachbarn
Benutzung von Sequenznummern	ja	ja	ja
Benutzung von "Hallo" Nachrichten	ja	nein	ja
Kritische Knoten	nein	ja, Clusterknoten	ein
Routingmaß	kürzester Pfad	kürzester Pfad	kürzester Pfad

Legende:	
N=	Anzahl der Knoten im Netzwerk
d=	Netzwerkdurchmesser
h=	Höhe des Routingbaums
x=	Anzahl Knoten, die von einer Topologieänderung betroffen sind

Tabelle 1: Eigenschaftsvergleich der tabellenbasierten Routingprotokolle

keine Routinginformationen und nehmen auch nicht am Austausch der Routingtabellen teil. Falls ein Quellknoten den Weg zu seinem gewünschten Ziel noch nicht kennt, initiiert er einen Pfadsuchprozess, um den Zielknoten zu lokalisieren. Er sendet eine Weganfrage (Route request: RREQ) an seine Nachbarn, die wiederum an ihre Nachbarn dieselbe Anfrage stellen, bis entweder das Ziel selbst oder ein Knoten auf dem Weg dahin, der einen aktuellen Pfad dahin kennt, gefunden wurde. Abbildung 6a illustriert die Verbreitung der Weganfragen über das Netzwerk. AODV verwendet Zielsequenznummern, die sowohl einen schleifenfreien Weg als auch die aktuellsten Weginformationen sicherstellen. Jeder Knoten verfügt über eine eigene Sequenznummer und Broadcast-ID. Die für jede RREQ um einen Zähler inkrementierte Broadcast-ID zusammen mit der IP-Adresse der Quelle identifizieren jede Weganfrage eindeutig. Der Quellknoten schickt die Weganfrage zusammen mit seiner eigenen Sequenznummer und seiner Broadcast-ID mit der aktuellsten Sequenznummer des Zielknotens auf den Weg. Die auf dem Weg liegenden Knoten dürfen nur dann antworten, wenn sie einen Weg zum Ziel kennen, dessen korrespondierende Zielsequenznummer größer oder gleich der im RREQ aktuell gültigen ist.
Während des Weiterleitungsprozesses speichern die Knoten die Adresse des Nachbarn von dem sie die Anfrage als erstes erhalten haben und stellen dadurch einen Rückweg zur Verfügung. Später eintreffende Kopien der Anfrage werden ignoriert. Sobald das Ziel, oder aber ein Knoten der einen aktuellen Weg zum gewünschten Ziel kennt, erreicht wurde, antworten diese mit einer Wegantwort (Route reply: RREP) dem Nachbarn, von dem sie zuerst die Anfrage erhalten haben. (Vergleiche Abbildung 6b). Während die Antwort auf dem gespeicherten

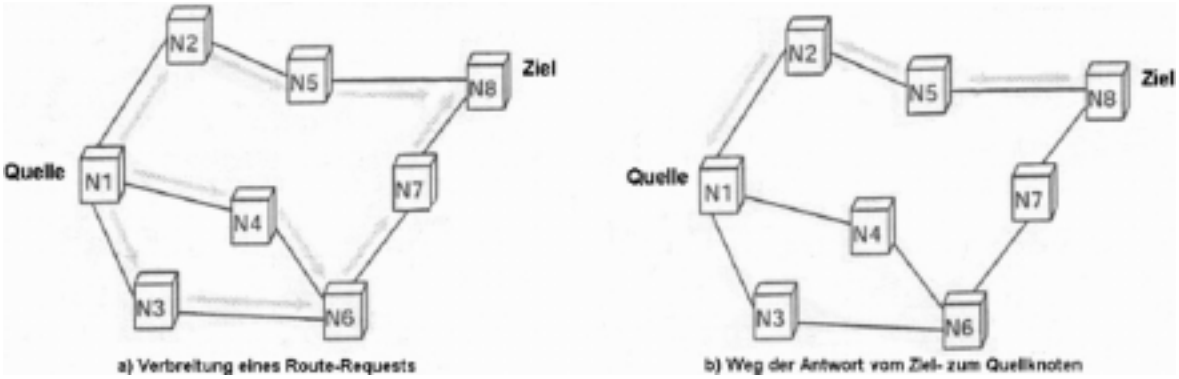

Abbildung 6: AODV Routenerforschung

Rückweg entlangläuft, speichern sich die Knoten auf diesem Pfad die Zieldaten in ihren Routingtabellen. Zusammen mit den Einträgen in diese Tabelle wird eine maximale Lebensdauer der Informationen generiert. Da das RREP auf dem vom RREQ geebneten Weg zurückläuft, unterstützt das AODV nur den Gebrauch symmetrischer Links.
Falls sich ein Quellknoten bewegt, kann er den Wegfindungsprozess erneut anstoßen, um einen eventuell neuen Weg zum Ziel zu finden. Sollte sich ein Knoten auf dem Weg zum Ziel bewegen, bemerkt sein Nachbar die Veränderung und informiert die weiteren Teilnehmer auf dem Weg zum Quellknoten mit einer *link failure notification message*, die diese wiederum bis zum Quellknoten weitersenden. Bei Bedarf kann der Quellknoten nun einen neuen Wegfindungsprozess einleiten.
Eine zusätzliche Option, die dieses Protokoll mit sich bringt, um die Existenz von Nachbarknoten sicherzustellen, ist die Fähigkeit, ebenfalls Hallo-Nachrichten versenden zu können. Hallo-Nachrichten können dazu verwendet werden, die lokale Anbindung eines Knotens im Netzwerk aufrechtzuerhalten. Knoten hören in ihrem Umkreis, ob Datenpakete weitergeleitet wurden, um sicherzustellen, dass der nächste Hop noch stattgefunden hat. Falls solch eine Weiterleitung nicht vernommen wurde, kann der Knoten eine Vielzahl von Techniken, einschließlich der Hallo-Nachrichten, anwenden, um festzustellen, ob der nächste Hop in Kommunikationsreichweite ist. Die Hallo-Nachrichten sind nun in der Lage andere Knoten aufzulisten, von denen ein Mobilgerät gehört hat, um sich dadurch ein größeres Wissen über die Teilnehmer im Netzwerkverbund anzueignen.

2.2.2 Dynamic Source Routing

Das Dynamic Source Routing Protocol (DSR) ist ein on-demand Routing Protokoll, dass auf dem Konzept der Quelleninitiierung basiert. Die mobilen Knoten müssen Speicher anlegen, um Weginformation zum Zielknoten zu unterhalten. Die Einträge in den Speichern werden ständig aktualisiert, sobald neue Wege bekannt werden. Das Protokoll besteht aus zwei Hauptphasen: der Routensuche und der Routenerhaltung. Falls ein Knoten noch keinen bekannnten Weg zu seinem erwünschten Ziel kennt, sendet er ein *route request packet* aus. Dieses Paket enthält die Adresse der Quelle, die Adresse des Ziels und eine eindeutige Identifikationsnummer. Jeder Knoten, der das Paket erhält, überprüft, ob er einen Weg zum erwünschten Ziel kennt. Falls nicht, hängt er seine eigene Adresse an das Paket mit an und schickt es an seine Nachbarn weiter. Um die Anzahl der Pakete zu reduzieren, verschickt ein Knoten nur dann die Anfrage weiter, falls seine eigene Adresse noch nicht enthalten ist.
Ein Wegeantwortpaket (route reply) entsteht, wenn das Ziel selbst oder ein Knoten, der einen gültigen Weg zum Ziel kennt, erreicht wird und enthält die bisher vollzogenen Schritte zum Ziel. Vergleiche Abbildung 7a. Falls der Zielknoten selbst die Wegeantwort generiert, listet er alle auf dem Weg zu ihm passierten Knoten in dem Antwortpaket auf. Sendet ein anderer

Knoten die Antwort, fügt er die zu ihm benötigten Schritte sowie die ihm bekannte Schritte zum Ziel hinzu. Dieses Antwortpaket kann entweder über eine im eigenen Cache gespeicherte Route zurückgeschickt werden oder falls im Cache keine Informationen darüber vorliegen, schickt er es mit Hilfe, der im Anfragepaket gespeicherten Weginformationen, weiter. Dieses Szenario ist nur möglich, wenn symmetrische Links vom Protokoll unterstützt werden. Ansonsten muss der Sendeknoten eine neue Weganfrage zum Quellknoten generieren und das Antwortpaket mit dem Huckepackverfahren (piggybacking) anhängen. Abbildung 7b verdeutlicht den Rückweg über den gleichen Hinweg. Die Routenerhaltung wird über Fehlermeldun-

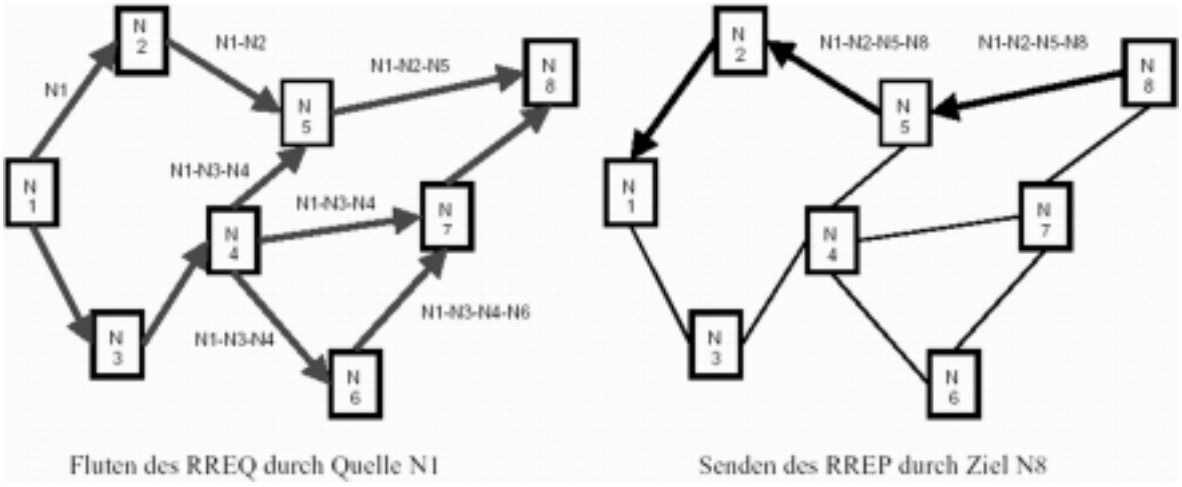

Abbildung 7: DSR Routenerforschung

gen und Bestätigungen überwacht. Fehlermeldungspakete werden an dem Knoten generiert an dessen Datenübertragungsschicht fatale Übermittlungsprobleme aufgetreten sind. Bei Erhalt des Fehlermeldungspakets wird der fehlerhafte Schritt zu dem Knoten aus dem Cache gelöscht und alle Wege, die diesen Schritt enthalten, werden an diesem Punkt abgetrennt. Zusätzlich zu Fehlermeldungspaketen werden Bestätigungen zur Verifizierung eines gültigen Wegesystems verwendet. Diese Bestätigungen enthalten auch passive Bestätigungen, die dadurch entstehen, dass mobile Knoten die Weiterleitung des Pakets durch den Nachbarknoten mitbekommen

2.2.3 Temporally Ordered Routing Algorithm

Das TORA Protokoll ist ein hoch anpassungsfähiger, schleifenfreier Routingalgorithmus, der auf dem Konzept der Linkumkehrung basiert. TORA ist besonders gut geeignet in einer hoch dynamischen, mobilen Netzwerkumgebung zu arbeiten. Er ist quelleninitiiert und bietet viele Routen für alle möglichen Quell-Ziel-Paare an. Das Schlüsselkonzept von TORA liegt darin, Kontrollnachrichten über eine sehr kleine Gruppe von Knoten zu lokalisieren, in derer näheren Umgebung sich topologische Veränderungen erreignen. Um das zu bewerkstelligen, müssen die Knoten Routinginformationen über ihre Nachbarknoten (über nur eine Stufe) verwalten. Das Protokoll stellt drei Basisfunktionen bereit:

- Wegeherstellung
- Wegeinstandhaltung
- Wegeauslöschung

Während der ersten beiden Phasen nutzen Knoten eine „hohe" Metrik, um einen direkten azyklischen Graphen (DAG: directed acyclic graph) auf dem Weg zum Ziel aufzubauen. Nach

diesem Schema bekommen Links eine Richtung zugewiesen (upstream oder downstream), die auf einer relativ hohen Metrik an Nachbarschaftsknoten basiert. Siehe Abbildung 8a. Der Prozess, diesen DAG aufzubauen, ähnelt sehr dem Nachfrage/Antwort-Prozess des Lightweight Mobile Routing(LMR)-Modells. Wenn sich Knoten bewegen, bricht der DAG-Weg zusammen und eine Weginstandhaltung wird notwendig, um DAG wieder zum selben Ziel aufzubauen. Wie in Abbildung 8b zu sehen ist, generiert ein Knoten nach der fehlgeschlagenen letzten Downstream-Verbindung, eine neue Referenzschicht. Diese resultiert in der Ausbreitung der Referenzschicht der Nachbarknoten, um eine effektiv strukturierte Reaktion auf dieses Fehlverhalten abzugleichen. Die Verbindungen werden in Anpassung an die neue Referenzschicht umgekehrt, um die Veränderungen umzusetzten. Falls ein Knoten keine Downstream-Verbindung mehr unterhält, hat dies denselben Effekt, wie wenn man eine oder mehrere Verbindungen umdreht.
Die Zeiteinteilung ist für TORA ein sehr wichtiger Bestandteil, weil die „hohe" Metrik von der

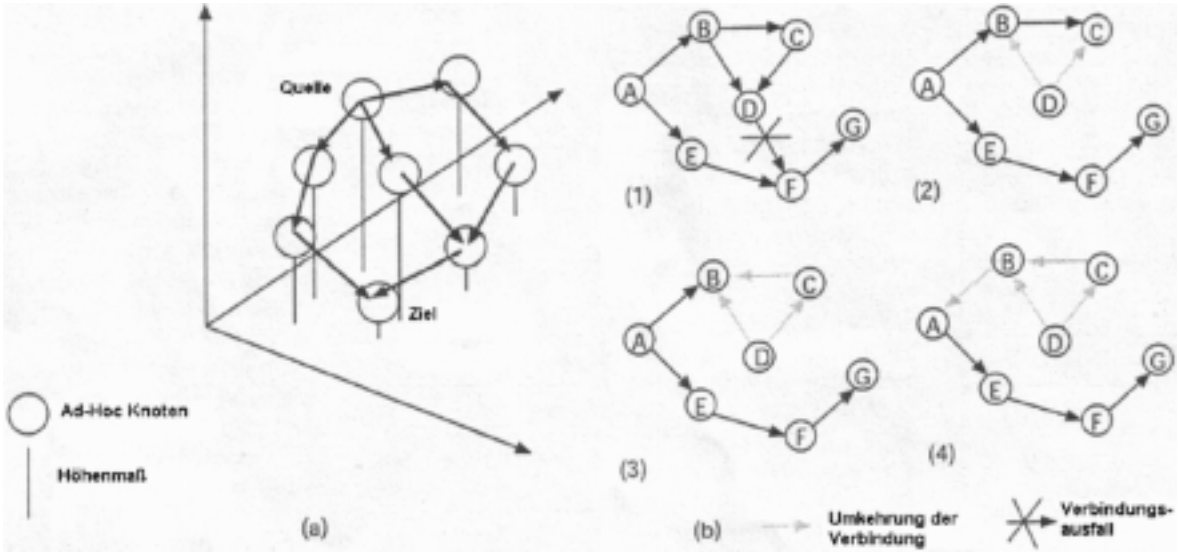

Abbildung 8: TORA: a) Wegeherstellung und b) Wegeinstandhaltung

logischen Zeit eines Verbindungsfehlers abhängt. TORA setzt deshalb voraus, dass alle Knoten synchronisierte Uhren verwenden (dies wird dadurch erreicht, dass externe Zeitsysteme, wie das GPS, benutzt werden). Die Metrik von TORA umfasst fünf Elemente, namens:

- Logische Zeit eines Verbindungsfehlers
- Die eindeutige ID des Knotens, der die neue Referenzschicht definiert
- Reflektion Indicator Bit
- Propagation Ordering Parameter
- Die eindeutige ID des Knotens

Die ersten drei Elemente repräsentieren die Referenzschicht. Eine neue Referenzschicht wird jedesmal dann definiert, wenn ein Knoten seine letzte Downstream-Verbindung aufgrund eines Verbindungsfehlers verliert. Die Wegeauslöschungsphase von TORA umfasst im wesentlichen ein Clear-Paket, das durch das Netzwerk gesendet wird, um ungültige Wege zu löschen.
Im TORA besteht hohes Potential für die Entstehung von Schwingungen, speziell dann, wenn mehrere Gruppen von koordinierenden Knoten zur gleichen Zeit Partitionen endecken, Wege auslöschen und neue Routen aufbauen, die aufeinander basieren. Da TORA Koordination

zwischen Knoten betreibt, ist sein Instabilitätsproblem dem des „count-to-infinty-Problem" im DVR-Protokoll sehr ähnlich, außer dass solche Schwingungen zeitlich begrenzt sind und sich die Wegekonvergenz einstellen wird.

2.2.4 Eigenschaftsvergleich der vorgestellen On-Demand Protokolle

Gravierende Unterschiede zwischen den Protokollen kann man anhand der Routinginformationen im Overhead feststellen. Der Overhead von DSR ist potentiell größer als der von AODV, da jedes DSR Paket die gesamten Routinginformationen enthält, bei AODV ist lediglich die Zieladresse im Overhead enthalten. Genauso muss auch bei DSR jeder Knoten in den Wegeantworten enthalten sein, bei AODV wieder nur die IP-Adresse und Sequenznummer des Zielknotens. Auch der Speicheroverhead ist bei DSR leicht größer, da die gesamten Weginformationen in ihm enthalten sind und AODV nur die Weginformation des nächsten Schrittes speichert. Ein weiterer Vorteil von AODV ist die Unterstützung von Multicastanwendungen. Kein anderes in dieser Ausarbeitung vorgestellte Protokoll unterstützt diese Fähigkeit. AODV kann nur auf symmetrischen Links basieren, wohingegen DSR auch mit asymmetrischen Verbindungen zurechtkommt. Der DSR Algorithmus eignet sich für Netzwerke, in denen sich die mobilen Geräte mit moderater Geschwindigkeit in Bezug auf die Verzögerung der Paketvermittlung bewegen. DSR kommt im Vergleich zu anderen Protokollen mit geringerer Bandbreite und niedrigerem Energieverbrauch aus, da es auf periodische Routingnachrichten verzichtet. Wenn keine Änderungen im Netzwerk stattfinden, sind somit auch keine Routingaktualisierungen nötig. Weiterhin hält ein Knoten mehrere mögliche Wege zum Ziel in seinem Cache bereit, so dass beim Ausfall eines Knotens ein anderer gültiger Weg gewählt werden kann. Dadurch ist DSR im Vergleich zu anderen On-Demand Protokollen zeit- und ressourcensparender. Jedoch ist DSR nur für kleinere Netzwerke geeignet, da es in seiner Skalierbarkeit begrenzt ist.
TORA ist ein „verbindungsumkehrender" Algorithmus, der besonders für Netzwerke mit einer hohen Knotendichte geeignet ist. Ein Teil der Erneuerung in TORA ist sicherlich die Erzeugung der oben beschriebenen DAGs, die zum Wegeaufbau beitragen. Ein entscheidender Vorteil des TORA Algorithmus ist unter anderem die Unterstützung von vielen Routen. TORA und DSR sind die einzigen hier vorgestellten Routingprotokolle, die mehrere Möglichkeiten für die Wegewahl eines einzigen Quellen/Zielpaares ermöglichen. Ein anderer Vorteil TORAs ist die Multicastunterstützung, obwohl er, anders als AODV, diese Unterstützung nicht in sein Basisverfahren einbezieht, sondern mit dem zugrunde liegenden Protokoll für LAM (Lightweight Adaptive Multicast Algorithm) funktioniert und zusammen mit diesem Protokoll die Multicastfähigkeit verwirklicht. TORAs Vertrauen in synchronisierte Uhren begrenzt seine Anwendungsfähigkeit in starkem Maße. Denn sobald ein Knoten nicht über GPS oder eine andere externe Zeitquelle verfügt, kann er diesen Algorithmus nicht angewenden. Außerdem stellt der Algorithmus seine Arbeit auch dann ein, wenn die externen Zeitquellen Fehler aufweisen. Weiterhin passiert der Routenaufbau nicht ganz so schnell wie bei den anderen Algorithmen, da das Potential für Schwingungen in dieser Periode stark zunimmt. Während auf die Festlegung der neuen Wege gewartet wird, kann dieses Oszillationspotential dazu führen, dass es dabei zu größeren Verzögerungen kommen kann. Aus Tabelle 2 sind die technischen und anwendungsorientierten Eigenschaften der Protokolle zu entnehmen.

2.3 Table-Driven im Vergleich zu On-Demand Routing Protokollen

Der Ansatz tabellenbasierter Routingalgorithmen ähnelt sehr der verbindungslosen Herangehensweise von Paketvermittlungen, ohne Bezug auf ein „Wann" und „Wie oft" die Wege benötigt werden. Er basiert auf zugrunde liegenden Routingtabellen, die mittels Update-Mechanismen auf dem neuesten Stand gehalten werden. Dafür ist eine permanente Tabel-

Parameter	AODV	DSR	TORA
Zeitkomplexität (Initialisierung)	O(2d)	O(2d)	O(2d)
Zeitkomplexität (nach Fehler)	O(2d)	O(2d)	O(2d)
Kommunikationskomplexität (Initialisierung)	O(2N)	O(2N)	O(2N)
Kommunikationskomplexität (nach Fehler)	O(2N)	O(2N)	O(2x)
Routing Topologie	flach	flach	flach
Schleifenfreiheit	ja	ja	ja
Multicastfähigkeit	ja	nein	nein
Benötigung von "Beacons"	nein	nein	nein
Möglichkeit vieler Routen	nein	ja	ja
Wegeinstandhaltung in	Routingtabelle	Routingspeicher	Routingtabelle
Benutzung von Verfallszeiten in Routingspeichern und	ja	nein	nein
Routen Umgestaltungsmethoden	Auslöschen von Routen; Quelle benachrichtigen	Auslöschen von Routen; Quelle benachrichtigen	Verbindungsumkehrung; Routenreparation
Routingmaß	neueste und kürzester Pfad	kürzester Pfad	kürzester Pfad
Legende:			
N=	Anzahl der Knoten im Netzwerk		
d=	Netzwerkdurchmesser		
h=	Höhe des Routingbaums		
x=	Anzahl Knoten, die von einer Topologieänderung betroffen sind		

Tabelle 2: Eigenschaftsvergleich der vorgestellen On-Demand Protokolle

lenaktualisierung nötig. Im Gegensatz dazu stehen die On-Demand Routing Protokolle, die für jede benötigte Verbindung einen Weg erforschen müssen. Der Quellknoten muss so lange mit der Verschickung seiner Daten zum Zielknoten warten, bis er eine Antwort auf seine Weganfrage erhalten hat. Andererseits stehen bei tabellenbasierten Protokollen sämtliche Routinginformationen zu allen im Netz befindlichen Knoten ständig zur Verfügung unabhängig davon, ob sie gebraucht werden oder nicht. Diese Eigenschaft ist zwar für Datagrammvermittlungen geeignet, sorgt aber für eine große Verkehrsbelastung und somit auch unnötiger Überbelastungen im Netz, sowie hohen Energieverbrauch. Da sowohl Bandbreiten als auch Energiekapazitäten in mobilen Computernetzen beschränkt sind, ist der Einsatz dieser Protokolle in mobilen drahtlosen Netzen problematisch. Tabelle 3 illustriert einige grundlegende Unterschiede zwischen Table-Driven und On-Demand basierten Routingprotokollen.

Eine andere Überlegung tut sich an dieser Stelle auf, ob flache oder hierarchische Netzwerkstrukturen eingesetzt werden sollen. Alle bisher besprochenen Protokolle in dieser Ausarbeitung, mit Ausnahme von CGSR haben sich für eine einstufige Netzwerkstruktur entschieden. Während eine einstufige Netzwerkstruktur unkomplizierter und dadurch einfacher zu handhaben ist, bestehen Zweifel bezüglich ihrer Skalierbarkeit. [2]

[2] Vgl. [Ba+ot97]

Parameter	On-Demand	Tabellenbasiert
Verfügbarkeit von Routinginformationen	Verfügbar, wenn benötigt	Immer verfügbar, unabhängig vom Bedarf
Routing Topologie	flach	Meistens flach, außer bei CGSR
periodische Routenupdates	nicht nötig	nötig
Bewältigung der Mobilität	Benutzt örtlich begrenzte Routenerkundung	Information anderer Knoten, um konsistente Routentabellen zu erhalten
erzeugter Signalisierungsverkehr	wächst mit steigender Mobilität der aktiven Routen	größer als beim On-Demand Routing
Quality of Service Unterstützung	Wenige unterstützen QoS, obwohl die meisten Protokolle kürzeste Pfade unterstützen	Hauptsächlich kürzester Pfad, wie die QoS-Metrik

Tabelle 3: Vergleich der Table-Driven und On-Demand Protokolle

3 Multicastprotokolle

In Datenkommunikationsnetzwerken bedeutet Multicasting, dass die Übertragung von Nachrichten an mehrere Empfänger gleichzeitig stattfindet. Multicast ist also ein Übermittlungsverfahren, bei dem eine Quelle an mehrere Ziele innerhalb einer Gruppe Daten versendet, was auf den ersten Blick dem Broadcasting ähnelt. Der Unterschied ist jedoch, dass beim Multicast eine spezifische Empfängerliste vorliegt und die entsprechenden Empfänger der Multicastgruppe gezielt kontaktiert werden. Die entscheidende Idee des Multicast ist also die nur einmalige Verschickung einer Datensendung an den Multicastrouter, der dann die Daten vervielfältigt und via Unicastprotokolle an die Gruppenmitglieder sendet. Dadurch werden Ressorcen auf dem Weg zwischen Quelle und Router eingespart und auf eine mehrfache Versendung der Daten verzichtet, die beim Unicasting von Nöten wäre.

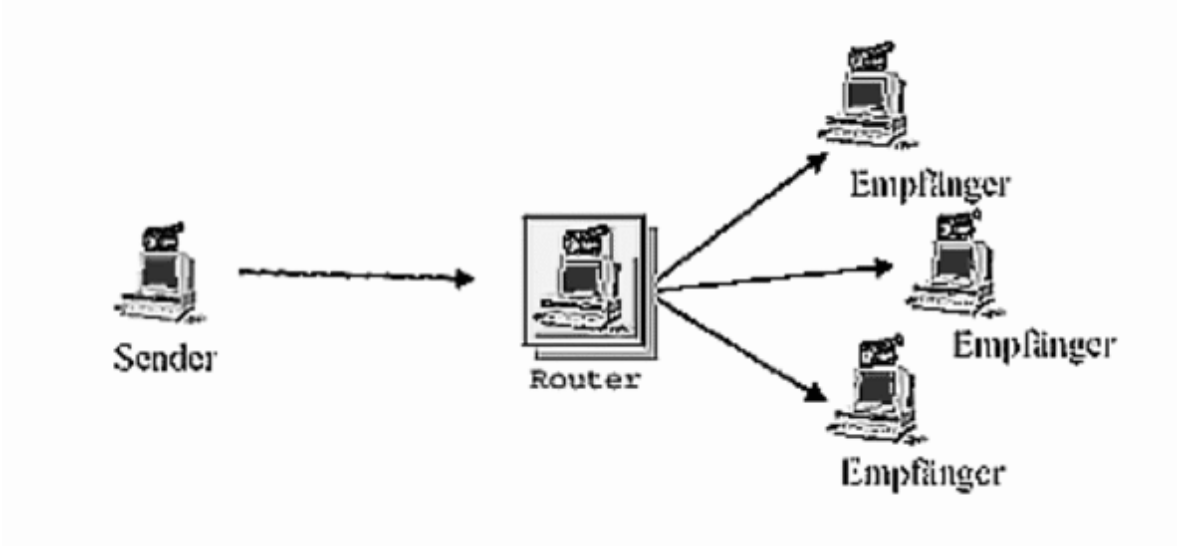

Abbildung 9: Vorgehensweise des Multicast

3.1 Ad-Hoc Multicast Routing Protokoll

Das Ad-Hoc Multicast Routing Protokoll (AMRoute) baut Verbindungen zwischen den Gruppenmitgliedern des Multicastnetzwerkes mittels einer Baumstruktur auf. Hierbei werden bidirektionale, gemeinsam genutze Multicastbäume durch Unicasttunneling gebildet. Jede Gruppe besitzt mindestens einen logischen Kern, der für die Mitglieder-und Baumwartung zuständig ist. Bei Initiierung des Netzwerkes deklariert sich jeder Knoten als sein eigener Kern, mit einer Gruppengröße von eins.
Jeder Kern sendet in periodischen Abständen eine „Join-Req", also eine Zusammenführungsanfragen, aus, um andere unzusammenhängende Netzsegmente zu finden. Vergleiche hierzu Abbildung 10. Wenn ein Knoten zwar aus der selben Gruppe, aber eines anderen Maschen-

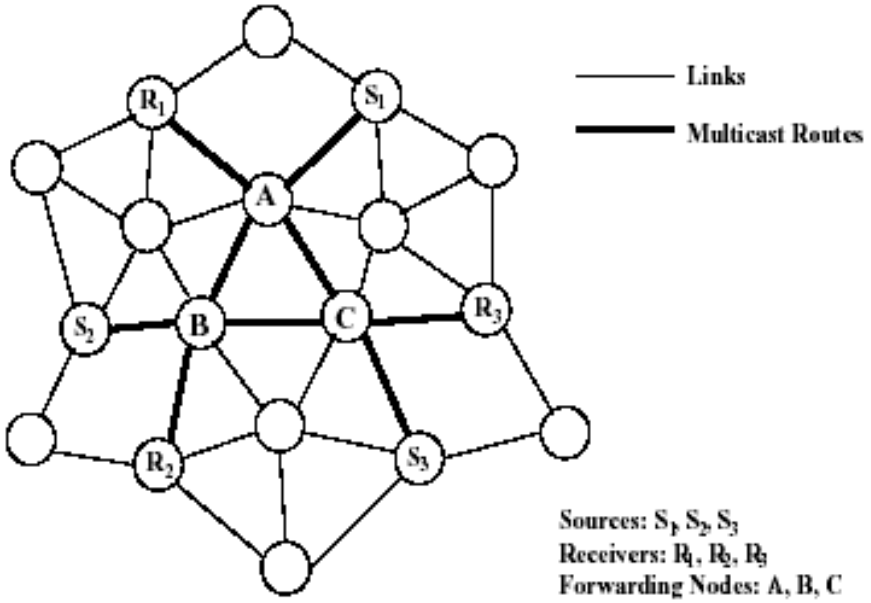

Abbildung 10: Maschenartige Netzwerkstruktur mit Multicastroutern

segmentes, eine solche Anfrage empfängt, antwortet er mit einer „Join-Ack", und merkt sich diesen Knoten als Maschennachbarn. Der Maschennachbar empfängt wiederum dieses „Join-Ack" und merkt sich den Absender als Maschennachbarn. Nach der Maschenbildung sendet jeder Kern in regelmäßigen Abständen „Tree-creates"-Pakete an Maschennachbarn aus, um gemeinsam benutzbare Bäume aufzubauen. Wenn ein Knoten eine solche neuartigen Anfrage erhält, schickt er es umgehend an alle seine Maschennachbarn weiter. Falls eine solche Anfrage schon einmal empfangen, also bereits bearbeitet wurde, wird ein „Tree-create-Nak"-Paket zurück an den Absender geschickt. Ein ankommendes „Tree-create-Nak"-Paket bewirkt eine Markierung des Pfades als Maschenpfad und nicht als Baumpfad.
Knoten, die die Gruppe verlassen wollen, senden ein „Join-Nak"-Paket, und leiten keine Datenpakete mehr an die Gruppe weiter.
Die Hauptcharakteristik des AMRoute-Protokolles ist die Verwendung von virtuellen Maschenpfaden zum Aufbau eines Multicastbaumes. Solange also Wege zwischen Baummitgliedern mittels Maschenpfaden existieren, muss der Baum auch bei Netzwerktopologieveränderungen nicht angepaßt werden. Nichtmitglieder des Baumes sind nicht an der Verschickung von Datenpaketen beteiligt und müssen auch nicht das Multicast-Protokoll unterstützen. Dadurch müssen nur die Mitgliedsknoten des Baumes den Overhead für Speicher und Prozess tragen.
Die Grundlage von AMRoute kann jedes beliebige Unicast-Protokoll sein, das den Kontakt

unter den Mitgliedsknoten herstellt.
Ein Hauptnachteil dieses Protokolles tritt bei Mobilität der Knoten auf, weil dann zeitweilig Schleifenbildung auftritt und nichtoptimale Bäume gebildet werden.

3.2 On-Demand Multicast Routing Protocol

Das On-Demand Multicast Routing Protocol (ODMRP) erschafft ein Netz aus Knoten (die weiterleitende Gruppe), das Multicastpakete durch Fluten innerhalb der Masche weiterleitet. Dadurch wird eine Pfadredundanz hergestellt. ODMRP ist ein on-demand Protokoll, so dass es, wie bereits bekannt, keine permanten Routinginformationen unterhalten muss. Es benutzt für die Gruppeninstandhaltung eine „soft state" Vorgehensweise. Mitgliedsknoten werden bei Bedarf aktualisiert und müssen keine expliziten Nachrichten beim Verlassen der Masche abschicken.
Beim ODMRP werden Gruppenmitgliedschaften und Multicastwege erst auf Anfrage erstellt und aktualisiert. Ähnlich wie bei On-demand Unicast Protokollen beinhaltet dieses Multicast-Protokoll Anfrage- und Antwortphasen. Sobald Multicastquellen Daten absenden möchten, aber noch keine Routing- oder Mitgliedschaftsinformationen haben, fluten sie ein „Join Data"-Paket. Vgl. Abbildung 11. Wenn ein Knoten ein nicht-dupliziertes „Join Data"-Paket erhält,

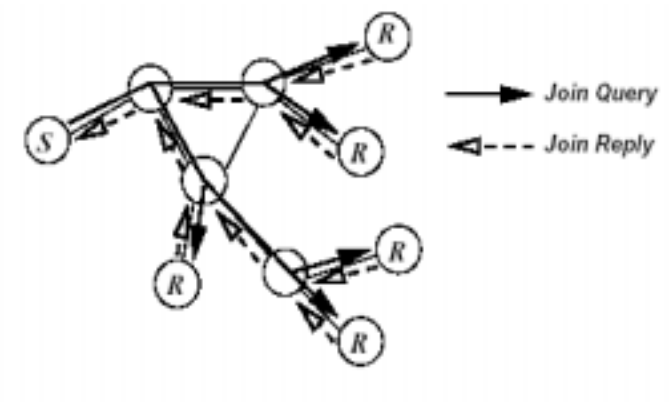

Abbildung 11: ODMRP: Einrichtung und Wartung der Struktur

speichert er die Upstreamknoten-ID (z.B. über Backward Learning) und schickt das Paket weiter. Falls das „Join Data"-Paket einen Multicastempfänger erreicht, erschafft dieser eine Jointabelle und sendet sie an seine Nachbarn. Sobald ein Knoten diese Jointabelle erhält, prüft er, ob in einer der nächsten Einträge seine Knoten-ID vorkommt. Ist das der Fall, realisiert der Knoten, dass er auf dem Pfad zur Quelle und so ein Bestandteil der Weiterleitungsgruppe ist. Er sendet wiederum seine eigene Jointabelle, die auf den zutreffenden Einträgen aufbaut, via Broadcast aus. Die Jointabelle wird auf diese Art und Weise von jedem weiterleitenden Gruppenmitglied solange verbreitet, bis sie die Multicastquelle über den kürzesten Pfad erreicht. Dieser Prozess konstruiert (oder aktualisiert) die Wege von der Quelle zu den Empfängern und bildet somit eine Masche von Knoten, die sogenannte „Weiterleitungsgruppe". Multicastsender frischen ihre Teilnehmerinformationen auf und aktualisieren die Wege, indem sie regelmässig „Join Data"-Pakete versenden. Siehe Abbildung 12. In Netzwerken, in denen GPS (Global Positioning System) verfügbar ist, kann ODMRP an Knotenbewegung angepasst werden, indem es Mobilitätsprognosen benutzt. Durch die Unterstützung von GPS für Lokalisierungs- und Mobilitätsinformationen kann die Wegerkundungszeit geschätzt werden und Empfänger können sich den Pfad auswählen, der die längste gültige Zeit erwarten lässt. Mit der Mobilitätsprognose-Methode können Quellen Wege, unter Beachtung von Wegausfällen, rekonstruieren. Auf diesem Wege wird das Protokoll in Hinblick auf die Mobilität

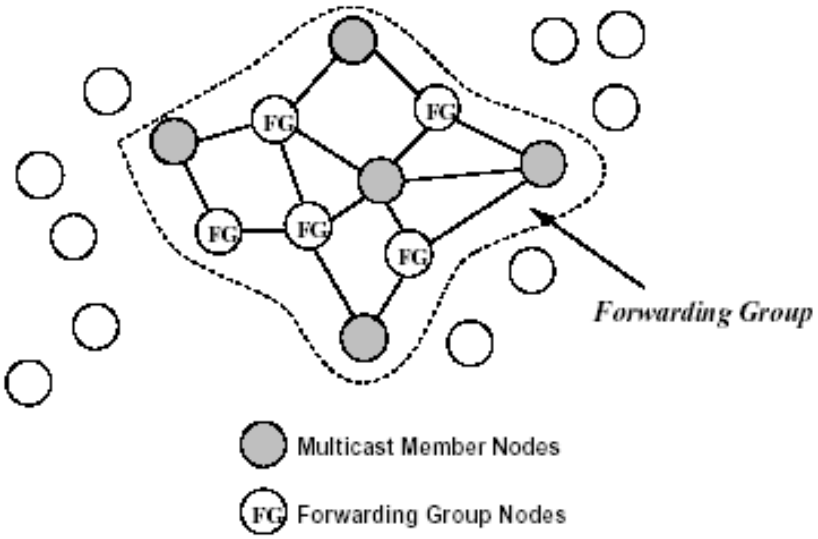

Abbildung 12: Das Prinzip der Weiterleitungsgruppe

immer stabiler. Der Nachteil vom Einsatz des GPS sind die hohen Kosten, sowie das zusätzliche Gewicht.
Die Datenübertragungsphase ist für beide Versionen identisch. Knoten leiten Daten nur dann weiter, wenn sie selbst Weiterleitungsknoten sind und die Pakete keine Duplikate sind. Wenn alle weiterleitenden Knoten die Daten weitergeben, können redundante Pfade (falls sie existieren) dazu beitragen, Datenpakete auszuliefern, falls der primäre Pfad aufgrund der Mobilität unterbrochen wird. Eine weitere einzigartige Eigenschaft des ODMRP ist seine Unicast-Fähigkeit. Es kann nicht nur neben einem Unicast-Routing-Protokoll einsatzfähig, sondern kann auch auf sehr effiziente Weise als solches fungieren. So brauchen Netzwerke, die mit ODMRP ausgestattet sind, kein separates Unicast-Protokoll.

3.2.1 Ad hoc Multicast Routing protocol utilizing Increasing id-numberS

Das Ad hoc Multicast Routing protocol utilizing Increasing id-numberS (AMRIS) errichtet ein gemeinsam genutztes Baumsystem für Multicastdatenweiterleitung. Jeder Knoten im Netzwerk bekommt eine Multicast-Session-ID-Nummer zugewiesen. Die Rangfolge der ID-Nummern wird dazu verwendet um den Fluss von Multicastdaten eine Richtung zu geben. Wie ODRMP braucht auch AMRIS keine separates Unicast-Routing-Protokoll. Am Anfang sendet ein spezieller Knoten, names SID, via Broadcast ein „New Session"-Paket aus. Dieses Paket beinhaltet auch die Multicast-Session-Member-ID (MSM-ID) von SID. Die Nachbarknoten, die darauf das Paket erhalten, berechnen ihre eigenen MSM-IDs, die größer als die im Paket bestehenden sind. Dadurch wachsen die MSM-IDs in zunehmenden Maße an. Die Knoten senden die „New Session"-Nachricht, mit der durch ihre eigene ersetzte MSM-ID, via Broadcast weiter. Jeder Knoten ist verpflichtet an seine Nachbarknoten sogenannte „Beacons" auszusenden. Die Beacon-Nachricht beinhaltet die Knoten-ID, die MSM-ID, den Mitgliedsstatus, sowie die registrierten Eltern- und Kinder-IDs, deren MSM-IDs und Partitions-ID. Ein Knoten kann an einer Multicast-Sitzung teilnehmen, indem er eine „Join-Req" absendet. Diese „Join-Req" wird an alle potentiellen Elternknoten mit einer kleineren MSM-ID als die eigene via Unicast gesendet. Der Knoten, der die „Join-Req" empfängt sendet daraufhin eine „Join-Ack"-Nachricht zurück, falls er schon ein Mitglied der Multicastsitzung ist. Andernfalls sendet er ein „Join-Req.Passive" zu seinen potentiellen Eltern. Falls ein Knoten es verfehlt hat

eine „Join-Ack" zu erhalten, oder eine „Join-Nak" erhält, nachdem er eine „Join-Req" gesendet hat, vollzieht er eine „Branch Reconstruction (BR)". Der BR-Prozess wird solange mit einer ausdehnenden Ringsuche ausgeführt, bis der Knoten an einer Multicastsitzung teilnehmen kann.
AMRIS entdeckt Verbindungsabbrüche mit einem „Beaconing"-Mechanismus. Falls keine Beacons in einem vordefinierten Zeitintervall gehört wurden, wird der Nachbarknoten als außerhalb der Sendereichweite eingestuft. Falls der frühere Nachbar ein Elternteil war, muss der Knoten in den Baum wieder eingegliedert werden. Dies geschieht durch ein ausgesendetes „Join-Req" an sein neues potentielles Elternteil. Falls die Eingliederung an der Sitzung misslingt, oder kein geeigneter Nachbar existiert, setzt er wiederum den BR-Prozess in Gang.
Die Datenweiterleitung wird durch die Knoten im Baum vollzogen. Es werden nur die Pakete von registrierten Eltern oder Kindern weitergeleitet. Wenn die Baumverbindung zusammenbricht, sind die Pakete solange verloren, bis der Baum wieder konfiguriert wurde.

3.3 Core-Assisted Mesh Protocol

Das Core-Assisted Mesh Protocol (CAMP) besteht aus einer gemeinsam genutzten Maschenstruktur. Alle Knoten im Netzwerk unterhalten Tabellen mit Mitglieds- und Routinginformationen, sowie Cachespeicher mit vorherigen Datenpaketinformationen und „unacknowledged" Mitgliedsanfragen. CAMP klassifiziert Knoten im Netz als Duplex- oder Simplexmitglieder oder aber als Nicht-Mitglieder. Duplex-Mitglieder sind vollwertige Mitglieder der Multicast-Masche, während Simplex-Mitglieder benutzt werden, um „one-way"-Verbindungen zwischen nur sendenden Knoten und der Multicast-Masche herzustellen. Kerne werden benötigt, um die Versendung von „Join-Requests" zu reduzieren.
CAMP besteht aus Maschenaufbau- und Erhaltungsprozeduren. Ein Knoten, der einer Masche beitreten möchte, befragt zuerst seine Mitgliedstabelle, um herauszufinden, ob schon evtl. ein Nachbarknoten Mitglied der gewünschten Masche ist. Falls dies der Fall ist, kündigt der Knoten seine Mitgliedschaft per „Camp update" an. Andernfalls schickt der Knoten ein „Join-Request"-Paket an einen der Multicastgruppen-Kerne, oder aber versucht einen Mitgliedsrouter über eine Ringsuche zu erreichen. Jedes Duplex-Mitglied der Masche kann mit einem „Join Ack"-Paket antworten, welches zur Quelle der Anfrage geschickt wird.
In regelmäßigen Abständen überprüft der antwortende Knoten seinen Datenpaketspeicher, um festzustellen, ob er Datenpakete von den Knoten erhalten hat, die auf dem kürzesten Rückweg zur Quelle liegen. Falls nicht, sendet er entweder eine „Heartbeat"- oder aber „Push Join"- Nachricht über den kürzesten Rückweg zur Quelle. Dieser Prozess garantiert, dass das Netz alle kürzesten Wege vom Sender zum Empfänger kennt. Die Knoten wählen in regelmäßigen Abständen ihre „Anker" und frischen diese auf, indem sie ihre Aktualisierung an das Multicastnetz via Broadcast aussenden. Diese „Anker" sind Nachbarknoten, die benötigt werden, um jegliche nicht-duplizierten Datenpakete, die sie erhalten haben, via Broadcast weiterzuleiten. Einem Knoten ist es erlaubt mit Nachbarknoten das Ankerverhältnis abzubrechen, die nicht in regelmäßigen Abständen ihre Verbindungen auffrischen. Er kann dann das Multicastnetz verlassen, wenn er nicht mehr an einer Multicast-Sitzung interessiert ist und ist dann nicht mehr gezwungen, als „Anker" für seine Nachbarknoten zu bestehen. CAMP ist auf das darunterliegende Unicast-Protokoll angewiesen, das korrekte Entfernungen zu allen Zielen in endlicher Zeit garantiert. Routing Protokolle, die auf dem Bellman-Ford-Algorithmus basieren können nicht zusammen mit CAMP benutzt werden. Um mit on-demand-Protokollen arbeiten zu können, muß CAMP zunächst erweitert werden.

3.4 Überblick der Multicast-Protokolle

Tabelle 4 fasst die Charakteristiken und Eigenschaften der Protokolle zusammen. Zu beachten ist, dass ODRMP periodische Nachrichten (Join-Data) nur dann benötigt, wenn Quellen Datenpakete verschicken wollen.

Protokolle	AMRoute	ODMRP	AMRIS	CAMP
Gestaltung	Baum	Masche	Baum	Masche
Schleifenfreiheit	nein	ja	ja	ja
Abhängigkeit von Unicastprotokollen	ja	nein	nein	ja
periodische Benachrichtigung	ja	ja	ja	ja
Kontrollpaket fluten	ja	ja	ja	nein

Tabelle 4: Überblick der Multicast-Protokolle

4 Ausblick

Aus Sicht vieler Wissenschaftler, die sich mit Protokollen und Algorithmen für Ad Hoc Netzwerke beschäftigen, sollte Multicast-Routing in mobilen Netzwerken auf Basis einer Unicast-Routing-Infrastruktur aufbauen. Diese Meinung basiert auf dem Glauben, dass sich das mobile Ad Hoc Netzwerk-Routing auf ähnliche Weise, wie das herkömmliche Internet-Routing entwickeln wird. Das ist der Grund, dass sich die meisten Forschungen auf Lösungen von Problemen des Unicast-Routing mit mobilen Endpunkten fokussiert haben. Viele mobile Netzwerke sind wegen ihrer Broadcastfähigkeit besser für Multicast- als Unicast-Routing geeignet und deshalb wäre der Ansatz, Multicast-Routing Probleme separat zu betrachten, effektiver.

Als Ziel weiterer Forschungen sollten Multicast-Routing und Paketweiterleitungsprotokolle für Ad Hoc Netzwerke die folgenden Eigenschaften sicherstellen:

- Robustheit gegen Effizienz: Viele Multicast-Routing Ansätze sind auf Router angewiesen, die über die Multicastgruppenmitglieder die Übersicht behalten müssen. Verbunden mit dem hohen Volumen an Routinginformationen, die ständig ausgetauscht werden, und der langsamen Konvergenz, machen traditionelle Multicast-Entwicklungen unhaltbar in hochdynamischen Ad Hoc Netzwerken. Weiterhin erschweren die niedrigen Energie- und Speicherkapazitäten der mobilen Geräte ein zufriedenstellendes Ergebnis. Deshalb müssen Techniken entwickelt werden, die sowohl die Geschwindigkeit als auch die Robustheit vorantreiben.
- Aktive Anpassungsfähigkeit: Mobile Geräte bewegen sich frei sowohl zwischen Ad Hoc-, Wireless LAN und verkabelten Netzwerken. Um sich schnell an Infrastrukturänderungen anpassen zu können, kann ein aktiver Netzwerkansatz eingesetzt werden: Geräte sollten sich in Echtzeit an neue Umgebungen anpassen können, indem sie die passenden Multicast-Mechanismen herunterladen.
- Unbegrenzte Mobilität: Einige bestehende Multicast-Lösungen gehen davon aus, dass es Perioden mit viel Bewegung aber auch Ruhephasen gibt. Andere geben vor, dass nur bestimmte Richtungen, Geschwindigkeiten oder die Anzahl der sich gleichzeitig bewegenden Knoten erlaubt sind. Im Kontrast dazu wird jedoch eine universelle, unbeschränkte Mobilität aller Netzwerkkomponenten gefordert.

- Integriertes Multicasting: Multicast-Lösungen unterscheiden sich in den meisten Fällen entscheidend von denen feststehender Netzwerke (eines der Hauptgründe ist der Unterschied in den Übertragungsraten). Um nahtlose und integrierte Multicast-Dienste anbieten zu können, müssen neue Mechanismen sowohl für feste als auch drahtlose Multicast-Lösungen entwickelt werden.

Die Forschung hat seit dem ersten drahtlosen Ad Hoc Netzwerk, das durch ARPA 1973 initiiert wurde und mit bis zu 138 Knoten über paketorientierte Vermittlung die einfachste Form eines Funknetzes darstellte, so große Fortschritte gemacht, dass Anwendungen, die vor kurzem nur auf Festnetzen möglich waren, heute und in naher Zukunft auf fast allen mobilen Geräten allgegenwärtig sein werden.

Literatur

[Ba+ot97] D. Baker und andere. *Flat vs. Hierarchical Network Control Architecture'*. 1997.

[Broc98] Johnson Hu Jetcheva Broch, Maltz. A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols. Technischer Bericht, Carnegie Mellon University, Pittsburgh, 1998.

[Cors99] Cirincione Corson, Macker. Internet-Based Mobile Ad Hoc Networking. Technischer Bericht, University of Maryland, 1999.

[Kati98] Gene Tsudik Katia Obraczka. Multicast Routing Issues in Ad Hoc Networks. Technischer Bericht, USC Information Sciences Institute, 1998.

[Lee01] Gerla Lee, Su. On-Demand Multicast Routing Protocol in Multihop Wireless Mobile Networks. Technischer Bericht, University of California, Los Angeles, 2001.

[Roye99] Toh Royer. A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks. Technischer Bericht, University of California, Santa Barbara, 1999.

[Schi00] Jochen Schiller. *Mobilkommunikation*. Addison-Wesley. 2000.

[Sung00] Julian Hsu Mario Gerla Rajive Bagrodia Sung-Ju Lee, William Su. A Performance Comparison Study of Ad Hoc Wireless Multicast Protocols. Technischer Bericht, University of California, Los Angeles, 2000.

[Tann96] Tannenbaum. *Computer Networks*. Prentice Hall. 1996.

[Zitt01] Prof. Dr. Martina Zitterbart. Next Generation Internet, 2001.

Abbildungsverzeichnis

1 Darstellung eines gleichberechtigten, einstufigen Ad-Hoc Netzwerks 61

2 Darstellung eines hierarchichen, zweistufigen Ad-Hoc Netzwerks 61

3 Vorgehensweise von Unicast 62

4 Kategorisierung von Unicastprotokollen 62

5 CGSR: Routing von Knoten 1 zu Knoten 8 64

6 AODV Routenerforschung 67

7 DSR Routenerforschung 68

8 TORA: a) Wegeherstellung und b) Wegeinstandhaltung 69

9 Vorgehensweise des Multicast 72

10 Maschenartige Netzwerkstruktur mit Multicastroutern 73

11 ODMRP: Einrichtung und Wartung der Struktur 74

12 Das Prinzip der Weiterleitungsgruppe 75

Tabellenverzeichnis

1 Eigenschaftsvergleich der tabellenbasierten Routingprotokolle 66
2 Eigenschaftsvergleich der vorgestellen On-Demand Protokolle 71
3 Vergleich der Table-Driven und On-Demand Protokolle 72
4 Überblick der Multicast-Protokolle 77

IP-basierende UMTS System Architektur

Steffen Kamuf

Kurzfassung

In dieser Ausarbeitung wird das Zusammenspiel des Mobilfunkstandards der dritten Generation (UMTS) und dem bewährten Internetprotokoll (IP) erläutert. Ausgehend von der historischen Entwicklung von UMTS stellen wir zunächst die allgemeine UMTS System Architektur vor. Diese Architektur stellt die Basis für alle UMTS Dienste dar. Ein wichtiges Merkmal der UMTS Architektur ist, dass UMTS keine vorgefertigen Services anbietet, sondern lediglich die Voraussetzungen für die unterschiedlichsten Dienste schafft. Nach der ausführlichen Diskussion der UMTS System Architektur gehen wir auf die möglichen Probleme bei der Entwicklung eines all-IP Basissystems ein. Im Anschluss daran berichten wir über die beiden aktuellen Projekte zu diesem Thema. Zum Abschluss wird ein mögliches Szenario für die Realisierung einer all-IP Umgebung vorgestellt.

1 Einleitung

Diese Ausarbeitung gliedert sich in folgende Teile: Zunächst wird die Entwicklung von UMTS als universeller Standard erläutert. Danach folgt eine Beschreibung der UMTS System Architektur, dabei gehen wir auch auf die beiden grundlegenden Netzwerke von UMTS ein. Daran schließt sich eine Diskussion über die Verwendung von UMTS in Verbindung mit IP an. Als letzter Teil folgt abschließend die Beschreibung eines möglichen Szenarios für die Verwendung von UMTS in einem all-IP Netzwerk.

2 Entwicklung von UMTS

UMTS steht für Universal Mobile Telecommunications Services. Das bedeutet, dass UMTS dafür entwickelt wurde, globale Telekommunikationsdienste anzubieten. Diese Dienste sollen weltweit unter einem einheitlichen Standard gemäß der Spezifikation ITU 2000 verfügbar sein.

Das Ziel von UMTS ist es, ein breites Spektrum von Diensten anzubieten, die bei der Entwicklung dieses Standards nicht festgelegt waren [OjPr98]. Damit soll der UMTS Standard sehr offen und flexibel für zukünftige Anwendungsformen sein. Um dieses Ziel erreichen zu können, definiert UMTS keine festgelegten Dienste sondern lediglich eine Serviceplattform, darauf aufbauend können die unterschiedlichsten Dienste angeboten werden.

Darüber hinaus ist es eine wichtige Anforderung an UMTS, dass die Dienste unter verschiedenen Verbindungsbedingungen angeboten werden können. Dabei soll das System so flexibel sein, dass es hohe Bandbreiten und asymmetrische Datenraten handhaben kann. Multimediaanwendungen und Datenübertragung in Echtzeit soll ebenfalls möglich sein.

Das UMTS Netzwerk wird es ermöglichen, dass Benutzer mit unterschiedlicher Ausstattung und Anforderungen nebeneinander existieren können. Die Resourcenzuteilung soll dabei für alle Benutzer je nach Bedarf auf eine faire Art erfolgen.

Bei dieser Zuteilung müssen allerdings jederzeit die zuvor abgegebenen Zusicherungen eingehalten werden und die Zuteilung muss auch für Echtzeitdaten, wie Telefongespräche, fair sein. Fair wird dabei im Sinne des Quality of Service definiert.

Das UMTS als mobiles Netz der dritten Generation soll zum Teil abwärtskompatibel zu den Mobilfunknetzen der zweiten Generation sein.

UMTS verwendet als Übertragungsverfahren das sogenannte wideband code division multiple access (WCDMA). Dieses Verfahren wurde urspünglich von der Association for Radio Industry and Business (ARIB), einer japanischen Standardisierungsorganisation in die Diskussion eingebracht.

Die Wahl der weltweit einheitlichen Übertragungsfrequenzen erfolgte im Jahr 1998. In den darauffolgenden Jahren wurden die Frequenzen in den einzelnen Ländern für zum Teil sehr hohe Summen versteigert.

Die erste Festlegung auf verbindliche Standards und Spezifizierungen erfolgte 1997. Damit stand einer engültigen Festlegung von UMTS bis Ende 2000 und einer geplanten kommerziellen Nutzung ab 2002 nichts mehr im Wege.

Eine wichtige Besonderheit von UMTS ist, dass bei UMTS keine festgelegten Services definiert werden, sondern dass lediglich Servicefunktionen angeboten werden. Dadurch ist das komplette System flexibler und kann sich besser den Bedürfnissen des Marktes anpassen.

3 Einführung in UMTS

In diesem Abschnitt werden die Besonderheiten von UMTS erläutert. Zunächst wollen wir auf die wichtigsten Unterschiede zwischen UMTS und GSM eingehen, bevor wir einige wichtige Architekturmerkmale von UMTS näher betrachten.

3.1 Unterschied zwischen UMTS und GSM

Im Unterschied zu GSM soll UMTS ein weltweiter Standard werden und somit wirklich global sein. Abgesehen von dieser Änderung gibt es weitere wesentliche Verbesserungen bei UMTS im Gegensatz zu GSM:

- Breitbandzugang: Höhere Bitraten öffnen Möglichkeiten für mobile Multimediaanwendungen. GSM erlaubt maximale Übertragungsraten von unter 10 Kilobit pro Sekunde während UMTS in seiner letzten Ausbaustufe Geschwindigkeiten bis zu 2 Megabit pro Sekunde ermöglichen soll.
- Mobile-fixed-Internet Konvergenz: Erlaubt dem Benutzer einen einheitlichen Zugang zum Internet unabhängig vom benutzen Medium.
- Flexible Systemarchitektur: Durch die Standardisierung von Servicefunktionen anstelle von Servicen kann sich UMTS flexibler an sich ändernde Bedürfnisse der Benutzer anpassen.

Diese Änderungen lassen sich unter dem Versuch zusammenfassen, ein möglichst flexibles Netzwerk zu entwickeln, dass nicht nur heutigen Anforderungen genügt, sondern auch künftige Bedürfnisse befriedigen kann.

3.2 Das VHE Konzept

Das virtual home environment (VHE) ist eine der wesentlichen Neuerungen von UMTS gegenüber GSM. Mit dem VHE kann ein Benutzer seine persönlichen Services unabhängig vom benutzten Netzwerk in Anspruch nehmen.

VHE ist eigentlich eher eine Eigenschaft eines Kommunikationsmediums und hat mit dem Mobilfunknetzwerk direkt wenig zu tun. Mit VHE wird versucht, alle Kommunikationsformen zusammenzufassen und dem Anwender damit umfassende Kommunikationsmöglichkeiten zu bieten.

Die vom Anwender nutzbaren Netzwerke können vom Breitbandinternetzugang bei der Arbeit über ISDN-Internetverbindungen am Computer zu Hause bis hin zu mobilen Verbindungen mittels Handy und PDA reichen. Die angebotenen Dienste werden je nach Bandbreite und der Übertragungsqualität an die Gegebenheiten angepaßt.

Ein Beispiel für einen solchen Service ist die automatische Benachrichtigung über eingehende Emails. Das VHE erkennt automatisch, mit welchem Mittel der Benutzer momentan am besten erreichbar ist und sendet die Meldung an die entsprechende Stelle. So kann z.B. eine Email mit einem Videofilm als Anhang auf dem Computer mit Breitbandinternetzugang direkt angezeigt werden, während auf einem PDA eventuell lediglich gemeldet wird, dass die entsprechende Nachricht eingetroffen ist.

Mit dem VHE soll der Anwender überall die gleichen Kommunikationsmöglichkeiten haben, unabängig davon, ob er sich im Büro, zu Hause oder unterwegs befindet.

Das VHE ist ein wichtiges neues Konzept auf dem Weg hin zu mehr Benutzerorientierung und einem intelligenteren System, dass besser an die zukünftigen Bedürfnisse der Benutzer angepasst werden kann. Mit diesem neuen Konzept sollen Kommunikationsdienste in Zukunft verstärkt für den einzelnen Benutzer personalisiert werden können.

4 Überblick über die UMTS System Architektur

UMTS wurde modular aufgebaut, so dass verschiedene Servicefunktionen ermöglicht werden können, ohne dass das zugrunde liegende Netzwerk geändert werden muss. Services werden von sogenannten service capability servers (SCS) bereitgestellt. Die Schnittstellen dieser Server sind so definiert, dass auch Drittanbieter diese Services bereitstellen können.

In der VHE Spezifikation werden folgenden SCSs genannt:

- UMTS call control server: Bietet grundlegende Dienste zur Anrufkontrolle an
- Home location register: Enthält alle Informationen über alle Benutzer die zu dem entsprechenden Netzwerk gehören
- Mobile execution environement server: Liefert die Umgebung, in der Anwendungen ablaufen können. Dieser Server kommuniziert mit dem Terminal mit Hilfe von WAP. Das Endgerät dient dabei nur als Terminal, die Anwendungen laufen auf dem mobile execution environement server
- SIM application toolkit server: Mit Hilfe dieses Servers können Anwendungen und Informationen direkt in die SIM (subscriber identity module) geladen werden
- Customized Application for Mobile Networks Enhanced Logic server: Erlauben die Benutzung intelligenter und vorbezahlter (prepaid) Dienste, wie z.B. SMS und GPRS

Die verschiedenen Server stehen nicht notwendigerweise miteinander in Verbindung und je nach gewünschter Anwendung werden nur einzelne Server benötigt.

Die UMTS System Architektur wurde sehr offen und standardisiert gestaltet, damit auch zukünftige Dienste ermöglicht werden können. Durch die offene Standardisierung der Architektur erhofft man sich außerdem, dass Drittanbieter spezielle Dienste anbieten werden. Weiterhin soll durch den festgelegten Standard der UMTS System Architektur die Entwicklungszeit für neue Dienste möglichst gering gehalten werden.

4.1 Einteilung der Architektur

Das UMTS Netz glieder sich in zwei unterschiedliche Teile, das sogenannte Kernnetz, dass für die Basisdienste zuständig ist und in das Funknetz, dass für die Einbindung der mobilen Endgeräte in das Kernnetz sorgt.

4.1.1 Das Kernnetz

Das Kernnetz verbindet die einzelnen Sendestationen des Funknetzes und sorgt für die Übertragung von Sprache und Daten. Es ist ebenfalls für die Übergänge zu anderen Telefonnetzen wie ISDN und zu Datenleitungen ins Internet zuständig.

Ursprüngliche Planung war, für das UMTS Kernnetz das schon vorhandene Netz von GSM zu übernehmen. Diese Überlegung beruht vor allem auf wirtschaftlichen Grundlagen, damit die für GSM installierte Infrastruktur auch weiterhin verwendet werden kann.

Mit dem neuen UMTS Release 2000 wurde allerdings eine neue Spezifikation verabschiedet, die genaue Spezifikationen für ein all-IP Kernnetz enthält. Es bleibt abzuwarten, inwiefern diese Release in die Tat umgesetzt werden wird.

Das Release enthält zwei Vorschläge für die Umsetzung eines all-IP Kernnetzwerkes. Ein Vorschlag geht davon aus, dass paketvermittelte und kanalvermittelte Dienste nebeneinander im System vorhanden sein werden. Bei diesem Ansatz könnte die vorhandene Infrastruktur weiterverwendet werden.

Der andere Ansatz geht von einem kompletten all-IP Netzwerk aus, der wenig Gemeinsamkeiten mit der bisherigen Infrastruktur besitzt.

Welcher der beiden Ansätze sich durchsetzen wird, bleibt abzuwarten.

4.1.2 Das Funknetz

Das Funknetzwerk von UMTS ist eine komplette Neuentwicklung, dabei wurden keine Kompromisse gemacht, um bestehende Infrastruktur weiterhin verwenden zu können.

Das Funknetz besteht aus vielen Sendestationen, die für die Verbindung der mobilen Endgeräte zum Kernnetz sorgen. Die größe der überwachten Funkzelle kann dabei je nach Bedarf variieren.

Die Zellen besitzen nicht nur unterschiedliche Größen, sie sind zudem hierarchisch angeordnet. Die verschiedenen Hierarchiestufen unterscheiden sich hauptsächlich in Zellengröße und Übertragungsgeschwindigkeit. Die maximale Übertragungsrate von knapp 2 Megabit pro Sekunde wird somit nicht flächendeckend zur verfügung stehen sonder nur an sogenannten “Hot Spots”, an denen diese Übertragungsrate auch gebraucht wird. Beispiele für “Hot Spots” sind Flughäfen und Bahnhöfe.

Durch den hierarchischen Aufbau der Zellen ist es auch möglich, verschiedenen Anwendungen unterschiedliche Hierarchiestufen zuzuordnen.

Somit könnten zum Beispiel Telefongespräche über das vorhandene GSM Netzwerk oder über die entsprechende Hierarchiestufen abgewickelt werden, wohingegen Datenübertragung von einer anderen Hierarchiestufe übernommen wird. Dabei kann es vorkommen, dass die für eine bestimmte Anwendung (z.B. Multimedia) benötigte Hierarchiestufe nicht flächendeckend vorhanden ist und die Anwendung deshalb nicht überall im Netzbereich verwendet werden kann.

4.1.3 Quality of Service

Zum Abschluss der Betrachtung des Funknetzwerkes möchten wir kurz auf die unterstützten Klassen des Quality of Service eingehen. Quality of Service ist ein sehr breites Thema, deswegen werden wir hier lediglich einen Überblick über die zur Verfügung gestellten Möglichkeiten geben.

Allerdings ist gerade für UMTS die Quality of Service ein wichtiges Kriterium, da in diesem mobilen Netzwerk sowohl Sprachanwendungen als auch paketbasierte Anwendungen gleichberechtigt nebeneinander existieren sollen.

Das UMTS Funknetzwerk unterstützt vier verschiedene Klassen des Quality of Service [Daum01]:

1. *Conversational:* Sprachübertragung und Ähnliches mit Übertragung in beide Richtungen in Echtzeit
2. *Streaming:* Übertragung in eine Richtung wie Musik oder Videos
3. *Interactive:* Übertragung in beide Richtungen mit geringeren Echtzeitbedingungen und weniger Bandbreite, Beispiele dafür sind Web-Browsen und Datenbankabfragen
4. *Background:* Übertragung von Daten im Hintergrund, wie z.B. das Herunterladen von Dateien aus dem Internet oder E-Mail-Postfach überprüfen

Mit diesen vier Qualitätsklassen hofft man, alle Anforderungen an ein mobiles Kommunikationsnetzwerk bewältigen zu können.

4.2 Zusammenfassung

Das UMTS Netzwerk besteht aus zwei verschiedenen Netzen, dem Basisnetzwerk und dem Funknetzwerk.

Das Basisnetzwerk ist für die Basisdienste und die Verbindung zu anderen Telekommunikationseinrichtungen wie andere Telefonnetzwerke und Internetdatenleitung zuständig.

Für die Anbindung der mobilen Endgeräte und die Funkschnittstelle ist dagegen das Funknetzwerk verantwortlich. Die wichtigste Neuerung beim Funknetzwerk von UMTS im Gegensatz zu GSM ist die Einteilung der Funkzellen in verschiedene Hierarchieebenen, mit unterschiedlichen Anforderungen und verwendeten Datenraten.

5 UMTS und IP

In diesem Abschnitt diskutieren wir die mögliche Verwendung von IP als Protokoll für UMTS. Wir gehen dabei zuerst auf die Vorteile dieses Ansatzes ein, bevor wir uns den Problemen zuwenden.

5.1 Warum ist IP sowohl für Anwender als auch Anbieter interessant?

Es stellt sich die Frage, warum für ein neues Netzwerk für mobile Kommunikation das doch schon recht alte Konzept des Internet Protokolls (IP), das ausgehend von anderen Voraussetzungen entwickelt wurde, verwendet werden soll.

Trotz dieser Bedenken hat jedoch die Verwendung von IP einige entscheidende Vorteile gegenüber anderen möglichen Varianten.

Zum einen können bestehende Anwendungen direkt über UMTS verwendet werden, ohne dass dazwischen eine Protokollumsetzung stattfinden muss oder die Anwendungen eventuell speziell für UMTS entwickelt werden müssten.

Desweiteren erlaubt IP eine gute Ausnutzung der vorhandenen Resourcen. Gerade für Mobilkommunikation mit eventuell vielen verschiedenen Benutzern auf einem stark begrenzten Medium ist aber die effiziente Verwendung der Resourcen von entschiedener Wichtigkeit.

IP liefert diese tollen Eigenschaften aber nicht nur in der Theorie. Durch den jahrelangen Einsatz von IP konnte man sich schon umfassend von seiner Praxistauglichkeit überzeugen.

Bei der Verwendung von IP als Protokoll für UMTS können neue Fortschritte und Entwicklungen bei IP direkt in UMTS einfliessen. Gerade im Bereich des Quality of Service (QoS) wird im Bereich von IP momentan einiges geforscht. Diese Entwicklung könnte dann ohne grössere Probleme in UMTS aufgenommen werden. Quality of Service ist gerade für ein Mobilfunknetz ein sehr wichtiges und heikles Thema, deswegen werden wir später nochmal darauf zurückkommen.

Ein sicherlich ebenfalls wichtiger Punkt ist die relative Simplizität von IP. Durch die Einfachheit dieses Protokolls kann es leicht und schnell implementiert werden. Damit sind die Kosten für den Entwurf integrierter Systeme geringer. Durch die Verwendung eines einfacheren Protokolls können eventuell sogar einfachere Bausteine zur Implementierung verwendet werden. Die schlägt sich dann nicht nur in der Abmessung der Geräte nieder, sondern eventuell auch im Stromverbrauch. Beide Punkte sind gerade für Mobilfunknetze von hoher Wichtigkeit.

Als mögliche Zukunftsperspektive wurde sogar angedacht, dass bei UMTS Voice over IP verwendet werden kann. Mit Voice over IP wird ein Verfahren bezeichnet, bei dem normale Telefongespräche nicht über eigene Kanäle und Protokolle abgewickelt werden, sonder - wie anderer Datenverkehr auch - über IP. Dieses Verfahren hätte den Vorteil, dass durch die Möglichkeit der Kompression eine effizientere Bandbreitennutzung möglich wäre. Durch die Verwendung von Voice over IP würde ein Gespräch dann nicht mehr einen Kanal für sich alleine Beanspruchen, es könnten sich also eventuell mehrere Gespräche einen Kanal teilen. Damit könnte eine bessere Auslastung der Netzwerkinfrastruktur erreicht werden.

Bevor aber Voice over IP verwendet werden kann, muss zuerst das Problem gelöst werden, wie Echtzeitdaten mittels IP übertragen werden können. An diesem Problem wird momentan gearbeitet und es gibt schon verschiedene Lösungsansätze, allerdings wurde dabei noch keine endgültige Lösung gefunden.

Voice over IP wäre sicher nur der Anfang neuer Anwendungsmöglichkeiten in einem echtzeitfähigen IP System. Eine Reihe weiterer Echtzeitanwendungen sind ebenfalls denkbar, so z.B. Videokonferenzsysteme, personalisierte Verkehrsleitsysteme und Netzwerkspiele.

Wenn das Problem der Übertragung von Echtzeitdaten über IP gelöst ist, dann ist es sogar denkbar, dass IP als grundlegendes Protokoll für UMTS verwendet wird.

Wie wir gesehen haben, gibt es einge Menge Gründe die dafür sprechen IP als Protokoll für UMTS zu nutzen. Wir haben allerdings auch gesehen, dass zuvor noch einige Fragestellungen zu klären sind.

Als nächstes wollen wir uns überlegen, welche weiteren Bedingungen an ein all-IP Basisnetzwerk gestellt werden.

5.2 Bedingungen an ein all-IP Basisnetzwerk

All-IP bedeutet in diesem Zusammenhang ein Netzwerk, dass als grundlegendes Protokoll IP verwendet. Im letzten Abschnitt sind wir bereits auf einige Probleme gestossen, die ein solches Netzwerk aufwerfen könnte. In diesem Abschnitt wollen wir überlegen, welche Bedingungen an ein all-IP Netzwerk gestellt werden, speziell falls dieses Netzwerk als Grundlage für ein Mobilfunknetz der dritten Generation dienen soll.

Wir haben die wichtigsten dieser Bedingungen in folgender Liste zusammengefaßt [BoLe01]:

- Unterstützung von Roaming und Überleitung zu Netzen der zweiten Generation
- Unterstützung von verbindungsoriertierten Endgeräten in einem all-IP Netzwerk, also paketorientierten UMTS Basisnetzwerk mit kompletter Abwärtskompatibilität auch für nicht paketorientierte Endgeräte
- Unterstützung sowohl neuer als auch bereits existierender Dienste, wie zum Beispiel Sprachübertragung und SMS

Diese Vorbedingungen kann man darunter zusammenfassen, dass auch alte nicht IP fähige Endgeräte benutzt werden können und dass sichergestellt ist, dass die momentane Hauptanwendung, nämlich telefonieren, auch in einem neuen Netz reibungslos und mit höchster Priorität verwendet werden kann.

5.3 Entwicklung eines all-IP Netzwerkes

Zur Entwicklung eines all-IP Netzwerkes basierend auf UMTS wurden zwei Partnerschaftsprogramme gestartet, die von unterschiedlichen Voraussetzungen ausgehend die Vision eines all-IP Netzwerks genauer spezifizieren sollten:

1. Das 3rd Generation Partnership Project (3GPP), das G3 Standards basierend auf GSM entwickelt
2. Das 3rd Generation Partnership Project 2(3GPP2), das G3 Standards basierend auf CDMA Systemen entwickelt

Wie man aus den beiden Projekten sehen kann, wird bei dem Projekt 3GPP der europäische Mobilfunkstandard (GSM) verwendet, während das Projekt 3GPP2 ausgehen vom amerikanischen Standard arbeitet. Durch diesen Ansatz sollen die Vorteile beider bisherigen Standards in die neue Technologie einfliessen.

Ein Hauptaugenmerk bei der Entwicklung der neuen Technologie war die Datenrate. Die Mobilfunknetze der zweiten Generation hatten nur relativ geringe Datenraten zu bieten. Gerade

im Bereich mobiler Anwendungen, die über das reine Telefonieren hinaus gingen, war man mit den Datenraten dieser Netze schnell an die Grenzen gestoßen. Deswegen war es für die Entwicklung der Mobilfunknetze der dritten Generation sehr wichtig, dass sie höhere Datenraten als die Netze der zweiten Generation liefern konnten.

Denn gerade auf dem Gebiet der mobilen nicht-Telefonanwendungen wurden die größten Zuwächse und Entwicklungschancen für ein neues Mobilfunknetz gesehen.

Das Ergebnis dieser Entwicklung ist eine Technik, die Datenraten von bis zu 2 Megabit pro Sekunde liefern kann und dabei den ITU IMT 2000 Anforderungen genügt [PaDe00]. Die Entwicklung der IP Technik beider Projekte führte zu zwei verschiedenen Architekturen die vereinheitlich werden müssen, bevor ein all-IP Basisnetzwerk Wirklichkeit werden kann.

Als nächstes stellen wir die entwickelten Architekturen beider Projekte vor.

5.3.1 3GPP Netzwerk Architektur

Die von 3GPP entwickelte Architektur basiert auf GPRS. Die Basis von GPRS wurde jedoch so erweitert, dass sie neben der IP Funktionalität auch Anrufkontrolle und Gatewayfunktionen durchführen kann. Diese beiden Funktionalitäten werden besonders für Voice over IP benötigt. Ein sehr wichtiger Aspekt dabei ist die Bereitstellung von Quality of Service Funktionalität. Ohne diese Möglichkeit wird sich ein all-IP Netzwerk kaum durchsetzen lassen, da eine wichtige Anwendung von UMTS sicher Sprachkommunikation und eventuell Multimediaanwendungen sein werden.

Der verwendete Ansatz zielte darauf ab, IP eine zusätzliche Komponente hinzuzufügen, die für die Sicherung des Qualitätsstandards verantwortlich ist. Diese Komponente sollte dafür sorgen, dass auch Realzeitanwendungen in einem all-IP Netzwerk funktionieren und das insbesonders Voice over IP möglich ist.

5.3.2 3GPP2 Netzwerk Architektur

Als Grundlage für die Architektur des Netzwerkes verwendete 3GPP2 das bestehende mobile IP. Mobile IP ist der momentane de facto Internetstandard für die Verwendung von IP über mobile Netzwerke. Der Vorteil dieses Ansatzes liegt darin, dass der Übergang zu bestehenden IP Netzwerken leicht möglich ist und dass privater Netzwerkzugriff mit Hilfe von IP security ermöglicht werden kann.

Die ursprüngliche Architektur sah keine Multimediaanwendungen vor, jedoch wurde sie überarbeite, so dass inzwischen auch Multimediaanwendungen möglich sind. Dabei wurde ebenfalls die Problematik von Quality of Service angegangen. Mit dieser neuen Architektur soll Voice over IP zuverlässig möglich sein.

5.4 Zusammenfassung

Die Ansätze der beiden Projekte sind sehr verschieden und es wird sicher noch eine Weile dauern, bis sich beide Projekte auf einen einheitlichen Standard geeinigt haben. Dieser Standard ist sehr wichtig, da UMTS ein universales System sein soll, das weltweit einsetzbar ist.

Das wichtigste Thema bei diesem kommenden Standard ist sicher die Frage, wie die Problematik von Echtzeitanwendungen und Quality of Service gelöst werden kann. Dieser Problempunkt ist wesentlich für die Akzeptanz und den Erfolg eines all-IP Netzwerkes.

Nachdem wir zuerst ein wenig über die Entwicklung von UTMS und die Dienstmöglichkeiten berichtet haben und zum Abschluss auf die weitere Entwicklung in Richtung eines all-IP Netzwerkes eingegangen sind, wollen wir als nächstes ein mögliches Szenario für ein mobiles IP Netzwerk vorstellen. Dabei möchten wir besonders auf die Probleme beim Routing der IP Pakete eingehen.

6 Beschreibung eines möglichen Szenarios

Zum Abschluss dieses Berichtes stellen wir kurz ein mögliches Szenario für ein mobiles all-IP Netzwerk vor. Dieses Netzwerk basiert auf GPRS und entspricht somit der Entwicklung der 3GPP [BeVE99]. Da es aber gleichzeitig auch Elemente von mobile IP beinhaltet, könnte es ein mögliches Szenario für die Einigung der beiden Projekte sein.

Im vorgestellten Szenario wird neben mobile IP für die Makromobilität auch HAWAII eingesetzt, ein Protokoll das die schnelle und einfache Realisierung der Mikromobilität gewährleisten soll.

6.1 Überblick über GPRS

GPRS steht für general packet radio services und wurde urspünglich entwickelt, um eine paketorientierte Verbindung über Mobilfunknetze der zweiten Generation (GSM) zu ermöglichen. GPRS bietet gegenüber GSM zwei wesentliche Verbesserungen bei der Übertragung von paketorientierten Daten:

1. schnellerer Verbindungsaufbau
2. höhere Datenübertragungsraten

Darüber hinaus bietet GPRS die Möglichkeit der paketorientierten Abrechnung, die bei GSM nicht möglich ist. GSM bietet nur die Möglichkeit kanalorientiert Daten zu übertragen. Bei Telefonaten ist diese Technik vorteilhaft, da bei einem Telefongespräch normalerweise stetig Daten zu übertragen sind.

Für eine paketorientierte Anwendung wie z.B. das Browsen im Internet ist eine kanalorientiere Verbindungsweise dagegen weniger vorteilhaft, da es starke Schwankungen im aufkommenden Datenverkehr geben kann. Da die beim Browsen übertragenen Daten auch weniger zeitkritisch als beim Telefonieren sind, ist es sinnvoll für die nicht zeitkritische Datenübertragung einen eigenen Dienst bereitzustellen. Dieser Dienst wird mit GPRS realisiert. GPRS erlaubt die paketorientierte Abrechnung von Verbindungen, d.h. es wird nach übertragenem Volumen abgerechnet und nicht nach der Verbindungszeit wie beim Telefonieren.

GPRS wurde von dem europäischen Institut für Telekommunikationsstandards (ETSI) seit 1995 standardisiert. Die ersten deutschen Netzbetreiber ermöglichten die Verwendung von GPRS Anfang 2001.

6.2 Vorbemerkungen

Das vorgestellte Szenario geht von der Voraussetzung aus, dass alle mobilitätsabhängigen Funktionen in Schicht 3 (IP Schicht) geregelt werden sollten. Das erlaubt den Entwurf eines einheitlichen IP-basierten Netzwerkes. Weiterhin ermöglicht es die Übernahme von Schicht 4 Protokollen ohne Änderungen, da diese Protokolle nichts über die darunterliegende Struktur wissen müssen.

6.3 Vorstellung des Szenarios

Als grundlegendes Protokoll wird mobile IP verwendet. Dieses Protokoll ist der momentan verwendete Internetstandard. Weiterhin wird eine Erweitung dieses Protokolls, genannt HAWAII verwendet.

Das Standardprotokoll mobile IP wird verwendet um die Makromobilität zu gewährleisten, für die Mikromobilität und Pagingfunktionalitäten wird HAWAII verwendet. Dieser Ansatz bietet verschiedene Vorteile, unter anderem, dass durch den Einsatz von HAWAII die Unterbrechungen durch Überleitung und Updates minimiert werden [RPST+00].

Die Pagingfunktionalität von HAWAII erlaubt ein verbessertes Powermanagement auf Seite des mobilen Gerätes. HAWAII "merkt" sich den letzten Standort eines Gerätes und weiss so bei einem ankommenden Paket für das Endgerät den ungefähren Aufenthaltsort. Das Endgerät kann dann mittels Paging ausfindig gemacht werden und das Paket wird zugestellt. Diese Funktion erlaubt es den Engeräten sich lediglich ab und an bei einer Basisstation zu melden und die restliche Zeit passiv zu hören ob es angepagt wird. Dadurch lässt sich ein Powermanagement einrichten, mit dem die mobilen Endgeräte Strom sparen können.

Durch die Verwendung von HAWAII behält ein Gerät die gleiche Netzwerkadresse (solange er innerhalb einer Domäne bleibt), dadurch wird die Verwendung von Quality of Service Funktionen erleichtert.

Da sowohl mobile IP als auch HAWAII im Rahmen dieses Seminars schon vorgestellt wurden, werden beide Protokolle hier lediglich kurz vorgestellt.

6.4 Mobile IP

Mobile IP ist der de fakto Standard um Makromobilität in IP Netzwerken zu implementieren. Zu diesem Zweck werden bei mobile IP zwei Entitäten definiert, ein Home Agent und ein Foreign Agent.

Der Home Agent wird dem mobilen Gerät basierend auf seiner permanenten IP Adresse zugewiesen, der Foreign Agent wird dem mobilen Gerät aufgrund des momentanen Aufenthaltsortes zugewiesen. Der Foreign Agent wird mit einer IP Adresse assoziiert, der sogenannten care-of Adresse.

Pakete für das mobile Endgerät werden zunächst an den Home Agent geschickt. Dieser tunnelt die Pakete zum Foreign Agent, der sie an den Host weiterleitet. Der Foreign Agent ist immer die Einheit, die dem Host am nächsten ist.

Wenn sich der mobile Host in den Einflussbereich eines anderen foreign agents begibt, wird eine Updatenachricht zum Home Agent geschickt, damit der Home Agent die Daten zum neuen Foreign Agent tunneln kann.

In mobilen Netzen können diese Wechsel der Foreign Agent oft geschehen, wenn sich ein mobiles Endgerät zum Beispiel im Grenzbereich zwischen zwei Basisstationen befindet. Durch die häufigen Wechsel der Foreign Agents würden viele Updatenachrichten erzeugt und da der Update Zeit braucht, bis er beim Home Agent angekommen ist und dieser den neuen Foreign Agent aktiviert hat, auch relativ viele Datenpakete verloren gehen.

Dieser Verlust einer größeren Zahl von Datenpaketen führt besonders dann zu Problemen, wenn Echtzeitdaten wie Telefongespärche oder Multimdiaanwendungen übertragen werden sollen.

Ein weiteres Problem von mobile IP ist, dass Paging nicht unterstützt wird. Deswegen muss sich ein Host permanent im Netz melden, um erreichbar zu sein. Dadurch ist kein sinnvolles Powermanagement möglich, wenn das mobile Endgerät errreichbar sein soll.

Letzendlich wird durch die Änderung der care-of Adresse eine neue IP-Adresse ausgehandelt. Die meisten Ansätze von Quality of Service basieren aber darauf, dass während einer Verbindung die IP-Adresse fest bleibt. Somit müßte bei jeder Änderung der IP-Adresse eine neue Quality of Service ausgehandelt werden, was zu zusätzlichem Overhead führt und ebenfalls in Verlust von Paketen resultieren kann.

Gerade bei häufigem Wechsel des Foreign Agents (wenn sich der Host zum Beispiel im Randgebiet zwischen zwei Basisstationen bewegt) führen die häufigen Änderungen des Foreign Agent zu merklichen Leistungseinbrüchen und einer hohen Paketverlustrate die gerade für zeitkritische Anwendungen wie telefonieren oder Multimediaanwendungen nicht annehmbar sind.

Die meisten der durch mobile IP im Bereich der Mikromobilität entstehenden Probleme werden durch die Verwendung von HAWAII beseitigt.

6.5 HAWAII

HAWAII wurde entwickelt, um die mit mobile IP auftretenden Probleme beim Handoff innerhalb einer Domäne zu umgehen. Bei HAWAII wird das Netz in verschiedene Domänen unterteilt, genau so wie das Internet als solches in Domänen unterteilt ist.

Der Zugang vom Internet aus zu jeder Domäne geschieht über einen sogenannten Domain Root Router. Es wird angenommen, dass jedes mobile Endgerät eine IP-Adresse und eine Heimatdomäne hat. Diese Adresse kann entweder statisch dem Host zugeordnet sein oder mittels DHCP während des Startvorgangs vergeben werden. Die Verwendung von DHCP zur Ermittlung der IP-Adresse und der Heimatdomäne hat den entscheidenden Vorteil, dass der anfängliche Ort des Hosts gleich seiner Heimatdomäne ist.

Der Domain Root Router der Heimatdomäne des mobilen Endgerätes (im weiteren auch mit Host bezeichnet) wird automatisch der Home Agent des Hosts. Bei der Verwendung von DHCP bedeutet dies, dass sich der Host anfangs automatisch in seiner Heimatdomäne aufhält.

So lange sich der Host innerhalb seiner der Heimatdomäne aufhält, ist der Domain Root Router gleichzeitig auch der Foreign Agent. Für die Bestimmung der Domänen und den Datenaustausch zwischen Home Agent und Foreign Agent wird mobile IP verwendet.

Innerhalb einer Domäne wird mittels HAWAI und dem sogenannten Path Setup Scheme ein Pfad etabliert, der vom Domain Root Router über eventuell vorhandene andere Router innerhalb der Domäne bis zu derjenigen Basissendestation führt, die dem mobilen Host am nächsten ist.

Das Path Setup Scheme ist ein wesentlicher Bestandteil von HAWAII und dient dazu, dass die Mikromobilität gewährleistet werden kann.

6.6 Implementierung der Mobilitäten

Bewegt sich der mobile Host, so müssen zwei Arten der Bewegung unterschieden werden, nämlich die Bewegung innerhalb der momentanen Domäne und Bewegung über Domänengrenzen hinweg.

Auf diese beiden Aspekte soll in den nächsten Abschnitten detaillierter Eingegangen werden. Bei der Bewegung innerhalb einer Domäne sprechen wir auch von Mikromobilität, dafür ist dann HAWAII zuständig. Für die Bewegung zwischen Domänen ist entsprechend mobile IP zuständig und dann ist die Rede von der sogenannten Makromobilität.

6.6.1 Bewegung innerhalb einer Domäne

Bei einer Bewegung innerhalb der Domäne greift nur der HAWAII Teil des Protokolls. Der Host behält seinen momentanen Foreign Agent und es müssen keine Updatenachrichten zum Home Agent des mobilen Hosts geschickt werden.

Durch das sogenannte Forwarding Path Setup Scheme wird lediglich der Pfad vom Domain Root Router zur dem Host am nächsten gelegenen Basissendestation angepaßt.

Bewegt sich der Host nur von einer Sendestation zur nächstgelegenen, so kann es vorkommen, dass sich die Änderung der Pfades nur auf den letzten Router bezieht. Dieser muss Pakete für den Host dann eventuell lediglich auf einen anderen Port legen.

Alle übrigen Router innerhalb der Domäne erfahren von dieser Änderung nichts. Somit wird die Zahl der Updates und Änderungen minimiert. Auch die Zahl der Paketverluste reduziert sich.

Das gesamte Verfahren ist darauf ausgerichtet, dass möglichst wenige Änderunge gemacht werden müssen, wenn der mobile Host seine Basisstation wechselt.

6.6.2 Bewegung über Domänengrenzen hinweg

Bewegt sich der Host über die Domänengrenzen hinweg, so greift der mobile IP Teil der Architektur des vorgestellten Szenarios.

Der Domain Root Router derjenigen Domäne, in der sich der mobile Host aufhält, wird dann Foreign Agent für den mobilen Host. Dem Home Agent des mobilen Hosts wird diese Änderung mitgeteilt und er wird Pakete an den mobilen Host zum neuen Foreign Agent tunneln.

Desweiteren wird mit dem Path Setup Scheme aus dem HAWAII Protokoll ein Pfad vom Domain Root Router zum mobilen Host gesetzt. Die Bestimmung dieses Pfades geschieht auf die gleiche weise wie bei der Bestimmung des urspünglichen Pfades innerhalb der Heimatdomäne des mobilen Endgerätes.

Die Änderung des Foreign Agents ist mit einem größeren Overhead verbunden und dauert länger. Dadurch kann es bei diesem Wechsel auch zu einigen Paketverlusten kommen. Da allerdings Domänenwechsel eher selten sind, wird sich der Verlust durch diesen Wechsel in Grenzen halten.

6.7 Zusammenfassung

Das vorgestellte Modell zeigt, wie mit einer Mischung aus mobile IP und HAWAII die Stärken beider Modelle im mobilen Einsatz verwendet werden können, ohne dass die Schwächen zum tragen kommen.

Aus Sicht der Domain Root Router wird bei diesem Szenario mobile IP verwendet, wobei allerdings nur Domain Root Router die Rolle von agents übernehmen. Innerhalb einer Domäne sieht es für die Router aus, als ob lediglich HAWAII zum Einsatz kommen würde. Die Router erhalten die Pakete und Informationen darüber, auf welchem Weg sie den mobilen Host erreichen können.

Damit benutzt das Szenario nicht nur die besten Teile beider Protokolle, es minimiert zudem noch die Komplexität der einzelnen Bestandteile. Lediglich die Domain Root Router müssen sowohl über mobile IP als auch HAWAII bescheid wissen, da sie zum einen den Home / Foreign Agent aufsetzen müssen und zum anderen auch dafür zu sorgen haben, dass es einen gesetzten Pfad von ihnen zum mobilen Endgerät gibt.

7 Abschlußbemerkung

In dieser Ausarbeitung wurde zuerst die Entwicklung von UMTS beschrieben. Danach folgte die Erläuterung der UMTS System Architektur. Eine Diskussion über UMTS und IP folgte, die mit der Beschreibung eines möglichen Szenarios endete.

Das vorgestellte Szenario stellt eine Verbindung aus mobile IP und HAWAII dar. Bei diesem Szenario wurde mobile IP für die Makromobilität verwendet, während HAWAII für die Realisierung der Mikromobilität eingesetzt wurde. Durch die Verwendung dieser beiden Protokolle konnten die Stärken beider Protokolle gut ausgespielt werden, wohingegen die Schwächen durch den Einsatz des jeweils anderen Protokolls ausgeglichen werden konnten.

Literatur

[BeVE99] Christian Bettstetter, Hans-Joerg Voegel und Joerg Eberspaecher. GSM PHASE 2 + General Packet Radio Service GPRS: Architecture, Protocols, and Air Interface. *IEEE Communications Surveys* 2(3), Third Quarter 1999.

[BoLe01] Lieve Bos und Suresh Leroy. Toward an All-IP-Based UMTS System Architecture. *IEEE Network*, January / February 2001.

[Daum01] Ralf Daum. UMTS - Universal Mobile Telecommunication System. Wintersemester 2000/2001.

[OjPr98] Tero Ojanperae und Ramjee Prasad. An Overview of Third-Generation Wireless Personal Communications: A European Perspective. *IEEE Personal Communication*, December 1998.

[PaDe00] Girish Patel und Steven Dennett. The 3GPP and 3GPP2 Movements Toward an All-IP Mobile Network. *IEEE Personal Communications*, August 2000.

[RPST+00] Ramachandran Ramjee, Thomas F. La Porta, Luca Salgarelli, Sandra Thuel, Kannan Varadhan und Li Li. IP-Based Access Network Infrastructure for Next-Generation Wireless Data Networks. *IEEE Personal Communications*, August 2000.

Clustering für Ad-hoc Netze

David Divisek

Kurzfassung

Ziel dieser Arbeit ist die Vorstellung von Methoden zur Bildung von Clustern in Ad-hoc-Netzwerken. Desweiteren werden Verfahren zur Erzeugung von Cluster-Hierarchien und zur Aufrechterhaltung der Cluster demonstriert.

1 Einleitung

Feste und drahtlose Kommunikations-Netzwerke verändern ihren Zustand häufig in einer Weise, die nur sehr schwer vorauszusagen ist. Daher sollte die Kontrollinstanz des Netzwerkes idealerweise zwei Ziele verfolgen:

- Eine Veränderung im Netzwerk möglichst schnell und richtig zu verarbeiten
- den Verbrauch der Ressourcen, wie Sende-, Speicher- und Rechenkapazität, zu minimieren

Es ist schwierig, diese Ziele gleichzeitig zu erfüllen, da sie gegensätzlich sind, und daher stellt diese Aufgabe eine große Herausforderung für die Planung der Netzstruktur dar. Bei Festnetzen wird der Designprozess von dem Netzwerkmanager offline durchgeführt. Für mobile drahtlose Netze mit einer ständig wechselnden Topologie muss eine andere Strategie überlegt werden. Beispiele für den Einsatz solcher Netze sind Schlachtfeldszenarios, Katastrophen-Hilfsprogramme und kurzzeitige Ereignisse wie öffentliche Veranstaltungen. In einer höchst dynamischen Umgebung ist das Fluten ein wirkungsvolles Mittel zur Versendung von Datenpaketen. Allerdings wird hierdurch ein großer Netzverkehr erzeugt, da die Daten auf mehreren, zum Teil überflüssigen Wegen versendet werden. Da die Anzahl der Verbindungskanäle und der Energievorrat knappe Ressourcen darstellen, sollten effizientere Methoden ausgearbeitet werden, die eine überflüssige Verschwendung dieser Ressourcen vermeiden. Diese Methoden benötigen die aktuellen Standorte aller Knoten. Da der Preis für Speicherkapazität Jahr für Jahr sinkt, kann diese nicht unbedingt als knapp angesehen werden. Die Haupteinsparungen der Bandbreite und Energie resultieren in erster Linie daraus, dass der Umfang von Übertragungen auf Knoten beschränkt wird, die zu einem bestimmten Zeitpunkt spezielle Informationen benötigen. Daher sollte der Overhead von Übertragungen reduziert werden, die die Aufgabe haben, nach topologischen Änderungen die Routing-Tabellen zu aktualisieren. Hier empfiehlt sich eine sich selbst organisierende Kontrollstruktur, welche die einzelnen Teilnehmer in Zusammenarbeit aufbauen und aufrecht erhalten. Die Wahl einer adäquaten Struktur hängt hauptsächlich von den Aufgaben der Kontrollinstanz, der Größe des Netzwerkes und der Häufigkeit und Ausmaße der Veränderungen des Netzwerkes ab.

Die Aufteilung in Cluster ermöglicht eine effiziente Nutzung der Ressourcen. Dies geschieht durch eine Transformation des physikalischen Netzwerkes in ein virtuelles Netzwerk von miteinander verbundenen Clustern. Jedes dieser Cluster hat einen oder mehrere Controller, der Entscheidungen für die Mitglieder seines Clusters trifft und dieses nach außen hin repräsentiert. Dieser Controller wird im Folgenden als Clusterhead bezeichnet.

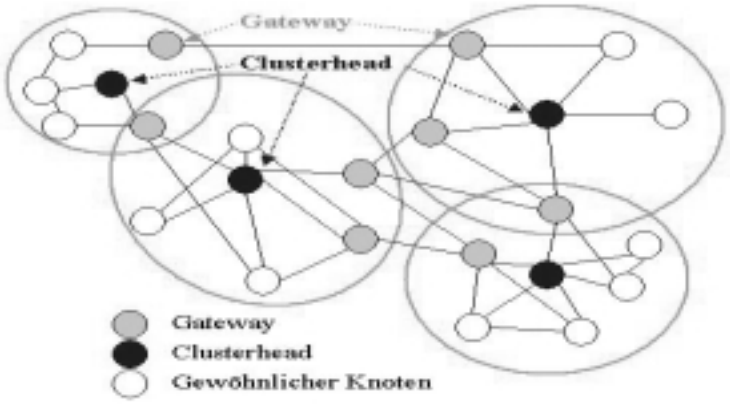

Abbildung 1: Clustering zur Verwaltung von Übertragungen

2 Ziele des Clusterings

Das Clustering dient in erster Linie zum Zwecke einer verbesserten Ausnutzung der vorhandenen Ressourcen. Dies wird durch folgende Methoden gewährleistet.

2.1 Das Verwalten von Übertragungen über das Medium Luft

Die Luft als Übertragungsweg stellt ein geteiltes Medium dar. Falls nun mobile Geräte miteinander kommunizieren wollen, entsteht ein Wettbewerb um die begrenzte Anzahl freier Kanäle. Hier hilft das Clustering diesen Wettbewerb zu reduzieren, indem der Zugriff auf das Medium durch eine Kontrollinstanz verwaltet wird.

Durch die Verwendung von Zeit, Raum, Frequenz oder Code-Multiplex können Interferenzen zwischen benachbarten Geräten vermieden werden. Durch das Clustering können diese Verfahren koordiniert werden, so dass die durchschnittliche Bandbreite des Netzwerkes erhöht wird. Hier bietet sich die Clustering-Methode der „Link-Cluster“-Architektur an.

Diese Architektur bietet eine natürliche Struktur eines Routing Backbone durch die Verwendung von Clusterheads und Gateways. Da allerdings der gesamte Verkehr über die Clusterheads zu laufen hat, resultiert dies in einer Überlastung der betreffenden Knoten. Daher wird diese Struktur nicht als Routing-Grundlage verwendet, wie Gerla [GeTs95] und Lin [LiGe97a] ausführen, sondern sehen das gesamte Netzwerk als Multihop-Struktur und jeder Knoten verwaltet eigene Routing-Informationen. Es werden also die Cluster nur für das Management der Übertragungen benutzt und nur sekundär, um einen Backbone zu erzeugen.

2.2 Aufbau einer virtuellen Netzwerkarchitektur

Will man die Cluster-Struktur einsetzen, um effiziente Routing Protokolle zu verwenden, ist es nötig, eine virtuelle auf Cluster basierende Netzwerkarchitektur aufzubauen, die dabei hilft, den Netzwerkdurchmesser zu reduzieren. Hierfür muss die Reichweite der Sender angepasst werden. Einige Knoten des gesamten Netzwerkes bilden den Backbone und erhöhen ihre Sendereichweite durch eine höhere Energie. Außerdem erfolgt noch eine Trennung der Frequenzen in die des Backbones und die der anderen Übertragungen. Es existieren also zwei Frequenzen, eine für den Backbone und eine andere für die Intracluster-Kommunikation. Hierfür existieren zwei Ansätze:

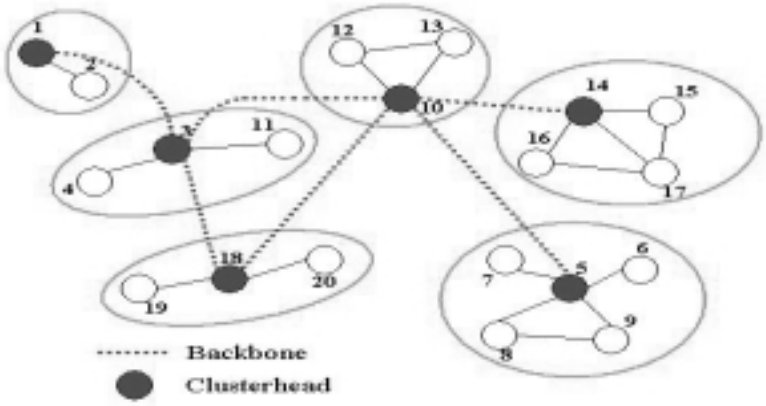

Abbildung 2: Near-Term-Digital Radio-Network

2.2.1 Near-Term-Digital-Radio-Network [Zavg97]

Hier arbeiten die Clusterheads als Gateways und bilden den Backbone. Alle Mitglieder des Clusters haben einen Hop Abstand zum Clusterhead und zwei zu jedem anderen Mitglied. Die Kommunikation ist nur über den Clusterhead möglich. Folglich wird dieser sehr beansprucht und stellt eine hohe potentielle Fehlerquelle dar. Daher existiert ein ständiger Wettbewerb um die Rolle des Clusterheads, da jeder Knoten in der Lage ist, seine Funktionen auszuführen. Ein Nachteil dieses Ansatzes ist, dass bei einem hochdynamischen Netzwerk jeder Knoten ständig zwischen dem Status Clusterhead und gewöhnlicher Knoten hin- und herpendelt.

Die Clusterheads halten sich gegenseitig über Veränderungen der Netzwerkstruktur auf dem Laufenden. Diese Informationen werden über den Backbone geflutet und anschließend Routing-Informationen berechnet. Diese werden aufgrund einer Metrik berechnet, die eine Messung der Interferenz beinhaltet, die in späteren Übertragungen erwartet wird.

Bei jeder Änderung, die das Routing im Backbone beeinflusst, muss die neue Information wieder im Backbone geflutet werden, was bei einem hochdynamischen Netzwerk zu einem großen Volumen an Update-Verkehr führt und die Übertragungskapazität überlasten könnte.

2.2.2 Virtual Subnet Architektur [Shar96]

In dieser Architektur existieren mehrere voneinander getrennte Backbones, um die Fehlerrate zu verringern und den Netzwerkverkehr besser zu verteilen. Das Netzwerk wird in eine Gruppe von getrennten Clustern aufgeteilt, die als physikalische Subnetze bezeichnet werden und auf der Position der Knoten beruhen. Mitglieder von verschiedenen physikalischen Subnetzen bilden wiederum virtuelle Subnetze, welche idealerweise alle physikalischen Subnetze abdecken. Durch eine Anpassung der Übertragungsstärke können die Mitglieder des virtuellen Netzes miteinander kommunizieren. Benachbarte physikalische Netze und alle virtuellen Netze benutzen unterschiedliche Frequenzen. Die Anzahl der Subnetze wird vorbestimmt. Jeder Knoten gehört genau einem physikalischen Subnetz und keinem oder einem virtuellen Subnetz an. Bei ständigen Verbindungen zu anderen virtuellen Netzen kann ein Knoten auch zu mehreren virtuellen Netzen gehören. Die Adresse eines Knoten beinhaltet zum einen seine physikalische und zum anderen seine virtuelle Adresse.

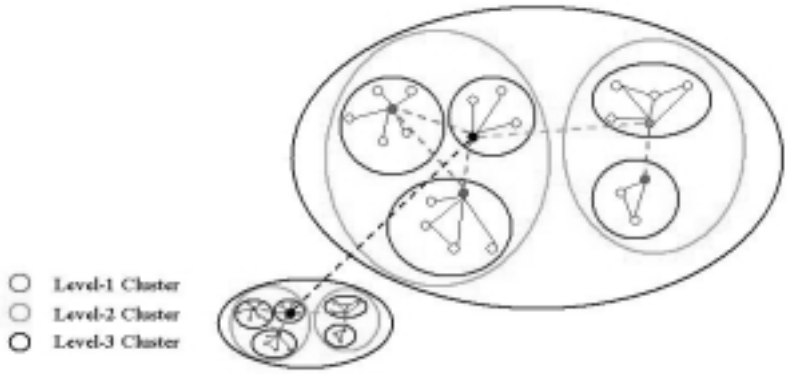

Abbildung 3: Clustering für ein effizientes Routing

2.3 Clustering für ein effizientes Routing

Zur Erreichung eines effizienten Routings bei sehr vielen mobilen Teilnehmern ist es nötig, eine abstrakte Abbildung der Netzstruktur zu schaffen. Dies ermöglicht es, die Datenmenge für Routing-Informationen zu minimieren, da nun nicht mehr jeder Knoten die genaue Position der anderen Knoten verwalten muss.

Die inhärente dynamische Struktur von Ad-hoc-Netzen stellt eine Herausforderung an das Design von effizienten Kontrollalgorithmen dar. Hierarchische, auf Clustern basierende Kontrollstrukturen ermöglichen es, die Informationen über die Netzstruktur erheblich zu verringern. Dies gilt insbesondere für Routing-Informationen, die überall im Netz verteilt, verarbeitet und gespeichert werden müssen. Um eine derartige Reduktion zu erreichen, muss ein hierarchisches Routing-Schema in der Lage sein, jedes Cluster in einer abstrakten Form darzustellen, die die Details der Verbindungen und die angebotenen Dienste innerhalb eines Clusters verbirgt. Um eine geeignete Abstraktion eines Clusters zu finden, muss zwischen dem Umfang und dem Detaillierungsgrad dieser Abstraktion abgewogen werden.

Es existieren zwei Ansätze für Routing in geclusterten Netzwerken. Das strikte hierarchische Routing [KlKa77][WTNi89] verwendet innerhalb des Clusters das „tier“-Routing und zwischen den Clustern das „link-state“-Routing. Mit anderen Worten wird der kürzeste Pfad zwischen den Clusterheads zuvor berechnet und das Paket immer erst zum Clusterhead und anschließend von Cluster zu Cluster gereicht. Im zweiten Ansatz, dem quasi-hierarchischen Routing [KlKa77], wird das „tier“-Routing insofern erweitert, als zusätzlich Informationen über die minimale Entfernung zu anderen Knoten im Cluster und zu den anderen Clustern verwendet werden. So wird das Paket direkt unter Verwendung der verfügbaren Informationen zum Cluster des Zielknotens geschickt. Im Cluster des Zielknotens angekommen, wird das Paket einfach zum Zielknoten weitergeleitet.

Die zahlreichen Routing-Protokolle unterscheiden sich in dem Ansatz, wie neue Routen gesucht und bekannte Routen im Falle einer Knotenbewegung angepasst werden. Der „shortest path“-Algorithmus eignet sich nicht für Netze, bei denen manche Knoten ausfallen können. Derartige Netze benötigen verteilte Algorithmen, mit deren Hilfe ein Knoten allein mit der Kenntnis seiner Nachbarn die Routing-Entscheidung treffen kann. Eine dieser Methoden ist GEDIR (geographic distance routing) [LiSt98], die auf den Ortsinformationen von GPS (geographic postioning system) beruht. Sie wird von Lin und Stojmenovic [LiSt98] vorgeschlagen und soll laut Chen und Stojmenovic [G. C] der leistungsfähigste unter den verteilten Algorithmen sein. Es existieren zwei Versionen für geclusterte und ungeclusterte Netzwerke. Wenn ein Knoten A eine Nachricht zum Knoten D senden will, benutzt er die Informationen über die

Position von D. Anschließend wählt er einen seiner Nachbarn C, der unter all seinen Nachbarn am nächsten zu D steht. Das Paket wird dann an C gesandt und die Auswahl des nächsten Knotens erfolgt auf die gleiche Weise, bis der Knoten D, falls überhaupt möglich, erreicht wird. Der Algorithmus stoppt, falls die Auswahl des nächsten Knotens auf den Knoten fällt, von dem die Nachricht kam. Dieser Knoten wird als konkaver Knoten bezeichnet. Dieser Mechanismus macht den Algorithmus frei von Schleifen. GPS-Karten werden in naher Zukunft in jedem Auto und möglicherweise in jedem mobilen Rechner vorhanden sein [Kapl96][NaIm97]. Beispielsweise hat das NAVSTAR Global Positioning System eine Genauigkeit von 50-100 Metern und das Differential GPS eine Genauigkeit von einigen Metern.

3 Methoden für das Clustering

Die Erzeugung und Aufrechterhaltung von cluster-basierten Kontrollstrukturen benötigt Algorithmen, um die Knoten zuerst in Cluster einzuteilen und anschließend die Cluster anzupassen, falls Knoten hinzukommen, das Netz verlassen oder sich innerhalb des Netzes bewegen. Folglich lassen sich die Clustering-Algorithmen in zwei Phasen unterteilen.

- Erzeugung der Cluster
- Aufrechterhaltung der Cluster

In diesem Kapitel werde ich auf die Erzeugung der Cluster näher eingehen. Im folgenden Kapitel werden Methoden zur Aufrechterhaltung der Cluster vorgestellt.

Die Idee des Clusterings existiert bereits seit dem Entstehen der Ad-hoc-Netzwerke. In den Arbeiten [BaEp81][DJBFl84][Ephr83][AEBa87] wird eine „fully distributed linked Cluster“ Architektur eingeführt, die in erster Linie das hierarchische Routing unterstützt und eine gute Anpassungsfähigkeit des Netzes an Veränderungen demonstriert.

Mit dem Auftauchen von Multimedia-Kommunikation entwickelt Gerla et al. [GeTs95][LiGe97a] eine Netzstruktur, die dem großen Bedarf an Ressourcen und der Dynamik des Netzwerkes Rechnung trägt.

Die beiden oben genannten Phasen bezeichnet Basagni [Basa98] als „Set up“ und „Maintenance“. Durch die drahtlose und mobile Natur von Ad-hoc-Netzen sollten folgende Bedingungen von den Clustering-Algorithmen erfüllt werden:

- Jeder Knoten hat mindestens einen Clusterhead als Nachbarn (dominance property)
- Jeder Knoten verbindet sich mit dem „besten“ Clusterhead
- Es gibt möglichst keine benachbarten Clusterheads, so dass diese gut verteilt sind (independence property)

Die k-Cluster-Architektur zeichnet sich dadurch aus, dass jeder mobile, drahtlose Teilnehmer, von nun an als Knoten bezeichnet, einen Abstand von höchstens k Hops zum Clusterhead hat. In einem 1-Cluster ist jedes Mitglied eines Clusters direkt mit dem Clusterhead verbunden und hat einen Maximalabstand von 2 Hops zu jedem anderen Cluster-Mitglied. Nachbarknoten sind Knoten, zu denen eine bidirektionale Verbindung besteht. Sie werden durch Broadcast-Beacons entdeckt.

Der Algorithmus zur Formung der Cluster besteht aus folgenden zwei Schritten:

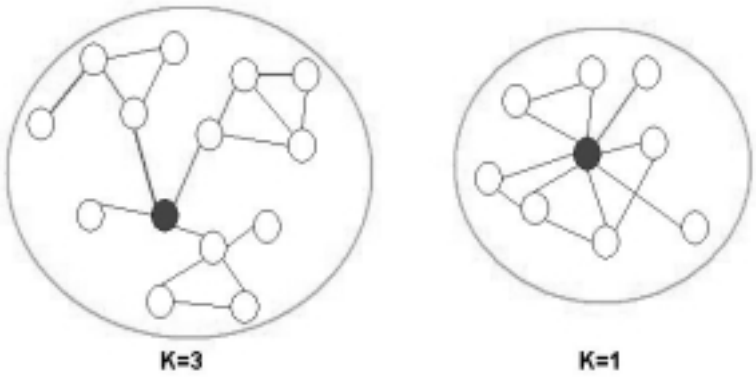

Abbildung 4: k-Cluster

1. Wahl des Clusterheads und Erzeugung der Cluster
2. Abstimmung über Gateways zwischen Clustern

3.1 Wahl des Clusterheads, Erzeugung der Cluster und Gateways

Einige Knoten müssen gewählt werden, die den Prozess des Clusterings koordinieren. Diese Knoten werden als Clusterheads bezeichnet. Die Nachbarn des Clusterheads bilden anschließend mit diesem das Cluster. Ein Nachbar ist jeder Knoten, der innerhalb der Übertragungsreichweite eines anderen Knotens liegt oder bei einem k-Cluster höchstens einen Abstand von k Hops hat. Alle Knoten, die kein Clusterhead sind, werden als gewöhnliche Knoten bezeichnet. Ein Clusterhead eines 1-Clusters ist also mit den Knoten seines Clusters direkt verbunden, aber mit einem anderen Clusterhead nur indirekt. Daher ist ein Knoten im Falle von 1-Clustern entweder selbst ein Clusterhead oder hat eine direkte Verbindung zu einem oder mehreren.

Die Aufgabe eines Clusterheads [GeTs95] ist es, den Zugang zum Medium zu kontrollieren, unter Benutzung einer Kombination von TDMA innerhalb des Clusters und CDMA zwischen den Clustern, den Energieverbrauch zu messen, die Synchronisation der Zeitrahmen aufrecht zu erhalten und Garantien für Bandbreite für Echtzeitverkehr zu gewährleisten. Jeder Knoten des Netzwerkes unterhält eine Liste von all seinen Nachbarn, eine Liste der Cluster, eine Liste der Gateways und eine Routing-Tabelle für den nächsten Hop für jedes Ziel. Für große Netzwerke ist die Menge der Daten, die bei einer Veränderung des Netzes aktualisiert werden muss, beträchtlich und verursacht einen großen Overhead, der die Bandbreite beeinträchtigt. Für die Wahl des Clusterheads stelle ich in den nächsten Abschnitten Algorithmen vor, die sich aufgrund des Kriteriums unterscheiden.

3.1.1 Clustering basierend auf dem Identifier

In [DJBFl84][AEBa87][LiGe97a] wird vorgeschlagen, den Knoten eindeutige Identifier (ID) zuzuweisen. Clusterhead wird der Knoten mit dem geringsten oder höchsten ID. Anschließend formen er und seine Nachbarn ein Cluster. Die Prozedur wird mit den übrigen Knoten wiederholt. Ein Knoten, der Übertragungen von zwei oder mehreren Clusterheads empfängt, schlüpft in die Rolle des Gateways. Hier unterscheidet man überlappende und getrennte Cluster. Bei getrennten Clustern formen zwei Knoten den Gateway, bei überlappenden nur einer.

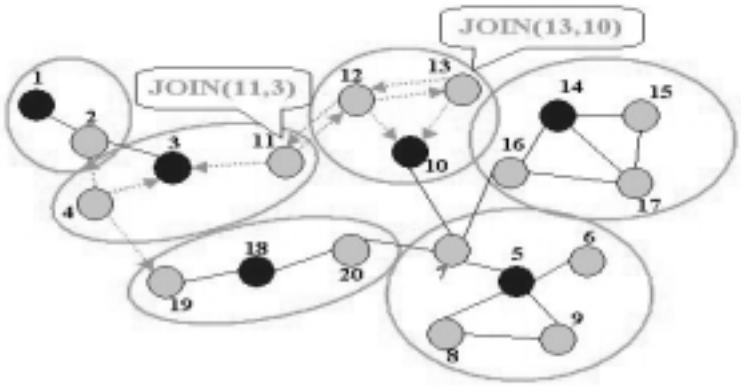

Abbildung 5: Clustering basierend auf dem Identifier

Hier wird wiederum das Kriterium des höchsten oder niedrigsten ID herangezogen, um die Wahl zu treffen.

Lin und Gerla [LiGe97a] beschreiben eine modifizierte Version, die einen geringeren Netzverkehr verursacht. Jeder Knoten gibt über einen Broadcast seine Entscheidung genau einmal bekannt. Der verteilte Clustering-Algorithmus wird von allen Knoten gestartet, die in ihrer Umgebung den geringsten ID aufweisen (local lowest ID nodes). Sie senden einen Broadcast an all ihre Nachbarn, mit der Information, dass sie ein Cluster bilden mit ihnen als Clusterhead. Jeder Knoten empfängt diesen Broadcast und wählt den benachbarten Clusterhead mit dem geringsten ID. Falls alle Nachbarn eines Knotens mit geringeren IDs ihre Zugehörigkeit zu anderen Clustern verkünden, so kann dieser Knoten selbst Clusterhead werden und gibt diese Entscheidung per Broadcast bekannt mit seinem ID als Cluster-ID. Ansonsten schließt er sich dem Cluster des benachbarten Clusterheads an und verkündet per Broadcast diese Entscheidung. Die Knoten treffen also erst ihre Entscheidung nachdem all ihre Nachbarn mit geringeren IDs dies getan haben. So findet also jeder Knoten genau ein Cluster und sendet nur eine Nachricht während des Prozesses. Dieser Algorithmus erzeugt Cluster, die sich nicht überlappen, da jeder Knoten genau eines wählt. Da er jedoch immer noch eine Verbindung zu einem anderen Clusterhead haben kann, ist eine Überlappung unabdingbar.

Um den Cluster-Algorithmus zu starten, muss die Initialnachricht geflutet werden, was allerdings einen weiteren Broadcast pro Knoten ausmacht [LiGe97a]. Zudem können Fehler auftauchen, wenn zwei benachbarte Knoten gleichzeitig ihre Broadcast-Übertragung starten. Um diese Problem zu umgehen, schlagen Chen und Stojmenovic [G. C] eine DFS-Traversierung aller Knoten des Netzwerkes vor, so dass diese ihre Übertragungen nach einer bestimmten Reihenfolge durchführen. Somit können sämtliche Kollisionen vermieden werden. 3.1.4

3.1.2 Clustering basierend auf der Anzahl von Verbindungen

Das Kriterium des ersten Algorithmus, der ID eines Knoten, sagt nichts über die Eignung eines Knotens für die Rolle des Clusterheads. Möchte man möglichst große Cluster erzeugen, sollte man einen Knoten wählen, der zu vielen Knoten eine Verbindung hat. Dieser Verbindungsgrad wird vom zweiten Algorithmus [GeTs95] als Kriterium verwendet. Dieser Grad steht für die Anzahl der Nachbarn eines Knotens, also die Anzahl der Knoten, deren Übertragungen er empfangen kann. Der Knoten mit dem höchsten Grad wird als Clusterhead gewählt. Dieser Algorithmus funktioniert recht gut in zufällig erzeugten Graphen, wie in [GeTs95] berichtet wird. Für Intervallgraphen (z.B. Autos auf der Autobahn) oder Dreiecksgraphen (z.B. der

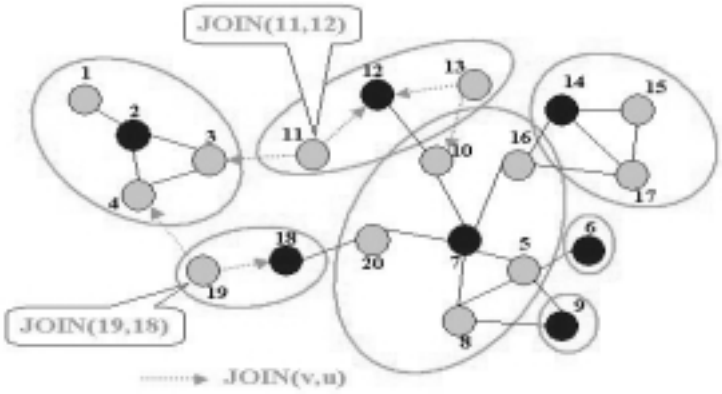

Abbildung 6: Clustering basierend auf der Anzahl von Verbindungen

Graph der Basisstationen im Mobilfunknetz) liefert er keine guten Ergebnisse. Daher sollte der Algorithmus um zusätzliche Entscheidungsregeln erweitert werden, wie im folgenden Abschnitt gezeigt wird.

3.1.3 Ein kombinierter Clustering-Algorithmus

Chen und Stojmenovic [G. C] stellen einen Algorithmus mit einer Kombination aus den zuvor beschriebenen Kriterien vor. Sie beziehen sich auf den Algorithmus von Lin und Gerla (1-lowestID) [LiGe97a], den ich bereits oben vorgestellt habe, und verallgemeinern ihn, um k-Cluster zu bilden (k-lowestID). Einer der Knoten startet den Clustering-Prozess, indem er eine Anfrage zum Clustering im gesamten Netzwerk flutet. Es wird angenommen, dass alle Knoten über ihre Nachbarn Bescheid wissen, die höchstens k Hops entfernt sind. Ein Knoten, der den niedrigsten ID im Umkreis von k Hops aufweist, wird Clusterhead. Er teilt diese Entscheidung per Broadcast den anderen k Hops entfernten Knoten mit. So wird seine Entscheidung weitergesendet, bis all seine Nachbarn im Umkreis von k Hops erreicht sind. Falls alle Nachbarn eines Knotens, die einen niedrigeren ID aufweisen, sich nicht dazu entscheiden, Clusterhead zu werden, kann dieser Knoten selbst Clusterhead werden und teilt dies per Broadcast seinen Nachbarn mit und erzeugt das Cluster mit seiner ID als Cluster-ID. Ansonsten wählt er den benachbarten Clusterhead, der höchstens k Hops entfernt ist und den geringsten ID hat, und verkündet dies per Broadcast. So entscheidet sich jeder Knoten erst dann, wenn sich alle benachbarten Knoten mit geringerem ID entschieden haben. Jeder Knoten wählt also genau ein Cluster und sendet nur eine Nachricht per Broadcast während des Algorithmus.

Da hierbei nicht die Verbindungen eines Knotens berücksichtigt werden, entstehen möglicherweise mehr Cluster, als benötigt werden. Der Algorithmus, der wiederum allein auf der Anzahl der Verbindungen basiert, arbeitet nicht einwandfrei, da in zahlreichen Fällen keine eindeutige Entscheidung getroffen werden kann. Daher schlagen Chen und Stojmenovic vor, den Grad eines Knotens als primäres und den ID als sekundäres Kriterium zu verwenden. Sie verallgemeinern zudem den Grad eines Knotens, indem sie nicht nur die direkten Nachbarn zählen, sondern auch die Nachbarn, die höchstens k Hops entfernt sind. Falls also nun zwei Knoten den gleichen Grad aufweisen, so entscheidet der ID die Wahl des Clusterheads. Dieser Algorithmus wird als k-CONID (k-hop connectivity ID) bezeichnet und funktioniert wie folgt. Jedem Knoten wird das Paar did=(d,ID) zugeordnet, das zum einen den Verbindungsgrad d und zum anderen den ID des Knotens enthält. Mit Hilfe des did wird nun der Clustering-

Algorithmus wie oben durchgeführt, nur dass diesmal zuerst der Verbindungsgrad und im Falle der Gleichheit der ID zu Rate gezogen wird.

3.1.4 Kollisionsfreies Clustering basierend auf DFS

Alle beschriebenen Clustering-Algorithmen haben das Problem, dass es zu Kollisionen kommen kann. Falls zwei Knoten, die einen gemeinsamen Nachbarn haben, gleichzeitig eine Nachricht per Broadcast versenden wollen, kann es bei dem gemeinsamen Nachbarn zu einer Kollision kommen. Um dies zu verhindern, schlagen Chen und Stojmenovic [G. C] vor, eine DFS (depth first search) Traversierung im Netzwerk anzuwenden. Dies ist zu vergleichen mit einem Token, das durch das Netzwerk läuft. Nur der Knoten, der gerade das Token hält, darf etwas senden. Für jeden Knoten existieren drei Flags: „Clusterhead“, „bedeckt“ oder „unentschieden“. Das Flag „bedeckt“ ist gesetzt, wenn sich der Knoten für ein Cluster entschieden hat, während ansonsten das Flag „unentschieden“ gesetzt ist. Anschließend können sämtliche Kriterien zur Wahl des Clusterheads verwendet werden, die ich bereits in den vorigen Kapiteln vorgestellt habe. Chen und Stojmenovic nennen sie:

- k-basic Clustering-Algorithmus:

  Wenn ein Knoten besucht wird, der „unentschieden“ ist, wird er zum Clusterhead, und all seine k Hops entfernten Nachbarn sind „bedeckt“. Der Algorithmus wandert zum nächsten „unentschiedenen“ Knoten und macht diesen wiederum zum Clusterhead, usw. Da dieser Algorithmus Clusterheads im Abstand von k+1 produziert, tauchen hier viele Grenzknoten auf, das heißt Knoten, die in mehr als einem Cluster liegen.

- k-lowestID Clustering-Algorithmus:

  Hier wird der Nachbar von einem Knoten u zum Clusterhead gewählt, der höchstens k Hops entfernt ist, den Status „unentschieden“ und den geringsten ID aufweist. Der durchschnittliche Intercluster-Abstand wird bei diesem Algorithmus ansteigen und die Anzahl der Grenzknoten sinken.

- k-degree Clustering-Algorithmus:

  Hier wird das kombinierte Kriterium aus Verbindungsgrad und ID verwendet. Dieser Algorithmus wählt Knoten mit vielen Nachbarn als Clusterheads und somit wird die durchschnittliche Clustergröße erhöht.

- k-farthest Clustering-Algorithmus:

  Dieser versucht, einen größtmöglichen Abstand zwischen benachbarten Clusterheads zu erreichen. Er wählt den Nachbarn u vom Knoten v zum Clusterhead, der die größte Anzahl von Nachbarknoten im Umkreis von k Hops hat. Man zählt also die Nachbarn von u und nicht von v! Im Falle von Gleichheit wird wiederum der ID als Kriterium verwendet. Statt einer Wahl des Knotens, der den höchsten Verbindungsgrad aufweist, vermeidet es dieser Algorithmus, zwei Knoten als Clusterhead zu wählen, die viele gemeinsame Grenzknoten haben. Es werden also Knoten gewählt, die einen hohen Verbindungsgrad aufweisen und einen großen Abstand zum bereits vorhandenen Clusterhead haben. Allerdings hat die resultierende, geringere Anzahl von Grenzknoten auch ihren Preis, da viel mehr Informationen über die Umgebung der Nachbarknoten gesammelt werden müssen, was einen großen Ressourcen-Aufwand darstellt.

3.1.5 Clustering basierend auf dem Gewicht eines Knotens

Die meisten Algorithmen haben einen entscheidenden Nachteil: Sie fordern die Bedingung, dass sich die Knoten im Erzeugungsprozess der Cluster nicht bewegen. Dies ist eine bedeutende Einschränkung, da man gerade für Ad-hoc-Netze keine derartige Bedingung an die Mobilität stellen kann. Basagni [Basa98] stellt einen Algorithmus DCA (distributed clustering algorithm) vor, der keine Bedingung an die Mobilität stellt und zudem noch ein neues Kriterium benutzt. Er wählt den Clusterhead aufgrund eines Gewichtes aus, für welches eine reelle Zahl > 0 bestimmt wird. Je größer das Gewicht, desto besser eignet sich der Knoten für die Rolle des Clusterheads. Dieser Parameter wird aufgrund der Mobilität des Knotens bestimmt. In erster Linie eignet sich der DCA allerdings nur für quasistatische Netzwerke, in denen sich die Knoten nicht oder nur wenig bewegen. Doch auch für hochdynamische Netzwerke präsentiert Basagni den DMAC (Distributed Mobility Based Clustering), bei dem die Knoten bei einer Veränderung der Netzstruktur lokal über ihre Rolle als Clusterhead oder gewöhnliche Knoten entscheiden. Der DMAC erfüllt alle drei zu Anfang des Kapitels genannten Bedingungen und darüber hinaus bietet er folgende Möglichkeiten, die sonst in keiner anderen Lösung vorhanden waren:

- Die Wahl der Clusterheads basiert auf dem Gewicht eines Knotens. So ist es möglich, bestimmte Präferenzen zu definieren, die die Eignung eines Knotens zum Clusterhead ausdrücken.
- DMCA passt sich den Veränderungen im Netzwerk an, falls Knoten das Netz verlassen, hinzukommen oder sich im Netz bewegen. Die Knoten können sich sogar während des Clustering-Prozesses bewegen.
- Ein Knoten entscheidet seine Rolle allein auf Basis der Kenntnis seiner einen Hop entfernter Nachbarn.

Es gibt zwei Arten von Nachrichten, zum einen CH(v), um die Nachbarn von v zu informieren, dass v die Rolle des Clusterhead übernehmen will, und zum anderen JOIN(v, u), um den Nachbarn von v mitzuteilen, dass v sich dem Cluster anschließt, dessen Clusterhead u ist. Die Hauptidee des DCA und DMCA ist, dass ein Knoten erst entscheidet, welche Rolle er übernimmt, wenn all seine Nachbarn mit größerem Gewicht sich für eine Rolle entschieden haben. Der DMCA stellt zusätzlich diverse Methoden zur Verfügung, die der Dynamik des Netzes Rechnung tragen, also bei einer Bewegung, beim Ein- und Ausschalten von Knoten oder beim Hinzukommen oder Verlassen des Netzes.

3.2 Effizienz der Clustering Algorithmen

Die Effizienz eines Clustering-Algorithmus wird durch die durchschnittliche Anzahl der produzierten Cluster, durch die durchschnittliche Anzahl der Grenzknoten und durch die durchschnittliche Größe der Cluster bestimmt. Chen und Stojmenovic [G. C] definieren für ihre Messungen folgende Variablen: die Größe des Netzwerkes n (Anzahl der Knoten) und die Verbundenheit des Netzes (der durchschnittliche Verbindungsgrad eines Knotens), die mit der Übertragungsreichweite zusammenhängt. Sie verwenden zufällige Graphen. Um die Verbundenheit des Netzes zu erhöhen, wird einfach der Übertragungsradius jedes Knotens erhöht. Sie führen Tests mit einer Anzahl zwischen 50 und 1000 Knoten und einem durchschnittlichen Verbindungsgrad zwischen 4 und 12 durch. Für jede dieser Konfigurationen wurden 10 Zufallsgraphen erzeugt und einige der oben beschriebenen Clustering-Algorithmen angewandt. Für k, also die maximale Entfernung der Knoten zum Clusterhead, werden die Werte 1 und 2 eingesetzt.

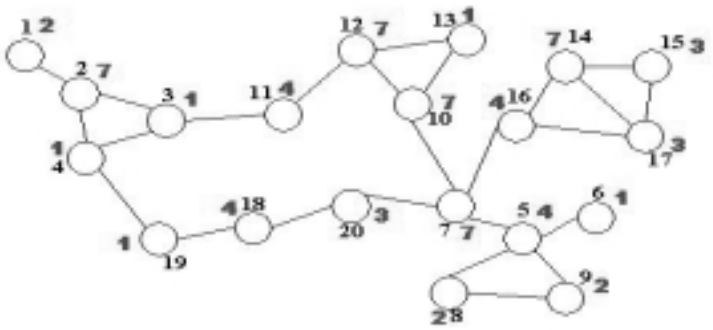

Abbildung 7: Clustering basierend auf dem Gewicht eines Knotens

Die Resultate zeigen einen klaren Vorteil der kombinierten Algorithmen gegenüber dem Algorithmus, der nur den geringsten ID als Kriterium benutzt. Man beachte, dass für die Untersuchungen Zufallsgraphen verwendet wurden. Generell kann man keine Dominanz eines Algorithmus feststellen. Je nach Topologie und Dynamik der Knoten können sich unterschiedliche Algorithmen in den jeweiligen Situationen besser eignen.

Eine intensivere Untersuchung der Ergebnisse zeigten einige lineare Beziehungen. Beispielsweise ist die Anzahl der erzeugten Cluster bei jedem Algorithmus eine lineare Funktion der Größe des Netzwerkes (falls der Verbindungsgrad konstant bleibt) oder eine lineare Funktion des Verbindungsgrades (wenn die Größe konstant ist).

3.3 Aufbau einer Cluster-Hierarchie

Eine hierarchische Netzstruktur ist ein effektiver Weg, ein Netzwerk mit einer großen Anzahl von Knoten zu organisieren. Speziell zum Zwecke des Routings muss jeder Knoten einige Informationen über den Aufenthaltsort der anderen Knoten sammeln. Falls aber die genauen Standorte bekannt sein müssten, so wäre es nötig, große Datenmengen über das Netz zu schicken, in den Knoten zu verarbeiten und in Listen zu speichern. Bei einer Hierarchie von Clustern können bestimmte Details versteckt werden, so dass eine viel geringere Datenmenge verarbeitet werden muss. Jeder Knoten kennt dann nicht mehr den genauen Standort jedes anderen Knotens, sondern nur das Cluster einer bestimmten Ebene, je nachdem wie weit der Knoten entfernt ist. Mit zunehmender Höhe der Ebene vereinen die Cluster immer mehr Knoten. In einer einstufigen Hierarchie werden die Knoten in Cluster eingeteilt, die Clusterheads haben können. Dies eignet sich für Netze mit einigen hundert Knoten. Eine mehrstufige Hierarchie [Laue95][Laue86][ShWe87] organisiert die Knoten in einer baumartigen Struktur mit mehreren Ebenen von Clusterheads. Eine dreistufige Hierarchie besteht aus gewöhnlichen Knoten, Clusterheads und Super-Clusterheads und eignet sich für Netzwerke mit einigen tausend Knoten. Die Cluster müssen derart gebildet werden, so dass alle Cluster verbunden sind.

Bei festen Netzen lassen sich die Clustering-Hierarchien bequem offline berechnen. In einem mobilen Netzwerk kann die Cluster-Hierarchie nur berechnet werden, wenn die Verbindungen zwischen den Knoten bekannt sind. In ad hoc Netzen trifft diese Bedingung nur selten dazu. Um die benötigten Rechenmittel zu reduzieren, werden Clustering Algorithmen für Ad-hoc-Netze im Gegensatz zu festen Netzen vereinfacht. Im Folgenden möchte ich zwei Beispiele für Clustering-Algorithmen beschrieben, die speziell für mobile Netze entwickelt wurden. Beide

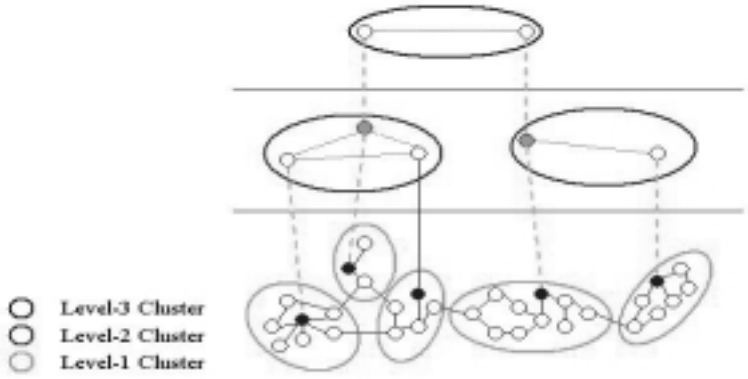

Abbildung 8: Aufbau einer Cluster-Hierarchie

Algorithmen haben allein das Ziel, verbundene Cluster zu erzeugen. Der zweite versucht auch noch der Beschränkungen der Cluster-Größe und der Veränderlichkeit der Netzstruktur Rechnung zu tragen.

Lauer [Laue86] beschreibt einen Algorithmus für SURAN-Netzwerke, bei dem alle Knoten in Zusammenarbeit einen Clusterhead und Super-Clusterheads wählen (es wird kein spezieller Wahlalgorithmus angegeben). Anschließend schließt sich jeder Knoten seinem nächsten Clusterhead an und zusammen bilden sie ein Cluster. Jeder Clusterhead schließt sich dann einem Super-Clusterhead an und bindet auch die Mitglieder seines Clusters mit ein. Alle Knoten, die zu einem bestimmten Super-Clusterhead gehören, formen ein Super-Cluster.

Ramanthan und Steenstrup [R. R98] geben einen Clustering-Algorithmus für MMWN (Multihop Mobile Wireless Networks), welche ich aus Gründen der Einfachheit nur für eine zweistufige Hierarchie beschreibe. Der Algorithmus benutzt Informationen über die Verbindungen von allen Netzwerkknoten und verwendet eine rekursive Halbierung, um verbundene Level-1-Cluster zu erzeugen, die eine obere und untere Schranke der Größe einhalten. Ein einzelner Knoten, der Cluster Leader, führt das Clustering durch. Der Cluster Leader ist der Knoten mit dem niedrigsten ID in einer Gruppe von sich gegenseitig erreichbaren Knoten. Allerdings ist jeder Knoten dazu in der Lage, die Rolle des Cluster Leaders zu übernehmen, wenn die Notwendigkeit besteht. Zu Anfang wählt der Cluster Leader zwei „Samen“-Knoten aus und führt eine Halbierung des ursprünglichen Initial-Clusters durch. Die folgende Heuristik für die Wahl der Samen hat die Aufgabe, einen gewissen Abstand dieser Knoten voneinander zu gewährleisten. Folglich hilft sie dabei, ausgeglichene Cluster zu erzeugen. Zuerst wird zufällig ein Knoten ausgewählt. Anschließend sucht der Algorithmus einen zweiten Knoten mit maximalem Abstand zu dem ersten. Dann wird ein dritter Knoten mit maximalem Abstand vom zweiten bestimmt. Der zweite und dritte Knoten sind die Samen für zwei neue Cluster. Um die beiden Cluster zu erzeugen, pendelt der Cluster-Leader zwischen den beiden Clustern hin und her und versucht, zu den Clustern weitere Knoten hinzuzufügen, die sich noch keinem der Cluster angeschlossen haben. Dies geschieht so lange, bis jeder Knoten Mitglied in einem der beiden Cluster ist. Falls die Größe eines Clusters die Schranke überschreitet, halbiert der Cluster-Leader wiederum sein Cluster mit dem gerade beschriebenen Algorithmus. Diese Prozedur wird so lange wiederholt, bis die Größen aller Cluster in den vorgeschriebenen Schranken liegen. Innerhalb jedes Clusters werden die Knoten von dem Cluster Leader über ihre Cluster-Identität informiert. Die Cluster-Identität entspricht der Identität des Cluster-Leaders. Knoten mit Links zu anderen Clustern versuchen, virtuelle Gateways aufzubauen. Jeder virtuelle Gateway besteht aus mindestens einem, vorzugsweise mehreren Links, die da-

zu beitragen, die Verbindungen zwischen den Clustern aufrechtzuerhalten. Cluster höheren Levels können gebildet werden, indem man die Cluster-Leader wiederum clustert.

4 Methoden zur Aufrechterhaltung von Clustern

In mobilen drahtlosen Netzwerken sind aufgrund ständiger Veränderungen die Verbindungen eine höchst vergängliche Ressource. Viele Umweltbedingungen beeinflussen die Qualität der Verbindung, wie z.B. der Abstand, die Anzahl und Stärke von benachbarten Übertragungen, andere Störterme wie Wetter, Terrain und Vegetation. Controller verbrauchen Ressourcen von Speicher, Bandbreite und Prozessorleistung, immer wenn sich das Netzwerk ändert, und sie diese Änderung aufzeichnen, darüber berichten und darauf reagieren müssen. Viele Reaktionen auf Veränderungen sind überflüssig, ineffizient, verschwenderisch und sogar kontraproduktiv. Die angemessene Sensibilität eines Controllers hängt von den vorhandenen Ressourcen, den benötigten Kontrollfunktionen, der Veränderlichkeit des Netzwerk-Status und dem erwarteten Ausmaß der Konsequenzen einer Veränderung ab.

Sind die Cluster erzeugt, entfällt die Bedingung, dass sich die Knoten nicht bewegen dürfen. Nun werden Techniken benötigt, die unter Berücksichtigung der Knotenmobilität die Cluster-Struktur aufrecht erhalten. In [BaEp81][DJBFl84] wird die Reorganisation periodisch durchgeführt. Es wird also die Methode zur Erzeugung der Cluster einfach erneut ausgeführt. Diese Möglichkeit ist nur bedingt einsetzbar, da sie bei häufigen Strukturänderungen äußerst aufwendig ist und sich somit nur bei quasi-statischen Netzwerken eignet. Alle Knoten müssen in der Lage sein, die Funktionen des Clusterheads und Gateways zu übernehmen. Die Algorithmen von Ephremides und Wieselthier [AEBa87], Gerla und Tsai [GeTs95] berechnen die Cluster-Zugehörigkeit und Clusterheads und Gateways immer dann neu, wenn ein Knoten sich aus einem oder in ein Cluster bewegt. Es ist offensichtlich, dass bei der Clustering-Methode, die auf der Verbindung eines Knotens basiert, sich der Clusterhead wesentlich häufiger ändert als bei der Methode, die auf dem ID basiert. Daher ist letztere Methode wesentlich stabiler.

Chiang [Chia96] präsentiert den Least Cluster Change Algorithmus, welcher nur unter zwei bestimmten Bedingungen den Status des Clusterheads ändert:

1. wenn zwei Clusterheads in die Nähe voneinander geraten, muss einer der beiden den Status als Clusterhead abgeben

2. ein Knoten hat keine Verbindung zu jeglichen Clustern und er wird selber Clusterhead

Diese Methode ist eine große Verbesserung in Sachen Stabilität im Vergleich zu den vorigen Algorithmen, deren Clusterheads bei jeder Veränderung der Netzwerkstruktur wechseln. Auch Lin und Gerla [LiGe97b][LiGe97a] schlagen einen Algorithmus vor, der nur kleine Anpassungen statt einer kompletten Restrukturierung vornimmt.

In [LiGe97a] wird eine Methode beschrieben, wie jeder Knoten selbst entscheidet, ob die durch ihn hervorgerufene Veränderung eine Reorganisation des Netzwerkes erfordert oder nicht. Dafür benötigt er nur die Kenntnis seiner einen und zwei Hops entfernten Nachbarn und Kenntnis der Topologie seines einen Hop entfernten Nachbarn mit dem höchsten Grad. Es ist also ein anderer Algorithmus als der zur Formung der Cluster und passt sich an die Mobilität der Knoten an.

Man unterscheidet vier Fälle , wie sich ein Netzwerk ändern kann. Die Anpassung der Knoten wird nach Chen und Stojmenovic [G. C] wie folgt vorgenommen:

- Ein Knoten wird angeschaltet und tritt in das Netz ein

  Er prüft, ob es in einer Entfernung von höchsten k Hops einen Clusterhead gibt, und schließt sich diesem gegebenenfalls an. Falls nicht, erzeugt er sein eigenes Cluster und lädt seine Nachbarn zum Beitritt ein.

- Ein Knoten wird ausgeschaltet und verlässt das Netzwerk

  Falls der Knoten kein Clusterhead war, wird nichts gemacht. Falls der Clusterhead ausfällt, muss ein neuer gewählt werden, und zwar nach dem Kriterium des höchsten Verbindungsgrades. Bei Gleichheit wird die Gesamtzahl der Verbindungen, also auch zu Knoten, die einem anderen Cluster angehören, geprüft. Falls immer noch keine Entscheidung getroffen werden kann, wird wiederum der ID als drittes Kriterium angewandt. Knoten, die nicht im neuen Cluster aufgenommen werden, wiederholen diese Prozedur, bis alle einem Cluster angehören. Dies kann zu Clustern mit nur einem Knoten führen. Daher erscheinen zusätzliche Funktionen zur Vereinigung und Neuorganisation von Clustern sinnvoll.

- Eine Verbindung wird unterbrochen

  Falls die beiden Knoten zu unterschiedlichen Clustern gehörten, geschieht wiederum nichts. Ansonsten werden alle Knoten innerhalb des Clusters informiert und der Clusterhead vergewissert sich, ob alle Knoten in seinem Cluster höchstens k Hops von ihm entfernt sind. Wenn dies zutrifft, passiert nichts. Wenn nicht, erzeugen die Knoten, die weiter entfernt sind, neue Cluster.

- Eine Verbindung wird geschaffen, indem sich zwei Knoten nähern

  In diesem Fall müssen mehrere Möglichkeiten betrachtet werden: Wenn keiner der zwei Knoten A und B ein Clusterhead war, wird die Struktur nicht geändert. Falls beide Clusterheads sind, müssen sie nach den oben genannten Kriterien entscheiden, wer die Rolle behalten darf. Die Knoten des anderen Clusters schließen sich dem Sieger dieses Vergleiches an, wenn sie höchstens k Hops entfernt sind, und erzeugen ansonsten eigene Cluster.

Diese vorgestellten Möglichkeiten zur Erhaltung, wie sie auch in [LiGe97a] beschrieben werden, erzeugen nach mehrfacher Anwendung eine schlechte Qualität der Cluster-Struktur. Die Qualität eines Clusters messen Chen und Stojmenovic [G. C] anhand seiner Größe (der Anzahl der Knoten) und der Anzahl der Grenzknoten. Sie schlagen drei mögliche Bewertungsergebnisse vor: exzellente, mittlere und schlechte Qualität. Die Schwellenwerte sollten durch Kriterien bestimmt werden, die von der Anzahl der Knoten, dem Mobilitätsmuster und der erwarteten Anzahl von Reorganisationen für die gewählten Schwellenwerte abhängen. Ein Beispiel für ein Cluster mit schlechter Qualität ist eines, das einen Clusterhead und nur Grenzknoten enthält. Falls ein Knoten nach einer Methode zur Erhaltung in einem schlechten Cluster liegt, wird er einen globalen Restrukturierungsprozess initiieren, also einen der Algorithmen zur Erzeugung der Cluster erneut aufrufen. Um den Datenverkehr nicht unnötig in die Höhe zu treiben, werden Knoten aus einem exzellenten Cluster diesen Aufruf zurückweisen und ignorieren. Dies beschränkt den Algorithmus allein auf die Nachbarschaft von schlechten Clustern. Die mittelmäßigen Cluster beteiligen sich am Restrukturierungsprozess in der Hoffnung, ihre Qualität zu verbessern. Einige der Cluster mit schlechter Qualität werden nicht in der Lage sein, ihre Qualität ohne die Mitarbeit der exzellenten Cluster zu verbessern. Sie werden dann den global besseren Status hinnehmen und keine weiteren Restrukturierungsanfragen stellen, bis eine weitere Veränderung in ihrem Cluster auftaucht.

Bei einer Bewegung von Knoten müssen zudem Informationen ausgetauscht werden, damit jeder Knoten des Netzwerkes über die neue Position dieses Knotens Bescheid weiss. Da dies

wiederum den Datenverkehr im Netz erhöht, wurden Methoden entwickelt, die den Knoten selbst die Entscheidung überlassen, ob sie durch ihre Bewegung ihre Adresse ändern müssen oder nicht. In einer Cluster-Hierarchie beispielsweise kann der Knoten innerhalb eines Clusters der untersten Ebene seine Adresse behalten. Wandert er jedoch über die Grenzen dieses Clusters hinaus, muss er seinen Clusterhead der untersten Ebene darüber informieren. Der Clusterhead der nächsthöheren Ebene braucht allerdings keine neuen Informationen, wenn sich der Knoten weiterhin in seinem Cluster befindet.

5 Schlusswort

Ich bin in meiner Arbeit der Frage nachgegangen, wie Kontrollstrukturen, die auf Clustern basieren, große dynamische Netzwerke dazu befähigen, ein geteiltes Medium zu managen, Routing Backbones zu erzeugen und eine vereinfachte Abbildung des Netzwerkes zu schaffen, und wie sie deswegen einen großen Einfluss auf die Netzwerk-Effizienz und letztendlich auch die Leistung ausüben. Obwohl die Idee von cluster-basierten Kontrollstrukturen bereits seit 30 Jahren existiert, warten sie bis heute auf ihren vollen Einsatz in jeglichen Kommunikationsnetzwerken. Nur einige wenige der vorgestellten Kontrollstrukturen wurden bisher in großen Netzen eingesetzt und erprobt. Die übrigen wurden nur durch rein mathematische Modellierung, Simulation oder beschränkte Experimente in kleinen Labortestnetzwerken eingesetzt.

Zwei Eigenschaft von Ad-hoc-Netzen machen sie zu den idealen Kandidaten für clusterbasierte Kontrollstrukturen: bewegliche Teilnehmer, was zu einer ständigen Veränderung der Netzwerkverbindungen führt und drahtlose Verbindungen, die sich das Medium teilen müssen. Mit der zunehmenden Verbreitung von drahtlosen mobilen Netzwerken, die heute als Ad-hoc-Netze bezeichnet werden, wird ein großes Interesse in den Studien und der Anwendung von cluster-basierten Kontrollstrukturen erweckt. Diese sollen dazu dienen, die Leistung dieser Netzwerke durch eine geringere Sensibilität bei kleinen Veränderungen und durch eine nur lokale Anpassung bei einer bedeutsamen Änderung des Netzwerkes zu verbessern.

Die Methoden zur Erhaltung der Cluster-Struktur sollten intensiver und umfassender studiert werden. Die Tests der Clustering-Algorithmen, die von Chen und Stojmenovic durchgeführt wurden, verwendeten für den Parameter k, der die Anzahl von Hops eines Clustermitglieds zum Clusterhead ausdrückt, nur die Werte 1 und 2. Größere Werte erzeugen einen erheblich größeren Datenverkehr im Erzeugungs- und Erhaltungsprozess der Cluster und werden nur durch eine große Anzahl von Knoten gerechtfertigt. Ab einer Anzahl von 1000 Knoten erscheint eine mehrstufige Hierarchie angebracht.

Einige der beschriebenen Algorithmen könnten weiter entwickelt werden. So wäre zum Beispiel laut Chen und Stojmenovic [G. C] eine (m, t)-Variante dieser Algorithmen denkbar. Jeder Knoten, der noch keine Entscheidung getroffen hat, prüft all seine Nachbarn im Umkreis von m Hops, die sich ebenfalls noch nicht entschieden haben und wählt denjenigen zum Clusterhead, der die größte Anzahl von Nachbarn in einer Entfernung von t Hops hat. Insbesondere die Fälle (1,2)-Clustering und (2,1)-Clustering sollten weiter untersucht werden.

Literatur

[AEBa87] J. E. Wieselthier A. Ephremides und D. J. Baker. A design concept for reliable mobile radio networks with frequency hopping signaling. *Proceedings of the IEEE* 75(1), 1987, S. 56–73.

[BaEp81] D. J. Baker und A. Ephremides. The architectural organization of a mobile radio network via a distributed algorithm. *IEEE Transactions on Communications* 29(11), 1981, S. 1694–1701.

[Basa98] S. Basagni. Distributed and mobility-adaptive clustering for ad hoc networks. *Technical Report UTD/EE-02-98, Erik Jonsson School of Engineering and Computer Science*, 1998.

[Chia96] C.-C. Chiang. Routing in Clustered Multihop, Mobile Wireless Networks. *Proceedings of ICOIN 11*, 1996.

[DJBFl84] A. Ephremides D. J. Baker und J. A. Flynn. The design and simulation of a mobile radio network with distributed control. *IEEE Journal on Selected Areas in Communication* 2(1), 1984, S. 226–237.

[Ephr83] A. Ephremides. Design concepts for a mobile-user radio network. *Computers & Electrical Engineering* 10(3), 1983.

[G. C] I. Stojmenovic G. Chen. Clustering and routing in mobile wireless networks.

[GeTs95] M. Gerla und J.T.C. Tsai. Multicluster, mobile, multimedia radio network. *Wireless networks* 1(3), 1995, S. 255–265.

[Kapl96] E.D. Kaplan. Understanding GPS: Principles and Applications. *Artech House, Boston, MA*, 1996.

[KlKa77] I. Kleinrock und F. Kamoun. Hierarchical routing for large networks. *Computer Networks* Band 1, 1977.

[Laue86] G. Lauer. Hierarchical routing design for SURAN. *Proc. ICC*, 1986.

[Laue95] G. Lauer. Packet-radio routing. *in: Routing in Communication Networks, ed. M. Steenstrup*, 1995.

[LiGe97a] C.R. Lin und M. Gerla. Adaptive clustering for mobile wireless networks. *IEEE Journal on Selected Areas in Communication* 15(7), 1997, S. 1265–1275.

[LiGe97b] C.R. Lin und M. Gerla. Multimedia Transport in Multihop Dynamic Packet Radio Networks. *IEEE ICNP*, 1997.

[LiSt98] Xu Lin und I. Stojmenovic. Geographic distance routing in ad hoc wireless networks. *IEEE Journal on Selected Areas in Communication*, 1998.

[NaIm97] J.C. Navas und T. Imielinski. GeoCast - geographic addressing and routing. *Proc. MOBICOM*, 1997.

[R. R98] M. Steenstrup R. Ramanathan. Hierarchically-organized, multihop mobile wireless networks for quality-of-service support. *ACM/Baltzer Mobile Networks and Applications Journal* 3(1), 1998, S. 101–119.

[Shar96] J. Sharony. An Architecture for Mobile Radio Networks with Dynamically Changing Topology Using Virtual Subnets. *ACM Mobile Networks And Applications (MONET)* 1(1), 1996, S. 75–86.

[ShWe87] N. Shacham und J. Westcott. Future directions in packet radio architectures and protocols. *Proc. of the IEEE* 75(1), 1987.

[WTNi89] W.K. Tsai W.T. Tsai, C.V. Ramamoorthy und O. Nishiguchi. An adaptive hierarchical routing protocol. *IEEE Trans. Commun.* 38(8), Januar 1989, S. 1059–1075.

[Zavg97] J. Zavgren. NTDR Mobility Management Protocols and Procedures. *Proceedings of the IEEE Military Communications Conference (MILCOM '97)*, 1997.

Abbildungsverzeichnis

1 Clustering zur Verwaltung von Übertragungen 96

2 Near-Term-Digital Radio-Network 97

3 Clustering für ein effizientes Routing 98

4 k-Cluster 100

5 Clustering basierend auf dem Identifier 101

6 Clustering basierend auf der Anzahl von Verbindungen 102

7 Clustering basierend auf dem Gewicht eines Knotens 105

8 Aufbau einer Cluster-Hierarchie 106