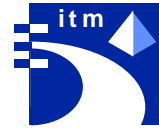


Universität Karlsruhe (TH)
Fakultät für Informatik
Institut für Telematik
76128 Karlsruhe



Klausurtagung des Instituts für Telematik



Schloss Dagstuhl, Wadern

29. März 2000 – 1. April 2000

Herausgeber:
Prof. Dr. Dr. h.c. mult. Gerhard Krüger
Roland Bless
Hartmut Ritter
Dr. Jochen Schiller
Rainer Ruggaber

Universität Karlsruhe (TH)
Institut für Telematik
<http://www.telematik.informatik.uni-karlsruhe.de/>

Fakultät für Informatik
Interner Bericht 2000-6
ISSN 1432-7864

Kurzfassung

Der vorliegende Bericht gibt einen Überblick über aktuelle Forschungsarbeiten des Instituts für Telematik an der Universität Karlsruhe (TH). Das Institut für Telematik ist in einem Teilgebiet der Informatik tätig, welches durch das Zusammenwachsen von Informatik und Kommunikationstechnik zur Telematik geprägt ist. Es gliedert sich in die Forschungsbereiche Telematik, Telecooperation Office (TecO), Cooperation & Management, Hochleistungsnetze und Netzwerkmanagement sowie dezentrale Systeme und Netzdienste.

Die Schwerpunkte des Forschungsbereichs „Telematik“ (Prof. Dr. Dr. h.c. mult. G. Krüger) liegen in den Bereichen „Dienstgüte“, „Mobilkommunikation“ und „Verteilte Systeme“. Gemeinsames Ziel ist die Integration heterogener Netze (Festnetze und Funknetze), Rechnersysteme (von Workstations bis zu PDAs) und Softwarekomponenten, um damit den Anwendern eine Vielzahl von integrierten Diensten effizient und mit größtmöglicher Qualität zu erbringen.

Das „Telecooperation Office“ (TecO, Prof. Dr. Dr. h.c. mult. G. Krüger) ist ein Institutsbereich, der in Zusammenarbeit mit der Industrie anwendungsnahe Forschungsthemen der Telematik aufgreift. Im Mittelpunkt steht die innovative Nutzung von Kommunikationsinfrastrukturen mit den Schwerpunkten Softwaretechnik für Web-Anwendungen, neue Formen der Telekooperation sowie tragbare und allgegenwärtige Technologien (Ubiquitous Computing).

Die Kernkompetenz des Forschungsbereichs „Cooperation & Management“ (Prof. Dr. S. Abeck) liegt im „prozessorientierten Netz-, System- und Anwendungsmanagement“. Es werden werkzeuggestützte Managementlösungen für Betriebsprozesse entwickelt und in realen Szenarien erprobt. Ein wichtiges Szenario stellt das multimediale Informationssystem „NEXUS“ dar, das als Plattform eines europaweit verteilten Lehr- und Lernsystems genutzt wird.

Der Forschungsbereich „Hochleistungsnetze & Netzwerkmanagement“ (Prof. Dr. W. Juling) befasst sich mit Technologie und Konzepten moderner leistungsfähiger Netzwerke sowie darüber hinaus mit sämtlichen Aspekten des Managements dieser zumeist ausgedehnten Netze. Um eine enge Abstimmung zwischen Forschungsaktivitäten und betrieblicher Praxis zu erzielen, werden insbesondere auch Synergien zwischen Institut und Rechenzentrum angestrebt.

Die Arbeiten des Forschungsbereichs „Dezentrale Systeme und Netzdienste“ (Prof. Dr. L. Wolf) befassen sich mit der Unterstützung verteilter Multimedia-Systeme, auch unter Berücksichtigung von Komponenten mit drahtlosem Zugang und den dafür geeigneten Architekturen und Infrastrukturen. Dabei werden vor allem Aspekte der Kommunikationssysteme wie Protokollmechanismen, Ressourcenverwaltung und adaptive und heterogene Systeme untersucht.

Vorwort

Der vorliegende Tagungsband entstand im Rahmen der achten Klausurtagung des Instituts für Telematik, die vom 29. März bis zum 1. April 2000 abgehalten wurde. Er dokumentiert in kurzen Beschreibungen die Forschungsgebiete der wissenschaftlichen Mitarbeiter des Instituts.

Die Kürze der einzelnen Beiträge erlaubt einen schnellen, aber prägnanten Einblick in die einzelnen Arbeitsgebiete. Der Preis dafür ist die Vernachlässigung konzeptioneller und technischer Detailinformationen, ohne die allerdings wissenschaftliche Arbeit nie praxisnahe Resultate liefern könnte. Dass auch diese Details existieren, zeigt ein Blick in die zitierten Veröffentlichungen, das Durcharbeiten der Studien- und Diplomarbeiten bzw. Dissertationen in den einzelnen Gebieten oder auch der persönliche Austausch mit den jeweiligen Autoren. Der vorliegende Band kann dabei eine zuverlässige Orientierungshilfe im weiten Feld der wissenschaftlichen Arbeit des Instituts sein.

Dieser Bericht verdankt seine Existenz wieder einmal der gemeinsamen Anstrengung aller Mitarbeiter des Instituts. Durch die Verwendung vorgegebener Schablonen konnte die Zusammenstellung der einzelnen Beiträge zu dem vorliegenden Tagungsband ohne größere Probleme durchgeführt werden. Ohne den Einsatz aller Mitarbeiter wäre das jetzige homogene Bild des Bandes nie zustande gekommen.

Die Tagung fand erstmals in den Räumen von Schloss Dagstuhl statt, einem sehr angenehmen Ort für derartige Veranstaltungen. Während Planung und Durchführung in den Händen von Mitarbeitern des „Institutsteils Krüger“ koordiniert wurden, engagierten sich alle Institutsteile an den verschiedenen Programmpunkten, angefangen von Tutorials über Podiumsdiskussionen bis hin zu Berichten aus der Forschungsförderung. Und nebenher dient eine solche Tagung natürlich auch dem besseren gegenseitigen Kennenlernen. Herrn Prof. Dr. Dr. h.c. mult. G. Krüger sei ausdrücklich gedankt für die Unterstützung bei der Planung und Umsetzung eines wieder etwas anderen Konzepts des Tagungsablaufs und nicht zuletzt für die Anwerbung von finanzieller Unterstützung, ohne die die Tagung nicht durchführbar gewesen und damit auch dieser Tagungsband nicht entstanden wäre.

Karlsruhe, im März 2000

*Roland Bless, Hartmut Ritter,
Rainer Ruggaber, Jochen Schiller*

Inhaltsverzeichnis

Kurzfassung	i
Vorwort	iii
<i>Marc Bechler:</i>	
Dienstgüteunterstützung in heterogenen Ad-hoc-Netzwerken mittels Multiplex-Mechanismen	1
<i>Michael Beigl:</i>	
Kommunikation in interaktiven Räumen	5
<i>Roland Bless:</i>	
Integrierte Managementarchitektur für Differentiated-Services-Netze	11
<i>Martin Gaedke:</i>	
Vorgehensmodell und Prozeßunterstützung für komponentenbasiertes Web Engineering	17
<i>Meng Gan:</i>	
Optimierung der Lastkontrolle in Kommunikationssystemen durch adaptive, CI-basierte Verfahren	23
<i>Hans-Werner Gellersen:</i>	
Context-Awareness Technologies for Situated Computing	27
<i>Guntram Graef:</i>	
Adaptive Web-Anwendungen	31
<i>Markus Lauff:</i>	
Unterstützung für Anwendungen in ubiquitären Systemen	35
<i>Christian Mayerl:</i>	
Eine integrierte Dienstmanagement-Architektur für die qualitätsgesicherte Bereitstellung von Netz- und Systemdiensten	39
<i>Daniel Müller:</i>	
Fragestellungen bei mobilen drahtlosen Ad-hoc-Netzwerken	43
<i>Frank Pählke:</i>	
FATIMA: Eine Architektur zur sicheren Mobilitätsunterstützung in IP-Netzen	47
<i>Hartmut Ritter:</i>	
Bedarfsorientierte Dienstgüteunterstützung durch adaptive Endsysteme . .	53

<i>Rainer Ruggaber:</i>	
Π²: Unterstützung mobiler und drahtlos angebundener Teilnehmer in ver-	
teilten CORBA-Architekturen	57
<i>Jochen Schiller:</i>	
Dienstgüte in mobilen, drahtlosen Kommunikationssystemen	63
<i>Albrecht Schmidt:</i>	
Implicit Human Computer Interaction	69
<i>Jochen Seitz:</i>	
Mobilität und Multi-Milliarden-Teilnehmersysteme: Neue Anforderungen	
an die Netzverwaltung	77
<i>Bernhard Thurm:</i>	
Mobile Agenten zur Unterstützung des Netzwerkmanagements	81
<i>Klaus Wehrle:</i>	
Skalierbare Dienstgarantien im Internet der nächsten Generation	85
<i>Hajo R. Wiltfang:</i>	
Herausforderung Netzwerkmanagement	89

Dienstgüteunterstützung in heterogenen Ad-hoc-Netzwerken mittels Multiplex-Mechanismen

Marc Bechler

1 Einleitung

Die Mobilkommunikation erlebt derzeit einen enormen Aufschwung, was sich nicht zuletzt anhand der aktuellen Verkaufszahlen von mobilen Endgeräten dokumentieren lässt – die Anzahl der Benutzer mobiler Endgeräte wächst im Moment weit schneller als die an das Internet angeschlossenen Rechner. Aufgrund der steigenden Verbreitung ergibt sich zum einen zunehmend der Wunsch, mit diesen mobilen Geräten von jedem beliebigen Ort auf das Internet zugreifen zu können. Die Mobilität der Teilnehmer wirft dabei bereits das Problem auf, dass die Protokolle aus dem Internet die Mobilität nicht oder nur sehr unzureichend unterstützen. Diese Fragestellungen werden derzeit intensiv in der *Mobile IP Working Group* der IETF (*Internet Engineering Task Force*) diskutiert. Auch findet das *Wireless Application Protocol* (WAP, <http://www.wapforum.org>) immer weitere Verbreitung, dessen Protokolle für den Internet-Zugriff in drahtlosen und mobilen Umgebungen optimiert sind. Weiterhin gilt es zu bedenken, dass ein Gerät unter Umständen über mehrere Kommunikationsmöglichkeiten verfügt (z.B. GSM-Adapter, Infrarot-Anbindung, WaveLAN, etc.) und somit der Internet-Zugang auf mehreren Wegen möglich ist.

Zum anderen zeigt sich auch zunehmend die Notwendigkeit, dass diese mobilen Geräte untereinander in sogenannten *Ad-hoc-Netzwerken* kommunizieren möchten, etwa beim spontanen Zusammentreffen für Netzwerkspiele (ohne zuvor aufwendig zu konfigurieren) oder aber auch in der Verkehrstelematik, wenn mehrere Fahrzeuge sogenannte *Cluster* zum Austausch von Informationen bilden können. Auch in diesem Bereich treibt die Industrie die Technik rasant voran, wie dies z.B. beim *Bluetooth-Konsortium* (<http://www.bluetooth.com>) der Fall ist. Ein wesentlicher Aspekt bei diesen Ad-hoc-Netzwerken ist hierbei die Wegfindung vom Sender zum Empfänger (also das Routing), welche aufgrund der Mobilität des Senders/Empfängers wesentlich komplexere Anforderungen an den Routingalgorithmus stellt. Diese Notwendigkeit wurde bereits von der IETF erkannt und daraufhin die MANET-Arbeitsgruppe (Mobile Ad-hoc-Networks) gebildet, die sich unter anderem mit dieser Thematik beschäftigt (<http://www.ietf.org/html.charters/manet-charter.html>).

Mithilfe von flexiblen Multiplex-Mechanismen soll nun versucht werden, eine Dienstgütebasierte Kommunikation in Ad-hoc-Netzwerken zu ermöglichen. Die Dienstgüteunterstützung bezieht sich dabei hauptsächlich auf die Ende-zu-Ende-Betrachtung, d.h. das Endgerät selbst, da gerade in Ad-hoc-Netzwerken aufgrund der inhärenten dezentralen Organisation jedes Endgerät über entsprechende Mechanismen zur Weiterleitung verfügen muss. Dieser Ansatz wird im weiteren Verlauf beschrieben.

2 Multiplexing zur Dienstgüteunterstützung

Neben der Dynamik eines Ad-hoc-Netzwerkes durch bewegte Sender und Empfänger ergeben sich weitere Aspekte, die bei der Wegefindung eine wesentliche Rolle spielen. Bei den Teilnehmern eines Ad-hoc-Netzwerkes handelt es sich um mobile Endgeräte, die durch einen hohen Grad an Heterogenität gekennzeichnet sind – sowohl bezüglich ihrer Leistungsfähigkeit als auch bezüglich ihrer Kommunikationsmöglichkeiten. Diese Unterschiede werden gerade im Bereich der Verkehrstelematik deutlich, wo verschiedene Dienste auf unterschiedlichen Infrastrukturen aufsetzen. Beispielsweise können in Fahrzeugen die folgenden Kommunikationseinrichtungen vereint sein:

- Informationen über den derzeitigen Straßenzustand oder das Wetter können via DAB (*Digital Audio Broadcast*) zum Fahrzeug gesendet werden,
- aktuelle Börseninformationen werden via SMS (*Short Message Service*) direkt auf das mobile Telefon gesendet werden,
- über ein GSM-Telefon werden Telefongespräche abgewickelt,
- über eine Fahrzeug-interne Infrastruktur, etwa ein CAN- oder MOST-Bus, sind elektronische Komponenten wie Autoradio, Autotelefon oder Navigationssystem miteinander vernetzt,
- im Fahrzeug selbst können sich Laptop, PTA (*Personal Travel Assistant*) und z.B. das Navigationsgerät über sogenannte PANs (*Personal Area Networks*) synchronisieren.

Dieses Beispielszenario, welches auch dem Sonderforschungsbereich 1791 („Selbstorganisation von Verkehrssystemen durch I+K-gestützte Dienste“) zugrunde liegt, bildet die Motivation für die Realisierung eines Software-Multiplexers zur Verdeckung der Heterogenität. In Abbildung 1 ist der Aufbau dieser Multiplex-Architektur dargestellt. Diese ist in der Transport-/Netzwerkschicht lokalisiert und beinhaltet zwei Multiplexer: einen Software-Multiplexer und einen Hardware-Multiplexer.

Der *Software-Multiplexer* ist dafür verantwortlich, ein entsprechendes Transportprotokoll auszuwählen, in Abhängigkeit von den spezifizierten Anforderungen der einzelnen Anwendungen. Beispielsweise definiert WAP die Anpassung an mehrere Übertragungstechniken – bei einem dynamischen Wechsel der Übertragungstechnik kann somit eine Adressumsetzung der IP-Adresse auf eine Telefonnummer notwendig sein. Ein solcher Multiplexer wurde bereits im Rahmen einer Studienarbeit entwickelt. Der *Hardware-Multiplexer* hingegen erlaubt es, dynamisch zwischen den unterschiedlichen Übertragungsebenen hin- und herzuschalten – falls diese das gleiche Transport- bzw. Netzwerkprotokoll unterstützen. Droht z.B. der Funkkontakt einer WaveLAN-Verbindung wegzubrechen, kann rechtzeitig eine GSM-Datenverbindung geöffnet werden, um die Verbindung anschließend über diesen Kommunikationskanal weiterzuführen. Dieser Mechanismus funktioniert dabei folgendermaßen: Beim Abbruch der (Funk-)Verbindung wird der Hardware-Multiplexer von diesem Ereignis benachrichtigt. Daraufhin wird (derzeit) anhand einer statisch angelegten Tabelle eine Alternative ausgewählt. Die einzelnen Pakete werden nun über diesen neuen Kommunikationsweg zum Empfänger getunnelt, d.h. die Pakete werden in neue IP-Pakete gepackt und über den alternativen Anschluss gesendet, was in Abbildung 1 durch den hellen Pfeil dargestellt ist. Auf der Empfängerseite werden diese Pakete wieder ausgepackt und nach oben weitergereicht. Dieser Protokollmechanismus entspricht von der Grundidee her dem Mobile IP (gemäß RFC 2002), umfasst allerdings noch nicht den kompletten Funktionsumfang. Zur vollen Mobilitätsunterstützung ist es geplant, Mobile IP vollständig in den Softwaremultiplexer zu integrieren.

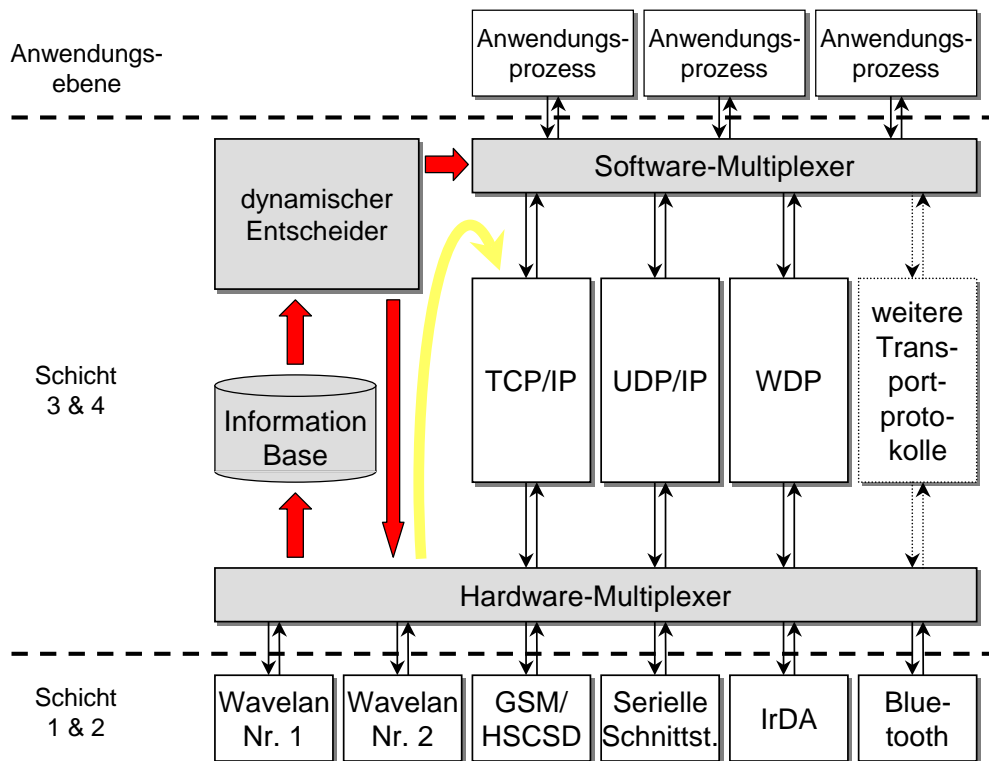


Abbildung 1: Die flexible Multiplex-Architektur

Diese Multiplex-Architektur soll im weiteren Verlauf dahingehend erweitert werden, dass in einer *Information Base* die Parameter der verfügbaren Kommunikationswege respektive den aktuellen Zuständen gespeichert werden. Der Hardware-Multiplexer ist dafür verantwortlich, diese Datenbasis aktuell zu halten, was durch den unteren dunklen Pfeil angedeutet ist. Diese Datenbasis wird von einem Modul genutzt, dem *dynamischen Entscheider*, der für die Dienstgüteunterstützung verantwortlich ist und die beiden Multiplexer entsprechend beeinflusst.

3 Erweiterungen und Ausblick

Die im vorangegangenen Kapitel beschriebene Multiplex-Architektur wurde teilweise bereits realisiert – allerdings fehlt bisher ein dynamisches Verhalten, welches im weiteren Verlauf sukzessive integriert wird. Neben diesen Erweiterungen ergeben sich weitere zentrale Fragestellungen und Forschungsansätze.

Eine zentrale Aufgabe in Ad-hoc-Netzwerken ist das Routing. Die derzeitigen Ansätze seitens der *MANET Working Group*, etwa DSDV (*Destination-Sequenced Distance-Vector*), AODV (*Ad Hoc On-Demand Distance Vector Routing*) [RoTo99], wurden bisher kaum untersucht bzw. simuliert und sind auch nicht für den Einsatz in heterogenen Umgebungen optimiert. Daher spielt die Dienstgüteunterstützung in solchen Netzwerken derzeit eine eher untergeordnete Rolle. Weiterhin stellt sich in diesem Zusammenhang die Frage, inwieweit Ansätze aus dem Internet für ein Dienstgüte-unterstütztes Routing Verwendung finden können. Gerade in Ad-hoc-Netzwerken ist es meist möglich, dass eine Station einen Übergang ins Internet ermöglicht. Daher müssen auch Ansätze zur Unterstützung von Dienstgüte im Internet, wie beispielsweise die *Differentiated Services-Architektur* (im RFC 2475 standardisiert), berücksichtigt werden, um die Dienstgüteunterstützung bis zum Endgerät zu ermöglichen. Erste Ansätze zu dieser Thematik wurden in [BeRS99] bereits vorgestellt und untersucht.

Eine zentrale Bedeutung kommt den in Abbildung 1 grau hinterlegten Komponenten zu. Es ergibt sich aufgrund der Heterogenität beispielsweise das Problem, dass die einzelnen Kommunikationseinrichtungen durch unterschiedliche technische Parameter charakterisiert sind. Diese Parameter und Zustände müssen in der Information Base hinreichend abstrakt modelliert werden, damit sie für die Entscheidung herangezogen werden können. Ebenfalls in diesem Zusammenhang steht die Frage, wie sich die unterschiedlichen Kommunikationskosten modellieren lassen und wie diese bei der Entscheidung berücksichtigt werden. Weitere Fragestellungen betreffen den *dynamische Entscheider*, dessen Aufgabe es ist, unter gegebenen Anforderungen (wie z.B. maximale Kosten, minimaler Durchsatz, begrenzter Jitter, etc.) und verfügbaren Ressourcen eine möglichst optimale Lösung zu finden – oder gegebenenfalls den Benutzer darüber zu informieren, falls ein Dienst nicht oder nur mit einer schlechteren Qualität erbracht werden kann. Hierbei gilt es zu überlegen, welche Parameter für die Entscheidung herangezogen werden müssen und wie die Optimierungsfunktion aussehen könnte.

Ein ebenfalls sehr wichtiger Aspekt ist das Zusammenspiel mit weiteren Protokollen, beispielsweise WAP oder HTTP, aber auch die Untersuchung der Auswirkungen der Mobilität (z.B. beim *handover* zwischen den verschiedenen Kommunikationstechniken) auf die Transportprotokolle und somit auch die Dienstgüteunterstützung auf Anwendungsebene.

Schließlich sollte die vorgestellte Multiplex-Architektur in eine Gesamtarchitektur zur Unterstützung von Dienstgüte in Endsystemen integrierbar sein. Diese Gesamtarchitektur ist in [BeRS00] beschrieben.

Literatur

- [BeRS99] Marc Bechler, Hartmut Ritter und Jochen Schiller. *Integration of a Traffic Conditioner for Differentiated Services in End-systems via Feedback-loops*. Broadband Communications '99, Hong Kong. 1999.
- [BeRS00] Marc Bechler, Hartmut Ritter und Jochen Schiller. *Quality of Service in Mobile and Wireless Networks: The Need for Proactive and Adaptive Applications*. 33rd Annual Hawaii International Conference on System Sciences, Maui. 2000.
- [RoTo99] Elizabeth M. Royer und Chai-Keong Toh. *A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks*. IEEE Personal Communications. April 1999.

Kommunikation in interaktiven Räumen

Michael Beigl

1 Zusammenfassung

Umgebungen nach Weisers Vision des Ubiquitous Computing erfordern neben der Ausstattung von Räumlichkeiten mit neuen Artefakten auch eine speziell auf solche lokalen Bereiche zugeschnittene Kommunikation. In diesem Artikel wird zunächst lokale Kommunikation analysiert und modelliert und dann mit dem RAUM-System (Relation von Anwendungsobjekten, die über Ubiquitäre Meldungen kommunizieren) ein Beispiel für ein lokales Kommunikationssystem vorgestellt.

2 Problemstellung

Interaktive Räume sind Orte, in denen sowohl die Umgebung mit Rechner- und Telekommunikationstechnologie ausgestattet ist, als auch die sich in ihnen befindlichen Gegenstände des Alltagsgebrauchs. Dieser Trend läßt sich an zwei Zahlen erläutern: In professionellen Anwendungen und Büroumgebungen besitzt inzwischen schon ein Mehrheit von Benutzern mehr als einen Computer [Reki98]. Im als Basistechnologie für solche Umgebungen wichtigen Bereich der eingebetteten Rechner (embedded computer) wird sich nach vorsichtigen Schätzungen die Zahl der Systeme verdoppeln [ReGr98]. Das Charakteristikum alltäglicher Gegenstände ist deren intuitive Benutzbarkeit, ohne daß dafür eine Einlernphase vorausgegangen wäre. Dadurch entsteht für den Menschen eine Umgebung mit zahlreichen Interaktionsmöglichkeiten, deren Ziel die Unterstützung des Menschen bei existierenden und zukünftigen Aufgaben ist [DeAS99]. Die Interaktion soll dabei unauffällig stattfinden, so wie dies von Mark Weiser als ubiquitäres Rechnen (ubiquitous computing) beschrieben wurde. Besonderes Kennzeichen einer solchen Umgebung ist, daß eine Verbindung im Ablaufschema der Interaktion und der lokalen Kommunikation besteht. Dieses lokale Kommunikationsschema unterscheidet sich von dem der Telekommunikation.

Existierende Rechnertechnologie ermöglicht zwar den Aufbau interaktiver Räume, allerdings ist die Grundlage für Kommunikation und Koordination der einzelnen Rechner untereinander und der Interaktion des Menschen zum Rechner noch kaum erforscht. Bestehende technische Ansätze versuchen durch unreflektiertes Übertragen von Telekommunikationstechniken interaktive Räume zu schaffen. Durch das Fehlen einer fundierten, den Interaktions- und Kommunikationsvorgängen innerhalb lokaler Umgebungen angepaßten Modellbildung, verfehlen diese Ansätze aber ihr Ziel: Sie können keine Unauffälligkeit und Intuitivität bei der Benutzung der Gegenstände zusichern. Die Modellbildung und ein darauf aufbauendes technisches Kommunikationssystem für interaktive Räume sind die beiden Kernprobleme, die für ein lokationsorientiertes Kommunikationssystem zu lösen sind:

- Kommunikation in lokalen Umgebungen unterscheidet sich grundlegend von der Telekommunikation. Es ist ein Modell zu finden, das die Gestaltung von Kommunikation

in lokalen Umgebungen beschreibt und dabei die Zusammenhänge von Kommunikation und Mensch-Rechner-Interaktion in interaktiven Räumen berücksichtigt.

- Es gilt, ein effizientes technisches Kommunikationssystem für Rechnergeräte in interaktiven Räumen zu beschreiben und zu implementieren. Dafür sind ein die Anwendung unterstützendes Kommunikationsmodell sowie entsprechende Kommunikationsprotokolle zu entwickeln.

3 RAUM-System

Das RAUM-System stellt ein neues Paradigma zur Lösung der oben aufgezeigten Problemstellung vor: Es muß bei der Kommunikation die physikalisch-räumliche Beziehung der beteiligten Objekte mit einbezogen werden. Das entspricht der natürlichen Interaktion des Menschen innerhalb eines lokal begrenzten Bereiches [KiMa94]. Die fachlichen Ansätze ergeben sich zunächst aus Beobachtungen und empirischen Studien. Diese werden dann systematisch auf ein Kommunikationsmodell übertragen. Das Modell wird in einer Architektur konkretisiert und dann in ein System umgesetzt. Als Grundlage für die Studien dienen mehrere Demonstratoren, die dem System entsprechend beispielhaft implementiert werden.

3.1 Implikationen der Interaktion auf lokale Kommunikation:

Kommunikation in interaktiven Räumen findet ebenso wie Interaktion lokal begrenzt statt und orientiert sich an der räumlichen Anordnung der Objekte. Dabei verschiebt sich der Fokus der Aufgaben der Kommunikation: Liegt bei der Telekommunikation das Hauptaugenmerk auf der Überbrückung der Kommunikationsdistanz, so ist bei einer lokal begrenzten Kommunikation die Flexibilität und Einfachheit besonders wichtig.

Für eine lokale Kommunikation sind Aspekte der direkten Mensch-Rechner-Interaktion wie Transparenz und Natürlichkeit der Kommunikation als entscheidende, den Ablauf der Kommunikation prägende Faktoren gefunden worden. Diese Aspekte betreffen nicht nur die Kommunikation selbst, sondern auch das Erstellen eines Kommunikationsraumes und die Auflösung von Kommunikationsbeziehungen. So stellt sich zum Beispiel die Frage, wie und mit welchem Computer in einem interaktiven Raum mit einer Vielzahl von Rechnern eine (kabellose) Tastatur kommunizieren soll. In den Arbeiten zum RAUM-System wird gezeigt, daß die räumliche Anordnung auch die Wahl der Kommunikationspartner festlegt: Der Rechner, vor dem die Tastatur liegt, erhält die Eingabeereignisse. Wie bei der zwischenmenschlichen Kommunikation die Interaktionspartner räumlich miteinander verbunden sein müssen (sich also zum Beispiel gegenüberstehen), so kommunizieren nur Objekte innerhalb eines bestimmten Bereiches miteinander. Als Kommunikationsmuster läßt sich ableiten, daß lokal begrenzte Kommunikation an die räumliche Anordnung der Kommunikationsobjekte gebunden ist.

3.2 Modellierung der Kommunikation:

Die sich aus dem Kommunikationsparadigma ergebende Frage, die in der Modellierung beantwortet werden soll, ist: Wie genau beeinflußt die räumliche Anordnung von kommunizierenden Objekten die Kommunikation? Es besitzen - vom Aspekt der Kommunikation aus - physikalische Rechnerobjekte und auf diesen Rechnern ablaufende Programme die gleiche Ordnung. Um diese Ordnung technisch zu erhalten, wird ein Modell zur Beschreibung der Position eines Objekts, ein sogenanntes Lokationsmodell, eingeführt, das die räumliche Umgebung vereinfacht beschreiben kann. In klassischen Telekommunikationssystemen werden die Geräte

nach den Identifikatoren (etwa Nummern oder Dienste), in räumlich orientierten Kommunikationssystemen dagegen nach ihrer räumlichen Position geordnet. Damit ein Programm in einem interaktiven Raum Kommunikationspartner auswählen kann, werden nicht die Partner identifiziert, sondern der räumliche Interessenbereich spezifiziert. Die Spezifikation von Interessenbereichen entspricht dem Kommunikations- und Interaktionsverhalten von Menschen. Die Gestaltung, Form und Ausdehnung von Interessenbereichen festzulegen, ist Aufgabe der Anwendung.

Das hier entwickelte Kommunikationsmodell, das die beschriebenen Bedingungen erfüllt, wird als RAUM-Modell bezeichnet [Beig99d]. Im RAUM-Modell werden neben den für die Interaktion wichtigen Artefakten auch Rechnersysteme zur Erhaltung der Kommunikationsinfrastruktur spezifiziert. Die Aufgabe der Infrastruktur ist es, zusätzliche Eigenschaften wie etwa gesicherte Auslieferung, Wahrung der Privatsphäre und Sicherheit, anzubieten. Die Gestaltung der Kommunikationsbeziehungen basiert auf Untersuchungen und Aufwandsberechnungen der Realisierungsmöglichkeiten.

3.3 Architektur eines Kommunikationssystems:

Das RAUM-Modell wird in meiner Arbeit in eine Kommunikationsarchitektur umgesetzt. In der Instanziierung der Architektur, dem RAUM-System, werden den verschiedenen im Modell spezifizierten Objekten konkrete Eigenschaften zugeschrieben. Weitere für die Funktion des RAUM-Systems notwendige Objekte werden ebenfalls eingeführt, das Positionsmodell wird in eine konkretere Beschreibung für das RAUM-System überführt.

Um die Kommunikation zwischen den Objekten genauer zu spezifizieren, bedarf es einer Schichtenarchitektur. Diese enthält die notwendigen Kommunikationsprotokolle, die dann für jedes der benannten Objekte innerhalb des RAUM-Systems spezifiziert werden. Die Architektur ist so entworfen, daß eine mögliche Implementierung auf einer großen Bandbreite von Rechnern, insbesondere auch auf sehr leistungsschwachen Microrechnern, umgesetzt werden kann.

3.4 Umsetzung des Kommunikationssystems und Beispielanwendungen:

Im RAUM-System werden als Komponenten mit Rechnertechnologie ausgestattete Gegenstände des Alltags sowie Geräte zur Aufrechterhaltung der Kommunikationsinfrastruktur spezifiziert. Hierbei werden Gesichtspunkte der praktischen Umsetzung eines solchen Systems berücksichtigt, da die resultierende Implementierung einen für die anschließenden Studien notwendigen hohen Praxisbezug aufweisen muß. Diese Implementierung dient dann als Basis für mehrere Beispielanwendungen, die Szenarien für die Evaluierung liefern.

4 Ergebnisse

Vergleicht man existierende Ansätze und Systeme zur Kommunikation in interaktiven Räumen, ist das vorgestellte RAUM-System das einzige, das Anwendungen eine räumlich abhängige Kommunikation erlaubt. Durch Implementierungen, Versuche und Simulationen konnten zahlreiche Vorteile der Umsetzung des Paradigmas der räumlich orientierten Kommunikation des Systems nachgewiesen werden. Die in dieser Arbeit vorgestellten Lösungen besitzen folgende Vorteile:

- Die Kommunikation der Geräte untereinander ist gegenüber dem Benutzer transparent, da Kommunikationsmuster verwendet werden, die an menschliches Interaktionsverhalten angelehnt sind. Ein Beispiele für ein solches Verhalten ist die räumliche Bezugnahme. Die Transparenz konnte in empirischen Studien unter anderem durch eine intuitivere Bedienbarkeit sowie durch eine Verbesserung der Anwendungserstellung nachgewiesen werden.
- Das System ist unauffällig, da es selbstregulierend ist und keine Eingriffe oder Einübung von Seiten des Nutzers erfordert. Diese, bei den durchgeführten Studien beobachtete Eigenschaft, resultiert aus der Robustheit des Kommunikationssystems.
- Das System ist effizient, da Modell und Architektur des RAUM-Systems für lokale Kommunikation optimiert wurden. Die Implementierung berücksichtigt darüber hinaus praktische Gesichtspunkte. Durch Studien und Leistungsberechnungen im Vergleich zu anderen Kommunikationssystemen konnte die Effizienz bestätigt werden.

Das vorgestellte System ist praxisrelevant und bildet die Grundlage zahlreicher demonstrierbarer Anwendungen wie den UbicompBrowser [BSLG98], ElectronicManual [Beig99b], MediaCup [GeBK99], AmbientTelepresence [Beig99a], AIDE [Beig99c] und SmartDoorPlate, welche zum Teil dauerhaft im Betrieb sind. Durch den Einsatz dieser Anwendungen konnte unter anderem der Anspruch der Arbeit verifiziert werden.

Literatur

- [Beig99a] Michael Beigl. Ambient Telepresence. *Proceedings of the Workshop on Changing Places*, April 1999, S. 63–67.
- [Beig99b] Michael Beigl. Electronic Manual. *Human-Computer Interaction : Communications, Cooperation and Application Design*, H-J Bullinger and J Ziegler ed. Band 2, Oktober 1999, S. 246–250.
- [Beig99c] Michael Beigl. Point and Click - Interaction in Smart Environments. *Proceedings of the 1th International Symposium on Handheld and Ubiquitous Computing (HUC99)*, September 1999, S. 311–314.
- [Beig99d] Michael Beigl. Using spatial Co-location for Coordination in Ubiquitous Computing Environments. *Proceedings of the 1th International Symposium on Handheld and Ubiquitous Computing (HUC99)*, September 1999, S. 259–273.
- [BSLG98] Michael Beigl, Albrecht Schmidt, Markus Lauff und Hans-Werner Gellersen. UbicompBrowser. *Proceedings of the 4th ERCIM Workshop on User Interfaces for All*, September 1998.
- [DeAS99] Anind K. Dey, Gregory D. Abowd und Daniel Salber. A Context-Based Infrastructure for Smart Environments. *In the Proceedings of the 1st International Workshop on Managing Interactions in Smart Environments*, Dezember 1999, S. 114–128.
- [GeBK99] Hans-Werner Gellersen, Michael Beigl und Holger Krull. The MediaCup: Awareness Technology embedded in an Everyday Object. *Proceedings of the 1th International Symposium on Handheld and Ubiquitous Computing (HUC99)*, September 1999, S. 308–310.
- [KiMa94] D. Kirsh und P. Maglio. On distinguishing epistemic from pragmatic action. *Cognitive Science* (18), 1994, S. 513–549.

-
- [ReGr98] Venture Development Corporation (VDC) Embedded Systems Industry Research und Consulting Group. The World Market for Embedded Operating Systems and Software Development Tools. *Journal*, 1998.
- [Reki98] Jun Rekimoto. Multiple-Computer User Interfacesl. *Proceedings of the First International Workshop CoBuild '98*, Februar 1998.

Integrierte Managementarchitektur für Differentiated-Services-Netze

Roland Bless

1 Einleitung

Das heutige Internet unterstützt in seiner derzeitigen Form keine Dienstgüte. Bereits Mitte der 90er Jahre wurden Mechanismen im Rahmen der „Integrated Services“-Architektur („Int-Serv“) innerhalb der Internet Engineering Task Force (IETF) entwickelt, um beispielsweise die Übertragung von Multimediadatenströmen mit einer vorhersagbaren Qualität zu ermöglichen [BrCS94]. Diese Mechanismen stellten sich jedoch als untauglich für die große Anzahl der im Internet auftretenden Datenströme heraus [BBBO⁺97], da die Garantien für jeden einzelnen Datenstrom innerhalb eines jeden Zwischensystems (häufig ein Router) überprüft und durch die Behandlung der Pakete mit besonderen Weiterleitungsmechanismen bereitgestellt werden. Dadurch folgt, dass im Kernnetz prinzipiell die Summe der Qualitätsparameter und Kontextdaten aller Datenströme verwaltet werden muss, was derzeit aufgrund der großen Menge nicht praktisch zu realisieren ist. Die Zuordnung von einem Datenstrom zu seinen Kontextdaten erfolgt üblicherweise durch Klassifikation eines jeden Pakets anhand des 5-Tupels bestehend aus Quell- und Ziel-IP-Adressen, sowie den Quell- und Zielpportnummern und der Protokollnummer (z.B. UDP oder TCP). Diese Klassifikation ist sehr aufwendig und wird in jedem Router erneut durchgeführt. Andererseits besteht nach wie vor der Bedarf nach besseren Diensten im Internet, sei es, um das Telefonieren über das Internet auch qualitativ attraktiv und damit wettbewerbsfähig zu machen, oder auch, um interaktive Anwendungen (z.B. Spiele mit mehreren über das Netzwerk gekoppelten Teilnehmern) zu ermöglichen, die eine garantierte obere Schranke für die Ende-zu-Ende-Verzögerung erfordern.

2 Differentiated Services

In dem neueren Ansatz „Differentiated Services“ (kurz „DiffServ“ oder auch DS abgekürzt), der lediglich unterschiedliche *Dienstklassen* bereitstellt, werden solche Skalierungsprobleme durch *Aggregation* umgangen, d. h. im Kernnetz werden nur noch aggregierte Datenströme betrachtet, denen jeweils eine Dienstklasse zugeordnet ist. Die Zuordnung von IP-Paketen zu Dienstklasse erfolgt über eine 6 Bit umfassende Kennung (den sogenannten Differentiated Services Codepoint – DSCP) im Paketkopf. Zur Übermittlung der Kennung wird das bisherige „ToS (Type of Service)“-Feld in IPv4 bzw. das „Class Field“ in IPv6 eingesetzt. Ein Wert wird in das Feld üblicherweise durch den ersten Router, an den das Endsystem angeschlossen ist (den sogenannten „First-Hop-Router“), anhand eines hinterlegten Verkehrsprofils eingetragen. Die Router im Netzzinnern inspizieren nur noch den DSCP und leiten das Paket mit den für diese Dienstklasse spezifizierten Mechanismen weiter. Dadurch wird die Anzahl der im Kernnetz von Zwischensystemen zu verwaltenden Zustände gegenüber dem „Integrated Services“-Ansatz drastisch reduziert. Pakete einer Übertragungsteilstrecke, die den gleichen

DSCP-Wert besitzen, werden daher auch als *Verhaltensaggregat* oder auch nur als *Aggregat* bezeichnet.

Komplexere Operationen wie die Klassifikation und Überwachung einzelner Datenströme müssen zwar auch bei „DiffServ“ durchgeführt werden, sie sind jedoch hauptsächlich in den Zugangs- und Randbereichen des Netzes lokalisiert (z. B. im First-Hop-Router).

Die vereinfachten Mechanismen im Datenpfad müssen durch ein Management ergänzt werden, um flexible Dienste auf Anforderung bereitstellen zu können. Ziel dieser Arbeit ist es, eine Management-Architektur zu entwerfen, die hinsichtlich der Skalierung im Kontrollpfad ähnliche Eigenschaften wie die Differentiated-Services-Basismechanismen im Datenpfad aufweist.

3 Bereitstellung von Diensten

Um Dienste auf Basis von DiffServ bereitzustellen, entscheidet ein Dienstbetreiber (Internet Service Provider – ISP) unter anderem, welche Mechanismen zur Weiterleitung der Pakete dieses Dienstes, welche zusätzlichen Maßnahmen zur Verkehrsbeeinflussung und welche zugehörigen Parameterwerte er dafür intern verwendet. Er schließt dann einen Vertrag (Service Level Agreement – SLA) mit seinen benachbarten Betreibern, falls der Dienst über den eigenen Zuständigkeits- und Verwaltungsbereich („Domäne“) hinaus erbracht werden muss. Eine Kunde kann somit einen Dienst bei seinem Betreiber zu bestimmten Zielsystemen (Endsysteme oder Teilnetzbereiche) anfordern, der dann einen bilateralen Vertrag mit dem nächsten involvierten Betreiber abschließt. Diese bilateralen Verträge setzen sich entsprechend fort bis zum Zielsystem [BBCD⁺98].

Ein recht einfacher aber nützlicher Dienst namens „Premium Service“ wurde bereits in einer frühen Phase der Entwicklung von Differentiated Services [JaNZ99] vorgeschlagen: er bietet einen garantierten Durchsatz, minimale Verzögerung sowie minimale Paketverluste und verkörpert somit eine Art „virtuelle Standleitung“, die beispielsweise zur Verbindung von virtuellen privaten Netzen innerhalb des Internets eingerichtet werden kann. Das *abschnittsweise Weiterleitungsverhalten* (Per-Hop Forwarding Behavior – PHB), der kleinste Dienstbaustein innerhalb der DiffServ-Architektur, lässt sich für diesen Dienst beispielsweise durch eine Ausgangswarteschlange im Router realisieren, deren Pakete vor allen anderen Paketen in den übrigen Warteschlangen bedient werden (Single Priority Scheduling). Kann man nun sicherstellen, dass die Summe der Raten aller (möglicherweise bereits aggregierten) eingehenden Datenströme dieser Dienstklasse, die zu diesem Ausgang weitergeleitet werden, nicht größer als die Bedienrate der Warteschlange ist, so wird sich die Warteschlange innerhalb eines betrachteten längeren Zeitintervalls nicht füllen, was u. a. in minimaler Verzögerung und minimalen Paketverlusten resultiert. Die Bedingung wird üblicherweise durch eine *Zugangskontrolle* sichergestellt, bei der vor Einrichtung eines Dienstvertrags die dazu notwendige Ressourcenverfügbarkeit geprüft wird. Es bleibt dem Dienstbetreiber überlassen, ob und an welchen Stellen er eine Verkehrsformung durchführt, um das Verhalten des Dienstes günstig zu beeinflussen. Des Weiteren wird im ersten Router der Domäne, an den der Kunde angeschlossen ist (First-Hop-Router), ein zugehöriges *Verkehrsprofil* eingerichtet, anhand dessen die Konformität des vom Kunden generierten Verkehrs überwacht und sichergestellt wird.

Auch wenn die Dienste auch in einer frühen Verbreitungsphase der DiffServ-Architektur anfangs manuell eingerichtet werden können, sollte die dynamische Einrichtung entsprechender Verträge vorgesehen werden. Dies impliziert den automatisierten Austausch von technischen Dienstspezifikationen zwischen Kunden und Dienstbetreiber sowie zwischen benachbarten Dienstbetreibern, ebenso wie die Einrichtung oder Änderung der Verkehrsprofile. Für die

Durchführung dieser Aufgaben wurde von Van Jacobson vorgeschlagen [JaNZ99], separate Verwaltungseinheiten (u. a. auch für die Zugangskontrolle) einzusetzen. Diese „Bandwidth Broker“ genannten Agenten vereinbaren die Verträge dynamisch mit den Instanzen der Nachbardomänen.

4 Integrierte Dienstmanagementarchitektur

Unter dem Begriff *DS-Domäne* wird ein Bereich bezeichnet, der von ein und derselben Organisation administriert wird und in dem sämtliche Systeme Differentiated-Services-konform sind. Der hierarchischen Aufteilung des Internets folgend, lassen sich auch mehrere DS-Domänen innerhalb eines *Autonomen Systems (AS)* vorstellen (siehe auch Abbildung 1). Letzteres zeichnet sich dadurch aus, dass die Netzwerkelemente innerhalb des AS nur von einer Organisation verwaltet werden und deren Struktur nach aussen hin nicht sichtbar ist, sowie dass intern ein oder mehrere verschiedene Intra-Domain-Routing-Verfahren (üblicherweise OSPF) eingesetzt werden.

In der aktuellen Forschungsarbeit wird ausgehend von der Idee der Bandwidth Broker eine integrierte Dienstmanagementarchitektur für Differentiated Services entwickelt, die auf dedizierten Managementinstanzen, den *Differentiated Services Domänen Managern (DSDM)* basiert. Je DS-Domäne existiert eine logische DSDM-Instanz, welche für die folgenden Aufgaben zuständig ist:

- *Unterstützung dynamischer Dienstanforderungen von Ende zu Ende (Signalisierung mit Dienstnutzern)* — Dies umfasst sowohl die sofortige Einrichtung eines Dienstes als auch die zeitliche geplante Einrichtung im Voraus.
- *Durchführung einer Zugangskontrolle anhand von Ressourcenverfügbarkeit und Betreiberpolitik* — Einige Dienste erfordern die Überprüfung der vorhandenen Ressourcenumenge, bevor ein neuer Teilnehmer zur Dienstnutzung zugelassen wird, weil sonst die Garantien der anderen Teilnehmer nicht mehr gewährleistet werden können.
- *Verwaltung logischer Ressourcen der Domäne* — Im Gegensatz zum Ressourcenreservierungsprotokoll RSVP, wird die Verfügbarkeit und Belegung einer Systemressource üblicherweise nicht beim System selbst angefragt und durchgeführt, sondern nur anhand einer zentralen Buchführung innerhalb des DSDM bewerkstelligt.
- *Aushandeln dynamischer Dienstverträge mittels Inter-Domänen-Signalisierung* — Zur Realisierung von Diensten über mehrere Domänen (insbesondere, wenn sie nicht von ein und derselben Administrationsorganisation verwaltet werden) hinweg, werden Mechanismen benötigt, um die Anforderungen und Zusicherungen für Dienste zu propagieren. Dadurch wird die Reichweite der Dienste vergrößert, die Einrichtung eines Dienstes beschleunigt und für die Netzwerkadministratoren vereinfacht.
- *Konfiguration der Router* — Wie schon zuvor erwähnt, besitzen einige Router (insbesondere an den Grenzen der Domäne) Verkehrsprofile zur Überwachung der vereinbarten Aggregatparameter. Diese Profile müssen, den Änderungen der einzelnen Dienstverträge entsprechend, eingerichtet bzw. angepasst oder gelöscht werden.
- *Authentisierung und Dienstabrechnung* — Es sollte sichergestellt werden, dass nur berechtigte Dienstanutzer Zugriff auf bestimmte Dienste haben und der Diebstahl von Diensten anderer Nutzer ausgeschlossen wird. Daher sollte eine Authentisierung und Integritätssicherung der Signalisierung vorgesehen werden. Differenzierte Dienste verlangen automatisch eine ebenso differenzierte Preisbildung und Abrechnung. Der DSDM

kann Verbindungsdaten aufzeichnen, die für Abrechnungszwecke verwendet werden können.

- *Unterstützung von mobilen Dienstnutzern und Gruppenkommunikation* — Eine Unterstützung mobiler Teilnehmer erfordert die Fähigkeit, deren aktuelle Dienstreservierungen zwischen zwei Domänen auszutauschen. Die Gruppenkommunikation benötigt eine besondere Unterstützung durch das Management, da sonst die Garantien für bereits bestehende Verbindungen nicht mehr gewährleistet werden können [BIWe99b, BIWe99a].

Die Architektur weist einige Vorteile gegenüber anderen möglichen Ansätzen (z. B. dem früheren RSVP oder auch [PaHS99]) auf: Der Kontrollpfad zur Signalisierung ist vom Datenpfad separiert, d. h. Zwischensysteme werden von Signalisierungs- und Ressourcenverwaltungsaufgaben entlastet und die Anfragen müssen von weniger Systemen bearbeitet werden. Über Ressourcen wird hauptsächlich nur Buch geführt, wodurch Zwischensysteme von Zugangskontrollanfragen und Aktivierung der Ressourcen unbehelligt bleiben. Ein Ausfall eines DSDM beeinträchtigt bereits etablierte Dienste nicht, weiterhin wird der Einsatz besonders leistungsfähiger und fehlertoleranter DSDM-Systeme möglich.

In Abbildung 1 sind einige DS-Domänen mit unterschiedlichen Router-Typen innerhalb verschiedener AS gezeigt. Üblicherweise verbinden Router des Backbone-Bereichs die DS-Domänen innerhalb eines AS miteinander und stellen die Verbindung zu benachbarten AS her. Die Ressourcen des Backbone-Bereichs werden unter allen DSDM zunächst gleichmäßig aufgeteilt, so dass jeder DSDM selbständig Ressourcen vergeben kann. Reichen die verwalteten Ressourcen nicht mehr aus, können bei den benachbarten DSDM innerhalb des AS mehr Ressourcen angefragt werden. Bei dieser Struktur bearbeitet in Transit-AS jeweils nur ein DSDM eine eingehende Dienstanfrage.

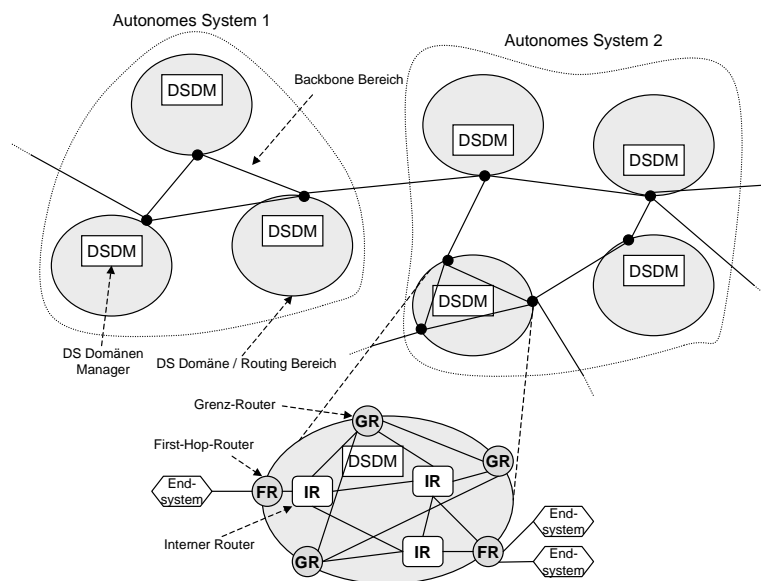


Abbildung 1: Netzstruktur von Differentiated-Services-Netzen im Internet

Garantierte Dienste von Ende-zu-Ende sind auch in DiffServ-Umgebungen nur mit einer Signalisierung von Ende-zu-Ende möglich, d. h. jede DS-Domäne auf dem Weg zwischen dem Initiator- und dem Zielsystem muss die entsprechende Zugangskontrolle durchführen. Würde die Anfrage bereits von einem DSDM entlang des Pfades früher entschieden, könnten die von ihm zugesagten Ressourcen beispielsweise in der letzten Domäne auf der letzten Teilstrecke nicht mehr zur Verfügung stehen. Somit muss für jeden einzelnen Datenstrom und für jedes

Aggregat eine Anfrage durchgeführt werden. Es gilt nun die beiden folgenden Probleme zu lösen:

- Die Signalisierungsnachrichten der Endsysteme summieren sich im Kernnetz und erzeugen dort unter Umständen ein erneutes Skalierungsproblem für die Managementinstanzen.
- Eine ständige Anpassung der Aggregate, d. h. mit jedem einzelnen Datenstrom, ist unerwünscht und auch gar nicht praktikabel, da die Änderung der Profile in den Routern (zeit-)aufwendig sein kann.

Beide Probleme lassen aber sich mit dem in der Forschungsarbeit neu entwickelten Konzept der dynamischen Aggregation lösen.

4.1 Dynamische Aggregation

Durch das Konzept der *dynamischen Aggregation* werden mittels der DSDM-Instanzen dynamisch Aggregate dort ausgebildet, wo sie am effektivsten sind. Vereinbaren zwei DSDM-Instanzen die Einrichtung eines Aggregats, so benötigen die zwischen diesen Instanzen lokalisierten DSDM keine Kenntnis über weitere Aggregate oder einzelne Datenströme, die innerhalb des umfassenden Aggregats enthalten sind. Daher können diese Systeme auch bei der Signalisierung übersprungen werden. Die Verarbeitung findet dann erst wieder in den Rand- und Zugangsbereichen statt, wodurch der Aufwand handhabbar bleibt.

Ein Aggregat kann immer dann gebildet werden, wenn mehrere Reservierungen (die gleiche Dienstklasse betreffend) einen gemeinsamen Teilpfad aufweisen. So könnte das Kriterium zum Erzeugen eines Aggregats die Anzahl bereits bestehender Reservierungen und die Lebensdauer solcher Reservierungen berücksichtigen. Bisherige Reservierungen werden im neu zu bildenden Aggregat zusammengefasst und die Systeme zwischen den beiden Aggregatendpunkten können die einzelnen Reservierungszustände der dann aggregierten Datenströme anschließend löschen.

Effektiv eingespart wird der Signalisierungsaufwand aber nur, wenn die Aggregate – gemessen an der Lebensdauer eines Datenstroms – relativ selten angepasst werden. Das lässt sich im Wesentlichen durch zwei Techniken erreichen: zum einen kann bereits bei Einrichtung oder „Erhöhung“ des Aggregats immer etwas mehr reserviert werden als momentan benötigt wird, um zukünftige Reservierungen ohne erneute Änderungen des Aggregats durchführen zu können. Zum anderen kann eine Hysterese beim Abbau von Verbindungen helfen, d. h. die Größe des Aggregats bleibt für eine gewisse Zeit auch nach Beendigung von Reservierungen innerhalb des Aggregats erhalten.

Nach einer gewissen Zeit sollten sich vor allem im Kernnetz entlang der vielgenutzten Strecken zwischen den Betreibern relativ stabile Aggregate ausbilden, die sich aber durchaus ändernden Verkehrs- und Lastsituationen anpassen können.

5 Zusammenfassung und Ausblick

In diesem Beitrag wurde das Management für Differentiated-Services-Netze motiviert und eine integrierte Managementarchitektur vorgestellt. Um ein entsprechend skalierbares Kontrollnetzwerk für Differentiated-Services zu schaffen, wurde das Konzept der dynamischen Aggregation entworfen, das zur Zeit in Simulationen genauer untersucht wird. Diese können einen ersten Aufschluss über noch zu bestimmende Parameterwerte und die Effektivität des

Verfahrens geben, wobei letztere auch relativ stark von der zugrundeliegenden Netztopologie abhängen wird. Deshalb wird schließlich versucht, Simulationsszenarien mit realen Topologiedaten des Internets zu untersuchen.

Literatur

- [BBBO⁺97] Fred Baker, Bob Braden, Scott Bradner, Michael O'Dell, A. Mankin, Allyn Romanow, Abel Weinrib und Lixia Zhang. Resource ReSerVation Protocol (RSVP) Version 1 Applicability Statement – Some Guidelines on Deployment. RFC 2208, September 1997.
- [BBCD⁺98] Steven Blake, David Black, Mark Carlson, Elwyn Davies, Zheng Wang und Walter Weiss. An Architecture for Differentiated Services. RFC 2475, Dezember 1998.
- [BIWe99a] Roland Bless und Klaus Wehrle. IP Multicast in Differentiated Services Networks. Internet-Draft – draft-bleess-diffserv-multicast-00.txt, September 1999.
- [BIWe99b] Roland Bless und Klaus Wehrle. Managementarchitektur zur Unterstützung von Gruppenkommunikation in Differentiated-Services-Netzen. In *Tagungsband des 1. GI-Workshops „Multicast – Protokolle und Anwendungen“*. Gesellschaft für Informatik (GI), Mai 1999.
- [BrCS94] Robert Braden, David Clark und Scott Shenker. Integrated Services in the Internet Architecture: an Overview. RFC 1633, Juni 1994.
- [JaNZ99] Van Jacobson, Kathleen Nichols und Lixia Zhang. A Two-bit Differentiated Services Architecture for the Internet. RFC 2638, Juli 1999.
- [PaHS99] Ping Pan, Ellen L. Hahne und Henning Schulzrinne. BGRP: A Tree-Based Aggregation Protocol for Inter-Domain Reservations. Computer Science Technical Report CUCS-029-99, Columbia University, Dezember 1999.

Vorgehensmodell und Prozeßunterstützung für komponentenbasiertes Web Engineering

Martin Gaedke

1 Einleitung

Anwendungen im World Wide Web basieren auf einem dokumentenbasierten Hypermedia-Implementierungsmodell. Das außerordentliche Wachstum dieser Anwendungen erfordert neue Verfahren einer weborientierten Softwaretechnik [Gaed98], für die neben der (Neu-)Entwicklung von Anwendungen im Web besonders die kontinuierliche Weiterentwicklung (Evolution) von zentraler Bedeutung sind. Für die Entwicklung und Evolution von Anwendungen im World Wide Web sind dabei zwei Aufgabenstellungen vorrangig zu betrachten [ScWZ00][ScRo98]:

- Entwicklung der Dokumente und ihrer Hypermedia-Verknüpfungen, welche die Funktionalität der Anwendung zur Verfügung stellen und in ihrer Gesamtheit im folgenden als Web-Anwendung bezeichnet werden.
- Evolution der Web-Anwendung, die durch Änderungen an den Dokumenten und ihren Strukturen sowie der Funktionalität bewirkt wird.

2 Web Engineering

Die klassische Softwaretechnik stellt für die Entwicklung von Anwendungen systematische Verfahren zur Verfügung [Fair85], die jedoch durch das grobgranulare Implementierungsmodell des World Wide Webs nur unzureichend oder gar nicht für die Entwicklung von Web-Anwendungen geeignet sind [BaSc98]. Softwaretechnische Ansätze für die Entwicklung von Web-Anwendungen entstehen erst in jüngerer Zeit in einem interdisziplinären Bereich [GeGa99], der sich als eigener Zweig der Softwaretechnik etabliert und als Web Engineering [Mura99][GeWG97] bezeichnet wird. Diese Ansätze, wie beispielsweise OOHDM [ScRB96] oder JESSICA [BaSc98], beschäftigen sich insbesondere mit der disziplinierten Abbildung feingranularer Artefakte (z.B. aus dem objektorientierten Entwurf) auf das dokumentenbasierte Implementierungsmodell, der hypermedialen Strukturierung von Web-Anwendungen (z.B. Navigation, Informationsgliederung und Darstellung) und der Sammlung von dedizierten Mustern zur Unterstützung der Entwicklung von Web-Anwendungen (z.B. Web-Entwurfsmuster, [GeCo00]). Bedingt durch den hohen technologischen Fortschritt im Web [Tura99] ist der Betrieb einer Web-Anwendung erheblich durch Anpassungen und Neuentwicklungen geprägt, so daß sich eine Web-Anwendung ständig in der Evolution befindet. Die Evolution einer Web-Anwendung durch die disziplinierte Wiederverwendung von Artefakten zu unterstützen (z.B. durch Wiederverwendung von Navigationsstrukturen oder Darstellungselementen), wird jedoch ebenfalls durch das dem Web zugrundeliegende grobgranulare Implementierungsmodell erschwert [GaTu99a].

Den Stand der Technik zusammenfassend, läßt sich feststellen, daß die Entwicklung und Evolution von Web-Anwendungen durch zwei Kernprobleme geprägt wird:

- Für eine disziplinierte Entwicklung von Web-Anwendungen mangelt es an Methoden, Modellen und Werkzeugen, die durchgängig und spezifisch die Entwicklung von evolutionsorientierten Lösungen unterstützen.
- Für eine systematische, wiederverwendungsgeprägte Unterstützung der Evolution von Web-Anwendungen ist das Implementierungsmodell des Webs ungeeignet, da auf feingranulare Entwurfsartefakte einer Web-Anwendung nach ihrer Abbildung auf die grobgranularen Entitäten des Implementierungsmodells nicht mehr zugegriffen werden kann.

3 Ein komponentenbasierter Ansatz für das Web Engineering

Kernstück des Ansatzes ist eine Komponententechnik zur Unterstützung der Entwicklung und Evolution von Anwendungen im World Wide Web. Im einzelnen umfaßt das neue Konzept die folgenden Ansätze zur Unterstützung im komponentenbasierten Web Engineering:

Komponenten. Durch einen auf Web-Standards aufbauenden Ansatz wird die technologieunabhängige Abstraktion von Artefakten einer Web-Anwendung, unter Berücksichtigung der Wiederverwendung, ermöglicht. Den Kern der entwickelten Lösung stellen Komponenten dar, die als *uniformes, generisches Modellierungskonzept* für Entwurfsartefakte in Web-Anwendungen dienen [GeWG97]. Die entwickelte Komponententechnik ermöglicht die Abbildung abstrakter objektorientierter Entwurfsmodelle auf das Komponentenmodell, das durch seine Implementierungsnähe eine automatische und noch im Detail manipulierbare Abbildung auf das Implementierungsmodell des World Wide Webs sicherstellt. Eine Web-Anwendung oder ganz allgemein eine zusammengesetzte Komponente wird als Komposition von Komponenten beschrieben. Für die Komposition von Komponenten stehen der Mechanismus der Prototyp-Instanz-Vererbung, die schnittstellenbasierte Referenzierung und weitere grundlegende Konzepte, wie sie aus der Objektorientierung bekannt sind, zur Verfügung. Das Modell macht im Gegensatz zum Implementierungsmodell des World Wide Webs die Entitäten, aus denen eine Web-Anwendung konstruiert wurde, zugreifbar und ermöglicht damit eine konsistente Evolution der Web-Anwendung bei Modifikationen der Komponenten [Gaed99].

Die im Rahmen der Forschungsarbeiten entwickelte *WebComposition Markup Language* [GaSG99], eine Anwendung der eXtensible Markup Language (XML) [Worl96], dient zur Beschreibung von Komponenten. Das dazugehörige Laufzeitsystem setzt das Komponentenmodell, das in einem sogenannten Virtual Component Store gespeichert ist, durch Nutzung grundlegender Standardtechnologie des World Wide Webs um, so daß die Komponenten in heterogenen Umgebungen genutzt werden können (vergleiche Abbildung 1). Durch den generischen Modellierungsansatz ist darüber hinaus die entwickelte Technologie nicht an den Stand der Technik gebunden [GBGS98].

Dediziertes Entwicklungsmodell. Die Entwicklung einer Web-Anwendung wird durch die Adaption des Spiralmodells [Boeh88], ein in der Softwaretechnik etabliertes Softwareentwicklungsmodell, systematisiert. Das Spiralmodell erlaubt die Berücksichtigung der Wiederverwendung durch die frühzeitige Identifikation von Implementierungsalternativen und der damit verbundenen Risiken. Die Wiederverwendung von Komponenten wird in dem entwickelten *WebComposition Vorgehensmodell* [GaGr00] durch die Integration einer expliziten Berücksichtigung der existierenden Komponenten erreicht. Der Entwicklungsprozeß wird dabei durch ein koordinierendes Wiederverwendungswerkzeug unterstützt. Das Vorgehensmodell

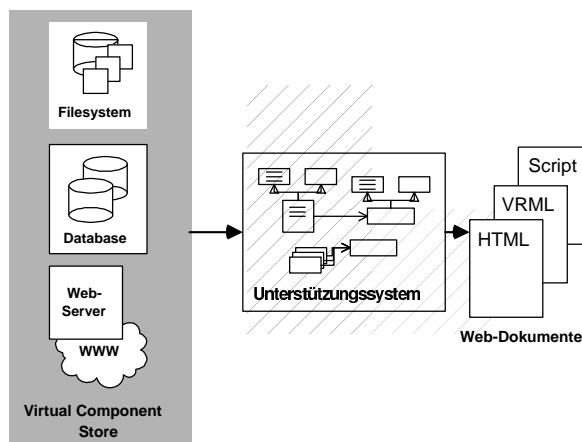


Abbildung 1: Arbeitsweise des Komponentensystems

wurde für die Entwicklung von Web-Anwendungen an eine im Rahmen der Forschungsarbeiten entwickelte, evolutionsorientierte und komponentenbasierte Rahmenarchitektur für Web-Anwendungen angepaßt, wobei die Prinzipien des World Wide Webs berücksichtigt wurden. Durch die streng aufeinander abgestimmte Integration von systematischer Vorgehensweise, disziplinierter Wiederverwendung und komponentenbasierter Rahmenarchitektur wird die Einarbeitungs- und Entwicklungszeit wesentlich verringert.

Wiederverwendung. Die Thematik der Wiederverwendung wird aus zwei Blickwinkeln betrachtet [Same97][Szyp97]. Zum einen wird die Entwicklung von Web-Anwendungen und Komponenten durch Wiederverwendung von Komponenten systematisiert (*Consumer Reuse*), zum anderen wird hierzu die disziplinierte Entwicklung von Komponenten für die Wiederverwendung (*Producer Reuse*) benötigt. Beide Sichtweisen werden durch ein koordinierendes Wiederverwendungswerkzeug, dem *WebComposition Reuse-Repository*, zusammengeführt [GaRG99]. Die Koordination orthogonaler Prozesse für die Entwicklung wiederverwendbarer Komponenten [McCl97] und die Web-Anwendungsentwicklung durch Wiederverwendung von Komponenten werden inhärenter Bestandteil aller Entwicklungsphasen im Entwicklungsprozeß einer Web-Anwendung. Die Entwicklung mit Wiederverwendung wird durch die Suche nach Komponenten unterstützt. Der Suche liegt dabei eine offene Architektur für die Integration beliebiger Beschreibungs- und Darstellungsmodelle zugrunde [GaRe00]. Die Offenheit der Architektur ermöglicht den Einsatz von beliebigen Werkzeugen zur Unterstützung beim Suchen sowie die potentielle Nutzung neuer Beschreibungsmechanismen. Die Entwicklung von Komponenten für die Wiederverwendung impliziert die Existenz unterschiedlicher, möglicherweise orthogonaler Entwicklungsprozesse. Die Koordination der Entwicklungsprozesse und ihrer Zustände wird durch die Manipulation der Beschreibungsmodelle ermöglicht, die dadurch sofort Einfluß auf die Suchergebnisse haben können.

Dienstleistungsorientierung. Die Evolution wird durch eine systematische und zweckorientierte Partitionierung der Web-Anwendungen in anwendungsspezifische Domänen unterstützt [GaTu99b]. Die Rahmenarchitektur für Web-Anwendungen basiert auf der Standardisierung domänenspezifischer Anwendungsbereiche. Hierzu wurde der *WebComposition Evolutionsbus* als Konzept für die Architektur komponentenbasierter Web-Anwendungen entwickelt [GaTu00]. Die Abbildung 2 zeigt eine vereinfachte Darstellung der Architekturidee. Die Evolution einer Web-Anwendung wird in die Bereiche Evolution von Anwendungsdomänen und domänen-spezifische Evolution gegliedert. Jede Anwendungsdomäne wird durch eine spezielle Komponente (*Servicekomponente*) repräsentiert, die eine strikte Trennung von Präsentation, Benutzerinteraktion, Information, Navigation und Verarbeitungsprozeß beschreibt. Die Evolution der Anwendungsdomänen vollzieht sich somit durch die Entwicklung einer zusätz-

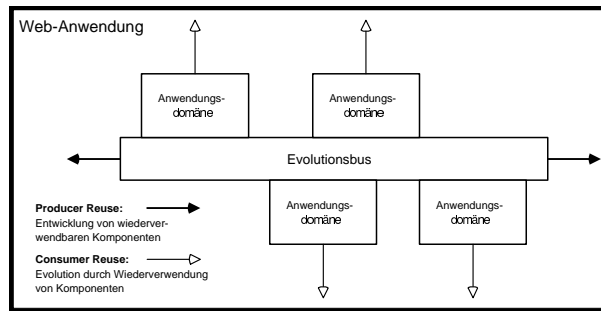


Abbildung 2: Vereinfachte Darstellung der Evolutionsbus-Rahmenarchitektur

lichen Servicekomponente, die eine neue domänenspezifische Dienstleistung anbietet und für die spezifische Evolution einer Domäne als standardisierender Prototyp zur Verfügung steht.

4 Zusammenfassung

Die in diesem Beitrag vorgestellten Methoden, Modelle und Werkzeuge stellen einen wesentlichen Fortschritt zur Entwicklung von Web-Anwendungen dar. Im Vergleich zu existierenden Ansätzen ist der neue Ansatz der einzige, der den Entwicklungsprozeß detailliert auf die Evolution einer Web-Anwendung ausrichtet und speziell die Wiederverwendung von Artefakten fördert. Darüber hinaus ist der entwickelte Ansatz systematischer und behandelt alle Phasen im Lebenszyklus einer Web-Anwendung.

Die Evaluierung des Ansatzes erfolgte im Rahmen eines Projektes mit der Firma Hewlett-Packard, bei der das umfangreiche europäische Intranet unter Verwendung der eingeführten Techniken nachimplementiert wurde und somit der realitätsnahen Bewertung im praktischen Einsatz diente. Neben den Kriterien der Erlernbarkeit und Benutzbarkeit des Ansatzes wurde insbesondere die Wiederverwendung und Evolution für die Evaluierung berücksichtigt.

Sehr kurze Entwicklungszeiten, hohe Wiederverwendungsraten von Komponenten und eine schnelle Evolution der Web-Anwendungen weisen auf eine effektive Lösung hin. Befragungen der Entwickler im Projekt sowie Nutzung des entwickelten Ansatzes im Rahmen von mehreren internationalen Forschungsprojekten belegen die Klarheit und leichte Erlernbarkeit des Ansatzes sowie die Anwendbarkeit im heterogenen Umfeld. Für die Evaluierung wurde der Ansatz auch für die beispielhafte Implementierung prominenter Web-Entwurfsmuster herangezogen.

Literatur

- [BaSc98] R.A. Barta und Markus W. Schranz. JESSICA: an object-oriented hypermedia publishing processor. *Computer Networks and ISDN Systems* 30(Special Issue on the 7th Intl. World-Wide Web Conference, Brisbane, Australia), April 1998.
- [Boeh88] B.W. Boehm. A Spiral Model of Software Development and Enhancement. *IEEE Computer* 21(5), 1988.
- [Fair85] R. E. Fairley. *Software engineering concepts*. McGraw-Hill. 1985.
- [Gaed98] Martin Gaedke. WebComposition: Ein Unterstützungssystem für das Web Engineering. *GI Softwaretechnik-Trends* 18(3), August 1998.

- [Gaed99] Martin Gaedke. Wiederverwendung von Komponenten in Web-Anwendungen. *GI Proc. 1. Workshop Komponentenorientierte betriebliche Anwendungssysteme (WKBA 1)*, Magdeburg, Germany, März 1999.
- [GaGr00] Martin Gaedke und Guntram Graef. WebComposition Process Model: Ein Vorgehensmodell zur Entwicklung und Evolution von Web-Anwendungen. *GI Proc. 2. Workshop Komponentenorientierte betriebliche Anwendungssysteme (WKBA 2)*, Wien, Austria, März 2000.
- [GaRe00] Martin Gaedke und Jörn Rehse. Supporting Compositional Reuse in Component-Based Web Engineering. *ACM Proc. of Symposium on Applied Computing (SAC 2000)*, März 2000.
- [GaRG99] Martin Gaedke, Joern Rehse und Guntram Graef. A Repository to facilitate Reuse in Component-Based Web Engineering. *International Workshop on Web Engineering at the 8th International World-Wide Web Conference (WWW8-WebE)*, Toronto, Canada, Mai 1999.
- [GaSG99] Martin Gaedke, Daniel Schempf und Hans-W. Gellersen. WCML: An enabling technology for the reuse in object-oriented Web Engineering. *Poster-Proceedings of the 8th International World Wide Web Conference (WWW8)*, Toronto, Canada, Mai 1999.
- [GaTu99a] Martin Gaedke und Klaus Turowski. Framework for Maintaining Evolution of E-Commerce Applications in the Web. *Proc. of 12th International Conference - Software and Systems Engineering and their Applications (ICSSEA '99)*, France, Dezember 1999.
- [GaTu99b] Martin Gaedke und Klaus Turowski. Resource Management in Web-based Business Applications. *GI Proc. of Workshop on Distributed Computing on the WEB (DCW99)*, Rostock, Germany, Dezember 1999.
- [GaTu00] Martin Gaedke und Klaus Turowski. Integrating Web-based E-Commerce Applications with Business Application Systems. *Netnomics Journal, Baltzer Science Publishers* Band Erscheinungsdatum N.N., 2000.
- [GBGS98] Martin Gaedke, Michael Beigl, Hans-Werner Gellersen und Christian Segor. Web Content Delivery to Heterogeneous Mobile Platforms. *Lecture Notes in Computer Science (LNCS)*, Springer Verlag 1552(Advances in Database Technologies), 1998.
- [GeCo00] D.M. German und D.D. Cowan. Towards a unified catalog of hypermedia design patterns. *IEEE Proc. of the 33rd Annual Hawai'i International Conference on Systems Sciences (HICSS-33)*, Januar 2000.
- [GeGa99] Hans-W. Gellersen und Martin Gaedke. Object-Oriented Web Application Development. *IEEE Internet Computing* 3(1), 1999.
- [GeWG97] Hans-W. Gellersen, Robert Wicke und Martin Gaedke. WebComposition: An Object-Oriented Support System for the Web Engineering Lifecycle. *Computer Networks and ISDN Systems* 29(Special Issue on the 6th Intl. World-Wide Web Conference, Santa Clara, USA), April 1997.
- [McCl97] Carma McClure. *Software reuse techniques: adding reuse to the system development process*. Prentice Hall. 1997.
- [Mura99] San Muragesan. Web Engineering. *ACM SIGWEB Newsletter* 8(3), 1999.

- [Same97] J. Sametinger. *Software engineering with reusable components*. Springer. 1997.
- [ScRB96] Daniel Schwabe, Gustavo Rossi und S. Barbosa. Systematic Hypermedia Design with OOHDM. *ACM Proc. of International Conference on Hypertext' 96*, März 1996.
- [ScRo98] Daniel Schwabe und Gustavo Rossi. An Object Oriented Approach to Web-Based Applications Design. *TAPOS - Theory and Practice of Object Systems* 4(4), 1998.
- [ScWZ00] Markus Schranz, Johannes Weidl und Johannes Zechmeister. Engineering Complex World Wide Web Services with JESSICA and UML. *IEEE Proc. of the 33rd Annual Hawai'i International Conference on Systems Sciences (HICSS-33)*, Januar 2000.
- [Szyp97] C. Szyperski. *Component software: beyond object-oriented programming*. ACM Press; Addison-Wesley. 1997.
- [Tura99] Volker Tura. Techniken zur Realisierung Web-basierter Anwendungen. *GI Informatik Spektrum* 22(1), 1999.
- [Worl96] World Wide Web Consortium. Extensible Markup Language (XML) Specification: W3C Working Draft 21 Apr 1996. <http://www.w3.org/XML>, April 1996.

Optimierung der Lastkontrolle in Kommunikationssystemen durch adaptive, CI-basierte Verfahren

Meng Gan

1 Einleitung

Lastkontrolle, insbesondere Überlastvermeidung spielt eine wichtige Rolle in Kommunikationssystemen. Selbst bei neuartigen Technologien wie ATM oder DiffServ kann man auf das Prinzip der Lastkontrolle nicht verzichten. [FlKa99] fordert nach wie vor eine Ende-zum-Ende Lastkontrolle für das Internet. Bisherige Entwurfsmethoden für Überlastvermeidung basierten meistens auf menschliche Erfahrung und Verfahren wie Versuchs-und-Irrtums-Methoden, da die genauen mathematischen Modellen des Datenverkehrs nicht nachvollziehbar sind. Computational Intelligence(CI), wie Fuzzy-Logik, Neuronale Netze und genetische Algorithmen, stellen ein sehr leistungsfähiges Werkzeug dar, um Lösungen für solche nicht-lineare Probleme zu optimieren [Gan98], [Ghos98].

2 GAFuCA-Architektur

In diesem Beitrag wurde ein neues Konzept zur Optimierung der Lastkontrolle, GAFuCA (Genetic-Algorithm-aided Fuzzy Logic System for Congestion Avoidance), entwickelt. Mit GAFuCA-Architektur kann man menschliche Kenntnisse und Optimierungsverfahren kombinieren, um die Algorithmen zu verbessern und zu optimieren [GaDS99].

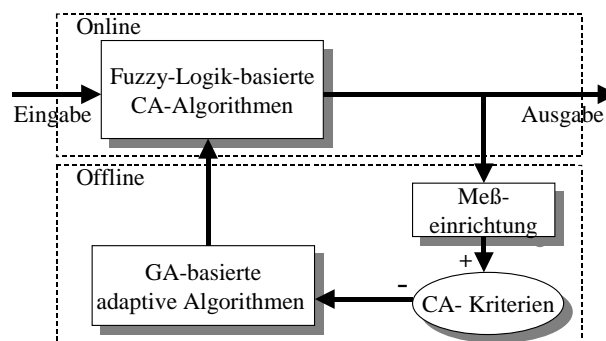


Abbildung 1: GAFuCA-Architektur

Die GAFuCA-Architektur ist in Abbildung 1 gezeigt. Sie besteht aus zwei Teilen. Der Offline-Teil ist für Entwurf und Optimierung zuständig, wobei der Online-Teil für den Einsatz in Netzwerken und die Auswertung verantwortlich ist. Die Algorithmen für Überlastvermeidung sind Fuzzy-Logik-basiert und werden so optimiert, daß sie die benutzerdefinierten Kriterien erfüllen können. Die Eingaben des Systems sind überwachende Parameter, wobei die Ausgaben des Systems die entsprechenden Entscheidungen für Lastkontrolle treffen müssen. Die Kriterien können leicht als Fitness-Funktionen für genetische Algorithmen formuliert werden,

welche das Optimierungsziel definieren. Während der Entwurfsphase ist eine Iterationszyklus zwischen Offline- und Online-Teil notwendig. Nachdem man eine optimierte Steuerkomponente für Lastkontrolle bekommen hat, kann man dann die Steuerkomponente umformulieren, z.B. um eine Tabelle konvertieren. Damit können die Anforderungen an die hohe Geschwindigkeit in Zwischenknoten erfüllt werden.

3 Implementierung und Entwicklungsrahmenwerk

Ein Entwicklungsrahmenwerk, das auf das GAFuCA-Konzept basiert, wurde entwickelt. Das Entwicklungsrahmenwerk dient dazu, daß man ein allgemeines leistungsfähiges Werkzeug verwenden kann, um eigene Algorithmen für Lastkontrolle nach dem GAFuCA-Konzept in Kommunikationssystemen zu entwerfen, optimieren und auswerten. Neben anderen besteht dieses Rahmenwerk aus folgenden Modulen:

- Lastkontrolle-Modul (Fuzzy-Logik-basiert)
- Optimierungsmodul (Genetische-Algorithmen-basiert)
- Netzwerksimulatoren (für TCP/IP [UCB], ATM [NIS], usw.)
- Benutzeroberfläche (vereinfachte Zugriffe auf andere Module, Festlegung der CA-Kriterien, usw.)

Mit diesem Rahmenwerk kann man umfangreiche Algorithmen entwerfen. FuRED und FuABR sind Beispiele für mögliche Einsätze des GAFuCA-Konzeptes in TCP/IP- und ATM-Netzen.

4 FuRED – Einsatz in TCP/IP-Netzen

FuRED ist ein Einsatzbeispiel der GAFuCA-Architektur in TCP/IP-Netzen.

RED(Random Early Detection)[FlJa93] ist eine Art aktiver Verwaltung der Warteschlangen in Routern der TCP/IP-Netze. Ein RED-fähiger Router überwacht die Warteschlange des Routers und verwirft Datenpakete bereits dann, wenn die Länge der Warteschlange größer als ein vordefinierter Schwellwert ist, wobei eigentlich noch keine echte Überlastung vorhanden ist. Die TCP-Verbindungen, die ein oder mehrere Datenpakete durch den Router verworfen bekommen, reduzieren ihre Übertragungsrate, damit andere TCP-Verbindungen nicht unnötigerweise auch ins Überlastverhalten gehen. Das Ziel von RED ist, die Verzögerung der Datenübertragung und Paketverlust zu minimieren und den Durchsatz der Datenübertragung zu maximieren. Die Festlegung einiger Parameter in RED müssen aber immer nach Erfahrung gemacht werden, was wegen der Vielfalt des Datenverkehrs meistens sehr schwierig ist. Dies führt oft dazu, daß die resultierende Leistung nicht optimal und die Bandbreitenverteilung zwischen TCP-Verbindungen unfair ist.

FuRED ist ein Fuzzy-Logik-basierter RED-Algorithmus. Hierbei wurde das Konzept des Fuzzy-Schwellwertes eingeführt. Im Gegensatz zum numerischen Schwellwert ist ein Fuzzy-Schwellwert eine Fuzzy-Variable, mit der der maximale und minimale Schwellwert für den RED-Algorithmus sehr flexibel definiert werden kann. FuRED implementiert RED als eine Fuzzy-Logik-basierte Steuerkomponente. Die Eingaben der Steuerkomponente sind die Länge der Warteschlange und die Änderung der Länge innerhalb einer gewissen Zeit. Die Ausgabe ist die Möglichkeit, daß ein Paket verworfen wird. Ein Simulator für TCP/IP-Netze, NS2[UCB],

wurde als Simulationswerkzeug genommen. Eine Optimierung für die Steuerkomponente wurden mit Hilfe von GAFuCA-Rahmenwerk durchgeführt. Die Kriterien der Optimierung sind beispielsweise der Paketverlust und die Ausnutzung der Leitungskapazität. Ein Vergleich zwischen FuRED und anderen Verfahren wurde durchgeführt und die Ergebnisse haben gezeigt, daß FuRED eine viel bessere Leistung erreichen kann [GaDS99].

5 FuABR – Einsatz in ATM-Netzen

FuABR ist ein Einsatzbeispiel der GAFuCA-Architektur in ATM-Netzen.

ABR (Available Bit Rate) ist eine Dienstklasse der ATM-Netze. Bei ABR fordert das Endsystem gar keine Bandbreite beim Verbindungsaufbau an, sondern das Netz teilt dem Endsystem mit, wieviel Bandbreite zur Verfügung steht. Ziel des ABR-Flußsteuerungsverfahrens ist es, die über ein ATM-Netz zu transportierenden Verkehrsströme so zu formen, daß gleichzeitig eine hohe Netzauslastung und eine geringe Zellverlustwahrscheinlichkeit erreicht werden können [ATM99]. Einige Algorithmen für ABR-Flußsteuerung wurden vorgestellt [GoSS98]. Die meisten davon basieren aber auf ungenauen Datenverkehrsmodellen und sind deshalb nicht optimal.

Die grundlegende Idee der FuABR ist es, daß man die ER(Explicit Rate)-Berechnung in ATM-Switch nach [ATM99] als Fuzzy-Logik-Steuerkomponente betrachten kann. Hierbei wurde eine Fuzzy-Logik-Steuerkomponente entwickelt, mit dem ER anhand von aktuellen Zuständen der ABR-Verbindung berechnet werden kann. NIST ATM Simulator [NIS] wurde als Simulationswerkzeug genommen und GAFuCA-Konzept wurde als Optimierungsverfahren für die ER-Berechnung eingesetzt. Die Kriterien für die Optimierung sind z. B. hohe Netzauslastung und gleichzeitig geringe Zellverluste. Simulationsergebnisse haben gezeigt, daß die trainierte Fuzzy-Logik-Steuerkomponente bessere Leistung als anderen Ansätzen erreichen kann. Ein ausführlicher Vergleich in Bezug auf Netzauslastung und Zellverlust wurde durchgeführt.

6 Zusammenfassung und Ausblick

In diesem Beitrag wurde ein neues Konzept zur Optimierung der Lastkontrolle entwickelt. Der Unterschied zwischen der GAFuCA-Architektur und bisherigen Ansätzen zur Überlastvermeidung liegt darin, daß die GAFuCA-Architektur menschliche Kenntnisse und Optimierungsverfahren kombinieren kann, um die Algorithmen zu verbessern und optimieren.

Ein Entwicklungsrahmenwerk für die Optimierung der Verkehrsüberlastvermeidung, das auf die GAFuCA-Architektur basiert, wurde implementiert. Hierbei wurden umfangreiche Optimierungsalgorithmen und benutzerfreundliche grafische Benutzerschnittstelle implementiert. Das Entwicklungsrahmenwerk dient dazu, als allgemeines Werkzeug für den Entwurf von Lastkontrollalgorithmen in Kommunikationssystemen eingesetzt zu werden. Die neu entwickelten Algorithmen FuRED und FuABR haben gezeigt, daß die GAFuCA-Architektur sowohl für TCP/IP-Netze als auch ATM-Netze geeignet ist.

Weitere Arbeiten umfassen die Bewertung der GAFuCA-Architektur und deren Robustheit. Optimierung der Lastkontrolle mit GAFuCA-Konzept in drahtlosen und mobilen Umgebung stellt eine weitere Arbeit dar.

Literatur

[ATM99] ATM Forum. *Traffic Management 4.1*, 1999.

- [FlJa93] S. Floyd und V. Jacobson. Random Early Detection Gateways for Congestion Avoidance. *IEEE/ACM Transaction on Networking* 1(4), August 1993, S. 397–413.
- [FlKa99] S. Floyd und F. Kall. Promoting the Use of End-to-End Congestion Control in the Internet. *IEEE/ACM Transaction on Networking* 7(4), August 1999, S. 458–472.
- [GaDS99] Meng Gan, Elmar Dorner und Jochen Schiller. Applying Computational Intelligence for Congestion Avoidance of High-Speed Networks. In *IEEE FTDCS'99*, 1999, S. 23–28.
- [Gan98] Meng Gan. Anwendung von Neuronalen Netzen und Fuzzy-Logik in der ATM-Verkehrslastregelung. In Institut für Telematik (Hrsg.), *Klausurtagung'98*, 1998.
- [Ghos98] S. Ghosh. A Survey of Recent Advances in Fuzzy Logic in Telecommunications Networks and New Challenges. *IEEE Transaction on Fuzzy Systems* 6(3), 1998, S. 443–447.
- [GoSS98] N. Golmie, Y. Saintillan und D. Su. ABR Switch Mechanisms: Design Issues and Performance Evaluation. *Computer Networks and ISDN Systems*, 1998, S. 1749–1761.
- [NIS] NIST, http://w3.antd.nist.gov/Hsntg/prd_atm-sim.html. *ATM/HFC Network Simulator*.
- [UCB] UCB/LBNL/VINT, <http://www-mash.cs.berkeley.edu/ns/ns.html>. *Network Simulator, V2*.

Context-Awareness Technologies for Situated Computing

Hans-Werner Gellersen

1 Introduction

It can be argued that pervasive access to information and services becomes more valuable when they are related and adapted to the real world situations of their use. We discuss sensor-based approaches to awareness of such situations based on experience from two projects, one investigating context-awareness as add-on technology for mobile devices, and the other exploring the use of everyday things as context suppliers.

Global and mobile networks make information and computer-based services available almost everywhere irrespective of location, embedded technologies let global information infrastructures stretch far into our everyday environments, and advances in information interfaces extend accessibility. Beyond the ultimate availability and accessibility an entirely new quality is added when information and services get related and adapted to the real world situations in which they are used. Situations of use have inherent locality, but there are many other aspects of situations that can serve as context for use of pervasive information.

Traditionally computer vision is applied to provide computer-based systems with perception of the surrounding real world. Advances in sensor technologies suggest alternative approaches to real world context acquisition based on embedded or body-worn sensor infrastructures [ScBG99]. In this position paper we relate our experience with technologies for sensor-based context-awareness, and discuss concepts for an architecture that we envision as platform for context-aware systems.

2 Experience with Technologies for Context-Awareness

We report on two projects we conducted over the last two years for exploration of sensor-based context-awareness. A fundamental idea underlying these projects is to use deliberately simple sensors and perception methods as opposed to expensive hardware and algorithms. The concept is to make do with the limited information obtained from individual sensors, and to achieve awareness of complex contexts through multi-sensor fusion (i.e. synthesis of features extracted from individual sensors). In comparison to computer vision it could be argued that this approach shifts complexity from the algorithmic to the architectural domain. The underlying motivation however is to yield at a perception technology suitable for devices and appliances with limited power and resources for computation.

2.1 TEA - Adding Context-Awareness to Mobile Devices

The TEA1 project is concerned with adding awareness of surrounding usage situations to personal mobile devices based on the following concepts:

- Association of multi-sensor data with specific situations such as for instance *being in a meeting*,
- *Integration of sensors and perception methods in an add-on component for mobile host appliances, for instance cellular phones and personal digital assistants*

The TEA system is based on a simple layered architecture that provides increasing levels of abstraction from sensors to adaptive application. The sensor layer is defined by an open array of sensors including both environmental sensors for perception of the real world and logical sensors for monitoring of conditions in the virtual world, for instance logical state of the host device. A second layer abstracts information from individual sensors to a number of so-called cues. The third layer provides for multi-sensor fusion based on synthesis of so-called contexts from cues. A fourth layer finally relates context to action to implement situated behavior.

For a feasibility study of the TEA approach we implemented a hardware/software prototype with an array of different environment sensors and embedded perception techniques (statistical analysis and AI methods). The prototype was used in conjunction with a mobile phone for a scenario in which otherwise manual selection of usage profiles (such as meeting, outdoors, in car etc, all with distinct notification and message forwarding behavior) was automated based on perceived context [SATT⁺99].

From the TEA feasibility study we gained relevant insights into design of sensor-based awareness technology, most importantly:

- Our study validated the two-step perception approach with feature extraction prior to sensor fusion. A similar study is reported, recommending likewise: „Don't forget to cook your sensors“ [GoLe99].
- In investigated „everyday situations“ we found audio, light and acceleration sensors to be most relevant for context acquisition, while others such as temperature and pressure contributed little or nothing.
- We realized placement of sensors has a large influence on performance. Tight packing of sensors as done in TEA compromises benefits that might be gained if sensors were placed in most appropriate positions on the body and connected via a body network to their host device.
- We found it makes sense to employ learning methods to improve context abstraction from available cues. This introduces the problematic issue of how to provide an interface between the TEA component and the user for feedback on performance of the awareness system.

2.2 MediaCup - Awareness Technology Embedded in Everyday Things

In another project on sensor-based context awareness we considered ordinary things we use in our everyday activities as source of context information. In contrast to TEA the idea was to embed awareness technology in everyday artifacts, to obtain context information from many such artifacts, and to make this context information available network-wide rather than to specific host. The model underlying this approach is:

- Artifacts have an awareness of their own local state
- Artifacts broadcast their state as context information into the surrounding information ecology

- Any application can process and use broadcast context information

For demonstration of this approach we created MediaCups which are ordinary coffee cups augmented with sensors, processing, and short-range communication in a proximity-based network [GeBe00]. MediaCups can sense movement, temperature and whether they stand or are lifted up. On board the augmented cup, more abstract context is computed from obtained sensor data, for example „cup is hot“, „cup is carried around“ and „cup is drunken from“. Every two seconds cups broadcast state information via an infrared network into the LAN and potentially the Internet. Any application can listen for context information, and derive synthesized context, for example for awareness of co-located hot cups, which may serve as indicator for a meeting of people.

We have a small number of MediaCups in everyday use in our research group, gaining experience with respect to usability and utility but also with respect to context-awareness, for example:

- The separation of context acquisition and context consumption is valuable: development of applications is simplified, and acquired context can be used by multiple applications independently. This relates to findings from building a context toolkit [SaDA99].
- In our system, synthesis of context information is responsibility of applications but we found that many applications have the same requirements for synthesis, for example to resolve co-location of artifacts.
- Applications often only need a subset of available context information, and we found that selection was most often based on evaluation of the location attribute of MediaCups rather than their ID.

3 Towards an Architecture for Provision, Synthesis and Use of Context

Based on our experience from the described projects we envision an architecture for context-awareness that takes further the separation of concerns (acquisition, synthesis and use of context), and that provides support for selection of available context:

- The architecture is centered on a context component that serves as mediator between context providers and context consumers.
- The context component keeps context information as set of so-called facts. Facts represent information obtained from logical and physical sensors, and provided by any kind of context provider, connected to the current time and the location.
- All facts have location and time as primary selection criteria: Fact=(location, time, value, description). Applications can define temporal-spatial regions of interest based on these selection criteria but this does not exclude selection by value or description.

The architecture is based on a blackboard model with three distinct interfaces:

- Context providers can write facts onto the blackboard.
- Fact abstractors can read facts for synthesis of new ones that they add to the board.

- Applications can read facts. The application can select facts of interest based on attributes.

The blackboard can be implemented efficiently as a database table with indices on time and location. Access to providers, abstractors and consumer can for example be realized over HTTP to facilitate existing infrastructure.

References

- [GeBe00] H.-W. Gellersen und M. Beigl. Experience with Design and Use of Computer-Augmented Everyday Objects. *Computer Networks*, 2000. Special Issue on Pervasive Computing.
- [GoLe99] A. Golding und N. Lesh. Indoor Navigation Using a Diverse Set of Cheap Wearable Sensors. In *Proceedings of the Third International Symposium on Wearable Computing (ISWC'99)*, San Francisco, CA, USA, oct 1999. IEEE Press.
- [SaDA99] D. Salber, A. Dey und G. Abowd. The context Toolkit: Aiding the Development of Context-Enabled Applications. In *Proceedings of CHI'99*, Pittsburgh, PA, USA, may 1999. ACM Press.
- [SATT⁺99] A. Schmidt, K.A. Aidoo, A. Takaluoma, U. Tuomela, K. Van Laerhoven und W. Van de Velde. Advanced Interaction in Context. In H.-W. Gellersen (Hrsg.), *Handheld and Ubiquitous Computing*, Nr. 1707 der Lecture Notes in Computer Science, Karlsruhe, Germany, sep 1999. Springer-Verlag, S. 89–101.
- [ScBG99] A. Schmidt, M. Beigl und H.-W. Gellersen. There is more to Context than Location. *Computer & Graphics* 23(6), Dezember 1999, S. 893–901.

Adaptive Web-Anwendungen

Guntram Graef

1 Einleitung

Ein Aspekt der Anwendungen im World Wide Web (Web) von PC-Anwendungen unterscheidet, ist die Möglichkeit Informationen über das Verhalten zahlreicher Benutzer an zentraler Stelle auszuwerten. Maes und Shardanand [ShMa95] haben gezeigt wie an zentraler Stelle gesammelte Informationen über Benutzerinteressen dazu verwendet werden können, Anwender mit Informationen zu versorgen die ihren Bedürfnissen entsprechen. Anstatt nur einzelne HTML Seiten mit Informationen, wie etwa Produktempfehlungen, dynamisch zu erzeugen, soll in diesem Beitrag die Anpassung von ganzen Web-Anwendungen an die Bedürfnisse einzelner Benutzer betrachtet werden.

2 Evolution und Adaption von Web-Anwendungen

Die automatische Adaption von ganzen Anwendungen auf der Ebene einzelner Codeprimitive ist mit vorhandener Technologie praktisch nicht realisierbar. Daher ist es notwendig, sinnvolle Abstraktionen und Komponenten als Anwendungsbausteine zu identifizieren um die Komplexität des Adaptionsprozesses einer Anwendung zu reduzieren und beherrschbar zu machen. Dazu soll die Konstruktion von Software in eine angebotsorientierte Mikroebene und eine nachfrageorientierte Makroebene unterteilt werden. Auf der Mikroebene werden funktionale Komponenten zur Verfügung gestellt, während auf der Makroebene Komponenten entsprechend der Benutzerbedürfnisse zu Anwendungen föderiert werden.

Die einzige Art von Codeabstraktion die im Web allgemein zur Verfügung steht sind Dokumentressourcen. Diese bieten jedoch sehr wenig Flexibilität bezüglich der Granularität von Codeartefakten und nur wenige Möglichkeiten Beziehungen zwischen Dokumenten zu beschreiben. Diese Beschränkungen können jedoch mit Hilfe eines Programmiermodells, wie z.B. WebComposition [GeGa99], umgangen werden.

Aufbauend auf einem solchen allgemeinen Programmiermodell dienen Services als Bausteine, um, auf einer höheren Systemebene, Funktionalität für eine bestimmte Aufgabe, wie z.B. die Konfiguration einer Bestellung oder die Annahme einer Reklamation, zu kapseln [GaTu99]. Im Gegensatz zu Komponenten auf einer niedrigeren Systemebene, kann die Funktionalität von Services ohne besondere technische Kenntnisse verstanden und beschrieben werden [GrGa99]. Service-Komponenten stellen die Grundlage für die Adaption und die Evolution einer Anwendung auf Makroebene dar.

Die Adaption einer Web-Anwendung auf der Makroebene wird durch einen besonderen Anwendungsservice realisiert. Dieser Service steuert, auf der Basis einer automatisierten Analyse der Benutzeranforderungen, in welcher Weise andere Services dem Benutzer präsentiert werden (Abbildung 1).

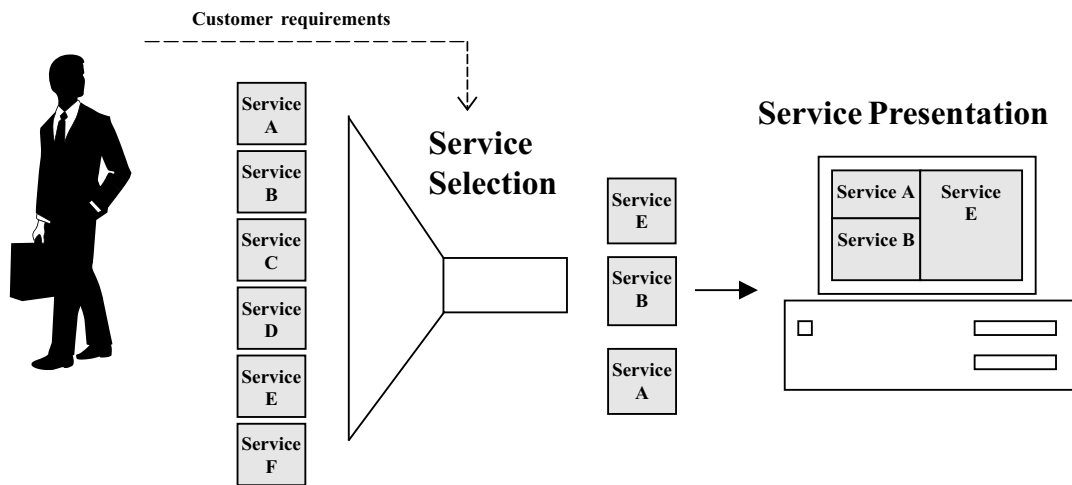


Abbildung 1: Adaptionsprozeß

3 Automatische Auswahl von Services

Eine Voraussetzung, um einen Adaptionsmechanismus für jede Art von Web-Anwendung zu realisieren, ist es, die Bedürfnisse einzelner Benutzer zu bestimmen. Im Rahmen dieser Arbeit wurden drei verschiedene Methoden untersucht, die alle auf der Analyse von Benutzerverhalten und der Verwendung von Services als funktionalen Bausteinen basieren:

Die erste Methode, hier bezeichnet als Conservative Service Selection (CSS), basiert ausschließlich auf dem früheren Verhalten desjenigen Benutzers, für den die Anforderungen bestimmt werden sollen. Diese Methode basiert auf der Annahme, daß Benutzer bestimmte, für sie wichtige, Tätigkeiten mehrmals ausführen. D.h. wenn ein Anwender bereits einen bestimmten Service einer Anwendung benutzt hat, wird es als wahrscheinlich betrachtet, daß er in Zukunft wieder auf den gleichen Service zugreifen möchte. Ähnliche Verfahren werden gewöhnlich für adaptive Anwendungen in Einbenutzerumgebungen, wie z.B. [ThKr93], eingesetzt. Leider ist CSS nicht in der Lage Benutzeranforderungen zu ermitteln die nicht auf wiederkehrendem Verhalten basieren.

Die zweite Methode soll hier als Social Service Selection bezeichnet (SSS) werden. Sie basiert auf der impliziten Auswertung wiederkehrender Verhaltensmuster bei Benutzern einer Anwendung. Verhaltensmuster basieren auf der Tatsache, daß Personen die in der Vergangenheit ein bestimmtes Verhalten A gezeigt haben, sehr wahrscheinlich in Zukunft ein bestimmtes Verhalten B ausüben werden. Früheres und zukünftiges Verhalten bilden dabei ein vollständiges Verhaltensmuster [Skin53]. Ein einfaches Verhaltensmuster kann z.B. sein, daß eine Person die bereits einen Service zur Installation von Telefonen genutzt hat, sehr wahrscheinlich (mit Wahrscheinlichkeit p) einen anderen Service zum Einkauf von Zubehör zu dem installierten Telefon benutzen wird. SSS basiert auf der impliziten Nutzung von Verhaltensmustern. Statt beobachtetes Verhalten direkt über Regeln mit Services zu verknüpfen, wird indirekt über Ähnlichkeitsbeziehungen zwischen Benutzern von beobachtetem Verhalten auf für einen Benutzer relevante Services geschlossen. Dieser Ansatz ähnelt existierenden Methoden zur dynamischen Selektion von Web Inhalten wie z.B. [ShMa95].

Die dritte Methode ist eine Kombination aus CSS und SSS, die im folgenden mit CSS/SSS bezeichnet wird. Dabei werden mit CSS und SSS ermittelte Resultate gleichgewichtet miteinander kombiniert.

4 Evaluation

Der hier vorgestellte Ansatz wurde in einem System für den elektronischen Handel angewandt. Es handelt sich dabei um Eurovictor II, das in einer Kooperation zwischen der Universität Karlsruhe und Hewlett-Packard entstand und heute europaweit im Einsatz ist (Abbildung 2).

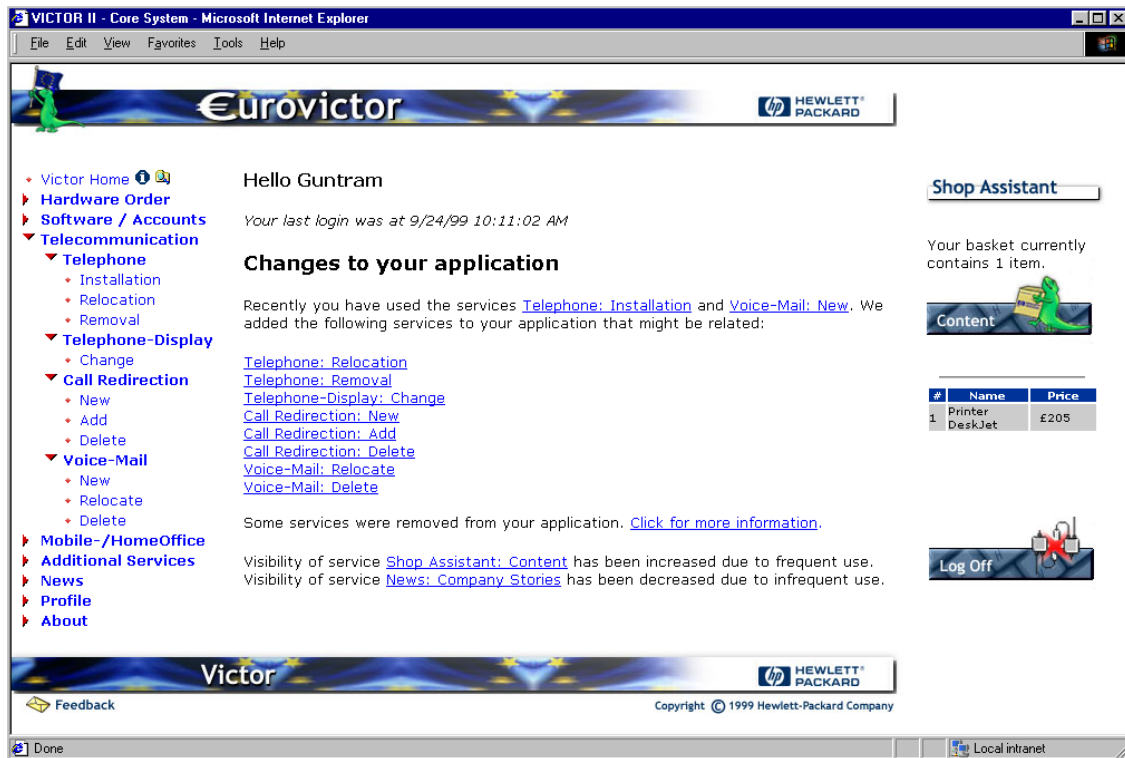


Abbildung 2: System Eurovictor

Eurovictor wurde mit Hilfe des WebComposition Programmiermodells entwickelt und seine Funktionalität liegt in der Form von Services vor, die zu unterschiedlichen Anwendungen föderiert werden können. Jede der drei oben genannten Methoden zur automatischen Service-Auswahl wurde getestet. Für die Evaluation wurden Daten über das frühere Verhalten von 1000 Benutzern verwendet und die daraus ermittelten Vorhersagen mit dem anschließenden tatsächlichen Verhalten der Anwender verglichen. In diesem Szenario wurden die besten Ergebnisse mit Hilfe der CSS/SSS Kombination erreicht, die 10% mehr Services korrekt ermittelt hat als die SSS Methode und 38% mehr als die CSS Methode.

5 Zusammenfassung

In diesem Beitrag wurde ein Konzept zur Erstellung selbstadaptiver Web-Anwendungen kurz beschrieben. Es basiert auf einem komponentenorientierten Programmiermodell für Web-Anwendungen sowie auf der automatisierten Analyse von Benutzerverhalten. Der Ansatz wurde in einem kommerziellen, auf Web-Technologie basierenden E-Commerce-System getestet und angewandt.

Literatur

- [GaTu99] M. Gaedke und K. Turowski. Generic Web-Based Federation of Business Application Systems for E-Commerce Applications. *Second International Workshop on Engineering Federated Information Systems (EFIS'99)*, Dezember 1999.
- [GeGa99] H.-W. Gellersen und M. Gaedke. Object-Oriented Web Application Development. *IEEE Internet Computing* 3(1), Januar 1999, S. 60–68.
- [GrGa99] G. Graef und M. Gaedke. An Evolution-oriented Architecture for Web Applications. *Second Nordic Workshop on Software Architecture (NOSA '99)*, September 1999.
- [ShMa95] U. Shardanand und P. Maes. Social information filtering: algorithms for automating 'word of mouth'. *Human factors in computing systems (CHI'95)*, Dezember 1995, S. 210–217.
- [Skin53] B. F. Skinner. *Science and human behavior*. New York. 1953.
- [ThKr93] F. C. Thomas und M. Krogsaeter. An adaptive environment for the user interface of Excel. *International Workshop on Intelligent User Interfaces (IUI)*, Dezember 1993, S. 123–130.

Unterstützung für Anwendungen in ubiquitären Systemen

Markus Lauff

1 Einleitung

Zu Beginn des neuen Jahrtausends erkennt man anhand von aktuellen Entwicklungen, dass sich die Art und Weise wie und wo verschiedenste Formen von Computern eingesetzt werden sehr stark verändert.

Als Schlagwort für diese Entwicklung finden sich Begriffe wie „Ubiquitous Computing“, „Calm Computing“ oder „Pervasive Computing“.

Die Abgrenzung der Begriffe ist nicht eindeutig möglich und ist zusätzlich oft durch individuelle Sichtweisen und Vorstellungen geprägt.

Als allgemeine Eckpunkte der im Rahmen dieser Arbeit betrachteten Veränderung von Computern und deren Anwendungen, lassen sich folgende Aspekte definieren:

- Die Anzahl der Computer in unserer Umgebung nimmt ständig zu. In immer mehr Geräten werden Mikrokontroller oder Mikrocomputer mit zum Teil beachtlicher Leistungsfähigkeit eingesetzt.
- Dem Benutzer stehen gleichzeitig mehrere Geräte mit unterschiedlichen Ein-/Ausgabemöglichkeiten zur Verfügung.
- Immer mehr Geräte verfügen über sehr unterschiedliche Kommunikationsmöglichkeiten. Zu diesen Kommunikationsmöglichkeiten zählen einfache Fernbedienungen, sowie auch komplexe Netzwerkprotokolle.
- Die Entwicklung von neuen Internettechnologien und die Vernetzung von Computern in allen Bereichen, erlauben transparente Zugriffe auf Informationen, unabhängig von Ort und Gerät.

Basierend auf diesen Punkten ist ein ubiquitäres System definiert als eine Menge von Umgebungen mit unterschiedlichen, heterogenen Ressourcen und Kommunikationsmöglichkeiten.

Die Entwicklung von Anwendungen für generische ubiquitäre Systeme scheitert bislang an der fehlenden Infrastruktur zur Verwaltung und Nutzung der unterschiedlichen Ressourcen und Umgebungen.

Die im Rahmen der hier vorgestellten Arbeit entwickelte Architektur, definiert eine solche Infrastruktur für ubiquitäre Systeme.

Als grundlegende Anforderungen an eine solche Architektur wurden folgende Punkte erarbeitet:

- Integrationsmöglichkeit existierender Systeme

- Verwaltung von Umgebungen, Ressourcen und deren Zustände
- Abstraktion bezüglich der Funktionalität und technischen Realisierung von Ressourcen

Die Architektur basiert auf zuvor erarbeiteten Modellen zur Definition der heterogenen Kommunikation, Ressourcen und Umgebungen.

Als Bezeichnung für diese Architektur wird im folgenden das Akronym UCMA für „Ubiquitous Computing Management Architecture“ verwendet.

2 Architektur

Die Architektur definiert drei Ebenen für Anpassungen, Dienste und Anwendungsunterstützung.

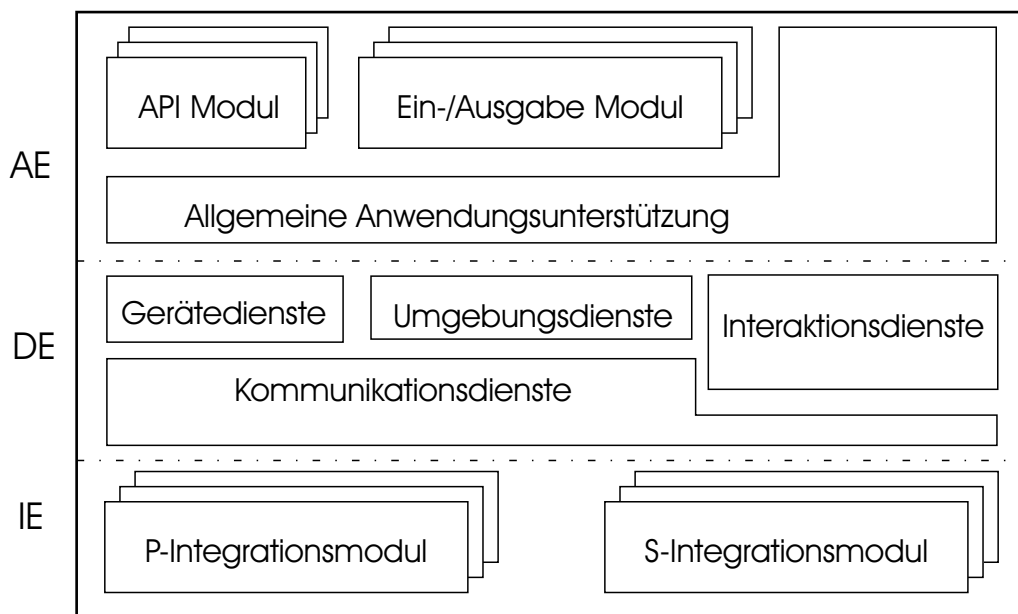


Abbildung 1: UCMA Architektur

2.1 Anpassungsebene

Die Anpassungsebene dient zur Integration der heterogenen Komponenten. Mit Hilfe von Anpassungsmodulen können die vorhandenen Kommunikationsprotokolle oder Automatisierungssysteme eingebunden werden. Die Module bilden damit die funktionale Komponente zum Aufbau des heterogenen Kommunikationsnetzes.

Beispiele für Systeme, die durch Anpassungsmodulen eingebunden werden können, sind X10, Jini, JetSend, Salutation oder Protokolle wie IrDA, Bluetooth.

Grundlegende Funktionen der Module sind

- die Umsetzung von Adressen zwischen den verschiedenen Systemen,
- die Umsetzung der Informationen in die unterschiedlichen Paketformate auf Anwendungsebene,

- die Überwachung des Netzes und die Bereitstellung von Informationen bezüglich des QoS,
- die An- und Abmeldung von Ressourcen beim entsprechenden Ressourcen- oder Kommunikationsdienst.

Zusätzlich zu der standardmäßigen Ende-zu-Ende Vermittlung, können Anpassungsmodulare auch Transit-Verbindungen erlauben. Mit Hilfe dieser Möglichkeit kann ein drittes Netz benutzt werden um zwei andere Netze zu verbinden.

2.2 Dienstebene

Die Dienstebene ist, wie auch die Anpassungsebene, modular aufgebaut. Zu den Standardmodulen der Dienstebene gehören die Dienste zur Verwaltung der Ressourcen, Umgebungen, Kommunikation und der Interaktion.

2.2.1 Ressourcendienste

Die Ressourcendienste sind für die Verwaltung und für die Ausführung der Funktionen der Ressourcen zuständig. Eine Ressource besteht innerhalb der Architektur aus einem Beschreibungsteil, einer Umgebungsinformation und einer Laufzeitkomponente.

Bei Aktivierung der Ressource wird die Laufzeitkomponente einem Interaktionssystem zur Ausführung übergeben. Die Laufzeitkomponente einer Ressource definiert ihre interne und externe Funktionalität.

Wechselt die Ressource in eine neue Umgebung, so wird sie bei der Laufzeitkomponenten der alten Umgebung ab und bei der Komponenten der neuen Umgebung wieder angemeldet. Bei Bedarf wird die Ressource auch einem neuen Interaktionssystem übergeben.

2.2.2 Umgebungsdienste

Hauptaufgabe der Umgebungsdienste ist die Abbildung der realen Umgebungen in ein hierarchische Modell.

Dieses Modell dient dann als Grundlage für die Zuordnung von Ressourcen zu Umgebungen. Die Zuordnung der Ressourcen zu ihrer Umgebung kann dabei auf unterschiedliche Art und Weise erfolgen.

Bei der expliziten Zuordnung einer Ressource zu einer bestimmten Umgebung, kann die die Ressource ihre Umgebung selbständig erkennen.

Bei der impliziten Zuordnung verwendet die Architektur Sekundärinformationen zur Eingliederung der Ressource in das Umgebungsmodell.

Zusätzlich zu diesen Funktionen zur Anmeldung von Ressourcen, stellt der Umgebungsdienst Funktionen zur Suche von Ressourcen und zur Modellierung von Umgebungen zur Verfügung.

2.2.3 Kommunikationsdienste

Der Kommunikationsdienst verwaltet die, durch Anpassungsmodulare zur Verfügung gestellten, Kommunikationsmöglichkeiten.

Elementare Aufgabe des Kommunikationsdienstes ist die Bereitstellung von Verbindungsinformationen mit Hilfe von dynamischen Routingtabellen.

Für die Berechnung der Routingtabellen verwendet der Kommunikationsdienst ausser den physikalischen Verbindungsinformationen auch Parameter über die benötigten QoS Eigenschaften und Kosten der Verbindung.

2.2.4 Interaktionsdienst

Der Interaktionsdienst stellt die Funktionen zur Verteilung der Laufzeitkomponenten auf Interaktionssysteme zur Verfügung.

Bei der An- oder Abmeldung von Ressourcen zu einem Interaktionssystem werden die aktuellen Zustände als Serialisierung dem Ressourcendienst übergeben.

2.3 Anwendungsunterstützungsebene

Die Anwendungsunterstützungsebene definiert die Schnittstelle (API) zu ubiquitären Anwendungen. Die API definiert Funktionen zum Zugriff auf das System. Die Funktionen sind unabhängig von der Realisierung der einzelnen Module und von den verwendeten Technologien.

Die Anwendungsunterstützung ist in die Bereiche Ressourcen, Umgebungen und Kommunikation gegliedert. Im Bereich Ressourcen finden sich Routinen zur An- und Abmeldung, zur Statusänderung, zur Lokalisation und zur Abfrage von Ressourcen. Der Bereich Umgebungen umfaßt die Routinen zur Verwaltung von Umgebungen. Die Routinen beinhalten zum Beispiel Funktionen zur expliziten und impliziten Zuordnung von Ressourcen zu Umgebungen. Zur Kommunikation gehören alle Routinen die zur kommunikationsprotokollübergreifenden Kommunikation innerhalb des ubiquitären Systems benötigt werden.

3 Zusammenfassung

Die UCMA Architektur erlaubt Anwendungen in ubiquitären Systemen den Zugriff auf verfügbare Ressourcen. Mit Hilfe einer funktionalen Abstraktion und der Verwendung von Umgebungsinformationen können Anwendungen in verschiedenen Umgebungen mit unterschiedlichen Ressourcen wiederverwendet werden.

Eine integrierte Dienstmanagement-Architektur für die qualitätsgesicherte Bereitstellung von Netz- und Systemdiensten

Christian Mayerl

Die Informationsverarbeitung sowie die zugrundeliegende technische Infrastruktur sind zu bedeutenden Wettbewerbsfaktoren von Unternehmen geworden. Netz-, System- und Anwendungskomponenten stellen zentrale Bestandteile der erforderlichen Infrastruktur dar. Aufgrund der Verteilung und der Vielfalt der zu betreibenden Komponenten ermöglicht erst der Einsatz von rechnergestützten *Managementwerkzeugen* einen effektiven und effizienten Betrieb dieser verteilten Informatiksysteme. Einen Ansatz für den effizienten Betrieb stellt die virtuelle Rezentralisierung von vernetzten Systemen (z.B. durch IT-Outsourcing) dar, durch den die Verantwortung für den Betrieb zentralisiert wird und die Systeme selbst verteilt bleiben. Dadurch entstehen dedizierte Netz- und Systemdienstleister, die mit Anforderungen an die Dienstqualität der vernetzten Informatiksysteme konfrontiert werden.

Ein Kunde definiert seine Qualitätsanforderungen in Form von *Dienstleistungsvereinbarungen*. Diese umfassen Anforderungen an die unmittelbar durch die Netz-, System- und Anwendungskomponenten realisierten Funktionen (*Kerndienste*) sowie an die zugehörigen Dienstleistungen wie eine Störungsbearbeitung oder eine koordinierte Änderungsdurchführung (*Zusatzdienste*). Bestehende Managementwerkzeuge unterstützen die notwendige technische Sicht auf Netz-, System- und Anwendungskomponenten. Diese reicht jedoch nicht aus, um die kundensorientierte Dienstsicht und damit eine Garantie der geforderten *Dienstqualität* hinreichend zu unterstützen. Steigende Qualitätsanforderungen sowie steigende Komplexität und Verteiltheit der Komponenten eines vernetzten Systems motivieren die Erweiterung bestehender Managementwerkzeuge um *fehlende Managementfunktionen*.

Ziel dieser Arbeit ist zum einen die Spezifikation einer integrierten *Dienstmanagement-Architektur (DMA)*, die Funktionen für den effektiven und effizienten Betrieb und für die qualitätsgesicherte Bereitstellung von Netz- und Systemdiensten integriert. Die DMA dient als Referenzarchitektur für die Implementierung von Managementlösungen, die individuelle Bedürfnisse des jeweils betrachteten Szenarios berücksichtigt. Da es aufgrund der unterschiedlichen Betreiberszenarien die Managementlösung bzw. das Managementwerkzeug nicht geben kann, soll zum anderen eine *Methode* entwickelt werden, mit der Betreiberanforderungen an Managementwerkzeuge spezifiziert werden können.

Als Grundlage dienen bestehende Managementarchitekturen und -werkzeuge, die diese Architekturkonzepte umsetzen. Eine Managementarchitektur stellt dabei ein Rahmenwerk für offene und integrierte Managementlösungen in heterogenen Umgebungen dar. Sie umfaßt folgende Teilmodelle: Das *Informationsmodell* beschreibt die zu verwaltenden *Managed Objects*. Das *Organisationsmodell* definiert Organisationsaspekte, Rollen und Kooperationsformen. Die für das Management von vernetzten Systemen benötigten Protokolle werden im *Kommunikationsmodell* spezifiziert. Notwendige Managementfunktionalität wird im *Funktionsmodell* strukturiert. Managementwerkzeuge implementieren Bausteine von Managementarchitekturen. Managementplattformen als die umfangreichste Art von Managementwerkzeugen stellen Referenzimplementierungen von (standardisierten) Managementarchitekturen dar.

Eine grundlegende Architektur stellt das *Management Framework for Open Systems Interconnection (OSI)* dar. Es führt alle vier Teilmodelle für den Betrieb von Telekommunikationssystemen aus. Dabei wird die management-relevante Information in einem objektorientierten Ansatz beschrieben. Im Organisationsmodell werden verwaltende und ausführende Rollen unterschieden. Das Kommunikationsmodell legt Mechanismen zum Austausch von Managementinformationen fest. Die Managementfunktionen werden in die Bereiche Konfiguration, Fehler, Leistung, Abrechnung und Sicherheit gegliedert. Die Ansätze des OSI-Managements werden vom *Telecommunication Management Network* aufgegriffen und um Funktionen für ein Dienst-, Kunden- und Unternehmensmanagement erweitert. Im Bereich der Datennetze dominiert das *Internet-/SNMP-Management*. Die Informationen werden hierbei in einem Verzeichnisbaum (Internet-Registrierungsbaum) mit Variablen als Blätter strukturiert. Der Austausch dieser Information erfolgt über das Simple Network Management Protocol (SNMP), das auf dem Internet-Protokoll basiert. Im Hinblick auf das Anwendungsmanagement und einem integrierten Management gewinnt die *Common Object Request Broker Architecture (CORBA)* als Managementarchitektur an Bedeutung. Für die Integration heterogener Arbeitsplatzrechner steht mit dem *Desktop Management Interface* der Distributed Management Task Force ein Managementansatz bis zum Kundenarbeitsplatz zur Verfügung. Ein flexibler Zugriff auf Managementinformationen und -funktionen wird durch *Web-basierte Managementarchitekturen* forciert. Proxy-Agenten überbrücken Schnittstellen zwischen unterschiedlichen Architekturen und proprietären Ansätzen.

Bestehende Managementwerkzeuge genügen den Anforderungen heutiger Netz- und Systemdienstleister nicht. Der Schwerpunkt dieser Werkzeuge liegt auf der Überwachung und Steuerung der Netz-, System- und Anwendungs-komponenten. Eine Dienst- bzw. Kundensicht wird nicht in hinreichender Form unterstützt. Als wesentliches Defizit ist der *Mangel an Kooperations- und Koordinationsfunktionen* für das Betriebspersonal anzusehen. Derartige Funktionen werden aber im Hinblick auf die Bereitstellung der Dienste und die Erfüllung der Dienstleistungsvereinbarungen benötigt. Die Individualität der Betreiberszenarien erfordert zudem ein methodisches Vorgehen für die Spezifikation der Managementwerkzeuge. Bestehende Methoden beschränken sich jedoch vorwiegend auf technische Teilaspekte, wie z.B. die Spezifikation einer individuellen *Management Information Base (MIB)*.

In dieser Arbeit ist ein Vorschlag für eine integrierte *Dienstmanagement-Architektur (DMA)* erarbeitet worden, die Managementfunktionen zur Überwachung und Steuerung eines vernetzten Systems mit Kooperationsfunktionen für eine qualitätsgesicherte Bereitstellung der Dienste in sich vereint. Dabei wird das Informationsmodell um Informationen über die bereitzustellenden Dienste und den kundenindividuellen und meßbaren Qualitätsinformationen erweitert. Diese Erweiterungen umfassen neben den Anforderungen an die Dienstgüte der *technischen Komponenten* ebenso *ablauforganisatorische Anforderungen* wie die Verfügbarkeit des Betriebspersonals, Bearbeitungszustände und Bearbeitungszeiten. Das Organisationsmodell wird um kooperierende Rollen innerhalb der Betriebsprozesse ergänzt. Neben der *vertikalen Kommunikation* zwischen dem Betriebspersonal und den vernetzten Systemen wird eine *horizontale Kommunikation* zwischen den Rollen innerhalb der Betriebsprozesse definiert. Für die Einhaltung der Dienstleistungsvereinbarungen werden Kommunikations-, Kooperations- und Koordinationsfunktionen spezifiziert, die die Betriebsprozesse anleiten (*Guideline Cooperation Concept*) und Managementfunktionen für vernetzte Systeme in diese Prozesse integrieren. *Assistenzfunktionen* unterstützen Rollen, innerhalb eines Betriebsprozesses effizient mit anderen Rollen zu kooperieren und die Dienstleistungsvereinbarungen zu erfüllen. Für den *Nachweis der Qualitätserfüllung* werden Funktionen für die Überwachung, Verdichtung und Auswertung von technischen und ablauforganisatorischen Kennzahlen bereitgestellt.

Für die betreiberindividuelle Implementierung der DMA ist in dieser Arbeit eine *Methode PRODEM (PROvider DEMands)* zur Spezifikation von Betreiberanforderungen an einzusetzende Werkzeuge entwickelt worden. Ausgehend von einer Bestandsaufnahme der Kundenan-

forderungen, der zu betreibenden Komponenten und bereits vorhandener Werkzeuge werden Defizite eines Betreibers identifiziert. Die Defizite führen zum Entwurf eines *prozeßorientierten Betriebskonzepts* und zur Spezifikation des Werkzeugeinsatzes. Für die Anwendung der Methode wird eine *Projektorganisation* vorgeschlagen, die i.W. aus den Rollen Betreiber, Berater und Entwickler besteht. Für die Beschreibung der Anforderungen werden *Ergebnismodelle* spezifiziert, die durch angepaßte und entwickelte *Modellierungswerkzeuge* erstellt werden können. Dabei ist das Konzept der *prozeßorientierten Managementmittel (POM)* entstanden, das bewußt die Sicht des Betreibers einnimmt, indem der Einsatz der Managementwerkzeuge im Kontext der Betriebsprozesse beschrieben wird.

Der Tragfähigkeitsnachweis der integrierten DMA erfolgte durch Anwendung der Methode PRODEM in Forschungs- und Industrieprojekten. Dabei wurden die *Dienststrukturen* in einer Datenbank implementiert, die sowohl technische als auch ablauforganisatorische Qualitätsanforderungen der Kunden dokumentiert. Zur Sicherstellung der Qualität von vernetzten Systemen wurden Managementwerkzeuge zur Überwachung und Steuerung der Netz-, System- und Anwendungskomponenten an das jeweilige Betreiberszenario angepaßt. Im Hinblick auf die Einhaltung der ablauforganisatorischen Anforderungen entstanden im Rahmen der Prozesse *Problem Management* und *Change Management prozeßorientierte Managementwerkzeuge*, die die Kommunikation und Kooperation des Betriebspersonals und damit die koordinierte Anleitung der Prozesse unterstützen. Die Anleitung der Betriebsprozesse wurde durch einen *Cooperation Assistant* ergänzt, der basierend auf Web-Technologien multimedial und in deskriptiver Form Richtlinien zur Durchführung der Betriebsprozesse beschreibt. Für den Nachweis der Qualitätserfüllung wurden Überwachungsfunktionen der Managementwerkzeuge genutzt. Ein *Generic Process Monitor* wurde entwickelt, der zur Transparenz der Betriebsprozesse beiträgt und Kennzahlen über die Durchführung der Prozesse liefert. Die Integration der prozeßorientierten Managementwerkzeuge und des Prozeßmonitors in die DMA schafften die Grundlage für einen dienstorientierten Betrieb von vernetzten Systemen. *Korrelationsmechanismen* zur Erhebung und Verdichtung von technischen und ablauforganisatorischen Kennzahlen dienten zum Nachweis für die Erfüllung der Dienstleistungsvereinbarungen.

Zur Unterstützung der Methode PRODEM wurde ein bestehendes *Prozeßmodellierungswerkzeug* an das Metamodell von PRODEM angepaßt. Ein weiteres Werkzeug wurde zur Definition von automatisierbaren Richtlinien (*Guided Definition Tool*) erstellt. Es erhält über eine XML-Schnittstelle Informationen aus dem Prozeßmodellierungswerkzeug und konfiguriert ein Workflow-Management-System zur Anleitung der Betriebsprozesse.

Die entwickelte Dienstmanagement-Architektur (DMA) beschreibt notwendige Funktionsbausteine sowie management-relevante Informationen eines Betreibers von vernetzten Systemen für die Erfüllung der Dienstleistungsvereinbarungen. Als ein zentraler konzeptioneller Beitrag ist die notwendige *Erweiterung der Architekturteilmodelle* des Managements anzusehen. Neben den technischen Informationen werden ablauforganisatorische Informationen benötigt. Die Funktionen zur Überwachung und Steuerung der Netz-, System- und Anwendungskomponenten werden um Funktionen zur Kommunikation, Kooperation und Koordination des Betriebspersonals ergänzt. Diese bilden die Grundlage für den Nachweis und die Garantie der zu erfüllenden Dienstleistungsvereinbarungen. Die Implementierung dieser Erweiterungen dienen dem Betreiber für eine *qualitätsgesicherte Bereitstellung* der Netz- und Systemdienste. Die Konzepte der DMA stellen daher nicht nur für den Entwickler von Managementwerkzeugen einen wertvollen Beitrag dar, sondern auch für den Betreiber, der diese Werkzeuge für den Betrieb von Informatiksystemen einsetzt. Die Methode PRODEM unterstützt ihn dabei, seine Anforderungen an die Werkzeuge zu spezifizieren und in regelmäßigen Intervallen an die Kundenanforderungen anzupassen. Die Anwendung der Methode PRODEM in Forschungs- und Industrieprojekten zeigte die Praxistauglichkeit auf. Der Einsatz der mit PRODEM spezifizierten Managementwerkzeuge hat dabei nachweislich zu Effizienzsteigerungen und zur Sicherstellung der Qualität geführt.

Fragestellungen bei mobilen drahtlosen Ad-hoc-Netzwerken

Daniel Müller

1 Einleitung

Die drahtlose Kommunikationstechnik hat mittlerweile starke Verbreitung gefunden, sowohl in Form von Mobiltelefonen, die noch hauptsächlich für Sprachkommunikation genutzt werden, als auch in Form von lokalen Datenkommunikationsnetzen (etwa nach IEEE 802.11). Die Kommunikation erfolgt dabei – auch wenn alle Kommunikationspartner eine drahtlose Netzanbindung nutzen – meist über eine feststehende (nicht mobile) Basisstation; diese stellt gegebenenfalls auch gleichzeitig den Übergangspunkt ins Festnetz dar. Auch bei der am weitesten fortgeschrittenen „Mobilitätsenerweiterung“ des Internet, Mobile IP, wird vorausgesetzt, dass mobile Systeme immer direkt – wenn auch evtl. drahtlos – an ein Festnetz an ihrem jeweiligen Aufenthaltsort angeschlossen sind.

Besonders in Umgebungen, in denen die Dichte mobiler, drahtlos angebundener Netzteilnehmer relativ hoch ist, bietet es sich an, den Umweg über eine Basisstation einzusparen, und direkt drahtlose Verbindungen zwischen den mobilen Kommunikationsteilnehmern einzusetzen. Auf eine feste Infrastruktur kann völlig verzichtet werden, wenn zwischen den mobilen Teilnehmern dynamisch ein so genanntes Ad-hoc-Netzwerk aufgebaut wird, dessen sämtliche Komponenten mobile Endgeräte sind. Insbesondere muss dazu die Weiterleitung von Daten zwischen kommunikationswilligen Teilnehmern, die sich physikalisch nicht innerhalb einer durch direkte Funkverbindungen überbrückbaren Reichweite befinden, ebenfalls von mobilen Netzteilnehmern übernommen werden.

Eine mögliche Anwendung solcher Netze ist die Schaffung einer Alternative zu zellularen Mobilfunkinfrastrukturen. Dies ist einerseits kostengünstig, da kein öffentlicher Dienstleister involviert ist, der Entgelte für die Benutzung der zur Verfügung gestellten Netzinfrastruktur berechnet, andererseits auch robuster, da keine Abhängigkeit von einer funktionsfähigen Infrastruktur besteht; dies kann entscheidend bei Rettungs- oder Katastropheneinsätzen sein, wo keine Infrastruktur vorhanden oder diese beschädigt ist. Weitere Anwendungsgebiete erschließen sich etwa unter den Schlagworten „Wearable Computing“ oder „Nomadic Computing“.

2 Eigenschaften von Ad-hoc-Netzwerken

Bedingt durch die Mobilität der Netzkomponenten und die Charakteristik drahtloser Verbindungsstrecken unterscheiden sich Ad-hoc-Netzwerke in einigen wesentlichen Eigenschaften erheblich von leitungsgebundenen Festnetzen.

- *Dynamische Topologie* — Die Topologie eines Ad-hoc-Netzwerks wird durch die drahtlosen Verbindungsstrecken zwischen den mobilen Routern gebildet und kann ein beliebiger (wegen der in drahtlosen Netzen häufiger auftretenden asymmetrischen Verbindungsstrecken evtl. auch gerichteter) Graph sein. Da die Router sich frei bewegen und beliebig

organisieren können, kann sich die Netztopologie schnell und unvorhergesehen ändern. Entsprechend ergeben sich gegenüber Festnetzen stark gesteigerte Anforderungen an Routing-Protokolle.

Da jederzeit mit dem Ausfall einzelner Verbindungsstrecken oder Netzknoten gerechnet werden muss, kann es in einem Ad-hoc-Netzwerk keinerlei zentrale Instanzen geben, alle Funktionalität muss verteilt und redundant erbracht werden. Die sich aus einer gegenüber Festnetzen im Allgemeinen höheren Anzahl möglicher Verbindungsstrecken zwischen den Netzknoten ergebende Redundanz kann sich durchaus als Vorteil erweisen.

- *Begrenzte Bandbreite* — Drahtlose Verbindungen besitzen momentan und aller Voraussicht nach auch in Zukunft eine wesentlich geringere Bandbreite als leitungsgebundene. Die Protokolle zur Netzwerkorganisation müssen deshalb mit dieser Ressource sehr sparsam umgehen, um möglichst viel davon für die Nutzdatenkommunikation freizuhalten.
- *Energieknappheit* — Da viele mobile Netzknoten auf Batterien angewiesen sind, stellt auch Energie eine knappe Ressource da, die entsprechend sparsam behandelt werden muss. Schon deshalb sind Ad-hoc-Netze im Allgemeinen kaum als Transitnetze geeignet, auch wenn sie mit einem Festnetz verbunden sind.

Durch die Notwendigkeit des Energiesparens begeben sich manche Endgeräte zeitweise in einen „Schlaf-Modus“, in dem sie nicht auf Netzwerkaktivität reagieren können.

Es wurde bisher (und wird auch im Folgenden) implizit davon ausgegangen, dass die Kopplung der mobilen Komponenten in der Schicht 3, der Netzwerkschicht des OSI-Basisreferenzmodells erfolgt. Dies ist jedoch nicht unbedingt erforderlich. Eine Kopplung auf Schicht 2, wie sie in leitungsgebundenen Netzen durch Brücken realisiert wird, kann beispielsweise ebenfalls sinnvoll sein. Ein Vorteil der Kopplung in der Netzwerkschicht ist die Möglichkeit, Heterogenität bzgl. physikalischer Medien abzudecken: Mehrere Netzwerkschnittstellen, die unter Umständen verschiedene physikalische Übertragungsverfahren einsetzen, können parallel verwendet werden, so dass ein Netz über verschiedene physikalische Medien hinweg entstehen kann. Eine Anpassung an neue, fortgeschrittenere Technologien auf physikalischer Ebene ist so ebenfalls auf einfache Weise möglich.

3 Fragestellungen

Stellvertretend für eine Vielzahl von Fragestellungen, die sich aus den im Vergleich zu herkömmlichen Fest- und Mobilfunknetzen verschiedenen Eigenschaften ergeben und die derzeit noch Gegenstand der Forschung sind, soll im Folgenden auf zwei Aspekte noch etwas näher eingegangen werden. Kaum eingegangen wird beispielweise auf die Berücksichtigung von Dienstqualitätsaspekten; der Beitrag von Marc Bechler (S. 1) enthält Näheres dazu.

3.1 Routing

Wie bereits angesprochen müssen Routing-Protokolle für Ad-hoc-Netzwerke einerseits auch bei einer hochdynamischen Netztopologie zuverlässig arbeiten und andererseits aufgrund der hohen Ausfallwahrscheinlichkeit für Verbindungsstrecken und Netzknoten verteilt, ohne zentrale Instanzen realisiert werden. Eine wichtige Vorgabe ist außerdem, dass durch das Routing-Protokoll möglichst wenig Kommunikationsaufwand entstehen sollte, um Energie zu sparen und die knappe Bandbreite für Nutzdaten freizuhalten.

Voraussetzung für das Routing in einem mehrere physikalische Medien überspannenden Netz ist die Verwendung einer Knotenkennung zur Identifizierung eines mobilen Netzknotens innerhalb des Multigraphen, der sich durch die Überlagerung der Topologien mehrerer physikalischer Medien ergibt [CoMa99].

Es gibt im Wesentlichen zwei grundsätzliche Strategien bezüglich der Aktualisierung von Routing-Informationen in Ad-hoc-Netzwerken:

- *Optimale Wegewahl* — Das Routing-Protokoll versucht, die Routing-Tabellen bei Änderungen so schnell wie möglich zu aktualisieren, so dass bezüglich einer bestimmten Metrik optimale Wege gewählt werden können.
- *Minimaler Overhead* — Das Protokoll versucht, mit möglichst wenig Kommunikationsaufwand funktionsfähige Wege zu bestimmen, die nicht unbedingt optimal sein müssen.

Als Metriken kommen zu den bei leitungsgebundenen Netzen üblichen Kriterien bei Ad-hoc-Netzwerken neue hinzu, etwa die zu erwartende Interferenz zwischen Funkübertragungstrecken, der Energieverbrauch oder die Erfüllung von Dienstgüteanforderungen unter engen Bandbreitenbeschränkungen.

Existierende Routing-Verfahren lassen sich nach der Art und Weise kategorisieren, auf welche die Router Routing-Information erhalten:

- *Tabellengesteuerte Verfahren* — Jeder Router hält Pfadinformation zu jedem bekannten Ziel vor (Beispiele: DSDV, WRP, WIRP und Least-Resistance-Routingprotokolle). Solche Verfahren arbeiten meist mit optimaler Wegewahl.
- *Bedarfsgesteuerte Verfahren* — Routing-Information wird nur für Ziele vorgehalten, die auch kontaktiert werden müssen. In einfachster Realisierung flutet der Router eine Suchnachricht ins Netz, wenn er ein gesuchtes Ziel noch nicht kennt. Verschiedene Ansätze (z.B. AODV, ABR, DSR, TORA, SSA, ZRP) unterscheiden sich in den Mechanismen, mit denen Suchnachrichten und Antworten verteilt, mitgehörte Antworten auf Anfragen anderer einbezogen, Kosten einzelner Links bestimmt oder die Existenz von Nachbarn geprüft werden. Bedarfsgesteuerte Verfahren verfolgen im Allgemeinen die Strategie des minimalen Overhead.

3.2 Sicherheit

Schon allein durch den offenen Zugang zur physikalischen Luftschnittstelle werden in drahtlosen Netzen im Allgemeinen Angriffe, die auf Abhören, Stören und Erzeugen von übertragenen Daten beruhen, stark vereinfacht. Bezüglich benutzerbezogener Kommunikation lässt sich dem häufig mittels kryptographischer Ende-zu-Ende-Sicherung, etwa mit Verfahren wie SSL bzw. TLS, begegnen. Da in Ad-hoc-Netzwerken aber außerdem die Netzstruktur ständig neu organisiert werden muss, bietet sich für Angreifer, die Protokollnachrichten beeinflussen oder sich selbst als aktive Protokollknoten etablieren können, die vielversprechende Möglichkeit, auf die Struktur des Netzes einzuwirken, um etwa die Verfügbarkeit des gesamten Netzes zu beeinflussen, einzelne Knoten abzuschneiden oder Daten gezielt über bestimmte Strecken zu leiten, wo sie z.B. besonders leicht abgehört werden können.

Da die Infrastruktur aus ständig wechselnden Knoten besteht, ist es unter Umständen schwierig oder unmöglich, bei Sicherheitsverfahren auf im Voraus vereinbarte Information zurückzugreifen. Aus den oben genannten Gründen sind außerdem keine zentralen Instanzen wie etwa Schlüsselverwaltungsdienste realisierbar. Einen interessanten Ansatz bilden hier kryptographische Verfahren, die auf der Mitwirkung einer Mehrheit aus einer bestimmten Menge von Teilnehmern beruhen, ausgehend von der Annahme, dass Angreifer im Allgemeinen keine Mehrheit darstellen [ZhHa99].

4 Ausblick

Bei der IETF gibt es eine Arbeitsgruppe, die sich mit dem Thema der Ad-hoc-Netzwerke befasst. Ziel ist dabei momentan primär die Entwicklung eines geeigneten Routingprotokolls für Unicast-IP-Verkehr. Ferner sollen Adressierung, Sicherheit und Zusammenarbeit mit anderen Schichten sowie langfristig z.B. auch Multicast und Dienstgüteunterstützung behandelt werden.

Den nächsten Schritt in eigenen Arbeiten stellt eine detaillierte Analyse der Problemstellungen und der existierenden Ansätze dar. Bezüglich Routing-Protokollen müssen dabei zunächst Kriterien zur Bewertung von Protokollen für Ad-hoc-Netzwerke zusammengestellt werden, um diese zu kategorisieren und Schwachstellen ausfindig zu machen. Auf dem Gebiet der Sicherheit ist eine Analyse der Möglichkeiten für Angriffe erforderlich, mit denen beispielsweise in das Routing eingegriffen werden kann. Anschließend kann mit der Entwicklung einer umfassenden Architektur begonnen werden, die eine sichere Organisation von Ad-hoc-Netzwerken ermöglichen.

Besondere Berücksichtigung soll dabei die Möglichkeit der Verwendung schichtenübergreifender Mechanismen finden; beispielweise bei der Behandlung von Fehlern könnten hier wertvolle Hinweise aus Informationen gewonnen werden, die anderen Schichten zur Verfügung stehen.

Literatur

- [CoMa99] S. Corson und J. Macker. *Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations*. IETF, Januar 1999. RFC 2501.
- [ZhHa99] L. Zhou und Z. J. Haas. Securing Ad Hoc Networks. *IEEE Network* 13(6), November 1999, S. 24–30.

FATIMA: Eine Architektur zur sicheren Mobilitätsunterstützung in IP-Netzen

Frank Pählke

1 Einleitung

Der derzeit verfügbare Standard zur Mobilitätsunterstützung in der IP-Schicht ist die „Mobile IP“-Architektur, welche in RFC 2002 [Perk96] spezifiziert ist. Dem Einsatz des Mobile-IP-Standards in realen Netzen stehen jedoch noch etliche Probleme im Wege, vor allem im Bereich der Sicherheit sowie bei der Interoperabilität mit einigen in der Praxis häufig eingesetzten Techniken. Insbesondere sind hier private, außerhalb eines Firmennetzes nicht sichtbare IP-Adressräume sowie Firewalls mit Paketfiltern und Adressumsetzungsmechanismen (network address translation, NAT) zu nennen.

2 Mobile IP

Im Folgenden wird zunächst der Mobile-IP-Standard vorgestellt. Danach wird auf die Defizite des Standards sowie auf aktuelle Ansätze zu deren Behebung eingegangen.

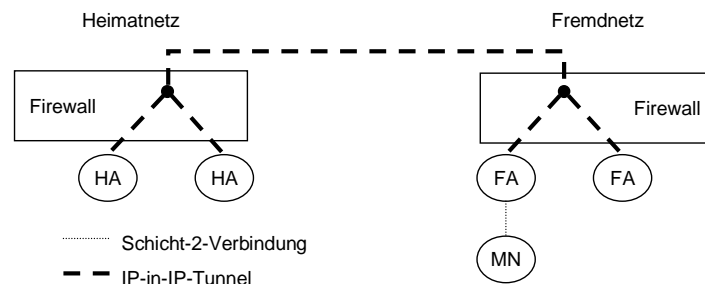


Abbildung 1: Die Mobile-IP-Architektur gemäß RFC 2002

Die generelle Architektur von Mobile IP ist in Abb. 1 zu sehen. Als Stellvertreter für einen abwesenden mobilen Rechner (mobile node, MN) fungiert in seinem Heimnetz ein sogenannter Heimatagent (home agent, HA). Dieser fängt während der Abwesenheit des MN alle an den MN gesendeten Datenpakete ab und schickt sie (üblicherweise durch einen IP-in-IP-Tunnel gemäß RFC 2003) an den aktuellen Aufenthaltsort weiter. Hierzu benötigt der HA Schicht-2-Konnektivität zum Heimat-Subnetz des mobilen Knotens, d.h. im Heimnetz muss pro Subnetz mindestens ein HA installiert werden.

Der aktuelle Aufenthaltsort eines MN, d.h. der Tunnelendpunkt für die umgeleiteten Datenpakete, wird als „care-of address“ (COA) bezeichnet. Die COA ist entweder eine IP-Adresse, welche dem MN in einem Fremdnetz zugewiesen wurde, oder die Adresse eines sogenannten „foreign agent“ (FA), welcher die getunnelten Pakete auspackt und an den MN weiterleitet.

Hierzu benötigt der FA Schicht-2-Konnektivität zum MN, sodass auch im Fremdnetz mindestens ein FA pro Subnetz installiert werden muss.

Vom MN gesendete Pakete müssen nicht den Umweg über den HA gehen, sondern können direkt an ihre Zieladresse geschickt werden, wobei als Quelladresse allerdings die Heimatadresse des MN und nicht die COA eingetragen wird.

2.1 Defizite und Erweiterungen

Einem breiten Einsatz von Mobile IP zur Mobilitätsunterstützung stehen noch mehrere Probleme entgegen, welche derzeit Gegenstand intensiver Forschungsarbeiten sind:

- Sobald auf der Strecke zwischen einem MN und einem Kommunikationspartner Firewalls installiert sind, werden die vom MN ausgesendeten Pakete auf Grund ihrer topologisch falschen Quelladresse i.d.R. verworfen. Als Abhilfe können auch diese Pakete zunächst zum HA getunnelt werden (sog. „reverse tunnelling“), was auf Grund des zusätzlichen Umwegs allerdings zu Effizienzeinbußen führt.

Noch größere Probleme ergeben sich im Falle von privaten Adressräumen und Adressumsetzungsmaßnahmen, da grundsätzlich eine vom Heimatnetz aus erreichbare COA gefunden werden muss.

- Jeder Tunnelendpunkt innerhalb eines Netzes (hierzu zählen z.B. auch Modem- und ISDN-Eingänge) ist eine potentielle Gefahrenquelle und sollte sorgfältig administriert werden. Da in jedem Subnetz ein HA bzw. FA installiert werden muss, untergräbt dies den wesentlichen Vorteil einer Firewall-Architektur, nämlich die zentrale Administrierbarkeit sicherheitskritischer Funktionen.
- Eine gegenseitige Authentisierung von MN und Fremdagent kann nur schwer erreicht werden, solange keine Infrastruktur zum Austausch der benötigten kryptographischen Schlüssel vorhanden ist.

Auch die gegenseitige Authentisierung von HA und FA ist selbst dann schwierig, wenn die Betreiber von Heimat- und Fremdnetz eng zusammenarbeiten, denn pro Subnetz ist ein HA bzw. FA vorhanden, und zwischen allen HA/FA-Paaren müssen geeignete Schlüssel vorhanden sein.

- Der Tunnel zum HA stellt für Datenpakete u.U. einen großen Umweg dar, etwa wenn der MN mit einem Partner im gerade besuchten Fremdnetz kommuniziert. Aus diesem Grunde wurden Mechanismen zur Routenoptimierung spezifiziert, bei denen der HA den Kommunikationspartnern eines MN dessen aktuelle COA mitteilt.
- Bewegt sich ein MN zwischen benachbarten Subnetzen desselben Fremdnetzes, so ist bei jedem Subnetz-Wechsel eine erneute Registrierung notwendig, wozu Kontrollnachrichten zwischen Fremd- und Heimatnetz ausgetauscht werden müssen. Kommt zusätzlich eine Routenoptimierung zum Einsatz, so müssen jedes Mal alle Kommunikationspartner des MN benachrichtigt werden.

In der Mobile-IP-Arbeitsgruppe der IETF (Internet Engineering Task Force) werden zur effizienten Unterstützung dieser sogenannten „Mikromobilität“ derzeit die konkurrierenden Ansätze „Cellular IP“ [CGWK⁺00], „HAWAII“ [RPTV⁺99] und „Hierarchical Mobile IP“ [Cast99] diskutiert.

3 Die Firewall-Aware Transparent Internet Mobility Architecture (FATIMA)

Ein weiterer Ansatz, um die geschilderten Probleme zu beheben, ist die derzeit am Institut für Telematik entwickelte „Firewall-Aware Transparent Internet Mobility Architecture“ FATIMA [MPSS00a]. Im Folgenden werden die Entwurfsziele von FATIMA erläutert sowie ein Überblick über die wichtigsten Bausteine der Architektur gegeben. Abschließend werden einige Vorteile des neuen Ansatzes herausgestellt.

3.1 Entwurfsziele

Aus der Analyse verschiedener bestehender Ansätze ergeben sich für FATIMA die folgenden Entwurfsziele [MPSS00b]:

- Sicherheitskritische Funktionen (z.B. die Registrierung von Mobilrechnern im Fremdnetz) müssen zentral administrierbar sein.
- Nachrichten, die zur Umkonfigurierung der Netzinfrastruktur (z.B. von Routing-Tabellen) führen, müssen grundsätzlich von Infrastrukturkomponenten stammen, nicht z.B. von anwesenden Mobilrechnern.
- Die Schnittstellen zwischen der Fremdnetz-Infrastruktur und besuchenden Mobilrechnern bzw. deren Heimatnetz sollten den bestehenden Standards entsprechen.
- Alle miteinander kommunizierenden Instanzen müssen sich gegenseitig sicher authentifizieren können. Hierzu werden hinreichend generische Schnittstellen zu Authentisierungs- und Schlüsselverwaltungsdiensten benötigt.

3.2 Bausteine der Architektur

Das wesentliche Merkmal der FATIMA-Architektur ist die Konzentration aller sicherheitskritischen Funktionen in einem Gateway, welches in die Firewall eines Netzes integriert wird. In den einzelnen Subnetzen werden anstelle der Heimat- und Fremdagenten von Mobile IP wesentlich einfachere Stellvertreteragenten (home/foreign agent proxies) eingesetzt. Zwischen allen Infrastrukturkomponenten werden die übertragenen Daten durch ESP-Tunnel (Encapsulating Security Payload, RFC 2406) authentisiert und optional verschlüsselt.

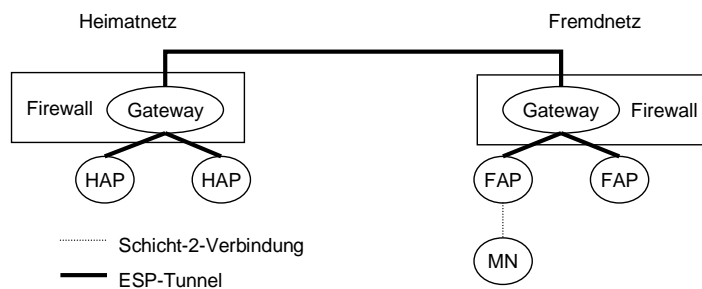


Abbildung 2: FATIMA-Komponenten: Minimale Konfiguration

Eine sehr einfache Konfiguration ist in Abb. 2 gezeigt; sie entspricht im Wesentlichen der Mobile-IP-Konfiguration aus Abb. 1. Um die Skalierbarkeit in großen Netzen zu verbessern, ist auch ein hierarchischer Aufbau mit zusätzlichen Zwischensystemen denkbar, worauf hier aus Platzgründen nicht weiter eingegangen wird. Die Funktionalität der in Abb. 2 gezeigten Komponenten teilt sich im Wesentlichen wie folgt auf:

- *FATIMA-Gateway*: Das Gateway ist die zentrale mobilitätsunterstützende Instanz in einem Netz. Es gehört zur Firewall des Netzes und ist von innen wie außen erreichbar. Alle sicherheitskritischen Funktionen (z.B. Registrierung von MNs und Entkapselung getunnelter Datenpakete) werden vom Gateway erbracht und können daher zentral administriert werden. Nach außen hin erscheint das Gateway als der einzige Fremdagent und gleichzeitig als der einzige Heimatagent des Netzes. Jedem besuchenden Mobilrechner wird als COA die Adresse des Gateways zugewiesen.
- *Foreign Agent Proxy (FAP)*: Alle Fremdagenten werden durch wesentlich einfachere Stellvertreter, die FAPs ersetzt. Gegenüber einem besuchenden Mobilrechner verhält sich der FAP exakt wie ein normaler Fremdagent. Er bearbeitet die Kontrollnachrichten z.B. bei der Registrierung allerdings nicht selber, sondern leitet sie an das Gateway weiter, welches die eigentliche Registrierungsfunctionalität erbringt. Die Antworten des Gateways werden umgekehrt an die Mobilrechner weitergeleitet, welche den FAP daher nicht von einem gewöhnlichen Fremdagenten unterscheiden können.
- *Home Agent Proxy (HAP)*: Alle Heimatagenten werden ebenfalls durch einfachere Stellvertreter ersetzt. Diese werden durch Kontrollnachrichten des FATIMA-Gateways aktiviert, fangen alle für abwesende Mobilrechner bestimmte Datenpakete ab und leiten sie an das FATIMA-Gateway des Heimatnetzes weiter. Das Gateway tunnelt die Pakete zur COA und wickelt den gesamten Mobile-IP-Kontrolldatenverkehr mit dem Fremdnetz ab.

3.3 Vorteile

Neben den Entwurfszielen aus Abschnitt 3.1 ergeben sich einige weitere Vorteile:

- Durch die Platzierung des FATIMA-Gateways innerhalb der Firewall ist der Einsatz auch dann problemlos möglich, wenn im Fremd- oder Heimatnetz private IP-Adressen verwendet werden.
- Die gegenseitige Authentisierung von Fremd- und Heimatnetzinfrastruktur wird vereinfacht, da nur noch zwischen je einem FATIMA-Gateway ein Schlüssel vorhanden sein muss und nicht zwischen zahlreichen FA/HA-Paaren.
- Da die COA bei Subnetz-Wechseln innerhalb eines Fremdnetzes gleichbleibt, werden im Falle einer Routenoptimierung zahlreiche Kontrollnachrichten zwischen dem HA und den Kommunikationspartnern eines MN eingespart.
- Ein FATIMA-Gateway bildet mit seiner zentralisierten Funktionalität einen guten Ansatzpunkt zur Implementierung effizienter Mikromobilitäts-Mechanismen.

4 Zusammenfassung und Ausblick

In den vorigen Abschnitten wurde erörtert, welches Defizite der derzeitige Mobile-IP-Standard aufweist und auf welche Weise diese Defizite durch die Mobilitätsunterstützungsarchitektur FATIMA behoben werden sollen. Derzeit laufende Arbeiten umfassen eine exakte Spezifikation von FATIMA sowie die Erstellung einer Simulationsumgebung, mit deren Hilfe Aussagen zu Effizienz und Skalierbarkeit der Architektur getroffen werden können. Von besonderem Interesse ist hierbei die Verhinderung potentieller negativer Einflüsse der aus Sicherheitsgründen notwendigen weitgehenden Zentralisierung der Mobile-IP-Funktionalität auf Effizienz und Skalierbarkeit der Architektur.

Literatur

- [Cast99] Claude Castelluccia. A Hierarchical Mobile IPv6 Proposal. In *Proceedings of AMOS ACTS Mobile Summit*, Sorrento, Italy, Juni 1999. Also published as INRIA technical report TR-0226, November 1998.
- [CGWK⁺00] Andrew T. Campbell, Javier Gomez, Chieh-Yih Wan, Sanghyo Kim, Zoltan R. Turanyi und Andras G. Valko. *Cellular IP*. IETF, Mobile IP Working Group, Januar 2000. Internet Draft, draft-ietf-mobileip-cellularip-00.txt.
- [MPSS00a] Stefan Mink, Frank Pählke, Günter Schäfer und Jochen Schiller. FATIMA: A Firewall-Aware Transparent Internet Mobility Architecture. In *Proceedings of ISCC 2000*, Antibes, France, Juli 2000.
- [MPSS00b] Stefan Mink, Frank Pählke, Günter Schäfer und Jochen Schiller. Towards Secure Mobility Support for IP Networks. Submitted to ICCT 2000, Beijing, China, August 2000.
- [Perk96] Charles E. Perkins. *IP Mobility Support*. IETF, Oktober 1996. RFC 2002.
- [RPTV⁺99] R. Ramjee, T. la Porta, S. Thuel, K. Varadhan und L. Salgarelli. *IP micro-mobility support using HAWAII*, Juni 1999. Internet Draft, draft-ietf-mobileip-hawaii-00.txt (expired).

Bedarfsorientierte Dienstgüteunterstützung durch adaptive Endsysteme

Hartmut Ritter

1 Einleitung

Die Integration verschiedener Dienste in ein einheitliches Kommunikationsnetz, wie es das Internet und seine mobilen Erweiterungen darstellen, stellt hohe Anforderungen an die Verwaltung der Betriebsmittel der beteiligten Kommunikationssysteme. Die Reservierung von Betriebsmitteln stößt jedoch in diesen Netzen an die Grenzen des technisch Umsetzbaren. Die Gründe dafür sind die hohe Heterogenität der Netze und die große zeitliche Varianz der verfügbaren Ressourcen. Letzteres ist leicht ersichtlich bei mobilen Netzen, bei denen die verfügbaren Ressourcen von schnell veränderlichen Faktoren wie Ausbreitungsbedingungen und momentaner Teilnehmerzahl abhängen, aber auch im Festnetz sind mit der Einführung neuer Techniken Schwankungen der verfügbaren Bandbreite um mehrere Größenordnungen möglich. Dies gilt insbesondere für das Zugangnetz, wo wichtige neuen Techniken wie xDSL, Datenkommunikation über Kabelmodem oder das Stromnetz, keine festen Frequenzbereiche und damit Bandbreiten mehr garantieren können.

Diese starken Varianzen in den Kommunikationsnetzen erfordern Kommunikationssysteme, die in der Lage sind, eine dynamische Anpassung an die verfügbaren Ressourcen vorzunehmen. Eine wichtige Forderung an derartige adaptive Endsysteme ist, dass die Anpassung bedarfsorientiert erfolgt, d.h. an den Anforderungen der Anwendungen und des Benutzers ausgelegt wird. Um Adaptivität der Endsysteme zu erreichen, wurde das im folgenden beschriebene Konzept der geschichteten Rückkopplungsschleifen entwickelt.

2 Konzept

Innerhalb des Endsystems wurden im Verlauf der Arbeit drei Einheiten identifiziert, die in die Adaptivitätsunterstützung einbezogen werden müssen: Netzwerksystem, Betriebssystemkern und Anwendungen. Die Schichten sind dabei durch die Größenordnungen der zeitlichen Änderungen voneinander getrennt. Die vorgeschlagene Architektur entkoppelt damit kurzfristige Änderungen auf der Ebene des Netzwerks von langfristigen Interaktionen auf der Ebene der Anwendungen und des Benutzers. Die Subsysteme bilden in der Architektur teilautonome Einheiten, die in der Lage sind, Schwankungen innerhalb der jeweiligen Größenordnung auszugleichen.

2.1 Netzwerksystem und Betriebssystemkern

Ein Beispiel für Adaptivität innerhalb eines Subsystems ist der im Netzwerksystem realisierte Verkehrsformer [BeRS99],[RiWe00]: Die Aufgabe des Verkehrsformers ist es, die An-

forderungen an typische Qualitätsparameter wie Datenrate oder Jitter (Varianz der Zwischenankunftszeiten von Paketen) zu erfüllen. Dabei werden oftmals statistische Anforderungen gestellt, z.B. dass ein gewisser Jitter zu 95 Prozent eingehalten werden muss. Diese statistischen Vereinbarungen ermöglichen es dem Verkehrsformer, einen kurzfristigen Mangel an Netzwerkressourcen durch Pufferung auszugleichen. Die Adaption an kurzfristige Schwankungen geschieht innerhalb des Netzwerksystems und ist damit für die Anwendungen transparent.

Ab einem gewissen Schwellwert kann das jeweilige Subsystem ein Signal absetzen, mit dem es mitteilt, dass die zur Verfügung stehenden Ressourcen langfristig nicht ausreichen, um die konkurrierenden Anforderungen an die Dienstgüte zu erfüllen. Das Signal realisiert eine Rückkopplung mit dem darüber liegenden Subsystem. Dieses muss nun die Konkurrenz um die Betriebsmittel des darunter liegenden Subsystems auflösen, indem es eine Priorisierung vornimmt und die Anforderungen an das darunter liegende Subsystem insgesamt senkt.

In dem gegebenen Beispiel würde ein solches Signal dann generiert werden, wenn die zur Verfügung stehende Bandbreite über einen längeren Zeitraum hinweg nicht verfügbar ist. In diesem Fall muss der Betriebsmittelverwalter des Betriebssystemkerns die Priorisierung vornehmen, d.h. die Verteilung der Rechenzeit auf die Prozesse wird beeinflusst. Indem eine Anwendung weniger Rechenzeit als die andere erhält, wird die Datenquelle gebremst und die Anforderungen an das Netzwerksystem sinken.

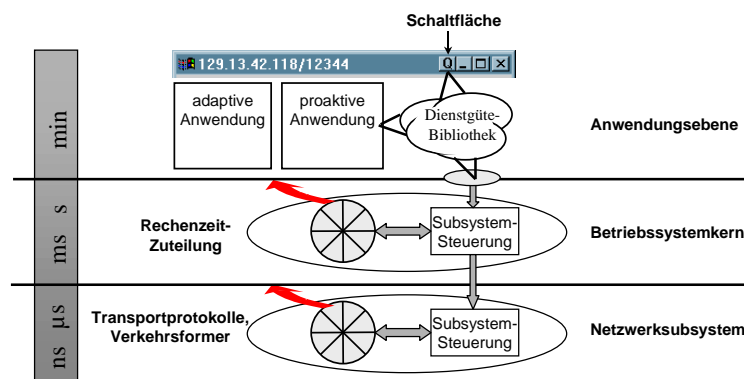


Abbildung 1: Architektur geschichteter Rückkopplungsschleifen

2.2 Betriebssystemkern und Anwendungsebene

Für die Rückkopplung von dem Betriebssystemkern zur Anwendungsebene hin ist kein explizites Signal erforderlich. Die Rückkopplung erfolgt direkt über eine merkliche Veränderung der Leistung der Anwendungen, deren Betriebsmittelzuteilung vom Betriebssystemkern geändert wurde. Bei Anwendungen mit hohem Kommunikationsaufwand erfolgt noch eine zusätzliche Rückkopplung über die Schnittstelle zum Betriebssystem, an der zu sendende Daten übergeben werden. Erhält eine Anwendung weniger Betriebsmittel, werden die Daten an dieser Schnittstelle gepuffert und bei Überlaufen des Puffers der sendende Zugriff blockiert. Um die Rückkopplungsschleife zwischen Betriebssystemkern und Anwendungsebene zu schließen, wurde eine neuartige Schnittstelle zum darunter liegenden Subsystem, dem Betriebssystemkern, geschaffen. Diese ermöglicht es, die Verteilung der Betriebsmittel zu Gunsten einer Anwendung zu beeinflussen. Neben die Klasse der adaptiven Anwendungen, die sich an wechselnde Ressourcen anpassen können, tritt die neue Klasse der sogenannten proaktiven Anwendungen, die zusätzlich in der Lage sind, von den darunter liegenden Subsystemen Ressourcen anzufordern [BeRS00]. Darüber hinaus wurde eine einfache graphische Schnittstelle geschaffen, eine Schaltfläche im Anwendungsfenster, mit der der Nutzer die Ressourcenverteilung

zugunsten einer Anwendung beeinflussen kann. Bei einer Ressourcenknappheit bedeutet das Aktivieren dieser Schaltfläche zwangsläufig, dass einer anderen Anwendung Ressourcen entzogen werden. Der resultierende Zustand entspricht jedoch am besten dem Benutzerwunsch angesichts beschränkter Ressourcen, wie in Messungen mit zwei Videosequenzen gezeigt werden konnte.

3 Anwendung des Konzeptes: Dienstgüteunterstützung im World Wide Web

Wie in [RiPW00] vorgestellt, kann das Konzept der geschichteten Rückkopplungsschleifen auch bei der Dienstgüteunterstützung im World Wide Web (WWW) Anwendung finden. Die Untersuchungen setzen dabei ein dienstgüteunterstützendes Netzwerk voraus, das mit der Einführung der Differentiated-Services-Architektur im Internet zu erwarten ist. Die Umsetzung des Konzepts der bedarfsorientierten Dienstgüteunterstützung im WWW erfordert die Einführung einer zusätzlichen Rückkopplung zwischen dem WWW Klienten und dem WWW Server. Die Hauptaufgabe bei der Dienstgüteeerbringung in dem für WWW-Kommunikation typischen Dienstnehmer-Dienstgeber-Paradigma liegt auf Seiten des Servers als Dienstgeber, der gleichzeitig Anfragen verschiedener Dienstnehmer (Klienten) entgegennehmen und bearbeiten muss. Mit dem unter Linux realisierten Konzept kann der WWW Server Anfragen mit unterschiedlichen Dienstgüteeanforderungen bearbeiten. Der Benutzer auf Seiten des Klienten muss dazu lediglich mit dem Aktivieren der beschriebenen Schaltfläche die Dienstgüte für den angewählten Inhalt auswählen. Auf Seiten des WWW Servers werden nun die vorhandenen Ressourcen entsprechend den Anforderungen aufgeteilt. Die Dienstgüteunterstützung erfolgt dabei sowohl im Netzwerksubsystem durch angepasste Verkehrsformung als auch im Betriebssystemkern durch adaptive Betriebsmittelvergabe.

4 Ergebnisse und weitere Untersuchungen

Starke Schwankungen der verfügbaren Ressourcen in den modernen Netzen, die auf physikalischen Gegebenheiten beruhen und auch in Zukunft nicht grundlegend beseitigt werden können, machen es notwendig, adaptive Kommunikationssysteme zu entwickeln, die in der Lage sind, auch unter stark wechselnden Bedingungen eine den Benutzer zufriedenstellende Leistung zu erbringen. Dazu wurde das Konzept der bedarfsorientierten Dienstgüteunterstützung entwickelt, das auf geschichteten Rückkopplungsschleifen beruht.

Die bisherigen Ergebnisse zeigen, dass dieses Konzept in realen Systemen umgesetzt werden kann und auch für ein wichtiges Anwendungsfeld, die Kommunikation im WWW, Anwendung finden kann. Dennoch sind noch Untersuchungen auf folgenden Feldern offen:

- Um das Konzept exakt evaluieren zu können, könnte sich ein Simulationsmodell als sehr hilfreich erweisen. Insbesondere das Langzeitverhalten ist dabei kritisch, das Auftreten von Schwingungen innerhalb der gekoppelten Rückkopplungsschleifen gilt es zu untersuchen. Gegenwärtig wird die Eignung des Simulationstools OMNeT++ für diese Aufgabe geprüft.
- Bei der Einbettung des Konzepts in eine reale Umgebung wie die Kommunikation im World Wide Web ist die Betrachtung von Faktoren interessant, die außerhalb der rein technischen Sicht liegen. Die Definition von Dienstgüte nach dem Standard ITU E.800

bezieht ausdrücklich die Zufriedenstellung eines Benutzers eines Dienstes in die Definition ein. Einen wesentlichen Faktor stellt dabei die Preisbildung dar, die mit der Einführung von verschiedenen Dienstgütern im Internet unausweichlich sein wird. Hierzu wurden bereits Arbeiten begonnen, in denen ein Rahmenwerk für ein flexibles Preismodell entwickelt wurde. Das vorgestellte Konzept der adaptiven Endsysteme ist damit weitgehend unabhängig von der konkreten Ausgestaltung des jeweils gültigen Preismodells.

- Die Beurteilung des Gesamtsystems muss neben simulativen Untersuchungen auch Erfahrungen aus dem praktischen Einsatz einbeziehen. Hier bieten sich standardisierte Verfahren zur Bewertung durch Testpersonen an, wie sie z.B. das Mean-Opinion-Score-Verfahren nach dem Standard ITU-T P.800 darstellt.

Literatur

- [BeRS99] Marc Bechler, Hartmut Ritter und Jochen H. Schiller. Integration of a Traffic Conditioner for Differentiated Services in End-systems via Feedback-loops. In *Broadband Communications. Convergence of Network Technologies. Proceedings of IF TC6 WG6.2 Fifth International Conference 10.-12. Nov. Hong Kong, Boston, 1999*. Kluwer Academic Publishers.
- [BeRS00] Marc Bechler, Hartmut Ritter und Jochen H. Schiller. Quality of Service in Mobile and Wireless Networks: The Need for Proactive and Adaptive Applications. In *Proceedings of the 33rd Annual IEEE Hawai'i International Conference on System Sciences (HICSS)*, Maui, Hawai'i, January 2000.
- [RiPW00] Hartmut Ritter, Thorsten Pastoors und Klaus Wehrle. DiffServ in the Web: Different Approaches for Enabling better Services in the World Wide Web. In *accepted at Networking 2000*, Paris, May 2000.
- [RiWe00] Hartmut Ritter und Klaus Wehrle. Traffic shaping in ATM and IP networks using standard end systems. In *accepted at ATM 2000, Conference on High Performance Switching and Routing, joint IEEE ATM Workshop 2000 and 3rd International Conference on ATM*, Heidelberg, June 2000.

Π^2 : Unterstützung mobiler und drahtlos angebundener Teilnehmer in verteilten CORBA-Architekturen

Rainer Ruggaber

1 Einleitung

Diese Arbeit hat zum Ziel mehrere unabhängigen Entwicklungen zu vereinen und effizient zu unterstützen:

- kürzere Time-to-Market-Zyklen, die zu einer stärkeren Verzahnung und Integration von Anwendungen, Prozessen und Daten führen
- hohe Anforderungen an die Mobilität von Mitarbeitern, die zu jeder Zeit und an jedem Ort Zugriff auf Unternehmensanwendungen und -daten benötigen
- heterogene drahtlose Zugangsnetze, die im lokalen Bereich sehr leistungsfähig sind, im Weitverkehrsbereich mit GSM jedoch nur eine geringe Leistungsfähigkeit aufweisen

Als wesentliche Technologie zur unternehmensweiten Integration von Prozessen, Anwendungen und Daten kommt in vielen Bereichen die Common Object Request Broker Architecture (CORBA) als Integrationsplattform zum Einsatz. Ein abgekoppeltes Arbeiten auf kopierten Daten und Anwendungen ist in diesem Anwendungsbereich selten geeignet, da Daten und Anwendungen ständigen Änderungen unterworfen sind. Mobilfunknetze, die einen Netzzugang im Weitverkehrsbereich zu realisieren, weisen eine geringe Leistungsfähigkeit (9.6 kBit/s und ca. 300 ms Verzögerung) auf, die sich auch durch neue Technologien wie HSCSD, GPRS oder UMTS nicht wesentlich ändern wird. Drahtlose oder drahtgebundene Netze im lokalen Bereich sind demgegenüber sehr leistungsfähig und stellen keinen zu behandelnden Engpass dar. Die Notwendigkeit über wechselnde Netzwerktypen Zugriff auf Anwendungen und Daten zu haben, ist eine charakteristische Eigenschaft dieser Arbeitsweise.

Für die Unterstützung dieses Szenarios ist eine weitgehende Transparenz bezüglich der genutzten Übertragungsnetze notwendig. Darüberhinaus darf die angestrebte Architektur nicht auf bestimmte Anwendungen zugeschnitten sein, sondern muss beliebige Anwendungen unterstützen, und transparent in diese integriert werden können. Die Lokation des Benutzers soll beim Zugriff auf bestehende Anwendungen weitgehend transparent sein, wohingegen im Falle neuartiger Dienste auch die Möglichkeit bestehen muss die Lokation explizit zugänglich zu machen. Die unterschiedliche Leistungsfähigkeit verschiedener drahtloser Netze kann jedoch nicht vor dem Benutzer verborgen werden. Eine Plattform, die die beschriebenen Anforderungen erfüllt, muss erweiterbar sein und die Möglichkeit zur Integration neuer Dienste bieten. Sie darf auch nicht zu einer zu großen zusätzlichen Verzögerung führen.

2 Konzept

Π^2 ist eine Architektur [RuSe99, KnRS00], die das beschriebene Szenario und die Anforderungen weitgehend erfüllt. Die zentrale Idee in Π^2 ist es, die Verbindung zwischen Klient und Server in mehrere Teilabschnitte aufzuteilen. Die Verbindung der Teilabschnitte wird von Stellvertretern vorgenommen, die durch Tunnels verbunden sind. Alle Aufrufe innerhalb der Architektur werden durch diese Tunnels geleitet.

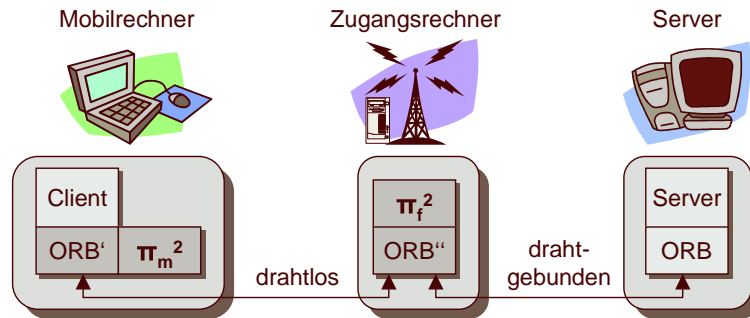


Abbildung 1: Integration von Π^2 in bestehende Anwendungen

Abbildung 1 zeigt die transparente Integration von Π^2 in bestehende Anwendungen. An der Klient-Anwendung und am Server (Server-Anwendung und Server-ORB) sind keine Änderungen notwendig. Der Stellvertreter auf dem Mobilrechner (Π_m^2) ist in den ORB integriert und wird automatisch mit der Klient-Anwendung gestartet. Der Stellvertreter im Festnetz (Π_f^2) ist der Zugangspunkt des Mobilrechners zum Festnetz und ist eine eigenständige Komponente, die vom Systemverwalter gestartet wird, und genau einem Netzwerktyp zugeordnet ist. Weitere Komponenten von Π^2 sind die auf dem Mobilrechner und dem Zugangsrechner installierten ORBs (ORB', ORB''), die modifiziert wurden, um einen Tunnelanfangs- oder -endpunkt zu realisieren.

Die Stellvertreter werden im CORBA-Kontext als Request-Level Halbbrücken bezeichnet. Sie ermöglichen die Nutzung proprietärer Protokolle zwischen den Halbbrücken. Auf den Proxys steht der entfernte Aufruf zur weiteren Bearbeitung in typisierter Form zur Verfügung. Die Typinformation wird auf Π_m^2 aus dem vorhandenen Stub abgeleitet. In Π_f^2 ist dies jedoch nicht möglich, da das anwendungsunabhängige Dynamic Invocation Interface (DII) zum Empfang der Aufrufe genutzt wird. Π_f^2 bezieht die zur Analyse des Aufrufs notwendige Typinformation daher vom Server oder einem Schnittstellenverzeichnis (Interface Repository, IFR).

Der ORB auf dem Mobilrechner übergibt die Aufrufe zunächst an Π_m^2 . Anschließend werden diese, unabhängig von der Ziel-Objektreferenz, über den Tunnel an Π_f^2 weitergeleitet. Π_f^2 empfängt den Aufruf, kann verschiedene Operationen darauf ausführen und leitet den Aufruf an den Server weiter. Das Ergebnis des Aufrufs nimmt den gleichen Weg durch Π^2 wie der Aufruf. Eine unveränderte Ende-zu-Ende-Aufrufsemantik (synchron, asynchron) kann durch den Einsatz derselben Aufrufsemantik auf den Teilstrecken zugesichert werden.

2.1 CORBA-Ebene

Eine standardkonforme CORBA-Implementierung muss das General Inter-ORB Protocol (GIOP) und die Abbildung von GIOP auf TCP/IP, das Internet Inter-ORB Protocol (IIOP) unterstützen. Der Standard bietet einerseits die Möglichkeit GIOP auf andere Transportprotokolle abzubilden, andererseits GIOP durch ein anderes Environment-Specific Inter-ORB Protocol (ESIOP) zu ersetzen.

Ziele bei der Standardisierung von GIOP waren eine Interoperabilität von CORBA-Implementierungen unterschiedlicher Hersteller zu erzielen, sowie gleichzeitig eine einfache Implementierung von GIOP zu erreichen. Die von GIOP zur Übermittlung von Aufrufen genutzte Transportverbindung erzeugt einen aufrufidentifizierenden Kontext. GIOP ist für Anwendungen im drahtlosen Umfeld nicht geeignet, da die Paketköpfe mehrere Byte Daten enthalten, die während einer Verbindung konstant bleiben. Darüberhinaus ist die Übertragungssyntax (CDR) so definiert, dass Objektreferenzen und Methodennamen unkomprimiert übertragen werden, und viele Datentypen nur auf vier-Byte-Grenzen beginnen dürfen, was insbesondere bei kleinen Datentypen zu einem großen Verschnitt führt.

Eine Implementierung, die einige der beschriebenen Nachteile von GIOP behebt, wurde durchgeführt. In dieser werden Elemente des Paketkopfes, die konstant bleiben, nur zu Beginn der Verbindung übertragen. Darüberhinaus wurde eine Paketgrößegröße (Message Size Size, MSS) eingeführt, die die Anzahl Bytes enthält, die für die Darstellung der Paketgröße im Paketkopf notwendig ist. Dies führt zu einer Reduzierung der GIOP-Paketkopf-Größe von 12 Byte auf zwei bis fünf Byte, abhängig von der Größe des Pakets. Ein Aufruf mit einer Größe von 64 kByte benötigt beispielsweise nur einen drei Byte großen GIOP-Paketkopf, was auf einer GSM-Verbindung zu einer Reduzierung der Verzögerung von 9,4 ms führt. Darüberhinaus werden Objektreferenzen und Methodennamen nur einmal pro Verbindung übertragen. Nachfolgende Übertragungen enthalten statt dessen einen Index, der einen Eintrag in inkrementell aufgebauten Tabellen in Π_f^2 und Π_m^2 referenziert.

2.2 Transport-Ebene

TCP/IP als Standard-Transportprotokoll zur Kommunikation in CORBA ist für den Einsatz in Mobilfunknetzen nicht geeignet. Der Verbindungsaufbau in TCP erfolgt über einen 3-Way-Handshake, der in Mobilfunknetzen, die eine große Verzögerung aufweisen, zu langen Verbindungsaufbauzeiten führt. TCP/IP überträgt im Paketkopf Daten, die während einer Verbindung konstant bleiben, und deshalb unnötig Bandbreite auf einem schmalen Link benötigen. Es existieren zwei grundsätzliche Möglichkeiten dies zu lösen. Zum einen könnte TCP/IP an den Einsatz angepasst werden. Zum anderen kann ein anderes Protokoll genutzt werden, das speziell für diesen Anwendungsfall entwickelt wurde.

Anpassungen von TCP/IP an drahtlose Links wurden beispielsweise von Stephen Pink durchgeführt. In den dort eingesetzten Verfahren werden beim Verbindungsaufbau, die im Paketkopf gleichbleibenden Daten einmal ausgetauscht, und bei nachfolgenden Paketen eingespart. Der Standard TCP/IP-Paketkopf kann somit auf vier Byte reduziert werden.

Das Wireless Application Protokoll (WAP) ist ein Protokoll, das speziell für die Kommunikation in Mobilfunknetzen entwickelt wurde. Wesentliche Vorzüge von WAP gegenüber TCP/IP ist die direkte Unterstützung eines für die Transaktionsverarbeitung ausgelegten Dienstes (WTP), der die Möglichkeit bietet schon im ersten Paket Daten zu senden, die Integration einer Sitzungsschicht (WSP), um unabhängig von Verbindungsabbrüchen unterliegender Schichten zu sein [DoRu99], und kleine Paketköpfe, um die geringe Bandbreite effizient auszunutzen.

2.3 Verbindungsübergabe

Die Mechanismen der Verbindungsübergabe werden benötigt, falls sich die Zuordnung von Π_m^2 und Π_f^2 ändert. Dies tritt ein, falls der Klient ausstehende Ergebnisse wartet und den Empfangsbereich des zugeordneten Netzes verlässt, sich über ein neues Netz wieder anbindet, in den Schlafmodus geht oder abgeschaltet wird. In diesen Fällen muss eine Verbindungsübergabe (Handover) bestehender Verbindungen vorgenommen werden. Verbindungsübergaben

innerhalb von Netzen sind transparent für die Anwendung und müssen nicht berücksichtigt werden. Verbindungsübergaben treten relativ selten auf, sind jedoch notwendig, um den Wechsel für den Benutzer transparent zu gestalten.

Die Initiative zum Abholen ausstehender Ergebnisse liegt bei Π_m^2 . Dieser kann ausstehende Ergebnisse zu einem späteren Zeitpunkt direkt über den ursprünglichen oder indirekt über einen anderen Π_f^2 beim ursprünglichen Π_f^2 abgeholt. Um Π_f^2 jedoch nicht mit der Speicherung vieler ausstehender Ergebnisse zu belasten, wurde eine neue Komponente in die Architektur eingeführt, die längere Zeit nicht abgeholte Ergebnisse zwischenspeichert. Fordert Π_m^2 das Ergebnis später an, wird der Aufruf transparent an diese Komponente weitergeleitet und von ihr beantwortet.

2.4 Filter

Filter sind Java-Objekte, die in den Proxy integriert werden können, um erweiterte Dienste anzubieten. Filter können sowohl statisch beim Start des Proxys oder zur Laufzeit integriert werden. Die Filter haben lesenden und schreibenden Zugriff auf alle Informationen, die auf dem Proxy vorliegen, insbesondere Aufrufe, die den Proxy passieren. Alle im Proxy registrierten Filter werden auf ankommende Aufrufe und Ergebnisse angewendet. Filter können beliebige Berechnungen durchführen, untereinander kommunizieren sowie einen Zustand über mehrere Aufrufe hinweg sichern. Ein Filter kann auch andere Filter nutzen, um seine Dienste zu realisieren.

Ein Anwendungsbereich sind beispielsweise Fehlertoleranzverfahren. In diesen Verfahren werden mehrere gleiche Aufrufe zu verschiedenen Servern geschickt. Die Antworten werden entsprechend gesammelt und verglichen. Auf Basis einer Mehrheitsentscheidung wird ein Ergebnis ausgewählt und an den Klient zurückgeliefert. Der Fehlertoleranzfilter kann den Weiterleitungsfiler des Proxys nutzen, um zusätzliche, selbst erzeugte Aufrufe an andere Server zu schicken. Sobald die Ergebnisse eintreffen muss eine Auswahl des richtigen Ergebnisses getroffen werden. Mit Hilfe der Filter ist es möglich diese Funktionalität transparent für den Klient in die Anwendung zu integrieren.

3 Implementierung und Messungen

Die Implementierung wurde mit Hilfe der CORBA-Implementierung ORBacus für Java durchgeführt, die für Forschung und Lehre ohne Lizenzgebühren inklusive Quell-Code von der Firma Object Oriented Concepts (OOC) frei verfügbar ist.

Die Leistungsfähigkeit von Π^2 wurde mit Messungen überprüft. Der Versuchsaufbau entspricht dem in Abbildung 1 gezeigten Szenario. Gemessen wurde die zusätzliche Verzögerung, die durch Π^2 erzeugt wird. Es wurden Messungen mit und ohne Π^2 durchgeführt. Um den Einfluss verschiedener Netzwerktypen zu untersuchen, wurden auf der Strecke zwischen dem Mobilrechner und dem Festnetzproxy verschiedene Netzwerktypen getestet. Auf allen Teilstrecken wurde TCP/IP und unveränderte CORBA-Protokolle eingesetzt. Tabelle 1 enthält die Messwerte des Experiments.

Die Messungen zeigen, dass durch Π^2 eine zusätzliche Verzögerung von 12-13 ms erzeugt wird. Der bei GSM gemessene Wert entsteht aus den zusätzlichen Daten (160 Byte), die für das Tunneln der Aufrufe zu jedem Aufruf hinzugefügt werden. Aufgrund der vergleichsweise geringe Bandbreite von GSM führen diese Daten direkt zu einer größeren Verzögerung (zusätzlich 13,3 ms). Jedoch ist die gesamte Verzögerung für weniger als 5% der Gesamtverzögerung verantwortlich.

Netzwerktyp	mit Π^2	ohne Π^2	Differenz
10 Mbit/s Ehternet	17,6 ms	5,5 ms	12,1 ms
10 Mbit/s WaveLan	19,6 ms	6,7 ms	12,9 ms
2 Mbit/s WaveLan	20,0 ms	7,7 ms	12,3 ms
9,6 kbit/s GSM	653,0 ms	622,0 ms	31,0 ms

Tabelle 1: Messung der Π^2 -Verzögerung

4 Ausblick

Im Rahmen eines anderen Projektes wurden Erfahrungen mit dem transparenten Zwischenspeichern von Objekten auf Klientseite gemacht. Die dort gemachten Messungen, obwohl im lokalen Netzwerk durchgeführt, zeigen eine mögliche Beschleunigung entfernter Aufrufe um etwa den Faktor 60. Der Einsatz der Zwischenspeicherung im beschriebenen Szenario verspricht größere Beschleunigungen, aber auch größere Probleme bei der Konsistenz der zwischengespeicherten Objekte, falls die Verbindung abbricht. Transaktionen aus dem Datenbankkontext sind das geeignete Mittel, um die Konsistenz sicherzustellen, und sollen deshalb bei der Integration der Zwischenspeicherung berücksichtigt werden. Die Nutzung verteilter Anwendungen ohne transaktionalen Kontext bietet auch bei leitungsgebundenen verteilten Systemen nicht genügend Garantien und kann deshalb vernachlässigt werden.

Erste Arbeiten zur Integration von WAP in die CORBA-Architektur wurden bereits gemacht [RuSS99] sollten jedoch unter Berücksichtigung jetzt verfügbarer Implementierungen und neuer Standards überarbeitet und in Π^2 integriert werden.

Literatur

- [DoRu99] Elmar Dorner und Rainer Ruggaber. A mobility aware CORBA Event Service for wireless end systems. In *10th IEEE Workshop on Local and Metropolitan Area Networks (LANMAN'99)*, Sydney, Australia, November 1999.
- [KnRS00] Michael Knapp, Rainer Ruggaber und Jochen Seitz. Π^2 : a Generic Proxy Platform for Wireless Access and Mobility in CORBA. In *19th Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (PODC)*, Portland, USA, Juli 2000.
- [RuSe99] Rainer Ruggaber und Jochen Seitz. A Generic Proxy Platform for CORBA Applications. In *Future Services for Networked Devices (FuSeNetD'99)*, Eurescom, Heidelberg, November 1999.
- [RuSS99] Rainer Ruggaber, Jochen Schiller und Jochen Seitz. Using WAP as the enabling technology for CORBA in mobile and wireless environments. In *7th IEEE Workshop on Future Trends of Distributed Computing Systems (FTDCS'99)*, Kapstadt, Süd Afrika, Dezember 1999.

Dienstgüte in mobilen, drahtlosen Kommunikationssystemen

Jochen Schiller

Dienstgüte, definiert als „die vollständige Wirkung der Leistungsfähigkeit eines Dienstes, die den Grad der Zufriedenstellung eines Benutzers desselben Dienstes bestimmt“, ruht aus Sicht der Datenkommunikation auf den drei Säulen Allgegenwart, Robustheit und Skalierbarkeit. Die *Allgegenwart* eines Kommunikationssystems beschreibt die Durchdringung aller Bereiche des täglichen Lebens, geographischer Regionen und Anwendungsgebiete. Unter den Begriff der *Robustheit* fallen Bereiche der Datensicherheit, Zuverlässigkeit der Diensterbringung und auch der Sicherheit gegenüber Personen. Die *Skalierbarkeit* schließlich ist die Eigenschaft eines Systems, dass es in beliebiger Dimensionierung hinsichtlich Ausdehnung, Teilnehmerzahlen oder beteiligten Komponenten verwirklicht werden kann, ohne seine wesentlichen Eigenschaften hinsichtlich der Diensterbringung zu verlieren.

Der Wunsch nach allgegenwärtigen Kommunikationsdiensten ist die Hauptmotivation mobiler, drahtloser Kommunikationssysteme, die sich heute schon in vielfältigen Ausprägungen weltweit finden (z.B. GSM, Satelliten, drahtlose lokale Netze, UMTS, Bluetooth). Die Attraktivität dieser Netze führte zu stärkeren Wachstumsraten als die des Internets, was sich bereits heute auch in der Zahl von über 350 Millionen Nutzern ausdrückt. Deutlich ist auch die Entwicklung zu erkennen, dass Techniken des Internets immer mehr Einzug in alle Bereiche mobiler Kommunikation nehmen, das Internet generell als die gemeinsame Basis heterogener drahtloser, mobiler Netze gesehen wird. Sehr große Teilnehmerzahlen, hohe Wachstumsraten und die Integration von Internet-Techniken, die ursprünglich nicht für ein drahtloses, mobiles Umfeld konzipiert wurden, stellen hohe Anforderungen an die Robustheit und Skalierbarkeit der Kommunikationssysteme und werden international als die Herausforderungen für Forschungsarbeiten im Telematik-Bereich der kommenden Jahre gesehen. Vielfältige *Anwendungen* für drahtlose, mobile Kommunikationssysteme finden sich beispielsweise auf den Gebieten der Touristen-Informationssysteme, des Internet-Zugriffs aus Fahrzeugen heraus, der Fernwartungs- und Ferndiagnosesysteme für Fahrzeuge, Flugzeuge oder Schiffe, der Unterhaltungsindustrie oder des „M-Commerce“, also der Erweiterung von „E-Commerce“ auf mobile Teilnehmer (z.B. mobile Bankanwendungen).

Die aktuellen Forschungsarbeiten beschäftigen sich mit der Frage, wie die Säulen der Dienstgüte, Robustheit, Skalierbarkeit und Allgegenwart, unterstützt werden können. Als notwendige Schlüsselfaktoren wurden hierbei die Anpassungsfähigkeit mobiler Systeme, die Integration von schichtenübergreifendem Wissen und die Integration von Sicherheitsfunktionen erkannt. Weitere Schlüsselfaktoren, welche zusätzlich zu untersuchen wären, sind beispielsweise die Hierarchiebildung und Mechanismen zur Fehlertoleranz.

1 Anpassungsfähigkeit

Festnetzrechner erfahren relativ geringe Schwankungen in der Dienstgüte auf der Netzverbindung zum nächsten Vermittlungsknoten (Switch, Router etc.). Die Bandbreiten liegen z.B. bei 100 Mbit/s, die Verzögerung bei wenigen Millisekunden. Mobile, drahtlose Kommunikationssysteme sind jedoch geprägt von einer hohen Heterogenität und stark schwankenden

Kommunikationsparametern. Die Bandbreiten variieren beispielsweise zwischen 2,4 kbit/s (Satellit) und 11 Mbit/s (drahtloses LAN). Verzögerungen können Werte bis zu einigen Sekunden erreichen, ebenso bringt die Nutzung unterschiedlicher Netze oft grundlegend unterschiedliche Kosten mit sich. Mobile Systeme müssen sich hinsichtlich ihrer Anwendungen, Betriebssysteme und Netzwerksysteme an stark schwankende Übertragungsbedingungen und Anwendungsanforderungen anpassen können.

Die Arbeiten zeigen, dass hierfür vielfältige *Rückkopplungsmechanismen* eingesetzt werden können. Dazu wurde zunächst ein Endsystem in drei Teile aufgeteilt: Netzwerksystem, Betriebssystem und Anwendungen. Alle Prozesse bekommen von einer zentralen Einheit im Betriebssystem (Scheduler) Betriebsmittel (im Wesentlichen Rechenzeit) zugeteilt. Genügt einer Anwendung diese Rechenzeit nicht, so kann sie sich über eine neue Schnittstelle zum Betriebssystem hin „beschweren“, also mehr Betriebsmittel anfordern. Ebenso kann ein Nutzer die mangelnde Qualität einer Anwendung feststellen und seinen Unmut über eine einfache Schnittstelle, einen Qualitäts-Knopf, gegenüber dem Betriebssystem äußern. Alle Wünsche nach mehr Betriebsmittel gehen bei einer zentralen Steuerung ein, welche nun zu entscheiden hat, wie die vorhandenen Betriebsmittel zu verteilen sind. Bei Ressourcenknappheit muss, je nach Präferenzen, einer Anwendung unter Umständen auch ein Teil ihrer Betriebsmittel entzogen werden, welcher dann einer anderen Anwendung zu Gute kommt. Diese Rückkopplung zwischen *Betriebssystem* und *Anwendungen* kann anhand von beispielsweise Videosequenzen gezeigt werden. Auf einem kleinen, mobilen Endgerät (z.B. eine Anzeige in einem Fahrzeug, auf einem PDA) werden zwei Videosequenzen gezeigt. Für den Nutzer sei in diesem Beispiel eine Sequenz wichtiger als die andere - das Kommunikationssystem kann dies natürlich nicht erkennen. Daher „klickt“ der Nutzer auf den Qualitäts-Knopf, das Betriebssystem teilt der entsprechenden Anwendung mehr Ressourcen zu und der Nutzer kann eine höhere Qualität sehen. Wesentlich sind hier die einfache Nutzerinteraktion und eine einfache Rückkopplungsarchitektur, welche unabhängig von der Plattform bzw. Rechenleistung ist.

Ebenso wurde eine Rückkopplung zwischen *Betriebssystem* und *Netzwerksystem* implementiert. Ist eine Anwendung nicht mit der Leistung zufrieden, kann also beispielsweise nicht genügend Bilder/s senden, so kann dies auch an der Kapazität der aktuellen Funkanbindung liegen. Stellt das Betriebssystem fest, dass es nicht selbst den Engpass darstellt, kann eine Anforderung auf Netzwechsel an das Netzwerksystem gegeben werden. Im Zusammenhang mit dem Wireless Datagram Protocol (WDP) wurde hier eine Architektur entworfen, welche ein Umschalten zwischen verschiedenen Adapters (z.B. GSM, IEEE 802.11, IS-95) erlaubt. Wesentlich ist hierbei eine noch zu verfeinernde Einheit zur Vorhersage und Entscheidung über den Umschaltzeitpunkt. Je nach Art des Systemwechsels kann ein relativ großer Aufwand mit einer Umschaltung verbunden sein, so beispielsweise für eine Adressumsetzung, eine Umleitung der Daten oder einen Verbindungsaufbau. Hier kann es besser sein abzuwarten, bis das aktuelle Netz wieder eine bessere Qualität anbieten kann.

Die *Allgegenwart* eines Kommunikationssystems wird mit diesem Ansatz durch den automatischen Schnittstellenwechsel unterstützt, wobei das Ziel ein möglichst „sanfter“ Wechsel durch gute Vorhersagen ist. Die *Robustheit* des Systems wird durch die Anpassung an Anwendungs- und Nutzeranforderungen, aber auch durch die Anpassung an Netzwerkparameter gesteigert. Zentrale Bedeutung besitzen bei diesen Arbeiten Rückkopplungsmechanismen.

2 Aufhebung der strikten Schichtentrennung

Die Zerlegung eines Kommunikationssystems in einzelne Schichten (z.B. Sicherungsschicht, Vermittlungsschicht) hat wesentlich zum Erfolg von vielen Kommunikationssystemen beigetragen. Durch die Errichtung von Schichten konnte eine gewisse Unabhängigkeit zwischen

diversen Produkten, eine Vielfalt an Netzen und dennoch eine Interoperabilität erreicht werden. Die Schichtenarchitektur kann auch weitgehend in den Implementierungen nachvollzogen werden. Während im Festnetzbereich die Trennung weitgehend Sinn macht, gibt es gerade im Bereich der drahtlosen und mobilen Kommunikation sehr gute Gründe, das Modell zu durchbrechen. Wissen, welches auf einer Kommunikationsschicht vorhanden ist, kann sehr hilfreich für eine andere Schicht sein. Beispielsweise bricht die Leistungsfähigkeit des Transportprotokolls TCP bei einer Verbindungsübergabe zwischen zwei Zugangspunkten zum Festnetz dramatisch ein. Hier kann ein „Einfrieren“ des TCP-Kontextes helfen, indem nämlich die Sicherungsschicht der Transportschicht signalisiert, dass sie ab sofort alle Parameter festhalten und auf ein „Aufweck“-Signal der Sicherungsschicht warten soll. Dies vermeidet, dass TCP etwas von Paketverlusten bei der Verbindungsübergabe mitbekommt, daraufhin von einem Stau im Netz ausgeht und die Senderate drosselt, um den vermeintlichen Stau aufzulösen. Hier kann also Schicht 2 helfen, die Datenrate auf Schicht 4 möglichst hoch zu halten. Ebenso können Informationen aus Schicht 1 bezüglich Interferenzen zur Wegewahl auf Schicht 3 herangezogen werden, da gerade bei mobilen Ad-hoc-Netzen nicht immer der kürzeste auch der beste Weg ist. Ein besseres Kriterium ist hier der Weg mit den geringsten Störungen.

Die beiden Beispiele bereits existierender Arbeiten zeigen lediglich die Signalisierung von gewissen Informationen von einer Schicht zu einer anderen. Neu entwickelt wurde ein Mechanismus, welcher alternativ auf verschiedenen Schichten implementiert werden kann. Traditionellerweise findet sich die Fehlerbehebung auf der Sicherungsschicht. Bei UMTS (Universal Mobile Telecommunications System), welches teilweise auf einer CDMA (Code Division Multiple Access)-Technologie beruht, ist neu, dass die Interferenzen innerhalb einer Funkzelle und die Größe der Funkzelle von der Sendeleistung aller Geräte innerhalb der Zelle abhängen - und umgekehrt. Zur Demonstration der schichtenübergreifenden Implementierung der Fehlerkorrektur in UMTS wurde eine Konfiguration mit einer Videoverbindung und vielen Sprachverbindungen simuliert.

Die klassischen *Maßnahmen gegen Paketverluste* finden sich auf Schicht 2: Vorwärtsfehlerkorrektur (FEC) und Übertragungswiederholung (ARQ). FEC ist bei Funksystemen praktisch immer vorhanden. ARQ kann je nach Anwendung nur eingeschränkt eingesetzt werden, da hierdurch die Verzögerung nichtdeterministisch erhöht wird. Eine neuartige Gegenmaßnahme auf Schicht 1, welche bei CDMA-Technologien möglich ist, ist die Erhöhung der Sendeleistung zur Beibehaltung einer gewissen Fehlerrate bei erhöhten Störungen oder zur Senkung der Fehlerraten bei gleichbleibenden Störungen. Das Problem besteht nun darin, dass FEC und ARQ die Last in einer Zelle erhöhen (in bit/s/km^2) und eine erhöhte Sendeleistung andere Stationen stört. Die Frage, die untersucht wurde, war nun, wie die Höhe der Paketverlustrate von Videodaten in Abhängigkeit der Verkehrslast in einer UMTS-Funkzelle unter Einsatz verschiedener Fehlerkorrekturverfahren aussieht. Dazu waren in einem Simulationsmodell ein UMTS-CDMA-Kanalmodell und H.263-Quellenmodelle gegeben. Die Datenrate der Videoquelle betrug im Durchschnitt 80 kbit/s bei Paketgrößen von 10-1200 byte. Die Fehlerrate auf dem Kanal wurde auf 0,4 % eingestellt (bezogen auf Übertragungsblöcke). Als Lastquellen dienten zwischen 5 und 25 Mobilsysteme mit 16 kbit/s Sprachdaten.

Es konnte gezeigt werden (siehe Abbildung 1), dass in dieser Konfiguration die H.263-Paketverlustrate bei kombiniertem Einsatz von Schicht-1- und Schicht-2-Mechanismen (Erhöhung der Sendeleistung plus ARQ) bis zu einer Last von 17 Stationen niedriger ist, als bei alleinigem Einsatz von Schicht-1-Mechanismen. Bei mehr als 17 Stationen stören die notwendigen Übertragungswiederholungen der Schicht 2 so stark, dass die Nutzung von ausschließlich Schicht-1-Mechanismen besser ist. Die ausschließliche Nutzung von Übertragungswiederholung ist in jedem Fall schlechter als die Kombination der Mechanismen. Wichtig bei dieser Arbeit sind nicht die exakte Anzahl an Stationen, ab denen umgeschaltet werden muss, oder die genauen Verlustrate. Wichtig war zu zeigen, dass die *Möglichkeit der automatischen Wahl* zwischen verschiedenen Mechanismen, welche alle das gleiche Ziel haben, insgesamt *gewinn-*

bringend ist. Durch die Möglichkeit des Umschaltens kann die Kapazität einer Funkzelle erhöht oder die Fehlerrate verringert werden. Damit trägt die Integration von Fehlerkorrekturmechanismen wesentlich zur *Robustheit* eines Systems und zur *Skalierbarkeit* hinsichtlich der Teilnehmerzahlen bei.

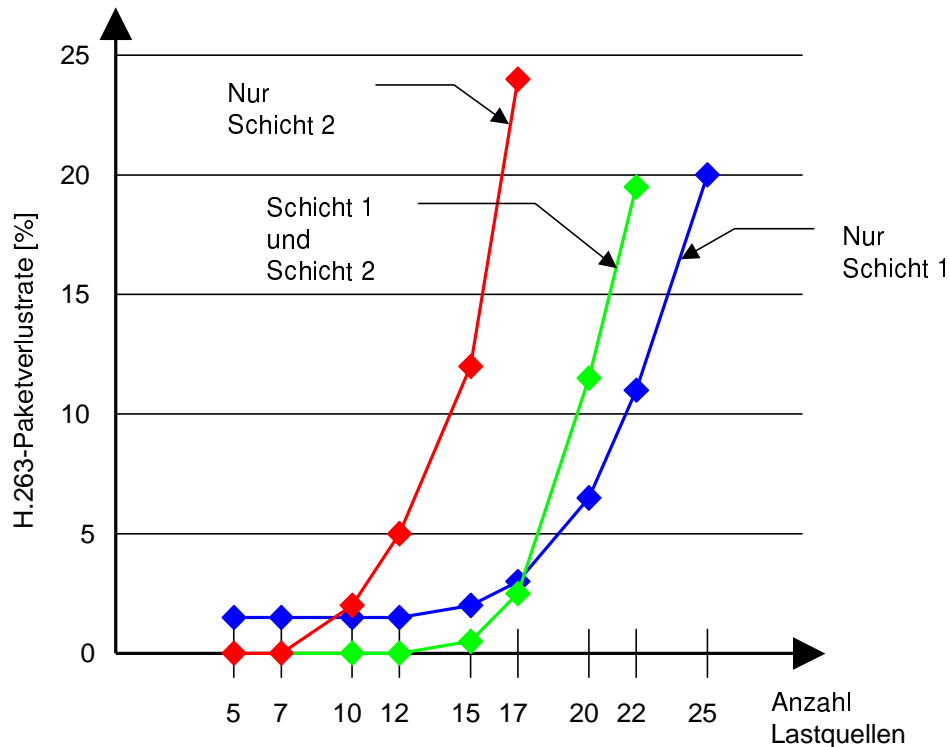


Abbildung 1: Paketverlustraten in Abhängigkeit von Korrekturmechanismen

3 Integration von Sicherheitsmechanismen

Für den Einsatz von Mobile IP in realen Netzen sind Sicherheitsmechanismen unabdingbar, da Mobilität im Internet ohne Sicherheit kaum akzeptiert wird. Die im Standard vorgesehenen Funktionen sind vollkommen unzureichend, in der aktuellen Version bietet Mobile IP sehr viele Möglichkeiten für einen Angriff an. Beispiele sind Attacken auf Fremd- und Heimatagenten, welche an der Registrierung von mobilen Teilnehmern beteiligt sind. Weiterhin können Bewegungsprofile erstellt und diverse Komponenten einem Nutzer vorgegaukelt werden, welcher sich dann fälschlicherweise an diese Komponenten an Stelle der echten Komponenten wendet. Manche Vorschläge sehen verteilte Sicherheitsmechanismen in allen Fremdagenten vor, dies ist jedoch ungeeignet, da Sicherheit zentral in Kombination mit Paketfiltern, Mechanismen zur Adressumsetzung oder auch Virenfiltern implementiert werden muss. Weiterhin bereitet die Paketweiterleitung in Mobile IP große Probleme in realen Netzen, da ein Mobilsystem in einem fremden Netz mit topologisch falscher Adresse sendet. Pakete dieser Art können Firewalls im Allgemeinen nicht passieren. Die hierzu im Internet vorgeschlagene Lösung, Tunnels durch eine Firewall zu legen, kann aus Sicherheitsgründen ebenso nicht zum Einsatz kommen. Schließlich ist Mobile IP hochgradig ineffizient, da bei jedem Wechsel eines Netzes eine erneute Authentifizierung der Komponenten durchgeführt werden muss. Wechselt also beispielsweise ein Besucher aus Japan in Deutschland von einem Netz in ein anderes (Mikromobilität), da er sich von einem Zimmer in ein anderes bewegt, so muss eine erneute Authentifizierung mit Komponenten im Heimatnetz (dem Heimatagenten in Japan) über das Internet hinweg durchgeführt werden.

Die Untersuchung zahlreicher Vorschläge zur Verbesserung dieser Situation brachte hervor, dass Vorschläge entweder auf eine effiziente Schlüsselverwaltung/Authentifizierung, eine sichere Durchquerung von Firewalls oder die Unterstützung von Mikromobilität abzielen. Kein Ansatz bietet Sicherheit, Effizienz und Transparenz gegenüber existierenden Anwendungen und Komponenten.

Die Entwicklung einer *transparenten, sicheren Mobilitätsunterstützung in Firewalls* bietet diverse Vorteile: *Transparenz* (es sind keine Änderungen in Endsystemen notwendig), *Zentralisierung der Sicherheitsmechanismen* (kleine Angriffsfläche, alle Überprüfungen in der Firewall), *Unterstützung privater Adressen* (Adressumsetzung ist in der Firewall möglich), *Unterstützung von Internet-Dienstgütearchitekturen* (längerfristige Beziehung zwischen dem Heimat- und Fremdnetz erhöht die Effizienz von z.B. Differentiated Services) und *Unterstützung von Mikromobilität* (keine erneute Authentifizierung bei lokaler Mobilität, wesentliche Optimierung der Datenpfade möglich). Durch diese Eigenschaften der Architektur kann sowohl die *Robustheit* eines Kommunikationssystems als auch dessen *Skalierbarkeit* wesentlich gesteigert werden.

4 Zusammenfassung

Insgesamt zeigen die Arbeiten deutlich auf, dass im Bereich der Mobilkommunikation der Begriff der *Dienstgüte* sehr weit gefasst werden muss und deutlich über die Angabe von einzelnen Parametern, wie z.B. Übertragungsraten oder Verzögerungszeiten, hinausgeht. Die *Robustheit* eines Kommunikationssystems kann wesentlich durch adaptive Systeme mit Rückkopplungsschleifen und der Möglichkeit von Schnittstellenwechseln, durch die Senkung von Fehlerraten auf Grund von Schichtenintegration und durch die Integration von Sicherheit in Mobile IP gesteigert werden. Die *Skalierbarkeit* wird in Mobile IP durch eine Hierarchiebildung und die Unterstützung von Mikromobilität unterstützt. Schließlich erbringen vielfältige Techniken der Mobilkommunikation eine *Allgegenwart* von Kommunikationsdiensten, welche maßgeblich durch einen möglichst „sanften“ Wechsel zwischen den Systemen unterstützt wird.

Die *Herausforderungen* für die Zukunft bestehen in der Ende-zu-Ende-Unterstützung vielfältiger Dienstgüteaspekte. Hierzu ist es notwendig, geeignete Modelle und Parameter für die Simulation zu finden (Datenquellen, Endsysteme, Mobilität), welche insbesondere schichtenübergreifend angelegt sein müssen. Die Beherrschbarkeit der Komplexität ist von besonderem Interesse, da gerade im Bereich der Mobilkommunikation eine sehr große Anzahl autonomer Einheiten zu erwarten ist (Verwaltung der Mechanismen, Stabilität des Gesamtsystems, Hierarchiebildung). Schließlich müssen angepasste Protokollmechanismen entworfen werden, welche effizient, energiesparend, adaptiv und schichtenübergreifend ausgelegt sind.

Implicit Human Computer Interaction

Albrecht Schmidt

1 Introduction

The way people interact with devices is vital for their success. Looking at HCI it is apparent that interaction techniques are limited by the technology available. Furthermore the anticipated user groups influence the interaction metaphors to a large extent. Considering the shift from punch cards to interactive text terminals and also the shift from command line interfaces to graphical user interfaces (GUI) this was observable.

Bearing in mind current and upcoming technologies, such as increased processing power (even on mobile devices), availability of sensors (ranging from simple temperature sensors to cameras), and the resulting perceptual capabilities as well as the fact that the main user group of current computing devices (e.g. mobile phones, PDAs, etc.) are non experts, we may observe yet another shift in HCI. Devices that have perceptual capabilities (even if they are very limited) will start the shift from explicit HCI towards a more implicit interaction with machines.

A vision of future devices

We will be able to create (mobile) devices that can see, hear and feel. Based on their perception, these devices will be able to act and react according to the situational context in which they are used.

In this paper it will be shown that this vision is not as far ahead as it seems. In our research we start with the perception of simple concepts and their exploitation. It is discussed how basic perception can enable a shift from explicit towards implicit HCI.

2 Implicit Interaction

Observing communication between humans we can see that a lot of information is only exchanged implicitly. The way people interact with each other and also the situation in which they interact carries information that is often implicitly exploited in the exchange of messages. While in a conversation the behavior of participants as well as what happens in the surrounding environment supplies valuable information that is often vital for the understanding of messages. In many cases the robustness of human- to-human communication is based on the implicitly introduced contextual information, such as gestures, body language, and voice. Another example is redundancy between body language (e.g. nodding) and spoken languages (e.g. the word 'yes'). This implicitly introduced knowledge is also used to disambiguate information, e.g. in a discussion with a student pointing at a computer the term 'sun' has a different meaning than the same term when on the beach together with friends.

2.1 Implicit vs. Explicit Human Computer Interaction

Considering current computer technology interaction is explicit – the user tells the computer in a certain level of abstraction (e.g. by command-line, direct manipulation using a GUI, gesture, or speech input) what she expects the computer to do. This is considered as explicit interaction.

Definition: Implicit Human Computer Interaction

Implicit human computer interaction is an action, performed by the user that is not primarily aimed to interact with a computerized system but which such a system understands as input.

The action of a user is always performed in a certain environment. Implicit interaction is based on the assumption that the computer has a certain understanding of our behavior in the given situation. This knowledge is then considered as an additional input to the computer while doing a task. A simple example is the garbage bin [Corp99] that scans in the bar code of products and reproduces the information for a suggested shopping list. The action performed by the user (e.g. throw away an empty can in a bin) is the same as with any other garbage bin. The recognition of the system (by scanning the bar code) and the built-in interpretation of the system (all things that go into the bin may be on the next shopping list again) make use of the action performed by the user. The user herself does not explicitly interact with the computer, thus the process describes an implicit interaction. As we see from the example implicit interaction is based on two main concepts: perception and interpretation.

For most applications implicit interaction will be used additionally to explicit interaction. There are other systems implemented that also facilitate the idea of implicit interaction on a rudimentary level, e.g. automatic light control (switches on the light when it is dark and someone is walking by) and active badge systems (automatically open a door when someone with appropriate permission likes to enter the building). In current computer systems we can observe that agent technology is used to build systems that have a certain ability to act proactively. These approaches are mainly based on user profiles and usage information. In these cases perception is limited to information gathered in the virtual space. If we look concepts that are needed to facilitate implicit interaction three basic building blocks can be identified:

1. the ability to have perception of the use, the environment, and the circumstances,
2. mechanisms to understand what the sensors see, hear and feel, and
3. applications that can make use of this information.

On a conceptual level (1) and (2) can be described as situational context. And (3) are applications that are context enabled. In the next section context is discussed in more detail.

2.2 What is Context

The notion of context is used in many different ways. In our work we propose to regard situational context, such as location, surrounding environment or state of the device, as implicit input to the system. We use the term situational context to describe implicit interaction fragments. This extends the concept of context beyond the informational context into real world environments. The word Context in general use has a multitude of meanings. Even within the field of computer science different disciplines, such as artificial intelligence, natural language processing, image recognition, and more recently mobile computing, have their very own understanding of what context is. In our work we found that very general descriptions of context as given by a dictionary and also synonyms found in a thesaurus come very close to our understanding. To illustrate this we like to provide the following definitions:

Context n 1: discourse that surrounds a language unit and helps to determine its interpretation [syn: linguistic context, context of use] 2: the set of facts or circumstances that surround a situation or event; „the historical context“ (Source: WordNet)

Context: That which surrounds, and gives meaning to, something else. (Source: The Free On-line Dictionary of Computing)

Synonyms Context: Circumstance, situation, phase, position, posture, attitude, place, point; terms; regime; footing, standing, status, occasion, surroundings, environment, location, dependence. (Source: www.thesaurus.com)

To build applications that have knowledge about their situational context it is important to gain an understanding what context is. Current research in context-awareness in mobile computing shows a strong focus on location [BrBC97]. Location is a concept that is well understood. Also the benefit of location- awareness is clearly given, at certain locations particular services are more important than others. An architectural approach, based on a smart environment is described by Schilit et. al. [ScAW94]. Other scenarios are using RF and GPS to determine the users location, e.g. [CBDF99]. But, as pointed out in [ScBG99] context is more than location. We use the term context considering mobile computing in a more general way to describe the environment, situation, state, surroundings, task, and so on. A wider view of context is also given by [SATT⁺99]. They suggest to consider the way a device is used (mobile phone in the users hand, on the table, in pocket, etc.) to be treated as context.

2.3 Applications in Context

Analyzing the way people use ultra-mobile devices (e.g. personal digital assistants, smart mobile phones, handheld and wearable computers) it becomes apparent that the periods of interaction are much shorter than in traditional mobile settings. Notebooks – considered as mobile computers - are mainly used in stationary setting, e.g. one takes a notebook to a meeting and takes note and a salesman takes a mobile computer to a customer for a presentation. In general in these scenarios the application is used in a stationary setting between several minutes and hours. Whereas considering the usage of ultra-mobile devices interaction periods are often much shorter e.g. looking up an address takes only a few seconds and making a note on a PDA is often in the range of several seconds up to some minutes. Also the fact that the applications are mainly used while doing something else or to carry out a certain task (like tools in the real world) calls for a reduction of the explicit human-machine interaction and creates the need to shift towards implicit HCI.

Knowledge about the situational context is of primary interest to the application, because we consider that the application will adapt to the context.

It can be observed that an application (mobile or stationary alike) is:

1. running on a specific device (e.g. input system, screen size, network access, portability, etc.),
2. at a certain time (absolute time e.g. 9:34 p.m., class of time e.g. in the morning)
3. used by one or more users (concurrently or sequentially),

4. in a certain physical environment (absolute location, type of location, conditions such as light, audio, and temperature, infrastructure, etc.),
5. social setting (people co-located and social role),
6. to solve a particular task (single task, group of tasks, or a general goal).

We consider the items (1) to (6) as the basic building blocks of context. For mobile applications especially (4) and (6) are of major interest. In mobile settings the physical environment can change while an application is executed e.g. making a phone call while walking from the office desk to the car park. The telephone application is running, while the noise level changes between office and outside.

2.4 HCI in Context

HCI for mobile devices is concerned with the general trade-off between devices qualities (e.g. small size, light-weight, little energy consumption, etc.) and the demand for optimal input-output capabilities. Here implicit HCI can offer interesting alternatives.

2.4.1 Output in Context

Over recent years the output systems for mobile devices became much better; features such as stereo audio output, high-resolution color screens for PDAs and even on mobile phones as well as display systems for wearable computers are commercially available. Also unobtrusive notification mechanisms (e.g. vibration) have become widely used in phones and PDAs. Still on the lower end devices with very poor display quality enter the market. Situational context can help to:

- adapt the output to the current situation (fontsize, volume, brightness, privacy settings, etc) [SATT⁺99].
- find the most suitable time interruption [SaSc97].
- reduce the need for interruptions (e.g. you don't need to remind someone to go to a meeting if he is already there.)

2.4.2 Input in Context

Considering very small appliances the space for a keyboard is very limited what results in bad usability. Other input systems, such as graffiti and handwriting recognition have been developed further but still lack in speed and accuracy. Advances in voice recognition have been made in recent years, but for non office settings (e.g. in a car, in a crowded place, sharing rooms with others, and in industry workplaces), the recognition performance is still poor. Also privacy and acceptance issues are a major concern. Implicit HCI does not solve these problems in general but can help to:

- adapt the input system to the current situation (e.g. audio filter, recognition algorithms, etc)
- limit need for input (e.g. information is already provided by the context and can be captured)
- reduce selection space (e.g. only offer appropriate options in current context)

2.5 Identifying Implicit Human Computer Interaction

To identify applications that can be improved by implicit HCI input and output of the application and the real world environment in which it is executed have to be analyzed. Then ways to capture the situational context must be assessed. Furthermore mechanisms for the interpretation of the situational context have to be found. Finally the reaction of the application has to be defined. The following questions help to identify these points:

1. What happens around an application while the application is in use? Are there any changes at all?
2. Do the surroundings (behavior, environment, circumstances) carry any valuable information for the application? Does it matter for the application?
3. Are there any means to capture and extract the information in a way that is acceptable for the application or device (processing cost, sensor cost, weight, etc.)?
4. How to understand the information? What interpretation and reasoning is possible and useful. What is an appropriate way for the application to react?

2.6 Modeling Implicit Human Computer Interaction

To specify applications that facilitate implicit HCI it is inevitable to have a specification language to describe situational context linked to events/change that occur in the application. In our recent work we found it helpful to use a notation that is human readable as well as easily to process using a computer. We decided to use a markup language that is specified in XML for this purpose. Extending the SGML based description model introduced by Brown in [BrBC97] we added two more concepts - grouping context with matching attributes and trigger attributes to make the description more expressive and suitable for our projects.

If contexts are composed of a number of components we found it very helpful to have a mechanism to bundle certain contextual variables in groups and select a matching semantic for each group description. For matching in a group we provided the following semantics: one (match one or more of the variables in the following group), all (match all variables in the following group), none (match none of the variables in the following group). All groups within the context description must evaluate to true to cause the trigger. We discriminate three different triggers: 'enter a context', 'leave a context', and 'while in a context'. The 'enter' and 'leave' triggers take a time value that specifies the time after which the action is triggered if the context stays stable over this time. For the 'while in a context' trigger the time indicates the interval in which the trigger is fired again.

2.7 Perception of Context

There are several ways to equip devices with perceptual capabilities. The range of complexity to consider is very wide, starting from simple sensors that know the way a device is held to complex audio and video analysis. We identified the following four basic approaches:

- device-databases (e.g. calendars, todo-lists, address books, profile, etc.)
- input to the application running (notepad - taking notes, calendar - looking up a date, etc.)
- active environments (active badges , IR-networks, cameras, audio, etc.)

- sensing context using sensors (TEA¹, Sensor Badges, e.g. [GoLe99], GPS, cameras, audio [SaSc97], etc.)

The perceptual capabilities can be located in the device itself, in the environment or in another device that shares the context over a network (e.g. body area network).

3 Conclusion and Further Work

Based on observations of new sensing technology, available sensors and anticipated users a new interaction metaphor is proposed. Implicit HCI is defined as an action, performed by the user that is not primarily aimed to interact with a computerized system but which such a system understands as input. It is further identified that perception and interpretation of the user, the environment, and the circumstances are key concepts for implicit HCI. Furthermore applications that exploit this information are required. Perception and interpretation are considered as situational context. Therefore we motivate a broad view of context, and also suggest that the context is described from the perspective of the application. To identify applications that can make use of situational context and thus can facilitate implicit HCI a number of questions are raised and an algorithm is suggested. It is based on the central questions: what happens around the application, how can this be sensed or captured, how to interpret this information, and how can applications make use of it. From current projects we learned that there is a need for a simple specification language for implicit HCI, based on situational context. We propose an XML-based markup language that supports three different trigger semantics. The language is easily human readable and also easy to process. Basic mechanisms of perception to acquire situational context are discussed. In a further section benefits of implicit interaction through situational context to HCI are discussed.

References

- [BrBC97] P.J. Brown, J.D. Bovey und X. Chen. Context-Aware Applications: From the Laboratory to the Marketplace. *IEEE Personal Communications*, Oktober 1997.
- [CBDF99] K. Cheverst, G. Blair, N. Davies und A. Friday. Supporting Collaboration in Mobile-aware Groupware. *Personal Technologies* 3(1), März 1999.
- [Corp99] NCR Corp. Mülleimer informiert Supermarkt. <http://www.heise.de/newsticker/data/anm-28.10.99-001/>, 1999.
- [GoLe99] A. Golding und N. Lesh. Indoor Navigation Using a Diverse Set of cheap wearable sensors. In *Proceedings of the third International Symposium on Wearable Computers, San Fransico*, Oktober 1999.
- [SaSc97] N. Sawhney und C. Schmand. Nomadic Radio: A Spatialized Audio Environment for Wearable Computing. In *Proceedings of the International Symposium on Wearable Computers, Cambridge, MA*, Oktober 1997.
- [SATT⁺99] A. Schmidt, K.A. Aidoo, A. Takaluoma, U. Tuomela, K. Van Laerhoven und W. Van de Velde. Advanced Interaction in Context. In *1th International Symposium on Handheld and Ubiquitous Computing (HUC99)*, September 1999.

¹<http://tea.starlab.org/>

-
- [ScAW94] B.N. Schilit, N.L. Adams und R. Want. Context-Aware Computing Applications. In *Proceedings of the Workshop on Mobile Computing Systems and Applications*, Santa Cruz, CA, Dezember 1994.
- [ScBG99] A. Schmidt, M. Beigl und H.-W. Gellersen. There is more to context than location. *Computer and Graphics* 23(6), Dezember 1999, S. 893–901.

Mobilität und Multi-Milliarden-Teilnehmersysteme: Neue Anforderungen an die Netzverwaltung

Jochen Seitz

1 Die drei Leitqualitäten der Telekommunikation

Heutige Kommunikations- und Rechnernetze sind derart komplex, dass sie ohne computergestützte Administration nicht betrieben werden können. Dabei sollen im Wesentlichen die drei Leitqualitäten der Telekommunikation unterstützt werden:

1. *Robustheit* — Das Thema der Robustheit behandelt insbesondere die Facetten Zuverlässigkeit (wie kann sichergestellt werden, dass ein Kommunikationsdienst unanfällig gegen Fehler oder Überlast erbracht wird) und Datensicherheit (wie kann der kommunizierte Inhalt gegen unberechtigten Zugriff Dritter geschützt werden).
2. *Skalierbarkeit* — Eine Netz- und Systemmanagementarchitektur muss dafür sorgen, dass das Netz mit den angebotenen Diensten skaliert, d.h. ungeachtet der geographischen Ausdehnung, der Teilnehmerzahl und der Anzahl und Heterogenität der beteiligten Komponenten funktioniert.
3. *Allgegenwart* — Schließlich muss im Zuge des mobilen Zugriffes auf unterschiedliche Netzdienste dafür Sorge getragen werden, dass dieser Zugriff überall (d.h. an unterschiedlichen Orten sowie in unterschiedlichen Anwendungskontexten) und zu jeder Zeit möglich ist.

Durch die rapide Fortentwicklung im Telekommunikationssektor sind besonders zwei Aspekte zu nennen, die Auswirkungen auf die Netzadministration haben. Zum einen sind mit der Mobilität die Leitqualitäten Robustheit und Allgegenwart betroffen. Zum anderen wirkt sich die steigende Zahl von Netzteilnehmern, die insbesondere immer mehr technischer Natur, also Geräte oder Softwareprozesse, sind, sehr auf die Robustheit und die Skalierbarkeit des Netzes aus.

Die Verwaltung von Kommunikationsnetzen und verteilten Systemen muss bei diesen neu entstehenden Anforderungen folgende Aspekte berücksichtigen:

- *Verwaltung der Netzkomponenten und Endgeräte*
Unter diesen Aspekt fallen die Überwachung der Funktionsweise, die Entdeckung von Leistungsengpässen sowie die Meldung und Behandlung von Fehlern.
- *Bereitstellung einer Dienstinfrastruktur*
Eine Managementarchitektur muss zudem allgemeine Infrastrukturdienste zur Verfügung stellen, auf die über wohldefinierte Schnittstellen zugegriffen werden kann.

Diese beiden Aspekte werden im Folgenden separat behandelt, wobei nach den neuen Anforderungen „Mobilität“ und „Multi-Milliarden-Teilnehmersysteme“ unterschieden wird.

2 Verwaltung der Netzkomponenten und Endsysteme

2.1 Einfluss der Mobilität

Automatisierte Konfiguration der Endsysteme — Prinzipiell mussten die netz- und systemspezifischen Einstellungen an Endgeräten von Hand vorgenommen werden. Dies ist bei mobilen Systemen nicht mehr praktikabel, da sich die Systemumgebungen aufgrund der Mobilität schnell ändern können. Insbesondere ist ein automatischer Wechsel der zugrundeliegenden Netzwerktopologie wünschenswert, der Benutzerpräferenzen ebenso wie momentane Erreichbarkeit von Netzwerken berücksichtigt. So kann ein mobiler Rechner, der innerhalb eines Gebäudes betrieben wird, dort eine WaveLAN-Installation nutzen. Bewegt sich der Benutzer mit dem Rechner aus dem WaveLAN-Bereich, ist möglicherweise immer noch eine DECT-Abdeckung verfügbar. Erst, wenn sonst keine anderen Netzwerke zur Verfügung stehen, kann der mobile Rechner GSM nutzen, um Internet-Konnektivität zu erreichen.

Übergabe der Verwaltungszuständigkeit — Traditionell wird jeder Netzkomponente und jedem Endgerät eine Verwaltungsinstanz statisch zugeordnet, die deren Funktionstüchtigkeit überwacht und bei Fehlern korrigierend eingreift. Mit der Möglichkeit zur Mobilität wechselt die Komponente jedoch auch den Zuständigkeitsbereich der ihr ursprünglich zugeordneten Verwaltungsinstanz. Das bedeutet, dass die Zuständigkeit an eine Instanz weitergegeben werden muss, die in der neuen Umgebung des mobilen Endsystems für das Management verantwortlich ist. Dabei ist zudem noch zu bedenken, dass oftmals eine absolute Übergabe der Zuständigkeit das mobile Endsystem ganz der neuen Verwaltungsinstanz ausliefert, die dann unter Umständen auch vertrauliche Daten aus dem mobilen Gerät gewinnen oder sogar ändern kann. Somit muss der Grad der Delegation der Zuständigkeit genauestens festgelegt werden, damit die neue Verwaltungsinstanz genügend Rechte hat, ihre Managementaufgaben zu erfüllen, aber in ihren Rechten dennoch derart eingeschränkt ist, dass sie keine unerlaubten Manipulationen am mobilen Endgerät durchführen kann.

2.2 Einfluss der Multi-Milliarden-Teilnehmersysteme

Dezentralisierte Verwaltung — Eine herkömmliche Managementarchitektur sieht eine zentrale Managementinstanz vor, die für alle Netzkomponenten in ihrer Domäne verantwortlich ist. Aufgrund der zunehmenden Teilnehmerzahl steigt auch die Zahl notwendiger Netzkomponenten, so dass ein zentraler Ansatz nicht mehr praktikabel ist. Vielmehr wird eine verteilte Verwaltungsarchitektur notwendig, wobei die folgenden Ausprägungen bereits Gegenstand der Forschung sind:

- *Hierarchische Verwaltung*: Einteilung in Unterverwaltungsbereiche.
- *Verteilter Ansatz*: Situationsbezogene Aufteilung der Verwaltungsaufgaben.
- *Verwaltung auf Basis mobiler Agenten oder aktiver Netze*: Verwaltungsfunktionalität vor Ort.

Kooperierende Dienstbringer — Mit der Anzahl der Teilnehmer und somit der möglichen Dienstnehmer steigt auch die Anzahl der Dienstbringer entsprechend, wobei diese sich immer mehr spezialisieren werden. Daher erfolgt die Dienstbringung dann meist über mehrere kooperierende Dienstanbieter, die somit unterschiedliche Rollen annehmen werden, die in Abbildung 1 dargestellt sind [Seit00].

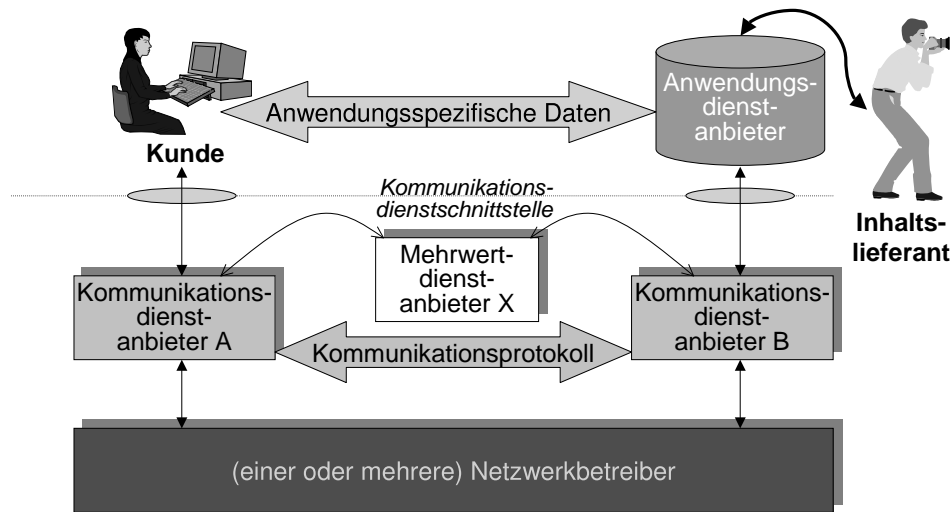


Abbildung 1: Übersicht über Rollen

Zwischen diesen kooperierenden Dienstbringern muss es zu einem Austausch von Managementinformation kommen. Dies betrifft insbesondere die Aushandlung der geforderten Dienstgüte in Form von sogenannten **Service Level Agreements (SLA)**, die daher notwendige Überwachung der Dienstgüte an den Dienstschnittstellen sowie eine Kooperation bezüglich Sicherheitsfragen und Abrechnung. Dieser Informationsaustausch muss sowohl Transparenz als auch Nachvollziehbarkeit für den Endbenutzer gewährleisten.

3 Bereitstellung einer Dienstinfrastruktur

3.1 Einfluss der Mobilität

Ortsabhängige Dienstewahl — Zum einen ist eine statische Dienstewahl aufgrund Mobilität der Dienstnehmer unvorteilhaft, zum anderen sind Dienste oft abhängig von der Lokation der Dienstausführung. Beispiele sind Auskunftsdienste, die auf eine bestimmte Region beschränkt sind. Daher muss es Mittel geben, den momentan geeignetsten Dienstgeber zu lokalisieren und auszuwählen. Dies führt zur gleichen Lösung, die in Abschnitt 3.2 angesprochen wird, nämlich der Dienstevermittlung.

Adaptivität bei Nutzung von Funkstrecken — Mobile Endteilnehmer nutzen häufig Funktechnologie zur Herstellung der Konnektivität. Solche Funkstrecken haben aber – verglichen mit der leitungsgebundenen Übertragung – relativ schlechte physikalische Eigenschaften, die sich in geringer Bandbreite, hoher Bitfehlerrate und Übertragungsverzögerung und somit schwankender Übertragungsdienstgüte ausdrücken. In der Dienstinfrastruktur sollte somit ein Mechanismus bereitgestellt werden, der die benötigte Übertragungsleistung auf dem drahtlosen Übermittlungsabschnitt anpasst, ohne dass andere Teilnehmer dadurch beeinflusst werden. Bisher benutzte Anwendungsprogramme sollten auf der anderen Seite möglichst unverändert übernommen werden können. Als Lösung bietet es sich an, in der Dienstinfrastruktur Stellvertreter für mobile Systeme zu integrieren, welche die Kommunikationsanforderungen an die jeweiligen Gegebenheiten anpassen können. Beispiele sind das Proxy-System RAPP [SCDE⁺99] zur Adaption von Multimedialströmen oder die Proxy-Plattform Π^2 [KnRS00] zur Unterstützung CORBA-basierter verteilter Anwendungen im mobilen, drahtlosen Umfeld.

3.2 Einfluss der Multi-Milliarden-Teilnehmersysteme

Dienstvermittlung — Aufgrund der Vielzahl an Teilnehmer wird sich ein breites Spektrum von zur Verfügung gestellten Diensten entwickeln, die sich nur in Nuancen unterscheiden. Dennoch sollte es jedem der Teilnehmer möglich sein, den ihm passenden Dienst auszuwählen. Dazu ist eine dezentrale Dienstvermittlung notwendig, an die sich die Endteilnehmer wenden können. Mit Hilfe einer zu spezifizierenden Dienstbeschreibung, welche sowohl Diensteigenschaften als auch Lokations- und Kontextinformation (siehe Abschnitt 3.1) enthält, kann dann der Dienst und somit der entsprechende Dienstanbieter vermittelt werden. Es existieren bereits eine ganze Reihe von Dienstvermittlungssystemen, sogenannten Trader, die jedoch einerseits nicht skalieren und für die es andererseits keine standardisierte Semantik einer Dienstbeschreibung gibt. Zudem fehlt die Möglichkeit, den Geltungsbereich eines Dienstes festzulegen. Dies ist vor allem dann aber sinnvoll, wenn die Inanspruchnahme eines bestimmten Dienstes nur innerhalb eines vorgegebenen Umfelds möglich ist.

Sicherheitsproblematik — Je mehr Teilnehmer Dienste in Anspruch nehmen können, desto höher werden die Sicherheitsanforderungen. Die Gefahr, unerkannt eine Dienstabwicklung mitzuhören oder sogar zu ändern, steigt mit der Anzahl der möglichen Dienstnehmer, da die Dienstabwicklung relativ offen gestaltet werden muss, um eine größtmögliche Akzeptanz seitens der Dienstnehmer zu erreichen. Andererseits kann eine umfangreiche Einführung des elektronischen Handels nur basierend auf fundierten Sicherheitsmechanismen erfolgen. Diese Sicherheitsmechanismen sollten in den eigentlichen Kommunikationsablauf integriert sein und sind somit Bestandteil des Netz- und Systemmanagements.

4 Zusammenfassung

Die obigen Abschnitte zeigten die neuen Anforderungen, welche durch die zunehmende Mobilität und die enorm wachsende Zahl der Endteilnehmer an die Netzverwaltung gestellt werden. Die Konzepte zur Bewältigung dieser Anforderungen sind — wie aufgezeigt — vorhanden, allerdings muss noch ein übergreifendes Rahmenwerk konzipiert werden, in welches die einzelnen Verfahren eingegliedert werden können. Die Definition eines solchen Rahmenwerks ist umso schwieriger, da sich für den traditionellen Bereich der Netz- und Systemverwaltung bislang auch noch keines durchsetzen konnte. Somit wird es zumindest in absehbarer Zeit keinen integrierenden Ansatz, sondern vielmehr Insellösungen geben.

Literatur

- [KnRS00] M. Knapp, R. Ruggaber und J. Seitz. Π^2 - a Generic Proxy Platform for Wireless Access and Mobility in CORBA. In *Nineteenth ACM Symposium on Principles of Distributed Computing (PODC 2000)*, Portland, Oregon, USA, July 16–19 2000. ACM.
- [SCDE⁺99] J. Seitz, K. Cheverst, N. Davies, M. Ebner und A. Friday. Management of Proxy Objects Providing Multimedia Applications in the Mobile Environment. In *IFIP/IEEE International Symposium on Integrated Network Management (IM'99)*, Boston, Mass., USA, May 10–14 1999. IFIP/IEEE.
- [Seit00] J. Seitz. A Modelling Technique for Inter Domain Management. In *Proceedings of the IEEE International Conference on Communications (ICC 2000)*, New Orleans, Louisiana, USA, June 18–22 2000. IEEE.

Mobile Agenten zur Unterstützung des Netzwerkmanagements

Bernhard Thurm

1 Einleitung

Die wachsende Popularität des Internets führt zu einer immer unübersichtlicheren Informationslandschaft – schon heute sind Benutzer oft überfordert, wenn es darum geht, bestimmte Ressourcen oder Informationen aufzufinden und deren Nutzen zu analysieren. Ein Rückgang des unbeschränkten Wachstums und der damit verbundenen Informationsüberflutung ist dabei zunächst nicht abzusehen. Eine seit Anfang der 90'er Jahre diskutierte Möglichkeit, dieser Problematik entgegenzusteuern, sind jedoch sog. „Software-Agenten“, im Wesentlichen eigenständig agierende Programme, die sich im Auftrag eines Benutzers im Netz bewegen können, um bestimmte Aufgaben – wie beispielsweise eine Flugbuchung – zu übernehmen.

Neben dem wachsenden Informationsangebot führt die steigende Akzeptanz des neuen Kommunikationsmediums Internet auch zur Einführung neuer Dienste (Videokonferenzen, Telelearning etc.), was gleichzeitig die Komplexität der gesamten Netztechnologie drastisch erhöht. Hierbei spielt zum einen der reine Größenfaktor eine wichtige Rolle, d.h. die rasch wachsende Zahl der Nutzer, die gleichzeitig unterschiedlichste (Netz-)Dienste in Anspruch nehmen, zum anderen aber auch die Qualität der geforderten Dienste, deren Charakteristika über einfache Bandbreitenanforderungen weit hinausgehen und stattdessen eine dedizierte Dienstgüte erfordern. In den seltensten Fällen ist hierbei von einer homogenen Netzstruktur auszugehen – die reibungslose Funktionsweise und Zusammenarbeit unterschiedlichster Komponenten und Technologien, seien es z.B. drahtlose oder drahtgebundene Netze, muss jedoch zu jedem Zeitpunkt vom Netzwerkmanagement sichergestellt werden.

Klassische Management-Ansätze wie CMIP¹ der ISO/OSI bzw. SNMP² der IETF besitzen dabei ein Reihe von Nachteilen, die sich aus dem zugrundeliegenden Client-Server-Modell ergeben. Zu erwähnen sind hier v.a. die schlechte Skalierbarkeit, die recht hohe Netzlast sowie die unbefriedigenden Reaktionsmöglichkeiten bei Netzausfällen. Um diese Nachteile zu kompensieren, wird seit einiger Zeit über den Einsatz mobiler Agenten im Bereich des Netzwerkmanagements nachgedacht.

2 Mobile Agenten im Netzwerkmanagement

Eine allgemeine Definition eines mobilen Agenten lässt sich wie folgt charakterisieren³ [BiPW98]:

¹Common Management Information Protocol

²Simple Network Management Protocol

³Die aufgezählten Eigenschaften sind dabei nicht ausschließlich zu verstehen, d.h. nicht auf jeden mobilen Agenten müssen alle Punkte zutreffen.

„Ein *mobiler Agent* ist eine Entität, die im Auftrag anderer arbeitet, autonom, proaktiv, reaktiv und lernfähig ist sowie die Fähigkeit zur Kooperation und Mobilität besitzt.“

Um eine Migration zwischen verschiedenen Hosts vornehmen zu können, ist ein mobiler Agent dabei auf eine einheitliche Agenten-Plattform angewiesen, die ihm eine entsprechende Ablaufumgebung zur Verfügung stellt. In diesem Bereich sind eine ganze Reihe von Produkten sowohl kommerziell als auch frei verfügbar [IBM00, Kine00, Mits00], die sich mittlerweile hauptsächlich auf die Programmiersprache Java konzentrieren. Tatsächlich verspricht Java eine weitreichende Einsetzbarkeit der Agenten aufgrund der Betriebssystemunabhängigkeit und der mittlerweile hohen Verbreitung der Java Virtual Machine – auch die Fähigkeiten zur Objektserialisierung, die den Migrationsvorgang recht einfach gestalten, sowie die Fähigkeiten zum entfernten Funktionsaufruf lassen Java als sehr geeignet erscheinen. Eine generische Architektur für solche Java-basierten Agenten ist in Abbildung 1 dargestellt [WoPM99]. Der sog. *Mobile Agent Server* ist dabei zuständig für die Ausführung der einzelnen Agenten. Seine Bestandteile sind der *Agent Manager* (zuständig für den Migrationsvorgang), der *Reliability Manager* (Sicherstellung des korrekten Empfangs eines Agenten), der *Security Manager* (Schutz vor böswilligen Agenten durch Authentifikation, Zugriffskontrolle und Verschlüsselung) sowie der *Communication Manager* (Kommunikation zwischen mobilen Agenten). Das *Application Gateway* bildet eine Schnittstelle zu herkömmlichen Diensten, die über den *Directory Manager* lokalisiert werden können. Die hauptsächlichsten Mängel liegen dabei in der nicht vorhandenen Kompatibilität unterschiedlicher Plattformen, in der Sicherheit (wie kann z.B. ein Agent vor einer böswilligen Plattform geschützt werden?), in der Ressourcenverwaltung (wieviel Rechenzeit darf z.B. einem Agenten zugeteilt werden?) und in der Dienstbereitstellung (wie kann ein Agent erkennen, wo ihm welche Dienste angeboten werden?).

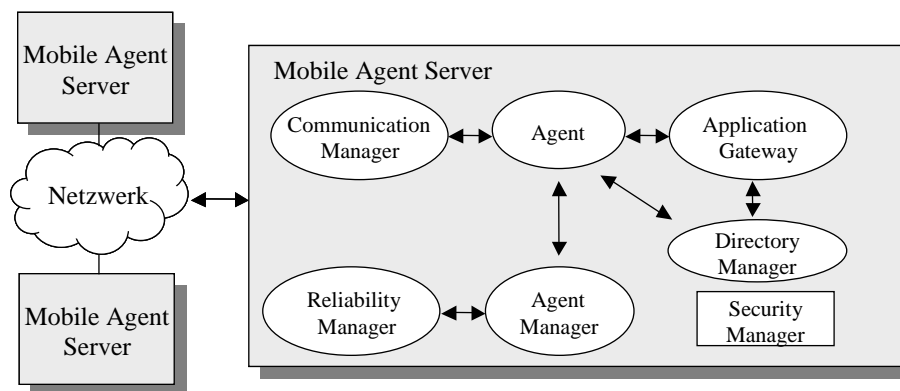


Abbildung 1: Architektur für Java-basierte mobile Agenten

Neben ihrer universellen Einsetzbarkeit, z.B. beim elektronischen Zahlungsverkehr, bei der Informationssuche oder als persönliche Assistenten bieten mobile Agenten gerade im Bereich des Netzwerkmanagements eine ganze Reihe von Vorteilen [GoWY98, KoGr99, LaOs99]:

- Zunächst wird eine *Reduktion der Netzwerklast* erzielt, da das im Netzwerkmanagement oftmals zum Einsatz kommende „Polling“ entfällt und nur noch der reine Programmcode (der meist um einen wesentlichen Prozentsatz geringer ist als die abzufragenden Daten) an das Zielgerät übertragen wird. Selten benötigte Programmfunktionen können von den Agenten auch erst beim Auftreten entsprechender Situationen nachgeladen werden, was die Netzlast weiter reduziert.
- Gerade drahtlose Netze, bei denen bei weitem noch keine so hohe Verfügbarkeit und Störungssicherheit wie im Festnetzbereich gegeben ist, profitieren von der *Asynchronität*

der Ausführung. Selbst bei einem Ausfall der Kommunikationsbeziehung zwischen mobilem Agenten und Auftraggeber kann der Agent weitgehend selbständig weiterarbeiten. Hierdurch wird zudem die *Robustheit und Störsicherheit* des Managements verbessert.

- Durch seine Fähigkeit zur *eigenständigen und dynamischen Anpassung an verschiedene Gegebenheiten* entfällt die Notwendigkeit eines zentralen Managers.
- Durch die Ausführung direkt innerhalb des zu verwaltenden Gerätes ist ein Agent vom aktuellen Netzzustand unabhängig. *Management- und Analysefunktionen* können *in Echtzeit* auf dem lokalen Host ausgeführt werden, was zu einer schnelleren Reaktion auf Problemsituationen führt.
- Mobile Agenten sind prinzipiell *unabhängig von verwendetem Betriebssystem und Transportprotokoll*, was momentan jedoch durch die verschiedenen Agenten-Plattformen noch nicht gegeben ist. Durch die breite Verwendung von JAVA scheint sich jedoch auch hier eine Lösung abzuzeichnen – auch im Zusammenhang mit der vereinfachten Sprachversion JINI, die in Kleingeräten in Zukunft weite Verbreitung besitzen wird.

Die vielfältigen Anwendungsbereiche mobiler Agenten beim Netzwerkmanagement sollen im Folgenden an einigen Beispielen verdeutlicht werden. Ein in heterogenen Netzen sehr wichtiges Gebiet ist beispielsweise die Erkennung und Klassifikation vorhandener Geräte und Dienste. [BiPW98] beschreibt hierzu ein Verfahren, bei dem mobile Agenten (*deglets*⁴) von Knoten zu Knoten wandern und die Ergebnisse zum Auftraggeber zurücksenden. Auch bei der Erkennung und Behandlung von Netzwerkfehlern sowie bei der Konfiguration von Netzelementen lassen sich mobile Agenten effizient einsetzen – eine wichtige Eigenschaft stellt dabei die autonome Vorgehensweise dar, die einen hohen Grad an Selbstverwaltung des Netzes ermöglicht und den Einsatz menschlicher Netz-/Systemadministratoren erst in Situationen erfordert, die nicht dem Standardfall entsprechen. Auch eine direkte Integration solcher Agenten in die Netzinfrastruktur, d.h. mobile Agenten als integraler Bestandteil neuer Netztechnologien, ist denkbar – modifizierte Agenten könnten dann vom Hersteller direkt in die zu verwaltenden Netze „gesendet“ werden. Eng verbunden mit Mechanismen zur Konfiguration des Netzwerks ist auch die Leistungsüberwachung. Hier sind durch die Lokalität der Agenten und den damit verbundenen Wegfall der Latenzzeiten zum einen genauere Messungen, zum anderen lokale Analysen der Messwerte und damit insgesamt verbesserte Reaktionsmöglichkeiten gegeben. Die Kriterien zur Migration einzelner Agenten bzw. ganzer Dienste, die als Agenten implementiert sind, lassen sich dabei zusätzlich mit der Leistungsüberwachung verknüpfen.

3 Zusammenfassung und Ausblick

Die grundlegende Idee mobiler Software-Agenten entstand bereits Anfang der 90'er Jahre, vornehmlich im Forschungsumfeld der künstlichen Intelligenz. Zunächst als eher abstraktes Konzept verstanden, herrscht mittlerweile in weiten Kreisen die Auffassung, dass das bisherige Client-Server-basierte Kommunikationsschema durch die neuen Möglichkeiten mobiler Agenten revolutioniert werden wird. Hauptaugenmerk gilt hier bisher dem Bereich des elektronischen Zahlungsverkehrs sowie der Beherrschung der allgegenwärtigen Informationsflut im weltweiten Internet. Wie gezeigt wurde, lässt sich dieses neue Paradigma jedoch auch im Bereich der Telekommunikation – und hier besonders im Netzwerkmanagement – sehr gewinnbringend einsetzen: Reduktion der Netzlast, verbesserte Behandlung von Fehlersituationen und die Verlagerung der „Intelligenz“ in die einzelnen Geräte lässt den Gedanken an Netzelemente, die sich selbst verwalten, nicht mehr nur als reine Utopie erscheinen.

⁴Die Autoren unterscheiden temporär existierende Agenten, die Delegation Agents (*deglets*) von längerfristig vorhandenen Agenten, den Network Agents (*netlets*).

Ansatzpunkte für weitergehende Forschungsaktivitäten ergeben sich beispielweise aus der Anwendung mobiler Agenten in der Leistungsüberwachung von Netzwerken, etwa bei der Planung und Koordination des Einsatzes einzelner Agenten bzw. der Kooperation zum Zwecke erweiterter Mess-/Analysefunktionen – auch Konzepte der Parallelverarbeitung zur Leistungssteigerung können hier zum Einsatz kommen. In Kombination hiermit kann die Möglichkeit gesehen werden, mobile Agenten zum Abwehren bzw. zum Aufspüren und Verfolgen von Netzattacken bzw. unbefugten Eindringlingen zu verwenden. Dies kann in Zusammenarbeit mit herkömmlichen Überwachungsmethoden erfolgen, durch die im weiteren Verlauf spezialisierte mobile Agenten erzeugt werden. Weitere Fragestellungen im Bereich Sicherheit behandeln den Einsatz von Verfahren zur Authentifikation und Verschlüsselung, um die Kommunikation der Agenten untereinander bzw. mit der jeweiligen Plattformen abzusichern. Sehr vernachlässigt in diesem Bereich ist bisher der Schutz des Agenten vor der jeweiligen Ablaufumgebung, in fast allen Produkten wird diese bisher als vertrauenswürdig erachtet – bei einer weltweit möglichen Migration kann dies jedoch durchaus nicht als gegeben angesehen werden. Desgleichen stellt die Verwaltung der mobilen Agenten ebenfalls ein wenig untersuchtes Aufgabenfeld dar, z.B. die Behandlung unterschiedlicher Zugriffsrechte für den eigentlichen Auftraggeber des Agenten bzw. die in „fremden“ Netzwerken vorhandenen Managementstationen. Da die Verwaltung der heutigen Netzkomponenten noch stark auf dem bisher favorisierten Client-Server-Prinzip beruht, sind zudem Strategien zu entwickeln, diese in effizienter Weise in das neuartige Management einzubeziehen. Wie schon in Abschnitt 2 erwähnt, eignen sich drahtlose Netz in besonderer Weise für den Einsatz mobiler Agenten – hier sind z.B. Agenten denkbar, die einem drahtlosen System beim ersten Kontakt zugeteilt werden und daraufhin für dessen Kontrolle sorgen.

Literatur

- [BiPW98] A. Biezcad, B. Pagurek und T. White. *Mobile Agents for Network Management*. IEEE Communications Surveys, Fourth Quarter 1998, Vol.1 No.1. 1998.
- [GoWY98] G. Goldszmidt, T. Watson und Y. Yemini. *Delegated Agents for Network Management*. IEEE Communications Magazine. Seite 66-70, März 1998.
- [IBM00] IBM. *Aglets Software Development Kit*. <http://www.trl.ibm.co.jp/aglets/index.html>. 2000.
- [Kine00] Kinetoscope. *Versatile Intelligent Agents*. <http://www.kinetoscope.com/via/default.htm>. 2000.
- [KoGr99] D. Kotz und R. Gray. *Mobile Agents and the Future of the Internet*. ACM Operating Systems Review. Seite 7-13, August 1999.
- [LaOs99] D. Lange und M. Oshima. *Seven Good Reasons for Mobile Agents*. Communications of the ACM. Seite 88-89, März 1999.
- [Mits00] Mitsubishi. *Concordia*. <http://www.meitca.com/HSL/Projects/Concordia/Welcome.html>. 2000.
- [WoPM99] D. Wong, N. Paciorek und D. Moore. *Java-based Mobile Agents*. Communications of the ACM. Seite 93-101, März 1999.

Skalierbare Dienstgarantien im Internet der nächsten Generation

Klaus Wehrle

1 Einleitung

Dieser Beitrag befasst sich mit der Integration von skalierbaren Mechanismen zur Dienstgüteunterstützung (Quality of Service – QoS) im Internet von morgen. Das Rahmenwerk basiert auf der neuen QoS-Architektur „Differentiated Services“ (DiffServ) der Internet Engineering Task Force [RFC 2475]. Beim deren Entwurf wurden – aufgrund von Erfahrungen aus früheren Ansätzen – die Mechanismen innerhalb des Netzes so einfach wie möglich gestaltet, um die Skalierbarkeit gewährleisten zu können. Komplexe Funktionen, wie Klassifizieren und das Speichern von Zustandsdaten wurden an die Netzgrenzen verlagert.

Somit wird ein skalierbares Rahmenwerk für Erbringung von „verschiedenen Diensten“ zur Verfügung gestellt. Jedoch gibt es bei diesem Rahmenwerk noch einige offene Punkte und Problemstellungen, mit deren Lösung sich die im Folgenden vorgestellten Arbeiten befassen.

2 Probleme der Differentiated-Services-Architektur

Die inneren Router von DiffServ-Domänen verfügen über keine Möglichkeiten zum Klassifizieren von Datenströmen und besitzen keine Zustandsinformationen. Sie können deshalb nicht überprüfen, ob eine Ressource tatsächlich reserviert wurde. Bei der Replikation eines Multicast-Pakets erfahren somit alle replizierten Pakete den gleichen (evtl. mehrwertigen) Dienst des Ursprungspaket – auch wenn ihnen keine Ressourcen zugeordnet wurden. Andere Datenströme können hierdurch enorm in ihrer Dienstqualität beeinträchtigt werden.

Für dieses Problem, welches als „Neglected Reservation Subtree Problem“ bekannt wurde, präsentiert [BIWe99] eine Lösung, die nicht nur sehr einfach ist, sondern auch die Skalierbarkeit der Architektur und der Dienstgarantien beibehält. Als weitere Vorteile ermöglicht sie nun auch erstmals die Verwendung von RSVP in DiffServ-Netzen.

3 Garantierte Dienstgüte

Wie bereits erwähnt, stellt die Differentiated-Services-Architektur ein Rahmenwerk zur Verfügung, mit dem Internet Service Provider in ihren Domänen nach eigenem Ermessen verschiedene Weiterleitungsmechanismen anbieten können. Aus der Konkatenation der Mechanismen in den betroffenen Domänen setzt sich schließlich die Ende-zu-Ende-Dienstgüte eines Datenstroms zusammen. Bisher gibt es kaum Modelle oder Untersuchungen, welche Aussagen über die tatsächlichen Ende-zu-Ende-Garantien von Datenströmen in Abhängigkeit der verwendeten Weiterleitungsmechanismen treffen können. Deshalb werden derzeit verschiedene Arbeiten in diesem Bereich durchgeführt, welche die resultierenden Ende-zu-Ende-Garantien mit drei verschiedenen Methodiken untersuchen.

3.1 Karlsruhe Implementation of Differentiated Services (KIDS)

Der erste Weg zur Untersuchung der erreichbaren Dienstgütern von DS-Weiterleitungsmechanismen ist deren praktische Implementierung unter dem Betriebssystem Linux. Nach einer ersten Implementierung der Basismechanismen [Wehr99b] wird derzeit eine sehr flexible und allumfassende Implementierung von Dienstgütemechanismen erstellt. Mit dieser Implementierung lassen sich, mit Hilfe kleiner Module, alle erdenklichen Kombinationen von Weiterleitungsmechanismen aufbauen.

Zusammen mit dem Differentiated-Services-Testnetz des Instituts für Telematik können nun mit der Implementierung KIDS ganze Domänen mit einer Größe von bis zu drei Dutzend Rechner nachgebildet werden. Weiterhin werden verschiedene Arbeiten durchgeführt, welche ein automatisches Konfigurieren der Versuchsumgebung und Durchführen von Versuchsreihen ermöglichen sollen. Das Ziel dieser Anstrengungen ist eine praktische Untersuchung der Dienstgütemechanismen.

3.2 Simulation

Neben den praktischen Untersuchungen werden derzeit auch Simulationsmodelle, sowohl für das Internet Protokoll, als auch die Differentiated-Services-Mechanismen erstellt. Diese Simulationsmodelle entsprechen in Aufbau und Funktionsweise genau der Architektur der Implementierung KIDS. Somit lässt sich nicht nur die Dienstgüte in einzelnen Domänen praktisch untersuchen, sondern mit den vergleichbaren Simulationsmodellen können Konkationen vieler verschiedener Domänen untersucht werden. Die Ähnlichkeit der praktischen Implementierung mit den simulativen Modellen erlaubt auch eine Überprüfung der Simulationsmodelle hinsichtlich ihrer wahrheitsgetreuen Beschreibung der realen Abläufe.

3.3 Analytische Modelle

Als dritter Weg zur Untersuchung der Ende-zu-Ende-Garantien werden analytische Modelle erstellt. Es können sowohl Abschätzungen bezüglich der zu erwartenden Werte von Dienstgüteparametern gegeben werden, als auch mögliche Obergrenzen für die Auslastung von Netzwerkverbindungen. Im Folgenden wird ein Modell beispielhaft erläutert:

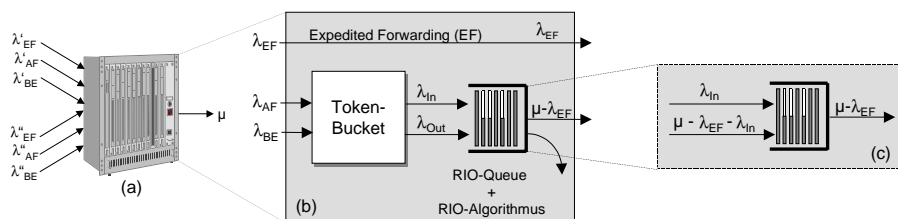


Abbildung 1: Modell zur Berechnung der garantierten AF-Bandbreite in einer RIO-Warteschlange

Abbildung 1 stellt den Grenz-Router einer DS-Domäne dar. Durch ihn fließen drei verschiedene aggregierte Paketströme: Expedited Forwarding (ein Weiterleitungsverhalten mit höchster Priorität), Assured Forwarding (bietet nur eine statistische Garantie der Bandbreite) und Best Effort (keine Dienstgarantien). Die statistische Garantie der AF-Bandbreite wird durch eine RIO-Warteschlange [RFC 2475] und den voran geschalteten Token Bucket erbracht. Je nachdem, wie die Parameter der Warteschlange gesetzt sind, wird eine andere Bandbreite für das AF-Aggregat garantiert. Leider gab es bisher keinerlei Relation zwischen der Bandbreite und den zu konfigurierenden Parametern. Durch das Aufstellen eines analytischen

Modells [Wehr99a] konnte die folgende Beziehung zwischen den Warteschlangenparametern (\min_{th} , \max_{th}) und der garantierten AF-Bandbreite (λ_{In}) gefunden werden:

$$\min_{th_{In}} \geq \frac{\lambda_{In}}{\mu - \lambda_{EF}} \cdot \max_{th_{Out}} \quad (1)$$

Ein praktischer Nachweis im Differentiated-Services-Testnetz zeigte die Korrektheit des Modells. Weitere analytische Untersuchungen befassen sich derzeit mit den Veränderungen von Teilströmen innerhalb aggregierter Datenströme bei Expedited Forwarding. Hierbei wurde beispielsweise festgestellt, dass es in bestimmten Routern zu erheblichen Verzögerungsschwankungen einzelner Ströme (innerhalb eines aggregierten Stroms) kommt, wobei die Qualität des Ende-zu-Ende-Dienstes aber nur geringfügig beeinträchtigt wird.

4 Nutzung der Dienstgüte im Internet von morgen

Mit Hilfe der Differentiated Services Architektur wird es zukünftig möglich sein, garantierte Dienste in der IP-Schicht zu erbringen. Jedoch sind die höheren Schichten im Internet-Protokollstapel derzeit noch nicht für solche Dienste vorbereitet. Deshalb werden verschiedene Arbeiten durchgeführt, welche die Nutzung von garantierten Diensten ermöglichen sollen.

4.1 Transportprotokolle mit Dienstgüteunterstützung

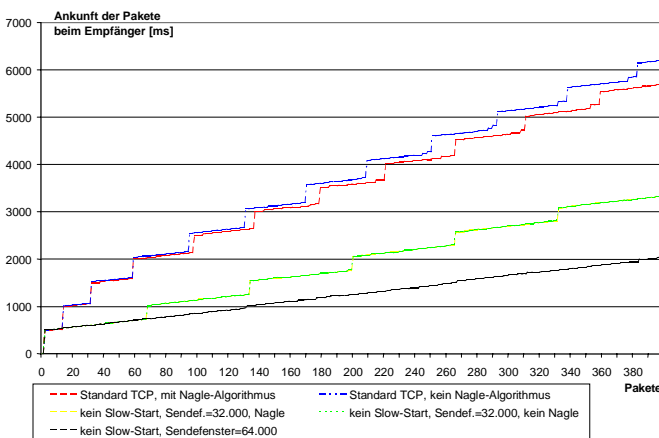


Abbildung 2: TCP-Durchsatz nach Anpassung an garantierte Bandbreite

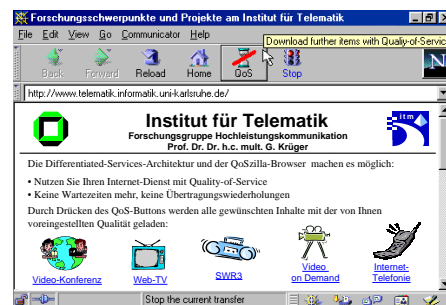


Abbildung 3: WWW-Browser mit QoS-Erweiterung

Vor allem TCP, welches das meistbenutzte das Transportprotokoll im Internet ist, ist derzeit nicht in der Lage garantierte Bandbreiten vollständig auszunutzen. Mit steigender Entfernung und Bandbreite verschlechtert sich diese Situation zunehmend (vgl. Abb. 2). Der eingebaute Staukontrollmechanismus, welcher bei garantierten Bandbreiten völlig unnötig ist, verhindert ein schnelles Senden der Daten. Desweiteren wird der Datenfluss oft durch ein zu kleines Sendefenster gebremst.

Nach einer Anpassung der TCP-Mechanismen konnte dessen Durchsatz um das dreifache gesteigert werden. Der Grad der Ausnutzung der garantierten Bandbreite liegt nahezu bei 100% und ist unabhängig von den Parametern Bandbreite und Verzögerung.

4.2 Signalisierung von Dienstgüteanforderungen im WWW

Um eine breite Resonanz im Internet zu bekommen, müssen sich die neuen Dienste im World Wide Web anwenden lassen. Bei der Migration von Differentiated Services und dem WWW treten jedoch einige Probleme auf, vor allem bei der Signalisierung der gewünschten Dienstgüteigenschaften vom Benutzer zum WWW-Server. In [RiPW00] wurden drei Ansätze vorgestellt, welche mögliche Lösungszenarien darstellen, die auf bestehenden Standards des World Wide Webs aufsetzen. Es wurde großen Wert auf die Einfachheit der Bedienung und Konfiguration der Dienste gelegt, um das Navigieren so einfach wie möglich zu halten. Abb. 3 zeigt das Beispiel eines erweiterten WWW-Browsers, der es ermöglicht die Dienstgüteanforderungen beim Anfragen einer HTML-Datei mitzusenden, ohne dass Erweiterungen am HTTP-Protokoll oder an HTML vorgenommen werden mussten.

Literatur

- [BlWe99] Roland Bless und Klaus Wehrle. Management Architektur zur Unterstützung von Gruppenkommunikation in Differentiated Services Netzwerken. In *Tagungsband des 1. GI Multicast Workshop, Braunschweig, Mai, 1999*.
- [RiPW00] Hartmut Ritter, Thorsten Pastoors und Klaus Wehrle. DiffServ in the Web. In *Tagungsband der Konferenz Networking 2000, Paris, Mai 2000*.
- [Wehr99a] Klaus Wehrle. Implementierung und Evaluierung neuartiger Dienste für das Internet der nächsten Generation. Diplomarbeit, Universität Karlsruhe (TH), Institut für Telematik, Januar 1999.
- [Wehr99b] Klaus Wehrle. Neuartige Dienstgütemechanismen für das Internet der nächsten Generation. In *Tagungsband der 1. GI Informatiktage, Bad Schussenried, November, 1999*.

Herausforderung Netzwerkmanagement

Hajo R. Wiltfang

1 Einleitung

Die ständig steigende Bedeutung der rechnergestützten Telekommunikation zeigt sich heute durch die Selbstverständlichkeit von Netzdiensten wie Email und Web, durch verstärkte online Abwicklung von Geschäftsprozessen in Form von Online-Banking oder Internet-Präsentationen sowie nicht zuletzt durch Schlagworte wie E-Commerce und E-Business. Einhergehend mit dieser Entwicklung wachsen zwangsläufig auch die Anforderungen an die Verfügbarkeit und Zuverlässigkeit von Netzdiensten, die wesentlich durch das ihnen zu Grunde liegende Netzwerk beziehungsweise den gesamten Netzwerkverbund bestimmt sind.

Der klassische Bereich des Netzwerkmanagements hat damit zur Aufgabe, über Mechanismen und Konzepte die Steuerung und Überwachung des Netzes derart umzusetzen, dass die erforderlichen Anforderungen gewährleistet werden können. Die Herausforderungen an das Netzwerkmanagement wachsen dabei sowohl mit der Größe des Netzwerks und der Anzahl seiner Benutzer als auch mit der ständig steigenden organisatorischen und funktionalen Komplexität der eingesetzten Netzwerktechnologien und der darauf basierenden Netzwerkkomponenten.

Hinzu kommt in jüngster Zeit die ständig zunehmende Konfrontation mit gezielten Angriffen oder ungewolltem Fehlverhalten, die beide situations- und technologieabhängig zu einer Beeinträchtigung der Netzverfügbarkeit führen können. Besonders moderne Hochleistungskomponenten wie Layer-3 und Layer-4 Switches oder Router zeigen hierbei noch einige Schwächen, vor allem wenn Sie zeitgleich mit einer großen Fülle neuer Verbindungsbeziehungen belastet werden.

Der vorliegende Beitrag soll in diesem Umfeld eine Reihe von Problemstellungen aufzeigen, deren Lösung sich die Forschungsgruppe „Hochleistungsnetze und Netzwerkmanagement“ in enger Zusammenarbeit mit dem Rechenzentrum der Universität Karlsruhe zum Ziel gesetzt hat. Hierzu wird zunächst genauer auf das Netzwerk der Universität Karlsruhe eingegangen, um anhand dieses großen Netzwerkbeispiels bedeutende Aufgabenstellungen näher aufzuzeigen. Im Anschluss daran erfolgt eine genauere Vorstellung erster erstellter Lösungsansätze sowie weiterer Planungen für zukünftige Arbeiten.

2 Universitätsnetz als Fallbeispiel

Prinzipiell gleich starke Betriebsanforderungen wie für kommerzielle Netze gelten auch für das Netzwerk der Universität Karlsruhe, das vom Rechenzentrum betrieben wird und allen Instituten der Universität den für die tägliche Arbeit benötigten Netzwerkzugang bereitstellt. Die Grundlage der Vernetzung an der Universität Karlsruhe bildet ein campusweites ATM-Backbone, das durch seine VLAN-Fähigkeit (Virtuelle LANs) die einfache Bildung von ortsunabhängigen Arbeitsgruppen ermöglicht. Nach dem Konzept des „Collapsed Backbones“

sorgen dabei nur wenige zentrale Router für die notwendige Verbindung und Vermittlung innerhalb dieser VLANs und zur Außenwelt.

Die zentrale Rolle der Router in derartigen Konfigurationen erfordert eine enorme Leistungsfähigkeit, die bei starker Überlastung des Netzwerks oder vor allem bei gezielten Angriffen schnell an ihre Grenzen gelangen kann. Beispielsweise ist es bei Hochleistungsroutern dieser Größenordnung üblich, die Weiterleitung von Daten direkt durch Hardware vornehmen zu lassen. Die Klassifikation von neuen Verkehrsbeziehungen und die Überprüfung ihrer Zulassungsvoraussetzungen erfolgt jedoch zuvor in einem initialen Schritt, der zumeist in Software von einer zentralen Einheit des Routers vorgenommen wird.

Im regulären Einsatz erscheint dieses Konzept völlig ausreichend und angemessen zu sein, jedoch birgt es Schwächen bei nicht normalem Netzwerkverkehr und gezielten Angriffen, wenn jeweils eine Unzahl an neuen Verkehrsbeziehungen gleichzeitig aufgebaut werden soll. Diese beispielhafte Problemsituation soll im Folgenden als Leitfaden dienen, wie mit Hilfe von Mechanismen des Netzwerkmanagements dennoch ein stabiler Betrieb großer Netzwerke erreicht werden kann.

3 Anforderungen an das Netzwerkmanagement

Wesentliche Anforderungen an das Netzwerkmanagement im beschriebenen Umfeld sind die kontinuierliche Überwachung des Betriebszustandes und die gleichzeitige Umsetzung eines geeigneten Sicherheitskonzepts.

Durch Überwachungsmechanismen ist dabei sicherzustellen, ausgehend von einer detaillierten Erfassung des Normalzustandes des Netzwerks möglichst früh jegliche Art von nicht normalem Verhalten im Netzwerk zu erkennen und zu lokalisieren. Spezielle Bedeutung kommt unter Berücksichtigung der potentiellen Angriffe der Aufgabe zu, bei erkanntem Fehlverhalten möglichst schnell die Ursache oder den Verursacher zu lokalisieren, um darauf aufbauend eine Behebung oder Ausgrenzung einzuleiten.

In Bezug auf die grundlegende Absicherung eines Netzwerks gegenüber möglichen Angriffen oder Fehlverhalten sind hingegen Aspekte wie interne Zugangsregelungen und externe Zugangsabsicherungen zu betrachten. Wichtige Hilfsmittel sind hierbei die bekannten Firewall-Konzepte oder auch einfache Filtermechanismen auf der Basis von Port-Listen.

4 Entwickelte Lösungsansätze

Als entwickelte Ansätze präsentiert der vorliegende Beitrag beispielhaft ein System für die Grundüberwachung des ATM-Backbones der Universität sowie ein Konzept für die Entdeckung von Netzangriffen.

4.1 Überwachung des ATM-Backbones

Ausgehend von der am Institut für Telematik entwickelten FuMA-Architektur [Wilt99] und darauf aufbauender Realisierungen [WiFo98, Thur99] ist in [Grün00] eine Erweiterung für den ATM-Netzwerkmonitor [Wilt98, Wilt99] entwickelt worden, die für physikalische Übertragungsabschnitte (Links) in ATM-Netzwerken völlig autonom Statistiken zur Auslastung und zu möglichen Fehlerfällen erfassen kann. Über flexible Konfigurationsmöglichkeiten in der erstellten ATM-QOS-LINK-MIB können dabei beispielsweise die zu überwachenden Links innerhalb der gesamten Netzwerktopologie genauer spezifiziert werden, ebenso lassen sich

Grenzwerte für zu beobachtende Werte festlegen, die bei Überschreiten dann zur eigenständigen Auslösung von Ereignissen führen.

Erste praktische Messungen in [Grün00] zeigen, dass die entwickelte Link-Überwachung im Universitätsnetz erfolgreich eingesetzt werden kann. Weitere Schritte bestehen nun in der Konfiguration einer kontinuierlichen Link-Überwachung für ausgewählte zentrale Bereiche des ATM-Backbones sowie in der gleichzeitigen Definition von Grenzwerten, um frühzeitig über auffällige Situationen bei der Netzwerkauslastung informiert zu werden. Parallel soll die bestehende Managementanwendung PVC-Manager [Thur99] um grafische Elemente erweitert werden, damit sie für die Konfiguration von Messungen der Link-Überwachung eingesetzt werden kann und darüber hinaus in der Lage ist, aktuelle Messergebnisse basierend auf der bereits vorhandenen Topologiedarstellung oder in neuartigen Darstellungsformen dem Benutzer zu präsentieren.

4.2 Konzept zur Entdeckung von Netzangriffen

Mit RMON (Remote Monitoring) und RMON-2 sind bereits zwei anerkannte Standards zur umfangreichen Überwachung des Verkehrsverhaltens in Netzwerken vorhanden [Schl97]. Für die Entdeckung von Netzangriffen bietet RMON mit seiner Vielzahl an Überwachungsmöglichkeiten gute Ansatzpunkte, wenn die verschiedenen Mechanismen innerhalb von RMON geeignet kombiniert werden.

Interessant ist dabei vor allem die „Packet Capture“-Gruppe, die eine kontinuierliche Aufzeichnung von Datenpaketen erlaubt, die aus dem Paketstrom des Netzwerks herausgefiltert wurden. Über die Definition von speziellen Aktivierungs- und Deaktivierungsereignissen kann zusätzlich erreicht werden, dass die Aufzeichnung von Paketen nur zu bestimmten Zeitpunkten beginnt oder endet. Somit ist es möglich, durch die kurzzeitige Aufzeichnung aller Pakete die Vorgeschichte oder Nachgeschichte zu einem vordefinierbaren Ereignis aufzuzeichnen, womit beispielsweise bei erkannten Netzproblemen die Entstehungsgeschichte dokumentiert und so eventuell die verursachenden Aktionen erkannt und lokalisiert werden können. Unter zusätzlicher Einbeziehung der netzwerkweiten Erfassungsmöglichkeiten von RMON-2 lässt sich bei böswilligen Angriffen so relativ genau der Ursprung identifizieren, der im Falle eines internen Rechners Rückschlüsse auf einen erfolgreichen Einbruch erlaubt und im Falle einer externen Herkunft Ansatzpunkte für zusätzliche Filtermechanismen liefern kann.

Wesentlich für eine erfolgreiche Umsetzung dieses RMON-basierten Überwachungskonzepts ist die Verfügbarkeit von geeigneten Managementanwendungen, die über angepasste Oberflächen die einfache Einstellung und Auswertung der konzipierten Überwachungsfunktionen erlauben. Dies bildet damit den Schwerpunkt der nächsten Arbeiten im betrachteten Bereich der Aufdeckung von Netzangriffen.

5 Zusammenfassung und Ausblick

Der vorliegende Beitrag führt in die zentrale Bedeutung des Netzwerkmanagements in heutigen Netzverbunden ein und belegt dies am Beispiel des Universitätsnetzwerks in Karlsruhe. Einer Reihe identifizierter wesentlicher Problemstellungen beim derzeitigen Management solcher Netzwerke werden exemplarisch zwei Lösungsansätze gegenübergestellt, die als erste Schwerpunkte innerhalb der Forschungsgruppe „Hochleistungsnetze und Netzwerkmanagement“ derzeit angegangen werden.

Bedeutende Vorteile bei der Entwicklung und Umsetzung erfährt die Forschungsgruppe dabei aus den Synergie-Effekten, die sich über die enge Zusammenarbeit zwischen dem Institut für

Telematik und dem Rechenzentrum ergeben. Zum einen bietet sich am Rechenzentrum ein ideales Umfeld, um Ergebnisse der Forschung anhand eines großen Netzwerks direkt in die Praxis umzusetzen, und zum andern können Anregungen und Anforderungen aus der betrieblichen Praxis am Rechenzentrum direkt in die Planung neuer oder vertiefender Forschungsschwerpunkte einfließen. Dies soll beispielsweise für das Themengebiet der Netzwerkabsicherung durch Firewalls sowie den dazugehörigen Konzepten und Architekturen in naher Zukunft einen weiteren Schwerpunkt der Arbeiten bilden.

Literatur

- [Grün00] Alex Grüner. Erfassen und Visualisieren aggregierter Link-Statistiken in ATM-Netzwerken. Diplomarbeit, Institut für Telematik, Universität Karlsruhe, Februar 2000.
- [Schl97] Gerhard Schlemm. Management von ATM-Netzwerken mittels Remote Monitoring (RMON). Diplomarbeit, Institut für Telematik, Universität Karlsruhe, Dezember 1997.
- [Thur99] Bernhard Thurm. Web-basiertes Management von ATM-Verbindungen. Diplomarbeit, Institut für Telematik, Universität Karlsruhe, Februar 1999.
- [WiFo98] Hajo R. Wiltfang und F. Fock. Topology Discovery for ATM Networks. In *Proceedings of the 9th IEEE Workshop on Local and Metropolitan Area Networks LAN-MAN'98*, Banff, Kanada, 17.–20. Mai 1998. The Institute of Electrical and Electronics Engineers (IEEE), 1998, S. 298–303.
- [Wilt98] Hajo R. Wiltfang. Topologie- und Dienstgüteefassung in ATM-Netzwerken. In Prof. Dr. Dr. h.c. G. Krüger, R. Bless und S. Dresler (Hrsg.), *Klausurtagung des Instituts für Telematik*, Nr. 10/98 der Interner Bericht. Universität Karlsruhe, Fakultät für Informatik, März 1998, S. 127–135.
- [Wilt99] Hajo R. Wiltfang. *Funktionsorientiertes Management heterogener ATM-Netzwerke*. infix, Sankt Augustin. ISBN 3-89601-457-9, 1999.