

Modellierung und Bewertung der Datenverwaltungskonzepte in UMTS

Von der Fakultät für Elektrotechnik
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades
eines Doktors der Ingenieurwissenschaften
genehmigte Dissertation

vorgelegt von
Dipl.-Ing. Damian Robert Lawniczak
aus Driesen

Referent: Univ.-Professor Dr.-Ing. Bernhard Walke
Korreferent: Univ.-Professor Dr. rer. nat. habil. Thomas Bemmerl

Tag der mündlichen Prüfung: 27. Juni 1995

AACHENER BEITRÄGE ZUR MOBIL- UND TELEKOMMUNIKATION

Herausgeber:

Universitätsprofessor Dr.-Ing. Bernhard Walke

Lawniczak, Damian Robert:
Modellierung und Bewertung der Datenverwaltungskonzepte in UMTS
1. Auflage Aachen:
Verlag der Augustinus Buchhandlung, 1995
(Aachener Beiträge zur Mobil- und Telekommunikation, Band 2)

ISBN 3-86073-381-8

© 1995 Damian Robert Lawniczak

Verlag der Augustinus Buchhandlung
Pontstraße 66/68
52062 Aachen

Druck: Copy Team, Aachen
Gedruckt auf chlorfrei gebleichtem Papier

"D 82 (Diss. RWTH Aachen)"

Vorwort

Die vorliegende Arbeit entstand während meiner Tätigkeit in der Forschungsabteilung der Ericsson Eurolab Deutschland GmbH in Kohlscheid. Ich möchte hiermit einen besonderen Dank an Frau Dr. Fiona Williams aussprechen, ohne deren Unterstützung diese Arbeit nie zustande gekommen wäre. Ferner danke ich meinen Kollegen sowohl in Kohlscheid als auch bei Ericsson Radio Systems AB in Stockholm, deren freundliche Zuneigung mir stets die Arbeit in positiver Atmosphäre erlaubte.

Ich möchte Herrn Prof. Dr.-Ing. B. Walke, dem Inhaber des Lehrstuhls für Kommunikationsnetze der Rheinisch-Westfälischen Technischen Hochschule Aachen, danken für die Betreuung der Arbeit und für seine wertvollen Anmerkungen zu dem Manuskript.

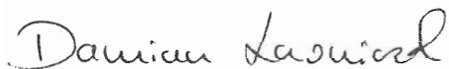
Herrn Prof. Dr. rer. nat. habil. T. Bemmerl danke ich für die freundliche Übernahme des Korreferates und für seine konstruktive und offene Haltung.

Ich danke ganz besonders Frau Jutta Eisbein und ihrem Vater Herrn Dr. Klaus Eisbein für ihren Einsatz bei der sorgfältigen Korrektur und Durchsicht der Arbeit.

Speziellen Dank möchte ich auch meinen Eltern für deren fortdauernde Unterstützung während des Entstehens dieser Arbeit aussprechen.

Aber vor allen Dingen danke ich dem Herrn, der mir das Talent, die Kraft und die Ausdauer gegeben hat, diese Arbeit in Angriff zu nehmen und bis zum Ende durchzuführen. Die etwaigen Fehler sind durch menschliche Unvollkommenheit zu begründen, denn nur der Herr ist vollkommen. Lobe den Herrn, von dem aller Segen fließt!

Ad maiorem Dei gloriam (A.M.D.G.)

A handwritten signature in dark ink, reading "Damian Lawisch". The script is cursive and fluid, with the first name "Damian" and last name "Lawisch" clearly distinguishable.

Aachen, im Juli 1995

Kurzfassung

Das allgemein in Europa als *Universal Mobile Telecommunication System* (UMTS) bekannte System der dritten Generation wird im Rahmen des europäischen Telekommunikationsprogramms *Research and development for Advanced Communication in Europe* (kurz RACE genannt) entwickelt und untersucht. Um die gewünschte Dienstqualität (z.B. eine weltweit einheitliche, personenbezogene Benutzernummer) zu gewährleisten, ist es notwendig, nicht nur moderne Technologien, sondern auch neue, anwendungsorientierte Konzepte der dienst- und benutzerspezifischen Datenhaltung einzusetzen. Da zu erwarten ist, daß die steigende Qualität der mobilen Netze ständig mehr Benutzer anziehen wird, ist mit einer hohen, dem Festnetz vergleichbaren Penetrationsrate zu rechnen. Beide zusammen - Realisierung der neuen Datenkonzepte und hohe Benutzerzahlen - stellen bedeutend höhere Anforderungen an die Datenspeicherung und -verwaltung, als man von den heutigen Systemen gewohnt ist.

Um den neuen Anforderungen an eine Datenspeicherung und -verwaltung gerecht zu werden, wurde im Rahmen des RACE-Projekts MoNet ein neues Konzept für die Benutzerverwaltung entwickelt, das auf einer verteilten Datenbank basiert. Die Daten der Benutzer werden hier gemäß einer bestimmten Strategie auf die Knoten der Datenbank verteilt. Um den Verwaltungsaufwand zu bewerkstelligen, der durch die Datenverteilung und infolge der Benutzermobilität in der Datenbank entsteht, wurde ein zusätzliches Verzeichnis von Hilfsknoten vorgesehen, kurz DIR genannt, die in einer baumähnlichen Struktur organisiert werden und eine Überwachung der Datenknoten ermöglichen.

Im Unterschied zu bekannten Anwendungsgebieten der verteilten Datenbanken ist die Problemstellung im UMTS-Netz differenzierter: einerseits wird die Datenreplikation und/oder -partitionierung auf maximal zwei Kopien und/oder zwei Knoten begrenzt, andererseits ändert sich der Datenspeicherungsort aufgrund der Benutzermobilität sehr oft. Die Leistungsfähigkeit eines Datenhaltungskonzeptes hängt wesentlich von der Zugriffshäufigkeit auf bestimmte Daten und dem Ort ihrer Speicherung ab. Um eine bezüglich der Leistungsfähigkeit optimale Lösung herauszuarbeiten, ist es notwendig, unterschiedliche, echtzeitfähige Ansätze der Datenhaltung und -verwaltung systematisch auszuarbeiten und anhand von Modellen des Benutzerverhaltens und der relevanten Datenbankankfragen quantitativ zu bewerten. Dies ist Hintergrund und Problemstellung der vorliegenden Arbeit.

Obwohl in der vergangenen Zeit unterschiedliche Datenhaltungsstrategien vorgeschlagen worden sind, wurde erst im Rahmen dieser Arbeit eine umfangreiche Systematisierung aller wesentlichen Datenbankalternativen durchgeführt. Es erfolgte eine Aufteilung der Alternativen in zwei Klassen, Architektur- und Verwaltungsalternativen, und es wurde der Einfluß jeder Alternative auf die Leistungsfähigkeit der Datenbank diskutiert.

Angesichts einer fehlenden UMTS-Implementierung, die ein Evaluieren unterschiedlicher Alternativen erlauben würde, mußten entsprechende Modelle hergelei-

tet werden. Das Hauptaugenmerk wurde dabei auf die Untersuchung der zwei wichtigsten Datenbank Anfragen gerichtet: Datenbankaktualisierung (aufgrund der *Benutzerbewegungen*) und Suche nach dem Aufenthaltsort des angerufenen Benutzers. Das zugrundeliegende Benutzerverhalten - Bewegung und Anrufverhalten - wurde formal beschrieben. Die Merkmale von Bewegung und Anrufverhalten wurden zueinander in Beziehung gebracht. Bezeichnend für die Modelle ist deren Flexibilität, d.h. die Möglichkeit, durch einige wenige Parameter ein geradezu beliebiges Benutzerverhalten darzustellen. Dies ist eine wichtige Voraussetzung für die Anwendbarkeit der Modelle zur systematischen Untersuchung aller relevanten Netzwerksituationen. Die beiden gewählten Anfragen wurden ebenfalls formal modelliert und in Relation zum Modell des Benutzerverhaltens gesetzt.

Auf der Grundlage der systematischen Analyse der Alternativen und unter Zuhilfenahme der Modelle wurde eine ausführliche quantitative Untersuchung der relevanten Netzwerksituationen durchgeführt. Diese Untersuchung ermöglichte es, fundierte Aussagen bzgl. der Leistungsfähigkeit der Datenbank in Abhängigkeit vom Benutzerverhalten und der gewählten Alternative zu machen. Darüber hinaus wurde aus der durchgeführten Untersuchung eindeutig klar, daß eine benutzerspezifische Verzeichnisarchitektur und eine den Bewegungen angepaßte Datenverteilung vorteilhaft gegenüber einem festen Schema sind. Dies ist eine in der Telekommunikationswelt neuartige Erkenntnis, die in bisherigen Arbeiten nicht beachtet wurde.

Die Arbeit enthält eine exemplarische Implementierung eines Transportprotokolls für die UMTS-Datenbank mit detaillierten Nachrichtenflüssen, die über eine formale Spezifikationssprache (SDL) erstellt wurde. Dabei wurde neben einer Abschätzung des programmtechnischen Aufwandes des Protokolls auch dessen alternativspezifische Komplexität bewertet. Beides ist von besonderem Interesse, wenn benutzerangepaßte Architekturen und Datenverteilungen angewendet werden sollen. Hier darf der Mehraufwand infolge der Implementierung von angepaßten Lösungen die Vorteile aufgrund der Anpassung zum Benutzerverhalten nicht überwiegen - die gewonnenen Erkenntnisse zeigen deutlich, daß eine flexible Protokollimplementierung nur eines geringfügigen zusätzlichen Aufwandes bedarf.

In der Arbeit wurde erstmalig eine systematische Ausarbeitung der relevanten Datenbankalternativen dargestellt. Unter Verwendung eines neuartigen Modellierungsansatzes wurde eine quantitative Bewertung der Alternativen in allen relevanten Netzwerksituationen durchgeführt. Die dabei gewonnenen Ergebnisse vermitteln interessante, auf quantitativen Rechnungen basierende Einsichten in die Effizienz unterschiedlicher Datenkonzepte. Die exemplarische Implementierung auf der Basis der formalen Spezifikation hat letztendlich die einfache Realisierungsmöglichkeit der dargestellten Konzepte (besonders der benutzerspezifischen Protokollanpassung) belegt.

Abstract

The third generation mobile system, known in Europe as *Universal Mobile Telecommunication System* (UMTS), was under development within framework of the European programme *Research and development for Advanced Communication in Europe* (RACE). The high expectation for service quality in future systems (e.g. personal and global user number) will require not only modern technology but additionally new, application oriented concepts for service and user specific data storage. Moreover the anticipated high quality of mobile networks will attract more and more users. Combining the required new data concepts and high penetration rate, place high demands on the data storage and data control mechanisms, both of which are unexplored areas in the network of today.

To fulfil these extraordinary difficult demands a novel concept for user control based on a distributed database was elaborated in the RACE project MoNet. In this concept the user data is distributed as data items among the *UMTS data nodes* which are administered following a particularly chosen strategy. In order to be able to control the data items which must change their data node location as users move around, a separated *UMTS directory* was introduced consisting of additional nodes organized in a tree-like manner.

Contrary to the well known applications where distributed databases are employed nowadays, the replication and/or duplication of data items in the UMTS database is reduced to only two copies. On the other hand the frequency of the data location updates is relatively high due to anticipated and frequent user movements. From this perspective the efficiency of a particular data distribution concept is dependant on the accessibility on data items and on their storage location. To draw some conclusions relating to the performance of a particular distribution strategy, numerous different real time approaches have to be systematically elaborated and evaluated. Models of users behaviour and relevant database procedures need to be derived for this evaluation. These issues lay the foundation and the challenging questions to be answered in my work.

With this work for the first time a comprehensive and systematic elaboration and analysis of all relevant database alternatives was performed. The alternatives were subdivided into two groups, architecture and control alternatives, and the impact of each of these alternatives on the database performance was investigated in detail.

In view of the absent UMTS implementation which would have enabled a founded evaluation of different alternatives, it was necessary to derive models describing relevant network procedures and actions to characterise users behaviour. As for the procedures themselves, two were chosen: location update on users movement and data searching for called party data during call set-up. The corresponding users actions, movements and call set-ups, were formally described and put into context to each other. The characteristic feature of the derived users models is their flexibility, in other words the capability to describe almost each users behaviour by tuning only a few parameters. This property make it possible to use the models above for a sys-

tematic investigation of all relevant network situations. The procedures considered (location update and searching) were also formally modelled and put into context with users behaviour.

On the basis of a systematic analysis of all alternatives and by using derived models, a sophisticated quantitative investigation of all relevant network situations was performed. This investigation made it possible to draw conclusions about database performance in dependency on users behaviour and alternative chosen. Moreover, from the investigations performed the conclusion which can be made is that a user-adapted directory structure and a data distribution strategy which is adapted to users movements offer a performance improvement over fixed schemes. This is a clear innovation in the telecommunication world of today.

Finally, a sample implementation of a transport protocol for a UMTS database including detailed information flows was elaborated by using the formal specification language SDL. An estimation of the programming efforts and complexity of the protocol in various solutions can be derived from this exercise. This information is of crucial importance when a user-adapted solution (as mentioned above) should be taken into consideration. In such cases the effort required for an implementation of various options shouldn't predominate over the advantages gained. Experience shows that a flexible implementation of protocol requires only a minor increase in programming effort.

Concluding the work, it can be stated that for the first time, all important alternatives of a UMTS database were systematically elaborated. New modelling approaches for both users behaviour and network procedures enabled a quantified analysis of those alternatives in all relevant network situations. The experienced attained enables founded conclusions concerning the efficiency of the various alternatives of a UMTS database. The implemented SDL protocol proves the principal simplicity of its realisation in all alternatives considered.

Inhaltsverzeichnis

1	Einführung und Problemstellung	1
2	Telekommunikation und Datenbanken	4
2.1	Grundlagen der Datenbanken	4
2.1.1	Zentrale Datenhaltung	4
2.1.1.1	Datenbankoperationen	4
2.1.1.2	Architektur der Datenbank	4
2.1.1.3	Komponenten der Datenbank	5
2.1.2	Verteilte Datenbanken	5
2.1.2.1	Strukturen der verteilten Datenbanken	6
2.1.2.2	Bearbeiten der Datenbankabfragen	7
2.1.2.3	Datenwiederherstellung	7
2.1.2.4	Datentransaktionen und Datenaustauschprotokolle	8
2.1.2.5	Gleichzeitigkeitsüberwachung	9
2.1.2.6	Datenbankverklemmungen	10
2.2	Datenbanken in der Telekommunikation	11
2.2.1	Allgemeines zur Datenverwaltung	11
2.2.2	Konzepte der Datenverwaltung	13
2.2.3	Zusammenfassung der Datenbankanwendungen	16
2.3	Mobile und personalisierte Systeme von heute	16
2.3.1	Global System for Mobile Communication	17
2.3.1.1	Allgemeine Übersicht	17
2.3.1.2	Datenkonzept	18
2.3.2	Intelligente Netze	19
2.3.2.1	Allgemeine Übersicht	20
2.3.2.2	Datenkonzept	21
2.3.3	Universal Personal Telecommunication	22
2.3.3.1	Allgemeine Übersicht	22
2.3.3.2	Datenkonzept	23
2.3.4	Personal Communication System	23
2.3.4.1	Allgemeine Übersicht	24
2.3.4.2	Datenkonzept	25

2.3.5	Personal Service Communication Space	25
2.3.5.1	Allgemeine Übersicht	25
2.3.5.2	Datenkonzept	26
2.3.6	Digital European Cordless Telephone	26
2.3.6.1	Allgemeine Übersicht	26
2.3.6.2	Datenkonzept	27
2.3.7	Andere Systeme	27
2.4	Universal Mobile Telecommunication Systems	28
2.4.1	Überblick	28
2.4.2	Das UMTS-Konzept	28
2.4.2.1	Das allgemeine Funktionsmodell	29
2.4.2.2	Das Datenbankmodell	30
2.4.3	Rolle der dezentralisierten Datenbank in UMTS	31
2.4.4	UMTS als System der dritten Generation	32
3	Einfluß der Datenbank auf die UMTS-Prozeduren	33
3.1	UMTS-Datenkonzept	33
3.1.1	Benutzerdaten	34
3.1.2	Endgerätedaten	35
3.1.3	Dienstanbieterdaten	36
3.1.4	Teilnehmerdaten	36
3.2	Modell der UMTS-Datenbank	37
3.2.1	Struktur der Datenbank	37
3.2.2	Funktionale Architektur	39
3.3	UMTS-Benutzerprozeduren	40
3.3.1	Mobilitätsbezogene Prozeduren	41
3.3.1.1	Location Update	41
3.3.1.2	Domain Update	41
3.3.1.3	Terminal Detach	42
3.3.1.4	Terminal Attach	42
3.3.1.5	Handover	42
3.3.2	Nichtmobilitätsbezogene Prozeduren	43
3.3.2.1	Session Set-Up / Session Release	43
3.3.2.2	Call Set-Up/Call Release	44
3.3.2.3	User Registration/Deregistration	45
3.3.3	Einfluß der Prozeduren auf die internen Transaktionen	45

3.4 Alternativen der UMTS-Datenbank	46
3.4.1 Strukturelle Alternativen	47
3.4.1.1 Vertikaler Aufbau der verteilten Datenbank	47
3.4.1.2 Organisation der DIR-Wurzelebene	48
3.4.1.3 Gleichmäßigkeit der Datenbankstruktur	49
3.4.1.4 Zeitvariabilität der Datenbankstruktur	50
3.4.2 Alternativen der Datenverteilung und Datenverwaltung	51
3.4.2.1 Datenverteilungsalternativen (Datenszenarien)	51
3.4.2.2 Interne Abfragetechniken	54
3.4.2.3 Zeigeroptionen in den DIR-Knoten	57
3.5 Bewertungskriterien der Leistungsfähigkeit	59
3.5.1 Anforderungen an die Datenbankknoten	60
3.5.1.1 Anforderung an die Prozessorkapazität	60
3.5.1.2 Anforderung an die Speicherkapazität	60
3.5.1.3 Knotenorganisation und dessen Speicherdurchsatz	61
3.5.2 Belastung der Verbindungen	64
3.5.3 Prozedurverzögerungszeit	65
3.5.4 Zuverlässigkeit und Fehlertoleranz	66
3.5.5 Zusammenfassung der Bewertungskriterien	67
4 Analytisches Modell	68
4.1 Allgemeine Bemerkungen	68
4.2 Mobilitätsmodell	68
4.2.1 Bewegungen innerhalb eines Teilgebietes	69
4.2.1.1 Allgemeine Annahmen	69
4.2.1.2 Stationäre Lösung	73
4.2.1.3 Nichtstationäre Lösung innerhalb eines Teilgebietes	79
4.2.1.4 Nichtstationäre Lösung zwischen den Teilgebieten	82
4.2.1.5 Überlagerung unterschiedlicher Lösungen	83
4.2.1.6 Ergänzung des Modells	84
4.2.2 Bewegungen zwischen Teilgebieten	85
4.3 Anrufverteilungsmodell	88
4.3.1 Allgemeine Bemerkungen	88
4.3.2 Das Modell	88
4.3.3 Einfluß der Mobilität auf das Anrufverhalten	90
4.4 Modellierung der ausgewählten Anfragen	91

4.4.1	Suchanfragen	92
4.4.1.1	Überblick	92
4.4.1.2	Formale Beschreibung	93
4.4.1.3	Leistungsfähigkeit der Suchanfragen	98
4.4.2	Aktualisierungsanfragen	99
4.4.2.1	Überblick	99
4.4.2.2	Formale Beschreibung	100
4.4.2.3	Leistungsfähigkeit der Aktualisierungsanfragen	105
5	Leistungsbewertung	107
5.1	Programmierungsumgebung	107
5.1.1	Matlab	107
5.1.2	Softwarestruktur	108
5.2	Grundkonfiguration	109
5.2.1	Topologie und Parameter	109
5.2.2	Präsentation der Ergebnisse	111
5.2.3	Suchanfragen in der Grundkonfiguration	112
5.2.4	Aktualisierungsanfragen in der Grundkonfiguration	115
5.3	Auswertung der Suchanfragen	117
5.3.1	Einfluß der Verteilungsparameter	117
5.3.1.1	Untersuchte Alternativen	117
5.3.1.2	Ergebnisse der Auswertung	118
5.3.2	Einfluß der Anrufparameter	120
5.3.2.1	Untersuchte Alternativen	120
5.3.2.2	Ergebnisse der Auswertung	121
5.3.3	Einfluß der DIR-Architektur	121
5.3.3.1	Untersuchte Alternativen	121
5.3.3.2	Ergebnisse der Auswertung	122
5.3.4	Einfluß der Gebietstopologie	125
5.3.4.1	Untersuchte Alternativen	125
5.3.4.2	Ergebnisse der Auswertung	125
5.3.5	Verzögerungszeit bei Suchanfragen	126
5.3.6	Folgerung bezüglich der Datenverteilungsszenarien	127
5.3.7	Folgerung bezüglich der Abfragetechnik	128
5.4	Auswertung der Aktualisierungsanfragen	128
5.4.1	Einfluß der Mobilitätsparameter	128

5.4.1.1	Untersuchte Alternativen	128
5.4.1.2	Ergebnisse der Auswertung	130
5.4.2	Einfluß der DIR-Architektur	132
5.4.2.1	Untersuchte Alternativen	132
5.4.2.2	Ergebnisse der Auswertung	133
5.4.3	Einfluß der Gebietstopologie	135
5.4.3.1	Untersuchte Alternativen	135
5.4.3.2	Ergebnisse der Auswertung	136
5.4.4	Folgerungen bezüglich der Datenverteilungsszenarien	136
5.4.5	Folgerungen bezüglich der Abfragetechnik	137
5.4.6	Einfluß des Vorwissens auf die Auswertung	137
5.5	Speicherbedarf bei verschiedenen Alternativen	138
5.5.1	Speicherbedarf in den DN-Knoten	138
5.5.2	Speicherbedarf in den DIR-Knoten	139
5.5.3	Numerisches Beispiel	139
5.6	Relation zwischen UMTS-Konfigurationen und GSM	140
5.7	Folgerungen und Empfehlungen	141
6	Applikationsprotokoll für die UMTS-Datenbank	144
6.1	Anforderungen an das Protokoll	144
6.1.1	Formale Informationsflüsse	144
6.1.1.1	Suchanfrage	144
6.1.1.2	Aktualisierungsanfrage	145
6.1.2	Flexibilität	145
6.1.3	Datenintegrität und Datenzuverlässigkeit	146
6.2	Protokollspezifikation	148
6.2.1	Grundelemente innerhalb eines Teilgebietes	149
6.2.1.1	Auslesen	149
6.2.1.2	Erstellen	149
6.2.1.3	Aktualisieren	150
6.2.1.4	Löschen	151
6.2.1.5	Bewegen	152
6.2.1.6	Verriegeln	153
6.2.2	Grundelemente zwischen den Teilgebieten	153
6.2.2.1	Auslesen	155
6.2.2.2	Erstellen	156

6.2.2.3 Aktualisieren	156
6.2.2.4 Löschen	157
6.2.2.5 Bewegen	157
6.3 Protokollimplementierung	157
6.3.1 SDL-Sprache	158
6.3.2 Funktionsblöcke des UMTS-Datenprotokolls	158
6.3.2.1 Anfragemanager (QH)	159
6.3.2.2 Datenverzeichnis (DS)	160
6.3.2.3 Datenspeicher (DA)	160
6.3.2.4 Administration und Simulator	160
6.4 Zusammenfassung	160
7 Schlußbemerkung und Ausblick	163
8 Literaturverzeichnis	165
Anhang A - Glossar	175
Anhang B - Numerische Ergebnisse	181

1 Einführung und Problemstellung

Die Entwicklung der Telekommunikationstechnologie auf dem Gebiet der Festnetze und die Fortschritte in der B-ISDN-Standardisierung in den letzten Jahren lassen die Erwartung wachsen, daß den künftigen Benutzern eine Reihe neuer Dienste mit bisher nicht vorhandener Qualität zur Verfügung gestellt wird. Andererseits beweisen die bisherigen Erfolge der neuen mobilen Systeme, z.B. des GSM-Netzes in Europa, daß die Endgerätmobilität zu einer unabdingbaren Eigenschaft der modernen Telekommunikation geworden ist. Die künftige, sog. dritte Generation der mobilen Systeme setzt sich zum Ziel, die Vorzüge der Dienstvielfalt in Festnetzen mit der globalen Mobilität zu verbinden.

Das allgemein in Europa als *Universal Mobile Telecommunication System* (UMTS) bekannte System der dritten Generation wird im Rahmen des europäischen Telekommunikationsprogramms *Research and development for Advanced Communication in Europe* (kurz RACE genannt) entwickelt und untersucht. Um die gewünschte Dienstqualität zu gewährleisten, ist es notwendig, nicht nur moderne Technologien, sondern auch neue, anwendungsorientierte Konzepte der dienst- und benutzerspezifischen Datenhaltung einzusetzen. Da zu erwarten ist, daß die steigende Qualität der mobilen Netze ständig mehr Benutzer anziehen wird, ist mit einer hohen Penetrationsrate zu rechnen. Beide zusammen - hohe Benutzerzahlen und Realisierung der neuen Datenkonzepte - stellen bedeutend höhere Anforderungen an die Datenspeicherung und -verwaltung, als man von den heutigen Systemen gewohnt ist.

Wenn man seinen Blick den heutigen mobilen Systemen der zweiten Generation zuwendet, dann ist zu erkennen, daß dort das Konzept der Datenspeicherung in einer Struktur realisiert wurde, die aus einem Heimknoten (HLR) und mehreren Besucher-knoten (VLR) besteht (siehe Kapitel 2.3.1). Dieses Datenkonzept kann nicht dazu verwendet werden, die entsprechenden Aufgaben des UMTS-Systems zu lösen. Abgesehen von den zu erwartenden Kapazitätsproblemen bei der Benutzung einer derartigen Struktur ist das UMTS-Konzept im Gegensatz zu dem GSM-Konzept auf einer personenbezogenen Benutzernummer aufgebaut (siehe Kapitel 2.4).

Um den neuen Anforderungen an Datenspeicherung und -verwaltung gerecht zu werden, wurde im Rahmen des RACE-Projekts MoNet ein neues Konzept entwickelt, das auf einer verteilten Datenbank basiert. Die Daten der Benutzer werden hier gemäß einer gewählten Strategie auf die sog. Datenknoten verteilt. Um den Aufwand zu bewerkstelligen, der durch die Datenverteilung und infolge der Benutzermobilität in der Datenbank entsteht, wurde ein zusätzliches Verzeichnisnetz, kurz DIR genannt, eingeführt. DIR besteht aus einer Reihe von Hilfsknoten, die in einer baumähnlichen Struktur organisiert werden und eine Verwaltung der Datenknoten ermöglichen. Dieses Konzept bietet Flexibilität der Datenverteilung und des Dateninhaltes und unterstützt eine globale Dienstversorgung (vgl. /Lawn95/).

Im Unterschied zu heutigen Anwendungsgebieten der verteilten Datenbanken ist die Problemstellung im UMTS-Netz differenzierter: einerseits wird die Datenrepli-

kation und/oder -partitionierung auf maximal zwei Kopien und/oder zwei Knoten begrenzt, andererseits sind die Daten hochmobil. Zusätzlich ist es wegen der Realzeitanforderungen der UMTS-Datenoperationen nicht möglich, die aus den Festnetzen bekannten Ansätze für baumähnliche Strukturen wie z.B. X.500 (vgl. /X.500/ und /Arms94/) anzuwenden. In diesem Zusammenhang erscheint es notwendig, eine UMTS-spezifische Lösung der Datenhaltung und -verwaltung auszuarbeiten, die auf dem im MoNet-Projekt vorgeschlagenen Konzept basiert.

Die Entwicklung einer optimalen Datenverteilungs- und -verwaltungsstrategie auf der Basis des Grundkonzeptes ist keine triviale Aufgabe. Verschiedene Vorschläge wurden bereits zum Teil ausgearbeitet (/Mone61/, /Lawn95/) und analysiert /Lawn94a/. Sie bieten jedoch keine endgültigen Lösungen und beschränken sich auf einige wenige Alternativen. Eine systematische und hinreichend fundierte Untersuchung fehlt bis jetzt. Im Rahmen dieser Arbeit wird eine umfangreiche Modellierung und Bewertung aller wesentlichen Architektur- und Datenverteilungsalternativen durchgeführt. Dabei wurden folgende Punkte ausgearbeitet:

- ein flexibles Modell des Benutzerverhaltens,
- eine wahrscheinlichkeitsbegründete Prozedurbeschreibung und
- SDL-basierte Protokollimplementierung unter Beachtung der Fehlertoleranz und Datenkonsistenz.

Im Kapitel 2 wird ein Überblick über den heutigen Stand der Entwicklung auf dem Gebiet der verteilten Datenbanken (DDB) gegeben. Die Konzepte der Datenspeicherung und -verteilung sowie die Anwendungen von DDB in der Telekommunikation werden der Reihe nach präsentiert.

Da das UMTS-Datenkonzept auf der Grundlage der bereits existierenden Ansätze entstanden ist, wird im Kapitel 2.3 eine Beschreibung der heutigen Systeme gegeben, wobei der Schwerpunkt auf die verwendeten Datenkonzepte und die daraus resultierenden Datenverwaltungen gelegt wurde. Die bestehende Vielzahl der heutigen Telekommunikationskonzepte, wie z.B. GSM, UPT oder PSCS, zeigt deutlich, daß sie das Gedankengut von UMTS sicherlich beeinflusst haben und weiterhin beeinflussen werden. Der Übersicht folgt eine allgemeine Präsentation des UMTS-Konzeptes und dessen Funktionalität.

Die komplexen Prozeduren, die im UMTS-System ausgeführt werden, verlangen nach einer regen Kommunikation mit der verteilten Datenbank. Im 3. Kapitel der Arbeit ist deshalb die Relation zwischen UMTS-Prozeduren und der verteilten Datenbank erläutert. In einem komplexen System, wie die verteilte UMTS-Datenbank es ohne Zweifel darstellt, gibt es eine Vielzahl von unabhängigen Architektur- und Datenverteilungsalternativen, die optimiert werden müssen. Die systematische Erfassung dieser Alternativen ist eine notwendige Voraussetzung für eine konsistente Systemauswertung. Die möglichen Alternativen, die einen Einfluß auf die Leistungsfähigkeit der UMTS-Datenbank haben, werden im Verlauf des Kapitels 3 detailliert dargestellt. Da im Rahmen dieser Arbeit eine quantitative Bewertung von unterschiedlichen Alternativen durchgeführt wird, ist das Ergebnis einer solchen

Bewertung von der Wahl angemessener Kriterien abhängig. Um die Analyse zu erleichtern, wird eine Spezifikation der gewählten Kriterien gegeben.

Die Untersuchung eines noch nicht existierenden Systems ist nur dann möglich, wenn ein Modell dieses Systems vorliegt. In der hier betrachteten Fragestellung einer optimalen Konfiguration des UMTS-Datenbanksystems ist es notwendig:

- das Verhalten der Benutzer, d.h. ihre Mobilität und Anrufverteilung, und
- die gewählten Prozeduren zu modellieren.

Die Mehrzahl der bereits bestehenden Ansätze, die das Benutzerverhalten beschreiben, betrachtet hauptsächlich die Mobilität auf der zellularen Ebene, wobei üblicherweise eine gleichförmige Benutzerbewegung als Grundlage genommen wird. Im Unterschied dazu ist es Ziel dieser Arbeit, das Verhalten der Benutzer im Gesamtnetz zu beschreiben. Das Benutzerverhalten wird in der vorliegenden Betrachtung durch zwei Komponenten repräsentiert, nämlich durch physikalische Bewegungen und durch Anrufe. Um die Bewegungen zu beschreiben, ist im Kapitel 4.2 ein formales Modell, auch Mobilitätsmodell genannt, entwickelt worden. Im folgenden Kapitel 4.3 ist ein Modell des Anrufverhaltens entwickelt worden, das die Benutzerbewegung und die von ihm ausgeführten Anrufe gleichzeitig betrachtet und in Relation setzt. Im Anschluß an diese beiden Modelle wurden die Prozeduren, die zu untersuchen sind, im Kapitel 4.4 exakt spezifiziert und dadurch einer quantitativen Auswertung zugänglich gemacht. Die entwickelten Spezifikationen des Benutzerverhaltens vermögen durch Parameteränderungen, verschiedene Situationen zu beschreiben und dadurch die Sensitivität einer speziellen Alternative zu testen.

Dem modelltheoretischen Teil folgt im Kapitel 5 eine Zusammenstellung der Ergebnisse ausgewählter Alternativen. Wegen Modifizierbarkeit und Schnelligkeit wurde die analytische Auswertung einer Simulation vorgezogen [Nieb91]. Die durchgeführten Untersuchungen erlauben einen Rückschluß auf bevorzugte Lösungen. Aufgrund dieser Ergebnisse, die vom Benutzerverhalten abhängig sind, wird es möglich, die UMTS-Datenbank entsprechend den erwarteten Benutzercharakteristika zu konzipieren.

Im 6. Kapitel dieser Arbeit wird ein Anwendungsprotokoll vorgeschlagen und dessen Implementierung beschrieben. Die geforderten Eigenschaften wie Flexibilität, Zuverlässigkeit und Datenkonsistenz werden anhand von Beispielen erläutert. Auf diese Art wird ein besserer Einblick in die Funktionsweise des implementierten UMTS-Applikationsprotokolls gegeben. Zwecks Validierung wurde dieses Protokoll in einer simulierten Umgebung getestet und seine Eigenschaften bezüglich der Fehlertoleranz und Datenkonsistenz erfolgreich unter Beweis gestellt.

Einige Schlußbemerkungen bezüglich der erzielten Ergebnisse sowie ein Ausblick auf verwandte Bereiche und offenstehende Fragen schließen die Arbeit ab.

Anmerkung:

Alle in dieser Arbeit verwendeten Abkürzungen sind im Anhang A erläutert.

2 Telekommunikation und Datenbanken

Im vorliegenden Kapitel werden die Grundlagen der Datenbanken präsentiert, wobei der Schwerpunkt auf verteilte Datenbankkonzepte gelegt wird. Darüber hinaus werden Konzepte der verteilten Datenbanken in der modernen Telekommunikation vorgestellt.

Ein kurzer Überblick über die Systeme von heute und eine Einführung in das UMTS-Konzept schließen das Kapitel ab.

2.1 Grundlagen der Datenbanken

2.1.1 Zentrale Datenhaltung

Obwohl in der vorliegenden Arbeit die Problematik eines einzelnen Datenknotens nicht explizit behandelt wird, so hat sie natürlich einen großen Einfluß auf die Funktionstüchtigkeit des Gesamtsystems. Die Organisation eines Knotens ist entscheidend z.B. für das Verhältnis zwischen den Verzögerungen beim Lesen und Schreiben. Darüber hinaus haben einige Aspekte der verteilten Datenbanken, wie z.B. Datensicherheit oder Gleichzeitigkeitsüberwachung, ihre Wurzeln in traditionellen Datenbanken, die aus einem Knoten bestehen.

2.1.1.1 Datenbankoperationen

Die Grundoperationen, die auf den Daten eines Datenbankknotens ausgeführt werden können, sind schon längere Zeit systematisiert. Folgende Grundoperationen können genannt werden /Voss91/:

- Lesen der Datenbankinformation
- Modifikation der Datenbankinformation
 - Hinzufügen eines neuen Eintrags
 - Ergänzen eines existierenden Eintrags um neue Daten
 - Aktualisieren von bestehenden Daten in einem existierenden Eintrag
 - Entfernen von Daten aus einem existierenden Eintrag
 - Löschen eines existierenden Eintrags

Diese Operationen werden vom Datenbank-Management-System (DBMS) durchgeführt, das auf der anderen Seite - d.h. zum Benutzer hin - eine Reihe von mehr oder weniger raffinierten Befehlen und eine Schnittstelle zur Verfügung stellt.

2.1.1.2 Architektur der Datenbank

Die Organisation der Daten auf dem physikalischen Medium ist gänzlich unterschiedlich zur Organisation der Daten aus der Sicht des Benutzers. Zum einen ist

der Benutzer normalerweise nicht an der physikalischen Konfiguration der Daten interessiert, und zum anderen ist es für ihn ohne jegliche Bedeutung, die physikalische Datenhaltung zu verfolgen. Dies hat zu einem mehrschichtigen Datenorganisationskonzept geführt, in dem die einzelnen Schichten durch Schnittstellen voneinander getrennt sind /Bell92/:

1. externe Schicht - sie repräsentiert die externe Benutzersicht
2. konzeptionelle Schicht - sie repräsentiert die logische Darstellung der Daten
3. interne Schicht - sie repräsentiert die physikalische Datenorganisation

In der praktischen Realisierung werden die 1. und die 2. Schicht sehr oft in einem gemeinsamen Modell realisiert (/Kort91/), z.B. in einem objektorientierten, relationalen, semantischen oder funktionalen Modell.

2.1.1.3 Komponenten der Datenbank

Die wichtigste Softwarekomponente der Datenbank ist das bereits erwähnte Datenbank-Management-System (DBMS). Sie besteht nach /Date85/ aus einer Implementierung von zwei Teilkomponenten - der Datendefinitionssprache (DDL) und der Datenmanipulationssprache (DML). Die Funktionalität der beiden Sprachen entspricht deren Namen.

Eine zuverlässige Implementierung der Datenbankfunktionalitäten verlangt eine Reihe von Zusatzdiensten, die ebenfalls im DBMS realisiert werden /Date91/:

- Interaktionen mit dem Dateimanager,
- Unterstützung der Datenintegrität,
- Unterstützung der Datensicherheit bei Vorliegen unterschiedlicher Zugriffsrechte,
- Anfertigung der Sicherheitskopien und Wiederherstellung verlorener Daten und
- Überwachung des gleichzeitigen Zugriffs.

Alle obigen Dienste werden durch das Managementsystem des Knotens zur Verfügung gestellt und deswegen in der vorliegenden Arbeit nicht untersucht, da hier das Ziel verfolgt wird, die Konzepte für die Datenverwaltung zu modellieren und zu untersuchen. Lediglich die Aspekte der Sicherheitskopien und der Zugriffsüberwachung werden näher betrachtet und sind deswegen im Kapitel 3.5.1 im Zusammenhang mit den Anforderungen an die Daten- und Verzeichnisknoten genauer beschrieben.

2.1.2 Verteilte Datenbanken

Die Entwicklung der verteilten Datenbanken, die sich unabhängig von der Telekommunikation vollzogen hat, ist eine konsequente Erweiterung des zuvor dargestellten zentralen Konzeptes und wurde durch folgende Umstände herbeigeführt:

- verteilte Datenzugriffsorte - aufgrund der immer stärkeren Ausbreitung von globalen und vernetzten Systemen sind die Teilnehmer, die ein gemeinsames System benutzen, geographisch verteilt,
- lokale Geltung und Verfügbarkeit der Daten - in den meisten Fällen einer verteilten Anwendung werden nur die lokalen Daten abgefragt /Bell92/,
- mangelnde Kapazität eines zentralen Knotens in einer zentralisierten Lösung und
- begrenzte Transportfähigkeit des heutigen Kommunikationsnetzes.

Die Vorteile eines funktionsfähigen verteilten Datenbanksystems liegen auf der Hand, wenn man die obigen Punkte genau analysiert. Andererseits bringt eine verteilte Datenhaltung neue Komplexität mit sich:

1. Die Datenknoten der verteilten Struktur müssen logisch (und natürlich auch physikalisch) miteinander verbunden werden.
2. Die Verzögerung der Datenabfrage ist abhängig von mehreren Faktoren und nicht nur von der Kapazität eines einzelnen Knotens, z.B. vom Algorithmus, der die Reihenfolge der Abfragen bestimmt.
3. Die Aspekte der Datenwiederherstellung sind durch die Verteilung über mehrere Knoten erheblich komplexer, da gleichzeitig mehrere Kopien und/oder mehrere Partitionen vorhanden sein können.
4. Die Datenaustauschprotokolle und -transaktionen müssen die Datenkonsistenz gewährleisten.
5. Die Aspekte möglicher simultaner Zugriffe und dadurch resultierende eventuelle Inkonsistenzen sind schwieriger zu unterbinden (Gleichzeitigkeitsüberwachung).
6. Die Problematik der Verklemmung (*deadlock*) muß neu konzipiert werden.

In folgenden Unterkapiteln werden die oben genannten spezifischen Aspekte genauer erläutert. Dabei wird im einzelnen darauf eingegangen, inwieweit diese Aspekte für die im Rahmen dieser Arbeit durchgeführten Untersuchungen eine Relevanz haben.

2.1.2.1 Strukturen der verteilten Datenbanken

Um eine netzweite Verfügbarkeit der Daten zu gewährleisten, müssen die einzelnen Datenbankstationen miteinander verbunden werden. Folgende Datenbanktopologien können unterschieden werden /Fort85/:

- voll integriertes Netz,
- teilweise integriertes Netz,
- baumartiges Netz,
- sternförmiges Netz,
- ringförmiges Netz und

- busähnliches Netz.

Eine zusätzliche Komplexität ergibt sich, wenn ein voll oder teilweise integriertes Netz eine spezielle logische, z.B. baumartige Struktur hat.

Will man beschreiben, wie die Daten auf die einzelnen Knoten zu verteilen sind, so ist eine aus den relationalen Datenbanken hervorgegangene Klassifizierung gebräuchlich, die zwischen horizontal und vertikal verteilten Daten unterscheidet /Lock93/. Die horizontale Verteilung gruppiert in einem Knoten die Daten, die eine bestimmte Eigenschaft (*attribut*) gemeinsam haben. Die vertikale Verteilung legt im Unterschied dazu Daten mit verschiedenen Eigenschaften zusammen. Beispielsweise speichert in GSM der HLR-Knoten permanente Daten von Benutzern, deren Nummern diesem Knoten angehören (horizontale Verteilung), und der VLR-Knoten speichert temporär Daten von all diesen Benutzern, die sich gerade im Gebiet des Knotens befinden, unabhängig von der Nummer des jeweiligen Benutzers (vertikale Verteilung)

In der Einführung in das UMTS-Datenkonzept im Kapitel 2.4 wird auf die dort vorgeschlagene Datenbankstruktur näher eingegangen. Es sei hier nur erwähnt, daß eine baumähnliche, zum Teil modifizierte Struktur die untersuchte UMTS-Lösung darstellt.

2.1.2.2 Bearbeiten der Datenbankabfragen

Im Gegensatz zu den Ein-Knoten-Datenbanken, wo das entscheidende Kriterium der Abfragezeitminimierung die Anzahl der Zugriffe auf das Speichermedium ist, sind im Falle der verteilten Datenbanken andere Faktoren ebenfalls von großer Bedeutung. Zu den zwei wichtigsten gehören hier /YuMe94/:

- die Möglichkeiten der Datenverteilung und die daraus resultierende verteilte Abfrageabarbeitung, die eine Variabilität der Verzögerung bedingt, und
- die Transportverzögerung.

Der Einfluß dieser beiden Faktoren, besonders des ersten, ist bei der Untersuchung von UMTS-Datenbankoptionen von großer Bedeutung. Die verschiedenen Datenverteilungsstrategien, später auch Datenszenarien genannt, werden in diesem Zusammenhang untersucht.

2.1.2.3 Datenwiederherstellung

Die Problematik der Datenwiederherstellung wird im Rahmen dieser Arbeit nicht explizit betrachtet. Zum einen ist die Art der Wiederherstellung in UMTS von der gewählten Datenverteilungsalternative unabhängig (vgl. Kapitel 6.4). Zum andern treten Datenbankfehler, die eine Wiederherstellung notwendig machen, relativ vereinzelt auf bzw. werden im zukünftigen System vereinzelt auftreten. Als solche haben sie keinen besonderen Einfluß auf die Gesamtbelastung der Datenbank.

2.1.2.4 Datentransaktionen und Datenaustauschprotokolle

Die Durchführung von Transaktionen verlangt nach *atomicity*, d.h. entweder wird eine Operation komplett durchgeführt oder gar nicht. Um dieser Anforderung gerecht zu werden, gibt es in verteilten Datenbanken in jedem Knoten die Funktion des Transaktionskoordinators (neben einem globalen Transaktionsmanager). Seine Aufgabe ist, die Datenabfrage in einzelne Teilabfragen aufzuteilen und an die entsprechenden Partnerknoten zu verschicken sowie die Abbruchroutine zu überwachen.

Die übliche Methode, wie die Transaktionskoordinatoren verteilte Transaktionen überwachen, ist das Bestätigungsprotokoll (*commit protocol*). Zwei Arten der Bestätigung werden in der Regel angewendet:

- Zweiphasenbestätigung und
- Dreiphasenbestätigung.

Die Zweiphasenbestätigung wird in zwei Phasen durchgeführt (siehe Abbildung 2.1):

1. Phase: Nachdem der Koordinator einer Transaktion von jedem beteiligten Knoten (sog. Partnerknoten dieser Transaktion) eine Information erhalten hat, daß diese Transaktion abgeschlossen worden ist (*complete*), wird eine Bestätigungsabfrage vom Transaktionskoordinator an alle Partnerknoten zurückgeschickt (*prepare*). Daraufhin sendet jeder der befragten Knoten eine Bestätigungsantwort, in der er dem Koordinator mitteilt, ob die durchgeführte Transaktion zu akzeptieren (Bereitantwort - *ready*) oder zu verwerfen (Abbruchsantwort - *abort*) sei.
2. Phase: Nachdem alle Bestätigungsantworten beim Koordinator eingetroffen sind (oder eine Zeitüberschreitung eingetreten ist), entscheidet der Koordinator, die Transaktion zu akzeptieren oder zu verwerfen. Eine Bestätigung (*commit*) bzw. ein Abbruch (*abort*) wird allen Partnerknoten mitgeteilt. Ein einziger Fehler oder die Zeitüberschreitung reichen aus, um die Transaktion als Ganzes abzulehnen.

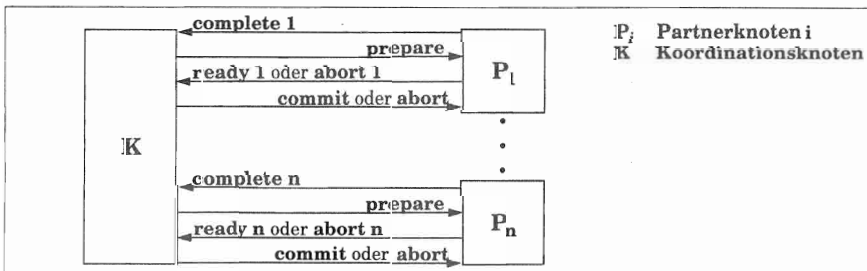


Abb. 2.1 Zweiphasenbestätigung

Das obige Verfahren arbeitet auch dann zufriedenstellend, wenn Fehler in den Partnerknoten auftreten. Der Nachteil des Verfahrens liegt darin, daß ein Fehler des Ko-

ordinationsknotens eventuell zu einem Blockierzustand der Partnerknoten führen kann /Rayn88/. Um diesem Nachteil zu entgehen, wird die Dreiphasenbestätigung angewendet /Bern87/:

1. Phase ist gleich mit der Zweiphasenbestätigung.
2. Phase: Wenn der Koordinationsknoten von allen Partnerknoten eine Bereit-Antwort erhalten hat, sendet er eine Vorbestätigung an alle beteiligten Partnerknoten. Daraufhin senden wiederum alle Partnerknoten eine Vorbestätigungsantwort an den Koordinator.
3. Phase: Der Koordinator wartet einen vordefinierten Prozentsatz von Vorbestätigungsantworten ab und sendet daraufhin endgültig eine Bestätigungsantwort an alle Partnerknoten.

In dieser Arbeit wird bei der Implementierung des Applikationsprotokolls auf die Bedeutung der Reihenfolge der Transaktionen für die Leistungsfähigkeit des Systems näher eingegangen, wobei die Aspekte der Transaktionsbestätigung nicht betrachtet werden. Diese gehören zur Eigenschaft der DDB und können somit als unabhängig von den jeweiligen Datenverwaltungsalternativen angesehen werden.

2.1.2.5 Gleichzeitigkeitsüberwachung

Wenn die Daten jeweils nur als Hauptkopie vorhanden sind, so ist die Gleichzeitigkeitsüberwachung auf die Prozeduren begrenzt, die aus den Einknotendatenbanken bekannt sind (siehe Kapitel 2.1.1).

Wenn die Daten jedoch dupliziert wurden, so müssen Methoden angewendet werden, um die Überwachung während der Modifizierungsvorgänge zu gewährleisten: Entweder wird ein zentraler Koordinator oder ein verteilter Koordinator eingesetzt.

Bei Anwendung des zentralen Koordinators hat dieser die Aufgabe, die Datenfreigabe und die Datensperrung zentral zu kontrollieren. Da in einer solchen Lösung alle Abfragen an den zentralen Koordinator geleitet werden, ist zu erwarten, daß er zum Engpaß des Systems wird. Deswegen wird diese Lösung nur selten angewendet /Bern87/.

Der Weg, die Restriktionen eines zentralen Koordinators aufzuheben (aber auch gleichzeitig die Verklemmungsgefahr herbeizuführen), ist die Anwendung eines verteilten Koordinators. Die Aufgaben des verteilten Koordinators können auf verschiedene Arten gelöst werden:

- der Mehrheitsansatz - wenn eine Modifizierungsoperation durchgeführt werden soll, wird die Mehrheit der Datenkopien gesperrt,
- der Präferenzansatz - wenn eine Leseoperation durchgeführt werden soll, wird die geteilte Sperrung angewendet; wenn eine Modifizierungsoperation durchgeführt werden soll, wird die exklusive Sperrung angewendet,
- der Hauptkopieansatz - wenn eine Modifizierungsoperation auf einer Kopie durchgeführt werden soll, wird immer nur die Hauptkopie gesperrt.

Im Zusammenhang mit der Gleichzeitigkeit sei hier noch auf zwei Methoden hingewiesen, wie in einer verteilten Umgebung eindeutige Zeitmarken spezifiziert werden. Die erste Methode bietet als Lösung eine Kombination aus einer logischen Knotennummer und der Zeit an, die zu einer globalen Zeitmarke zusammengefaßt werden. Die zweite Methode, die bei Zugriffen auf kritische Abschnitte verwendet wird, führt zusätzlich zu der Zeitüberwachung die sog. lokalen Zähler und Zeitstempel ein. Mit deren Hilfe werden dann die Prioritäten der Zugriffe während einer Transaktion ermittelt.

In der UMTS-Datenbank ist die Problematik der Gleichzeitigkeitsüberwachung von keiner essentiellen Bedeutung. Es werden zwar Bemühungen unternommen, die Konsistenz so weit wie möglich zu bewahren (Präferenzansatz). Es ist jedoch nicht das primäre Ziel, alle Möglichkeiten der Inkonsistenz auszuschließen. Aufgrund der hohen Anzahl der Modifikationsoperationen wäre der reguläre Aufwand, der für eine absolute Konsistenz notwendig ist, nicht durch den Auftritt einer wenig wahrscheinlichen Inkonsistenz zu rechtfertigen. Der aufgrund der eventuellen Inkonsistenz nötige Zusatzaufwand wird in Kauf genommen, um im Normalfall (bei konsistenten Daten) über eine Datenbank hoher Leistungsfähigkeit zu verfügen.

2.1.2.6 Datenbankverklemmungen

Zwei Arten von Verklemmungen sind nach /Kort91/ zu unterscheiden:

- lokale Verklemmungen und
- globale Verklemmungen.

Die lokalen Verklemmungen sind ein Resultat einer widersprüchlichen Information in einem Knoten und relativ einfach zu erkennen.

Die zweite Art der Verklemmungen resultiert aus einer global widersprüchlichen Information, die jedoch aus der Sicht eines einzelnen Knotens korrekt ist - die Überlagerung in mehreren Knoten führt dann zu einer Verklemmung /Rayn88/. Dieses Problem ist schwieriger zu lösen, und seine Erkennung bereitet erhebliche Schwierigkeiten, wenn nicht auf eine zentrale Überwachung zurückgegriffen wird. Bei einer verteilten Überwachung wird der Verklemmung dadurch vorgebeugt, indem man bei Transaktionen zwischen Knoten die externen Transaktionsrepräsentanten in den jeweiligen Knoten hinzufügt. Dadurch wird im Falle der gegenseitigen Abhängigkeit in jedem Knoten das Wissen über eine potentielle Zwischenknotenverklemmung existieren und eine Erkennung ermöglicht.

Bei den Untersuchungen der Leistungsfähigkeit der Datenbank werden die Aspekte der Verklemmung nicht explizit behandelt: die Protokolle, die im Kapitel 6 präsentiert werden, wurden in solcher Art und Weise aufgebaut, daß eventuelle Verklemmungen in einem fehlerfreien Fall nicht entstehen können. Zu einer Verklemmung kann es lediglich kommen, wenn eine *Information* falsch abgespeichert worden ist. Solche Fehler werden erkannt und durch eine erneute Aktualisierung ausgebessert. Bei der Leistungsbewertung werden jedoch solche Zwischenfälle ausgeschlossen.

2.2 Datenbanken in der Telekommunikation

Eines der Hauptprobleme der modernen Telekommunikation ist die Unterstützung der Personenmobilität, d.h. die Möglichkeit der logischen Trennung des aktuellen Standortes eines Benutzers und dessen Benutzernummer. Im gewöhnlichen Telekommunikationsfestnetz (PSTN - *Public Switched Telephone Network* oder ISDN - *Integrated Services Digital Network*) ist die Benutzernummer gleichzeitig die Leitwegnummer und als solche an das festinstallierte Endgerät gebunden. Dies soll im künftigen Konzept von UMTS, aber z.B. auch in UPT (/Alna94/ und /UPT1/) oder PCS (/T1P1/) fallengelassen werden. Die Benutzernummer wird zu einer persönlichen, rein logischen und nach Möglichkeit lebenslangen Nummer.

In diesem Kapitel werden die existierenden Datenkonzepte dargestellt, die sich des Problems der Personenmobilität und der personenbezogenen Benutzernummer annehmen. Andere Aspekte der künftigen mobilen Netze, wie die Personalisierung der Benutzerdaten oder der globalen Unterstützung der Endgerätemobilität, sind mit der Personenmobilität eng verbunden und können darauf aufbauend untersucht werden.

Die Personenmobilität und die personenbezogene Benutzernummer stellen folgende Anforderungen an das Netz:

- die Erreichbarkeit des Benutzers verlangt nach einer Möglichkeit, seinen aktuellen Standort während des Anrufes herauszufinden, und
- die durch die Mobilität bedingten Änderungen des Benutzerstandortes verlangen, daß das Netz die Überwachung des aktuellen Standortes beibehält.

Die dafür notwendigen Personendaten müssen im Netz gespeichert, während des Anrufes ausgelesen und in die physikalische Adresse übersetzt werden sowie nach einer Bewegung aktualisiert werden.

Im folgenden Unterkapitel werden grundsätzliche Aspekte der Datenverwaltung genauer diskutiert und im Kapitel danach verschiedene, heute gebräuchliche Konzepte vorgestellt.

2.2.1 Allgemeines zur Datenverwaltung

Die Datenverwaltung kann auf verschiedene Art gelöst werden, wobei zwei Punkte von besonderer Bedeutung sind (vgl. /Awer91/):

- die Informationsverwaltung und
- die Datenbankstruktur.

Aus der Sicht der Informationsverwaltung können drei Realisierungsmöglichkeiten unterschieden werden /Awer91/:

1. vollständige Information

Die Datenbank enthält eine vollständige Information über den aktuellen Standort aller Benutzer. Die Strategie verlangt bei einer verteilten Datenbank einen

erheblichen Aufwand, wenn alle Datenknoten die Bewegungen der Benutzer erfassen sollen. Der Aufwand für die Übersetzung der logischen Benutzernummer in die Leitwegnummer ist hier minimal.

2. keine Information

Bei dieser Möglichkeit ist die Speicherung der Standortdaten überflüssig, weil keine Übersetzung der Benutzer- in die Leitwegnummer durchgeführt wird. Der Aufwand für die Erfassung der Benutzerbewegung ist somit gleich Null. Dagegen ist der Aufwand, um einen Anruf aufzubauen, sehr groß, da der Angerufene erst gefunden werden muß, z.B. mit dem *paging*-Verfahren.

3. Teilinformation

Hier wird eine Mischung aus den beiden dargestellten Formen angewendet, die sich natürlich nur für verteilte Datenknoten eignet. Der Aufwand, der für die Erfassung der Benutzerbewegungen notwendig ist, ist geringer als bei der 1. Möglichkeit, aber dafür kann die Übersetzung der Nummer nicht in jedem Knoten durchgeführt werden. Im Vergleich mit der 2. Möglichkeit ist der Erfassungsaufwand größer, aber dafür kann die Übersetzung der Nummer in einigen Knoten durchgeführt werden.

Aus den obigen Erklärungen kann abgeleitet werden, daß sich sowohl die 1. als auch 2. Möglichkeit in bestimmten Situationen als vorteilhaft erweisen werden: Im Falle einer Lösung mit einem zentralen Datenknoten oder Szenarien mit sich selten bewegenden Benutzern ist die 1. Möglichkeit besser. Umgekehrt, im Falle von sich häufig bewegenden Benutzern und wenn die Anzahl der Benutzer (oder deren räumliche Verteilung) nicht durch einen zentralen Datenknoten unterstützt werden kann, ist die 2. Möglichkeit vorzuziehen. Aus dem Gesagten kann somit geschlossen werden, daß die 3. Möglichkeit einen Kompromiß darstellt, der sowohl in der einen als auch in der anderen Situation angewendet werden kann, ohne dabei das Optimum zu bieten.

Aus der Sicht der Datenbankstruktur können drei Techniken verwendet werden:

1. zentrale Datenbank

Die Benutzerdaten werden zentral gespeichert, was nicht notwendigerweise bedeuten muß, daß nur ein physikalischer Knoten vorhanden ist. Werden mehrere Knoten verwendet, so wird in jedem Knoten eine Kopie abgelegt. Alle Kopien müssen in solchem Fall synchronisiert werden.

2. verteilte Datenbank

Die Benutzerdaten werden unabhängig voneinander aufgeteilt und in mehreren Knoten abgelegt, z.B. nach geographischen Merkmalen oder nach den zuständigen Dienstleistern.

3. Hybriddatenbank

Es existiert ein zentraler Datenknoten, der die Information über alle Benutzer speichert. Zusätzlich sind Teile der Daten über das Netz verstreut.

Wenn man die oben aufgeführten Informationsverwaltungsmöglichkeiten mit diesen Techniken in Verbindung bringt, dann kann man feststellen, daß die vollständige In-

formation sich der zentralen oder hybriden Technik bedient. Das Konzept der Teilinformation bedient sich somit folgerichtig dem Konzept der verteilten Datenbank.

Das UMTS-Datenkonzept (vgl. Kapitel 2.4) gehört nach obiger Klassifikation zur letzten Kategorie, d.h. es bedient sich der Teilinformation unter Verwendung der verteilten Datenbank.

2.2.2 Konzepte der Datenverwaltung

Acht verschiedene Verwaltungskonzepte (im folgenden auch als Techniken bezeichnet) können heute in der Literatur gefunden werden. Sie werden hier kurz dargestellt, damit die Rolle und die Stellung des UMTS-Konzeptes, das in den zwei folgenden Kapiteln eingeführt wird, besser verstanden werden kann.

1. Broadcasting von einer zentralen Datenbank und lokale Nummeranpassung

Dieser Ansatz wurde von Bellcore in einem System, Datacycle genannt (/Bowe91/ und /Bowe92/), entwickelt. Er ist dadurch charakterisiert, daß nur ein zentraler Datenknoten existiert, der die Information über alle Benutzer speichert. Um die Verzögerung zu vermeiden, die sich aus der Überlastung durch die zentralen Operationen ergeben könnten, werden nur die Modifizierungsoperationen auf den zentralen Daten durchgeführt. Die Übersetzung der Benutzernummer und des Benutzerstandortes (Nummeranpassung) wird dagegen lokal durchgeführt, indem die zentrale Datenbank alle Daten regelmäßig, z.B. 2-3 mal pro Sekunde, an alle lokalen Knoten sendet und so mit der aktuellen Information versorgt.

Diese Lösung ist an folgende Anforderungen geknüpft:

- Eine Breitbandverbindung zwischen dem zentralen und den lokalen Knoten muß vorhanden sein.
- Die Datenmenge des zentralen Knotens darf nicht zu groß sein, da sonst das regelmäßige Senden zu einem erheblichen Aufwand führt.

Daraus folgt, daß diese Methode gut geeignet ist, wenn es sich nur um die Übersetzung von der Benutzer- in die Standortnummer handelt, aber nicht angemessen ist, wenn die Benutzerdaten mehrere kByte umfassen und die Größe des Systems einige Hunderttausende von Benutzern überschreiten sollte.

2. Anpassung im zentralen Knoten

Dieser Ansatz ist eine übliche Methode, wenn die Anzahl der Benutzer gering ist, so daß alle Übersetzungen zentral gehandhabt werden können, und/oder wenn die Benutzer nicht beweglich sind, so daß die Modifikationen sich lediglich aus der Aktualisierung der Daten infolge von statischen Änderungen ergeben.

Beispiele dafür sind die 800-Dienste in USA /Bell93/, die auf der Grundlage des SS7-Netzes arbeiten /Boyl90/, oder die 0130-Dienste in Deutschland, die sowohl mit PSTN als auch mit ISDN arbeiten.

Der Nachteil einer solchen Lösung ist wegen seiner mangelnden Skalierbarkeit offensichtlich - der Zentralknoten kann zum Engpaß des Systems werden.

3. Hierarchische und verteilte Adressierung

Eine hierarchische, auf einer verteilten Datenbank aufgebaute Technik wurde in den neueren Ansätzen der mobilen Netze der dritten Generation vorgeschlagen (siehe für FPLMTS /WanJ93/ und für UMTS /Mone14/).

Charakteristisch für diese Lösung ist die Unterteilung der Datenbankstruktur in zwei Knotentypen: die Datenknoten, die die aktuellen Benutzerdaten speichern, und die Verzeichnisknoten, die die Zeiger zu den Benutzerdaten in den Datenknoten verwalten. Die Gesamtstruktur hat einen baumähnlichen Charakter.

Vorteilhaft für diese Art der Architektur ist deren Skalierbarkeit und Möglichkeit, auch große Mengen der Benutzerdaten zu speichern. Zusätzlich schützt hier die Last- und Datenverteilung vor übermäßig langen Wartezeiten.

Die einzelnen Datenknoten werden klein und damit auch preiswert sein.

Abhängig davon, wie die Struktur des Verzeichnisbaumes spezifiziert ist, kann man hier vom Konzept der Vollinformation (ein globaler Baum wird angewendet) oder der Teilinformation (mehrere, kooperierende und unabhängige Bäume werden eingesetzt) sprechen.

4. Globale Datenbank mit vielen lokalen Datenknoten

In dieser Lösung gibt es zwei Arten von Datenknoten: den zentralen Hauptknoten und die lokalen Regionalknoten (vgl. /Chan93/). Der zentrale Knoten enthält die Information über alle Benutzer, wogegen die lokalen Knoten nur Informationen über die sich im lokalen Gebiet befindenden Benutzer enthalten. Somit enthält die Regionaldatenbank eine Untermenge der Daten aus dem Hauptknoten.

Die Übersetzung der Benutzernummer in die Standortnummer während des Anrufs erfolgt in zwei Schritten. Im ersten wird die Abfrage lokal verarbeitet und erst im zweiten Schritt, wenn die Daten lokal nicht zu finden sind, die Abfrage an den zentralen Knoten gerichtet. Eine solche Politik spart in den meisten Fällen die globale Übersetzung, da der Erfahrung nach die meisten Anrufe lokal durchgeführt werden. Es ist ebenfalls möglich, in den lokalen Knoten zusätzlich die Übersetzungstabellen von den am häufigsten angerufenen Benutzern von außerhalb der Region zu verwalten.

Wenn der Benutzer sich bewegt, wird eine Aktualisierung sowohl lokal als auch global durchgeführt.

Diese Technik, die auf dem Konzept der Vollinformation mit hybriden Datenbanken aufgebaut ist, bietet Vorteile, wenn die Anzahl der Bewegungen im Vergleich zur Anzahl der Anrufe viel kleiner ist.

5. Lokale Datenbankaktualisierung und Senden an alle Knoten zwecks Nummerübersetzung

Das Netz wird in mehrere Regionen aufgeteilt und jeder Region eine lokale Datenbank zugewiesen. Die lokale Datenbank speichert die Information über die Benutzer, die sich aktuell in dieser Region befinden /Mura89/.

Während eines Anrufes wird zuerst in der lokalen Datenbank überprüft, ob die Daten des Angerufenen sich dort befinden. Wenn nicht, dann wird eine Nachricht an alle sonstigen Datenbankknoten geschickt mit der Anfrage nach der Übersetzung. Der Datenbankknoten, der die geforderte Information enthält, führt die Übersetzung durch und schickt die Standortnummer an die anfragende Datenbank zurück.

Diesen Ansatz in einer realisierten Applikation benutzt das Konzept von TCP/IP (*Transmission Control Protocol/Internet Protocol*) Adressierung und wurde in /Come88/ beschrieben. Ein ähnlicher Ansatz wurde für das Mobilitätsmanagement für das Internet in /Ioan92/ vorgeschlagen.

Aus der prinzipiellen Sicht ist dieser Ansatz für hochmobile Benutzer gut geeignet, da hier die Aktualisierung sehr einfach durchzuführen ist. Der Prozeß des Suchens der Benutzerdaten ist jedoch unter Umständen recht aufwendig, besonders wenn die Anzahl der lokalen Knoten ansteigt.

6. Lokale Datenbankaktualisierung und Anpassung in einem entfernten, aber lokalen Datenknoten

Ähnlich wie im 5. Ansatz wird hier das Netz in Regionen aufgeteilt und jeder Region eine Datenbank zugewiesen. Die lokale Datenbank speichert jedoch nicht nur die Information über Benutzer, die sich aktuell in der Region befinden, sondern auch Daten der Benutzer, die sich dort permanent aufhalten, d.h. in dem Gebiet ihr Zuhause haben (/Awer91/ und /WanD93/). Die Benutzernummer enthält im Gegensatz zu den bisherigen Konzepten eine Information darüber, welche Datenbank den Permanentknoten des Benutzers darstellt.

Während eines Anrufes wird zuerst in der lokalen Datenbank überprüft, ob die Daten des Angerufenen sich dort befinden. Sind sie nicht dort vorhanden, wird eine Nachricht an den permanenten Knoten geschickt, der die Übersetzung der Benutzer- in die Standortnummer vornimmt und an die anfragende Datenbank zurückschickt. Dies ist offensichtlich vorteilhafter gegenüber der Technik, die im Ansatz 5. verwendet wird.

Die Aktualisierung erfolgt immer gleichzeitig sowohl im lokalen als auch im permanenten Knoten, was im Vergleich mit der 5. Technik mit einem höheren Aufwand verbunden ist.

Die Hauptschwierigkeit dieses Ansatzes liegt in der Annahme über die Relation zwischen der Benutzernummer und dem permanenten Knoten. Eine solche Lösung schließt die persönliche Nummer aus, da diese z.B. nach einem Umzug des Benutzers geändert werden müßte, um die Relation zur permanenten Datenbank wieder herzustellen.

7. Propagationstechnik

In dieser Technik, die sich besser für die Verwaltung von Computernetzen und der Mobilität zwischen ihnen eignet als für Telekommunikationsnetze, wird die Information über den Benutzerstandort im lokalen, permanenten Netz gespeichert. Wenn der Benutzer sich aus dem lokalen Netz herausbewegt, wird die Information darüber sowohl in das neue Netz als auch in alle Netzübergänge auf dem Weg von der alten zu der neuen Lokation eingetragen (/Tera91/ und /Tera93/).

Wenn jetzt ein Anruf gestartet wird, so wird die Übersetzung entweder lokal durchgeführt oder, wenn das lokale Netz keine Information über den Angerufenen besitzt, der Anruf aufgrund der Benutzernummer zum permanenten Standort geleitet. Auf dem Wege dorthin befindet sich mindestens ein Netzübergang, der darüber informiert ist, wo sich der Angerufene aktuell befindet. In diesem Übergang wird die gesuchte Benutzernummer erkannt und der Anruf von dort an die richtige Adresse weitergeleitet.

Der Nachteil dieser Technik ist wie im Konzept 6 durch die feste Beziehung zwischen der Benutzernummer und seinem permanenten Standort bedingt. Darüber hinaus verlangt diese Technik zusätzliche Speicherung der Information in den Netzübergängen.

8. Verteiltes HLR (*Home Location Register*) und VLR (*Visited Location Register*)

Der Ansatz, der sowohl im amerikanischen (IS-54) als auch im europäischen (GSM) digitalen Zellulernetz verwendet wird, stützt sich auf das Konzept von zwei Datenknotentypen:

- HLR-Knoten, der die Information über die permanenten Benutzer eines Gebietes enthält, und
- VLR-Knoten, der die Teilinformation über die lokalen Benutzer in einem Gebiet speichert.

Darüber hinaus kontrolliert ein HLR-Knoten mehrere VLR-Knoten, d.h., der HLR-Knoten enthält einen Zeiger auf den aktuell zuständigen VLR-Knoten, in dessen Versorgungsbereich sich der Benutzer befindet. Erst aus dem VLR-Knoten wird der aktuelle Standort des Benutzers ausgelesen (vgl. Kapitel 2.3.1).

Die hier angewendete Lösung ist für mobile Benutzer optimiert, erlaubt aber keine personenbezogene Nummer, da die Benutzernummer immer auf den jeweiligen HLR-Knoten bezogen ist. Darüber hinaus ist die Skalierbarkeit der Lösung nicht unbegrenzt, da die Anzahl der VLR-Knoten pro HLR-Knoten aus implementierungstechnischen Gründen nicht zu groß sein darf.

2.2.3 Zusammenfassung der Datenbankanwendungen

Die in diesem Kapitel vorgestellten Techniken geben einen Überblick über die Situation der Datenverwaltungskonzepte heutiger Kommunikationsnetze. Dies erleichtert dem Leser, das im Kapitel 3.1 dargestellte UMTS-Konzept mit existierenden Techniken zu vergleichen.

Abschließend werden die vorgestellten Techniken mit Hilfe der eingeführten Datenorganisationsoptionen in einer Tabelle dargestellt:

Tabelle 2.1 Datenverwaltungskonzepte

	Vollinformation	Teilinformation
zentrale Datenbank	(1), (2)	
verteilte Datenbank	(3)	(5), (6), (7), (8)
hybride Datenbank	(4)	

2.3 Mobile und personalisierte Systeme von heute

Die vorliegende Arbeit beschäftigt sich mit den Datenverwaltungsaspekten des UMTS-Systems. So ist es zweckmäßig, die heutigen Systeme aus dem Blickwinkel

der Datenverwaltungsaspekte kurz zu analysieren. Im einzelnen werden folgende Systeme und Ansätze näher betrachtet:

- **Global System for Mobile Communication** (GSM)
- **Intelligent Network** (IN)
- **Universal Personal Telecommunication** (UPT)
- **Personal Communication System** (PCS)
- **Personal Service Communication Space** (PSCS)
- andere

Die hier gegebenen Systembeschreibungen haben aus Platzgründen nur allgemeinen Charakter und beschränken sich auf die Grundeigenschaften der jeweiligen Systeme. Sie sollen lediglich die Ähnlichkeiten und Unterschiede zu UMTS erkennen lassen. Für detaillierte Information sei auf die entsprechende Literatur verwiesen.

2.3.1 Global System for Mobile Communication

Das GSM-System ist ein mobiles, zellulares System, das in seiner Spezifikation dem Benutzer eine uneingeschränkte, funkbasierte Sprach- und Datenkommunikation im Deckungsbereich bietet (/GSM-Spec/ und /Mou92/). Es kann als eine digitale Weiterentwicklung der früheren analogen Systeme, wie NMT und TACS (Europa), RTMS (Italien), C-450 (Deutschland und Portugal), ETACS (Großbritannien) angesehen werden (/Padg95/ und /EIA91/).

2.3.1.1 Allgemeine Übersicht

GSM hat folgende Grundeigenschaften:

- gute zellulare Deckung des Gebietes im Freigelände, jedoch für Innenanwendungen wegen ungeeigneter Leistungsparameter und aufwendiger Anrufübergabe unwirtschaftlich (nach /Mall88/ ist die minimale empfohlene Zellengröße ca. 300 m)
- ausgereiftes Anrufübergabekonzept im Freigelände, das auch bei hohen Bewegungsgeschwindigkeiten eingesetzt werden kann (z.B. beim Autofahren)
- begrenzte Benutzung der Datendienste neben der Sprache (Die Bandbreite des verfügbaren TDMA-Kanals beträgt insgesamt 25 kHz, wobei die Nutzbitrate 9.6 kbit/s beträgt)
- Möglichkeit von Kurzmeldungen unabhängig von der bestehenden Anrufverbindung
- gute Sicherheitsvorkehrungen und ausgereifte Authentifizierungsverfahren
- Benutzernummer ist an den HLR-Datenknoten des permanenten Aufenthaltes gebunden
- sehr begrenzte Personalisierung der Dienste möglich, z.B. eine Zugangsberechtigungsüberwachung mit Hilfe einer SIM-Karte ist vorgesehen

- keine Unterscheidungsmöglichkeit zwischen dem Teilnehmer und dem Benutzer (vgl. Kapitel 3.1.1)
- in wenig bevölkerten Gebieten ist eine sehr gute Deckung leicht erreichbar wegen der Möglichkeit, große Funkzellen anzuwenden (Zellen bis ca. 35 km sind nach /Moul92/ möglich)
- ein separates Teilnetz ist notwendig
- Schnittstelle zu den bestehenden festen und mobilen Netzen werden unterstützt
- gute Verwaltung des Funkkanals (TDMA mit Frequenzsprungverfahren)
- übersichtliche funktionale Architektur (siehe Abbildung 2.2)

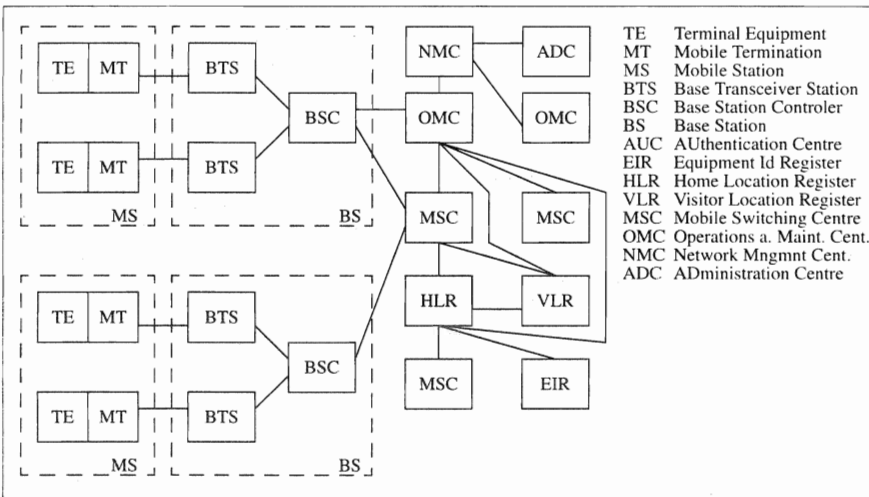


Abb. 2.2 GSM-Architektur /Stee92/

Es sei hier betont, daß GSM gleichzeitig ein Systemansatz und ein selbständiges Netz ist, das zusätzlich zu der eigenen Funktionalität dem mobilen Benutzer einen Zugang zum bereits vorhandenen Festnetz bietet.

2.3.1.2 Datenkonzept

Das GSM-System verfügt über einen sehr gut konzipierten Ansatz der Datenhaltung, der sich auszeichnet für die Mobilitätsproblematik eignet. Dieser Ansatz ist realisiert unter Anwendung von zwei im Prinzip hierarchisch angeordneten datenbankähnlichen Knotentypen, von denen im vorherigen Kapitel schon in einem anderen Zusammenhang die Rede war:

- **Home Location Register (HLR)** und
- **Visited Location Register (VLR).**

Der HLR-Knoten speichert permanent die Teilnehmerdaten (entspricht den Benutzerdaten, da hier der Teilnehmer dem Benutzer äquivalent ist), die die Benutzung der angebotenen Dienste ermöglichen. Der Status des Benutzers (z.B. die von ihm in Anspruch genommenen Dienste) und ein Zeiger zum jeweiligen VLR-Knoten, in dem sich der Benutzer aktuell befindet, sind dort beispielsweise gespeichert.

Im VLR-Knoten sind die temporären Informationen gespeichert, wie z.B. der Status des gerade durchgeführten Anrufes oder das Aufenthaltsgebiet, in dem sich der Benutzer befindet.

Die beiden Knotentypen, HLR und VLR, tauschen miteinander die Daten aus, wobei der HLR-Knoten den VLR-Knoten mit der notwendigen Information versorgt. Jedemal, wenn eine Netzaktion die Grenzen des Versorgungsgebietes eines VLR-Knotens überschreitet (z.B. ein Anruf von außerhalb oder eine Benutzerbewegung zwischen Gebieten zweier VLR-Knoten), wird der HLR-Knoten abgefragt und in den Informationsaustausch eingebunden. Wenn beispielsweise ein Benutzer angerufen wird, dann wird erst geprüft, ob der Angerufene sich im Bereich des eigenen VLR-Knotens befindet. Anderenfalls wird eine Abfrage im permanenten HLR-Knoten durchgeführt - dieser ist mittels der Benutzernummer eindeutig zugeordnet. Da der HLR-Knoten die Information darüber besitzt, in welchem Teil des Netzes sich der Benutzer gerade befindet, wird daraufhin der Anruf zum entsprechenden Vermittlungsknoten durchgeschaltet und für weitere Informationen (z.B. *paging*) der lokale VLR-Knoten abgefragt. Ähnliches tritt bei der Benutzerbewegung auf: Ändert sich während einer Benutzerbewegung der zuständige VLR-Knoten, werden nicht nur die entsprechenden Zeiger im HLR-Knoten aktualisiert, sondern auch die Daten von dem einen VLR zu dem anderen transferiert.

Obwohl GSM die Benutzer- und Endgerätemobilität unterstützt, eignet sich das Datenkonzept aus folgenden Gründen nicht für eine verfeinerte Personalisierung:

- Es gibt keine Personennummer.
- Mobilität wird mittels fester Relation zwischen dem HLR-Knoten und der Benutzernummer verwaltet.
- Eine Anpassung der Kommunikationsumgebung an die Benutzerwünsche wird nur beschränkt unterstützt, z.B. durch Blockieren unerwünschter Anrufe

2.3.2 Intelligente Netze

IN ist ein Konzept, das auf der Grundlage des konventionellen Telekommunikationsfestnetzes eine netzunabhängige Infrastruktur zur Dienststeuerung spezifiziert. IN ist im Rahmen von CCITT (heutiges ITU - *International Telecommunication Union*) SG11 und ETSI (*European Telecommunication Standards Institute*) NA 6 in einem kontinuierlichen Standardisierungsprozeß entstanden, der bis heute noch nicht endgültig abgeschlossen ist (SG11 und NA6 sind entsprechend Untergruppen von CCITT und ETSI). IN wurde entwickelt, um neue Mehrwertdienste in das bestehende Netz einführen und steuern zu können. Das Hauptmerkmal des Konzeptes ist eine konsequente Trennung der dienstspezifischen Elemente von der bestehenden

Kommunikationsinfrastruktur unter Einführung einer spezifischen IN-Infrastruktur (/Kung91/ und /Garr93/).

2.3.2.1 Allgemeine Übersicht

Das IN-Konzept ist durch folgende Merkmale charakterisiert (/LiNi93/ und /Chen93/):

- Unterscheidung von vier verschiedenen Abstraktionsebenen der Dienstspezifizierung (auch *planes* genannt):
 - die Dienstebene, die sich auf die Beschreibung der Eigenschaften eines Dienstes aus der Sicht des Benutzers konzentriert,
 - die allgemeine Funktionsebene, die sich auf die Struktur des Dienstes bezieht und dabei die dienstunabhängigen Bausteine (*Service Independent building Blocks* - SIB) benutzt,
 - die verteilte Funktionsebene, die sich auf die Verteilung der einzelnen, bereits spezifizierten logischen Funktionen im Netz konzentriert; die Funktionsebene wird den wachsenden Aufgaben mittels neuer *Capability Sets* (CS) systematisch angepaßt (vgl. Abbildungen 2.3 - 2.5); die CS-2 stellt gegenüber CS-1 offensichtlich eine bedeutend leistungsfähigere Beschreibung dar, in der die nicht-anrufbezogenen Aktionen (z.B. Anrufübergabe - *Handover*) ihren Platz finden; die Trennung zwischen der Verkehrs-, Anruf- und Dienstverbindung und ein erweitertes Endgerätkonzept stellen die nächste Stufe der IN-Weiterentwicklung dar,
 - die physikalische Ebene, die sich auf die Ressourcen und deren Benutzung im Netz konzentriert.
- eine relativ einfache Dienstspezifikation,
- Orientierung an den anrufbezogenen Netzereignissen, deshalb auch in seiner Ursprungsform nicht zur Mobilitätsunterstützung geeignet, sowie
- eine Reihe von Diensten, die eine vereinfachte Benutzung der Festnetzinfrastruktur zur Folge haben, z.B. Freianruf, logische Trennung der Benutzernummer und seiner physikalischen Lokation (auf einer permanenten Basis, d.h. nichtmobilitätsbezogen) sowie Einführung von VPN (*Virtual Private Networks*).

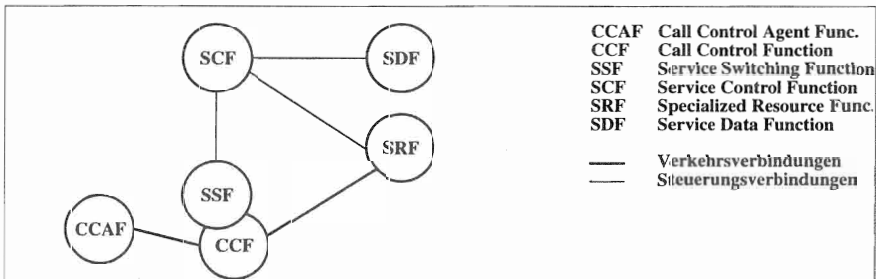


Abb. 2.3 IN CS-1 verteilte Funktionsebene (Managementfunktionalität ausgelassen) /Q.1200/

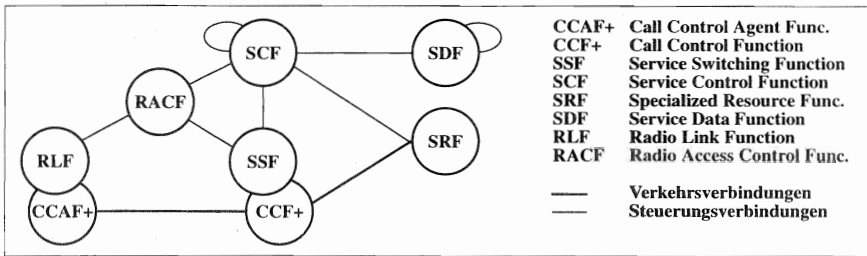


Abb. 2.4 IN/UMTS SMG5 funktionales Modell, vorgeschlagen als Ergänzung in CS-2 /IN95/

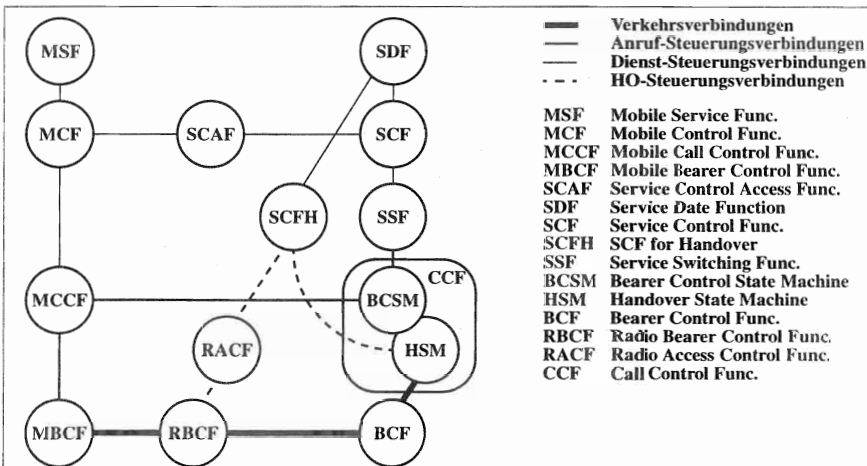


Abb. 2.5 IN CS-2 verteilte Funktionsebene (Managementfunktionalität ausgelassen) /CS-2/

Das IN-Konzept ist mit dem Ziel entwickelt worden, die Dienstverfügbarkeit im bestehenden Festnetz zu erweitern. Daraus resultiert auch die Tatsache, daß Mobilitätsaspekte im ursprünglichen IN-Konzept nicht betrachtet worden sind. Diese Einstellung zur Rolle von IN im Mobilitätsprozeß hat sich jedoch radikal geändert. Nach der heute vorherrschenden Meinung wird IN als Möglichkeit angesehen, in einer relativ konsistenten Weise sowohl die Mobilität (/Isbe94/ und /Maru94/) als auch eine Reihe von Mehrwertdiensten anzubieten (/Mart94/ und /Chen93/). Letztere sollen auch anwendbar sein, wenn der Benutzer sich in einem anderen mobilen Netz (z.B. GSM) befindet

2.3.2.2 Datenkonzept

Das Datenkonzept von IN kann bereits aus den obigen Bildern abgeleitet werden. Aus dem funktionalen Modell ist eine explizite Trennung zwischen den dienstspezifischen Funktionen (SCF) und den Daten (SDF) sichtbar.

Die Meinungen bezüglich der Aufgabe des SDF-Knotens, die in verschiedenen Literaturquellen vertreten werden, reichen von einer lokalen Datenbank, die lediglich die Funktionalität von SCF unterstützt, wie in CS-1 spezifiziert /Q.1200/, bis hin zu einem verteilten Datenbankkonzept (/Maru94/ und /CS-3/).

Generell ist jedoch festzustellen, daß IN über kein eigenes Datenkonzept verfügt. Die jeweils verwendete Lösung wird eher durch die Aufgabe (oder allgemeiner gesprochen: durch den Dienst) vorgegeben, die mittels IN gelöst werden soll.

In den letzten Jahren entwickeln die Fortschritte auf dem IN-Gebiet eine sehr starke Dynamik. Dies macht das IN-Konzept zu einem geeigneten Mittel für die Realisierung der personalisierten (/John94/, /Chon94/ und /LiNi93/), mobilen und flexiblen (/Isbe94/ und /Maru94/) Kommunikationsnetze der Zukunft.

2.3.3 Universal Personal Telecommunication

UPT steht für das Konzept einer fortschrittlichen Funktionalität in künftigen Kommunikationsnetzen und ist ein Ergebnis der Bemühungen in den Standardisierungsgremien: im NA7 innerhalb von ETSI und in einigen SG (*Sub-Group*) innerhalb ITU /UPT2/.

Folgende Schwerpunkte der UPT-Aktivitäten können beobachtet werden /UPT1/:

- Benutzermobilität,
- Personalisierung und
- Dienstmobilität.

Benutzermobilität in Unterscheidung zur Endgerätemobilität bedeutet für den Benutzer die Möglichkeit, auf der Basis der personenbezogenen Nummer den Zugang zum Netz zu erlangen sowie die Anrufe zu initiieren und zu empfangen (allgemeiner: die Dienste des Netzes in Anspruch zu nehmen). Die Kommunikation wird nach den Benutzervorgaben durchgeführt, die in einem Personalprofil abgelegt worden sind (vgl. /Gran93/).

UPT ist demnach eine Anforderungsspezifikation an das künftige Telekommunikationsnetz. Als Grundlage für die Realisierung des UPT-Netzes wird allgemein das bereits dargestellte IN-Konzept angesehen (vgl. /Laue94/).

2.3.3.1 Allgemeine Übersicht

Da UPT hauptsächlich durch eine Reihe von Dienstspezifikationen beschrieben wird, ist es angebracht, eine Übersicht über die spezifizierten Dienste zu geben. Es werden drei Gruppen von Diensten unterschieden:

1. Gruppe: Kerndienste
2. Gruppe: Zusatzdienste
3. Gruppe: Spezialdienste

Zur ersten Gruppe werden solche Merkmale gezählt, die als grundlegend für das UPT-Konzept und notwendig für die Langzeitperspektive der UPT-Entwicklung angesehen werden. Beispiele solcher Dienste sind das UPT-Dienstprofil, die personenbezogene Nummer, Netzunabhängigkeit oder Sicherheitsvorschriften. Eine vollständige Liste von Kerndiensten kann in /UPT1/ und /UPT3/ gefunden werden.

Zur zweiten Gruppe gehören solche Merkmale, die eine nicht notwendige Erweiterung von Kerndiensten bieten, wie z.B. Personalisierung der Dienstauführung oder erweiterte Benutzeradressierung (Name, Adresse usw.).

Die dritte Gruppe bildet ein Gerüst von Verfeinerungen mit solchen Merkmalen, die aus der Sicht der künftigen Benutzer wünschenswert sind, jedoch nicht den Kern des UPT-Konzepts darstellen. Dazu gehören Identifizierungs-, Anrufinitiierungs- und Anrufausführungsdienste sowie Anrufsteuerung. Eine komplette Liste der Spezialdienste kann in /UPT4/ gefunden werden.

Es wird erwartet /Chon94/, daß das UPT-Konzept mit Hilfe des IN-Ansatzes realisiert wird. Die praktische Einführung von UPT-Diensten soll in vier Phasen realisiert werden:

1. Phase: Begrenztes Kurzzeitszenario
2. Phase: Grundszenario
3. Phase: Erweitertes Szenario
4. Phase: Evolutionsszenario

Die einzelnen Phasen spezifizieren die Entwicklungsschritte der UPT-Funktionalität und berücksichtigen gleichzeitig den Anpassungsprozeß des Telekommunikationsnetzes an die UPT-Anforderungen.

2.3.3.2 Datenkonzept

Das Datenkonzept in UPT besteht hauptsächlich aus einer Spezifikation des Benutzerprofiles, seiner Zugreifbarkeit und Modifizierbarkeit.

Diese Spezifizierung unterstreicht das zugrundeliegende Konzept und ermöglicht eine benutzerangepaßte Dienstinanspruchnahme /UPT5/.

Trotz früherer Bemühungen, ein hierarchisches und verteiltes Datenkonzept zu entwickeln (/Fuji90/), ist bis heute nicht eindeutig spezifiziert, welche Lösungen angewendet werden sollen, um die UPT-Datenbank zu realisieren. Dies wird sicherlich in Zukunft zum Aufgabenbereich gehören, der durch das unterliegende Konzept gelöst werden muß, auf dem das UPT-Netz aufgebaut wird.

2.3.4 Personal Communication System

PCS ist ein Sammelbegriff, der verschiedene Formen der schnurlosen Kommunikation mit dem UPT-Ansatz zu verbinden versucht (/Woin94/) und gleichzeitig die Forde-

rung nach einer Multioperator- und Multidienstanbieterumgebung unterstützt. PCS fordert folgende Funktionalität vom künftigen Netz (vgl. /Lawn94b/):

- Dienstvielfalt und -integration:
 - Sprache
 - Daten
 - bewegte und unbewegte Bilder
- Personalisierung von Diensten im Sinne von Benutzerspezifikation, ähnlich wie in UPT,
- Endgeräte- und Benutzermobilität, die sowohl im zellularen (außen) als auch lokalen (innen) Bereich nach gleichen Prinzipien ausgeführt wird.

PCS ist im Standardisierungsprozeß eingebettet, der von zwei Institutionen betrieben wird: T1.P1 ist für die Festnetzaspekte und TIA TR 46 für den Funkzugang sowie die Frequenzvergabe zuständig (vgl. /Woin94/). T1 ist ein im Rahmen von ANSI (*American National Standards Institute*) gegründetes Gremium mit dem Ziel, Telekommunikationsstandards zu entwickeln. T1.P1 ist eine der Untergruppen, die sich mit den PCS-Aspekten beschäftigt. TIA TR 46 ist ein nationales Gremium, das sich mit der Vergabe der Funkfrequenzen in den USA beschäftigt.

2.3.4.1 Allgemeine Übersicht

Die Spezifikation von PCS wird ähnlich zum UPT-Konzept aufgebaut, d.h. als eine sehr gut ausgearbeitete Liste von Anforderungen an das künftige Netz. Im Prinzip soll das PCS-Netz eine schnurlose Kommunikation für jeden zu jeder Zeit und von jedem Ort mit verschiedenen Endgeräten und unter Benutzung von personalisierten Merkmalen ermöglichen /Wizg94/.

Wegen der Variabilität der Bandbreitenanforderungen durch unterschiedliche Dienste ist auf der Funkseite die Benutzung sowohl von Schmalband- als auch Breitbandkanälen denkbar /Woin94/.

Um dem Konzept der globalen Dienstverfügbarkeit gerecht zu werden, wird die Nutzung von Satellitensystemen sowohl in *low orbit* als auch *medium orbit* (/Wild94/, /Dini94/, /Dean94/ und /Loui94/) eine wichtige Rolle spielen.

Nach dem heutigen Stand der Standardisierung können die PCS-Dienste wie folgt klassifiziert werden:

- Teledienste - Sprache, Faksimile der 3. Gruppe, Datendienste (bis 144 kbit/s) und deren Zusammenarbeit,
- Spezialdienste - eine Reihe von Zusatzfunktionen, die aus der Benutzerperspektive eine Art von Mehrwertdiensten darstellen und eine einfache Handhabung und Personalisierung von Telediensten erlauben.

In Ergänzung zu den allgemeinen Anforderungen an das PCS-Konzept werden im Prozeß der Standardisierung solche Aspekte wie z.B. die Minimalfunktionalität eines PCS-Netzes festgelegt.

Die Spezifizierung eines einheitlichen funktionalen Netzmodells ist bis jetzt nicht durchgeführt worden. Beispielsweise sind zwei unterschiedliche physikalische Referenzarchitekturen von T1.P1 und von TR 46 vorgeschlagen worden. Die erste geht vom IN-Ansatz aus (/Fara94/ und /Hans94/), und die zweite nimmt als Basis die Ansätze, die aus dem zellularen Bereich bereits bekannt sind /Wizg94/.

2.3.4.2 Datenkonzept

Es ist zur Zeit kein selbstständiges Konzept für die PCS-Datenhaltung vorgesehen. Wie bereits gesagt, sind zwei unterschiedliche Ansätze vorgeschlagen worden. Es ist zu erwarten, daß die künftigen Datenkonzepte ebenfalls unterschiedliche Wege gehen und dabei Ähnlichkeiten mit den beiden Netzansätzen aufweisen werden.

2.3.5 Personal Service Communication Space

PSCS ist ein neues Konzept, das auf dem Gedankengut von UPT aufbaut und dieses erweitert. Der Benutzer wird hier in das Zentrum der Betrachtung gestellt. Der um ihn aufgebaute Kommunikationsraum (auch *Communication Space* genannt) soll dem Benutzer in seinen verschiedenen Rollen die Funktionalität bieten, nicht nur zu kommunizieren, sondern auch die Kommunikation nach seinen Präferenzen und unter Beachtung von Zeit, Raum, Transportmedium, Kosten, Integrität, Sicherheit, Qualität, Zugriffsmöglichkeit und Privatsphäre zu gestalten (/Mobi4/ und /Nieb94/).

Das PSCS-Konzept ist im Rahmen des RACE-Projekts Mobilise R2003 entstanden. Seine Weiterführung wird im Rahmen des ACTS-Programms der Europäischen Gemeinschaft erwartet. Darüber hinaus wird erwartet, daß die PSCS-Beiträge die Weiterentwicklung der UPT-Phase 4 positiv stimulieren werden.

2.3.5.1 Allgemeine Übersicht

Die Zukunft der Telekommunikation geht von einem rapiden Anstieg der angebotenen Dienste aus. Diese Dienste werden unabhängig vom unterliegenden Netz betrieben, d.h. sowohl in festen als auch in mobilen Netzen. Unter solchen Voraussetzungen nimmt PSCS folgende, zum Teil bereits aus UPT bekannte Aspekte in Angriff:

- Persönliche Mobilität,
- Anpassung der Kommunikation an die Präferenzen des Benutzers,
- Dienstinteroperabilität, d.h. die Möglichkeit, die Dienstübersetzung durchzuführen.

Die entwickelte PSCS-Architektur beachtet die gegenseitigen Relationen zwischen den verschiedenen Marktteilnehmern, d.h. zwischen dem Netzbetreiber, dem Dienstanbieter, dem Teilnehmer (Dienstkäufer), dem Benutzer und dem Endgeräteanbieter. Darauf stützend wurde eine IN-ähnliche, jedoch hierarchisch aufgebaute Architektur spezifiziert (/Mobi9/ und /Gunt93/).

Eine Reihe von spezifischen PSCS-Diensten wurde definiert, wie z.B. Personalisierung, Benutzeradaptation oder spezifische Authentifizierung /Mobi12/.

In Kürze sollen hier die wesentlichen Neuerungen von PSCS im Vergleich zu UPT aufgelistet werden (/Mobi12/ und /Mobi16/):

- erhöhte Flexibilität der Dienstmodifikation unter der Benutzung eines weiterentwickelten Dienstprofils,
- Möglichkeit, sich an die Dienste dynamisch anzuschließen, ohne sie davor zu abonnieren,
- bessere Anpassung des Dienstcharakters durch die Beachtung aller Marktteilnehmer im Prozeß der Dienstanpassung,
- Benutzerfreundlichkeit der Schnittstelle zwischen Benutzer und Netz und
- bessere Möglichkeit der Dienstanpassung an die Endgeräteigenschaften.

2.3.5.2 Datenkonzept

Im Rahmen von PSCS wurden bis jetzt lediglich die Aspekte der Datenspeicherung und Verwaltung innerhalb eines Knotens konzipiert und zum Teil implementiert. Die Datenverwaltung in einer verteilten Umgebung eines Telekommunikationsnetzes ist bis jetzt nicht endgültig ausgearbeitet worden. Die bisherigen Erfahrungen lassen jedoch den Schluß zu, daß die künftig benutzte Lösung eine starke Ähnlichkeit mit den UMTS- und den IN-Konzepten haben wird /Mobi94/.

2.3.6 Digital European Cordless Telephone

DECT ist die europäische Weiterentwicklung der herkömmlichen analogen Systeme, wie CT1+ und CT2 (*Cordless Telephony*), und bietet lokale Mobilität und eine schnurlose Kommunikation innerhalb von räumlich abgegrenzten Gebieten wie z.B. Büros, Privatgebäuden, Flughäfen usw. (/Nied93/, /Åker92/ und /DECT92/).

2.3.6.1 Allgemeine Übersicht

Das DECT-System bietet Dienste, die die Kommunikation für einen Benutzer, der sich innerhalb eines Gebietes langsam bewegt, sehr interessant machen:

- die Möglichkeit, neben der Sprache Daten mit 32 kbit/s zu übertragen,
- die Möglichkeit, die Einzelkanäle zusammenzulegen und damit eine höhere Bandbreite der Funkverbindung zu erzielen,

- die Übergabe des Anrufes, wenn der Benutzer sich zwischen den Zellen eines Systems bewegt,
- die Personalisierung der Dienste mittels einer PIN-Karte (*Personal Identification Number*) und
- die Unterstützung des Telepoint-Konzeptes.

2.3.6.2 Datenkonzept

Das Datenkonzept ist im DECT-Ansatz sehr vereinfacht und besteht im Prinzip aus einer zentralen Datenbank, die das gesamte Gebiet eines DECT-Systems steuert. Die Datenbank steuert gleichzeitig die sog. Festpunkte (*fixed parts*) des Netzes, die sich jeweils aus einer Steuerungseinrichtung und einer Antenne zusammensetzen.

Es besteht jedoch die Möglichkeit, die Daten zwischen zwei DECT-Systemen mittels eines DECT-Überganges zu übertragen und damit dem Benutzer die Mobilität über die Grenzen eines Gebietes hinaus zu gewähren.

2.3.7 Andere Systeme

Unabhängig von den beschriebenen Grundkonzepten gibt es eine ganze Reihe von Systemen, die eine lokale Bedeutung haben. Dazu zählen z.B.:

- PDC (*Personal Digital Cellular in Japan*) - eine japanische Variante von GSM, die sich von GSM hauptsächlich durch das fehlende VLR und durch mehrere MSC (*Mobile Switching Center*) je Aufenthaltsgebiet (*Location Area* - LA) unterscheidet /Res91/,
- ADC (*American Digital Cellular*) - eine amerikanische Variante von GSM, (auch IS-54 genannt), die sich von GSM hauptsächlich durch die notwendige Integration von VLR und BSC (*Base Station Controller*) in einem MSC unterscheidet /IS-54/,
- PCN (*Personal Communication Network*) - ein Derivat des GSM-Systems, das sich im Vergleich zu GSM durch höhere Funkfrequenz (im 1.8 GHz-Band) und dadurch auch kleinere Zellen, höhere Gesamtkapazität und niedrigere Sendeleistung auszeichnet /Rams94/,
- PHS (*Personal Handyphone System*) - eine japanische PCS-Variante, die sich besonders für sich langsam bewegende Benutzer eignet und auf dem PSTN/ISDN-Netz in Verbindung mit einem erweiterten DECT-Konzept aufgebaut ist (/Hato90/, /Ogaw94/ und /Res93/).

Die heute existierende Vielfalt an Systemen, die entweder bereits kommerziell vertrieben werden oder kurz vor der Marktreife stehen, läßt die Schlußfolgerung zu, daß die verwendeten Ansätze keine endgültigen Lösungen bieten, die ausgereift genug wären, um allen Benutzern von morgen die gewünschte Funktionalität bereitzustellen. Dies soll sich mit dem neuen Konzept der digitalen Systeme der dritten Generation, wie beispielsweise UMTS, ändern.

2.4 Universal Mobile Telecommunication Systems

UMTS ist die Bezeichnung für die europäische Entwicklung eines mobilen Systems der dritten Generation. Die Spezifikation des Systems entsteht als eine gemeinsame Anstrengung vieler Industrieunternehmen im Rahmen des RACE-Programmes und befindet sich gleichzeitig im Standardisierungsprozeß innerhalb von ETSI/SMG5 (/Hans92/). Ähnliche Anstrengungen, die auf die Spezifikation eines globalen mobilen Systems der dritten Generation ausgerichtet sind, werden ebenfalls im Rahmen von ITU unternommen, wobei dort der Begriff FPLMTS (*Future Public Land Mobile Telecommunication System*) geprägt worden ist (/Lyck92/ und /Call94/).

2.4.1 Überblick

Aus der Sicht des Benutzers sind folgende Eigenschaften des UMTS-Konzeptes als wesentlicher Fortschritt gegenüber den bekannten Systemen der zweiten Generation zu nennen (/Boer93/ und /R1D67/):

- eine breite Palette der unterstützten Dienste, von Sprache und Daten bis zum Video (mit einer maximalen Bandbreite von 2 Mbit/s)
- Unterstützung der Endgerätemobilität in verschiedenen Umgebungen - zellulär, zu Hause, im Büro und in sich bewegenden Transportmitteln,
- *Roaming* und *Handover* zwischen verschiedenen Netzen und verschiedenen Umgebungen,
- eine breite Palette an Endgeräten, von billigen und einfachen bis zu ausgereiften,
- eine globale Benutzernummer, die eine personalisierte Dienstangebot ermöglicht, und
- eine strenge Unterscheidung zwischen dem Benutzer und dem Teilnehmer.

Zusammenfassend handelt es sich um ein neuartiges Konzept einer globalen Mobilität, die auf der persönlichen Nummer basiert und gleichzeitig eine begrenzte Dienst Anpassung bietet, nach den PSCS-Maßstäben gemessen (vgl. /Jans90/).

2.4.2 Das UMTS-Konzept

Die Entwicklung von UMTS geht über den Spezifikationsprozeß von Anforderungen weit hinaus (vgl. /SSD94/). Bereits heute verfügt UMTS unter anderem sowohl über eine detaillierte Beschreibung des funktionalen Modells und der Referenzkonfiguration für unterschiedliche Prozeduren (vgl. /Keat94/) als auch über eine Protokollspezifikation (vgl. /Norp94b/). Die Fortschritte auf dem Gebiet der UMTS-Datenmodellierung haben auch die Grundlage für diese Arbeit gelegt.

2.4.2.1 Das allgemeine Funktionsmodell

Das entwickelte Funktionsmodell (vgl. /Broe93/) beinhaltet viele IN-ähnliche Aspekte, was auch im ETSI/SMG5 Report /SMG5-4/ zum Ausdruck gebracht worden ist. Dies ist auch sichtbar, wenn man das allgemeine funktionale Modell für UMTS, wie in der Abbildung 2.6 dargestellt, mit dem des IN-Systems der CS-2-Architektur vergleicht.

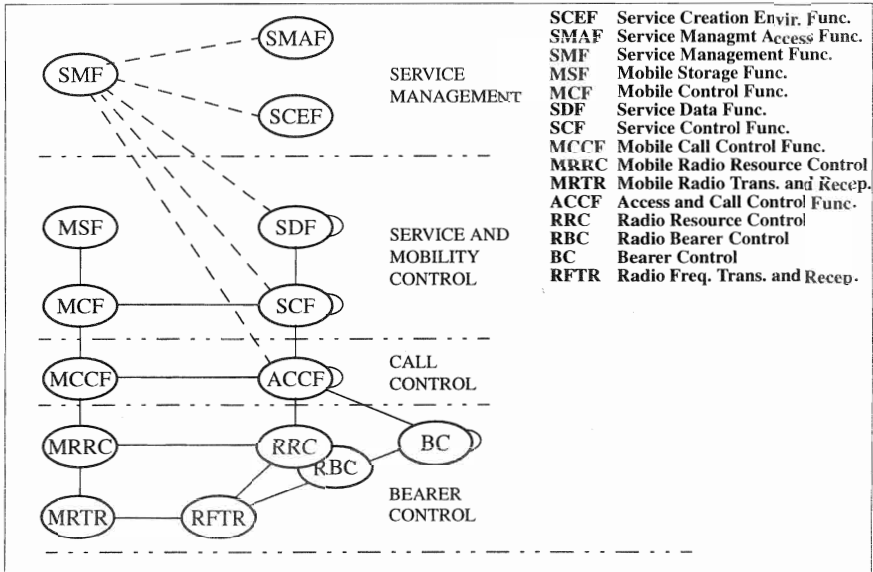


Abb. 2.6 UMTS-Funktionsmodell, ETSI SMG5 / Broe94/

Der obige Ansatz hat als erster die Trennung zwischen drei Steuerungsebenen postuliert:

- Trägersteuerung - verantwortlich für den Aufbau, Wartung und Freigabe der Funk- und Verkehrsverbindung,
- Verbindungssteuerung - verantwortlich für Anrufaufbau, -wartung und -freigabe und
- Dienststeuerung - verantwortlich für Ausführung aller UMTS-Dienste, die nicht mit der Anruf-, Funk- und Verkehrsverbindung in Zusammenhang stehen.

Die obige Aufteilung der Steuerung trägt im wesentlichen dazu bei, daß vieles von der geforderten UMTS-Funktionalität erst möglich geworden ist: Die Dienststeuerung kann unabhängig von der Verbindung und vom Träger aufgebaut werden. Dadurch ist eine viel größere Flexibilität der Dienstleistung gegeben.

Zur Erleichterung der später notwendigen Datenbankbeschreibung wird zusätzlich eine Aufteilung der Funktionen des UMTS-Systems in zwei Gruppen durchgeführt:

- Vermittlungsteilnetz (VTN), aus solchen Netzknoten bestehend, die alle Funktionen mit Ausnahme von SDF ausführen, und
- Datenteilnetz (DTN), ausschließlich aus solchen Netzknoten bestehend, die die SDF-Funktion durchführen.

Das zweite Untersystem stellt im wesentlichen den Schwerpunkt der vorliegenden Arbeit dar.

2.4.2.2 Das Datenbankmodell

Das allgemeine funktionale Modell geht nicht auf die Spezifikation der UMTS-Datenbank ein. Die Datenspeicherfunktion im Netz ist dort als eine funktionale Einheit dargestellt worden. Dies bringt auch die Sicht des Vermittlungsuntersystems zum Ausdruck, die gesamte Datenfunktionalität als eine *black-box* anzusehen. Die räumliche Verteilung der Daten und deren aktueller Ort ist eine interne Angelegenheit der Datenbank, die mittels einer Schnittstelle mit dem restlichen Vermittlungsuntersystem kommuniziert - eine Idee, die ihre Verankerung in der neueren Version von IN-Spezifikationen ihre Bestätigung fand /CS-3/.

Hier wird nur eine Einführung in das Datenbankmodell gegeben. Eine ausführliche Diskussion, die auf die in der Arbeit untersuchten Alternativen näher eingeht, liegt im Kapitel 3 vor.

Die vorgeschlagene UMTS-Datenbankarchitektur ist in zwei Ebenen aufgebaut, die der generellen Aufteilung des gesamten UMTS-Gebietes in Teilgebiete Rechnung trägt:

- erste Ebene: die *inter domain*-Ebene, die die Kommunikation und den Informationsaustausch zwischen den einzelnen UMTS-Teilgebieten unterstützt und eine globale Dienstleistung ermöglicht; durch die Teilung des Gesamtnetzes in Teilgebiete wird auch das Multianbieterkonzept im künftigen mobilen Netz unterstützt,
- zweite Ebene: die *intra domain*-Ebene, die die Kommunikation und den Informationsaustausch innerhalb eines einzelnen UMTS-Teilgebietes unterstützt.

Innerhalb der zweiten Ebene werden zwei Knotentypen unterschieden:

- Datenknoten (DN-Knoten), in denen die Daten gespeichert werden, und
- Verzeichnisknoten (DIR-Knoten), in denen die Verzeichnisse gespeichert werden, so daß es jederzeit möglich ist, die gewünschten Daten zu lokalisieren.

Diese Knoten werden in einer baumähnlichen Konfiguration logisch verknüpft mit den Datenknoten als Baumblätter. Die Verwaltung der Datenuntersätze mittels eines baumartigen Verzeichnisses kann durch zwei Annahmen begründet werden:

1. Die personenbezogene Nummer hat einen quasi-lebenslangen Charakter, d.h., nach einer hinreichend langen Zeit gibt es keine Relation zwischen der Benutzernummer und der Lokation der permanenten Daten.
2. Die Mobilität des Benutzers soll so unterstützt werden, daß der Verzeichnisbaum jederzeit einen Zeiger zu den aktuellen Daten aufrechterhält.

Es ist zu beachten, daß in der gesamten Arbeit die dienstbezogenen Steuerungsdaten und nicht etwa die Sensordaten (z.B. Statistik- oder Gebührendaten) oder die strukturellen Daten gemeint sind, wenn allgemein von Daten gesprochen wird.

Die betrachteten Daten werden dem UMTS-Konzept nach in einige Klassen und Unterklassen eingeteilt. Eine Unterklasse stellt die kleinste Dateneinheit dar, die unabhängig von anderen gespeichert und verwaltet werden kann. Diese Spezifizierung, die bereits in /Mone61/ durchgeführt wurde, bringt neue Anforderungen zum Ausdruck, die an das UMTS-System gestellt werden:

- Der Benutzer kann sich frei bewegen und die Dienste empfangen sowie auf verschiedenen Endgeräten und in verschiedenen Netzen initiieren, lediglich auf der personenbezogenen Nummer basierend.
- Der Benutzer und der Teilnehmer können verschiedene Personen sein.
- Mehrere Benutzer können auf dem gleichen Endgerät gleichzeitig angemeldet sein. Daraus folgt, daß die Relation zwischen dem Benutzer und dem Gerät einen temporären Charakter hat.

Verschiedene Arten der Verteilung und der eventuellen Duplizierung der Datenunterklassen werden im weiteren als Datenverteilungsstrategien bezeichnet (vgl. Kapitel 3).

Aus der Tatsache, daß sowohl verschiedene strukturelle Möglichkeiten für die Datenbank denkbar sind als auch die Daten und die Zeiger verschiedenartig verwaltet werden können, ergibt sich eine Fülle von Alternativen, deren Untersuchung von Bedeutung ist, wenn man sich als Ziel die Optimierung der Datenbankleistungsfähigkeit setzt. Im Kapitel 3 werden die Datenkonzepte ausführlich dargestellt und die sich daraus ergebenden Alternativen der Datenbankimplementierung im einzelnen diskutiert. Zusammen mit den Auswertungskriterien, die ebenfalls im Kapitel 3 eingeführt werden, bilden sie die Grundlage der quantitativen Analyse im Kapitel 4.

2.4.3 Rolle der dezentralisierten Datenbank in UMTS

Das beschriebene Konzept der Datenhaltung in UMTS stützt sich auf eine dezentralisierte Datenbank (im folgenden auch verteilte Datenbank genannt). Dieser Ansatz kann als Konsequenz aus folgenden Annahmen angesehen werden (/Mone14/, /Mone61/ und /Casa94/):

- Das Multianbieterszenario fordert eine beschränkte Datenpartitionierung.
- Die Anwendungen und durchgeführten Aktionen (z.B. Anrufe oder Bewegungen) haben meistens einen lokalen Charakter.

- Die Skalierbarkeit des Systems ist sehr schwierig zu erreichen, wenn man einen zentralen Knoten hat.
- Die Zuverlässigkeit des Systems nimmt mit der Verteilung der Daten auf mehrere Knoten zu.
- Die zentrale Lösung innerhalb eines großen Netzes von einem einzelnen Anbieter würde eine eventuelle Überlastung des Zentralknotens mit sich bringen (lange Wartezeiten) und zu einer hohen Belastung von Verbindungsleitungen führen (hohe Transportkosten).
- Eine unsichere Relation besteht zwischen den Kosten eines zentralen Hochleistungsknotens und den Kosten von mehreren kleineren lokalen Knoten.

Diese Annahmen sollen hier nicht als absolut betrachtet werden. Sie stellen jedoch eine Reihe von vernunftmäßigen Erklärungen dar, warum eine dezentralisierte Datenbank als die Ziellösung für die Datenhaltung in UMTS angesehen wird. Sie basieren sowohl auf den bereits genannten Quellen als auch auf eigenen Überlegungen.

2.4.4 UMTS als System der dritten Generation

In den vorgestellten Telekommunikationskonzepten im Kapitel 2.3 wurden verschiedene Tendenzen der heutigen Entwicklung dargestellt. Folgende Haupteigenschaften konnten dabei erkannt werden, die als Merkmale von Systemen der dritten Generation angesehen werden können (/Groe94/ und /Laih92/):

- uneingeschränkte Mobilität, basierend auf der personenbezogenen Nummer,
- Personalisierung der Kommunikationsumgebung,
- breite Palette an angebotenen Diensten,
- Zusammenwirken der Dienste und
- Variabilität des Netzzugriffes.

Eine eingehende Studie der UMTS-Anforderungen (/R1D67/, /Gius91/, /Chia92/ und /Gril92/) führt zur Schlußfolgerung, daß praktisch alle der obigen Anforderung, mit Ausnahme der extensiven Personalisierung der Kommunikationsumgebung, im UMTS-System erfüllt werden. Aus diesem Grunde erscheint es berechtigt, die anfangs in der Arbeit aufgestellte Gleichsetzung von UMTS und einem System der dritten Generation weiter zu verwenden.

Wenn an dem UMTS-Konzept etwas auszusetzen sei, dann ist es die fehlende Möglichkeit, die Umgebung des Benutzers nach dessen eigenen Anforderungen zu spezifizieren. Dies ist lediglich in einem sehr beschränkten Maße möglich. Eine Erweiterung des UMTS-Konzeptes, die dieser Anforderung Rechnung tragen würde, obwohl durchaus sinnvoll und angemessen, wird im Rahmen dieser Arbeit nicht näher untersucht, da sie den Umfang der Arbeit erheblich erweitern und vom anfangs gesetzten Ziel ablenken würde.

3 Einfluß der Datenbank auf die UMTS-Prozeduren

Nachdem ein Überblick über die Datenbanken und die Telekommunikationstrends gegeben und die grundlegenden Konzepte des UMTS-Systems eingeführt worden sind, wird jetzt der Schwerpunkt der Betrachtung auf die UMTS-Datenbank und deren Relation zum Vermittlungsteilnetz (VTN) gelegt.

Um eine bessere Übersichtlichkeit der Prozedurbeschreibung zu gewährleisten, wird zuerst das Anwendungsdatenkonzept beschrieben, das für das UMTS-System entwickelt worden ist. Danach werden die charakteristischen UMTS-Benutzerprozeduren vorgestellt und die zu ihrer Durchführung notwendigen Daten spezifiziert. Die Benutzerprozeduren werden im Vermittlungsteilnetz initiiert. Die Daten, die von diesen Prozeduren benötigt werden, sind im Datenteilnetz (in der verteilten Datenbank) abgelegt. Da die Anforderungen der Dienstversorgung nach regelmäßigen Datenzugriffen verlangen, kommt es zu regulären Interaktionen zwischen den beiden Teilnetzen. Diese Interaktionen werden genau diskutiert.

Da in dieser Arbeit eine Leistungsanalyse vorgenommen wird, ist es notwendig, die Bewertungskriterien zu spezifizieren, die beim Vergleich von verschiedenen Lösungsansätzen relevant sind.

Davon ausgehend und basierend auf dem UMTS-Datenkonzept werden Optionen vorgestellt, die die Leistungsfähigkeit der verteilten Datenbank im UMTS-System beeinflussen. Diese Optionen werden im folgenden als Alternativen bezeichnet. Die unterschiedlichen Klassen von Alternativen, wie z.B. Verzeichnisstruktur oder Datenverteilungsalternativen, werden wegen ihrer ausgeprägten Unabhängigkeit Dimensionen genannt.

3.1 UMTS-Datenkonzept

Grundsätzlich werden vier Klassen von Anwendungsdaten unterschieden, die zur UMTS-Dienstanbietung benötigt werden:

- Benutzerdaten (*User Profile* - UP),
- Endgerätedaten (*Terminal Profile* - TP),
- Dienstanbieterdaten (*Service Provider Profile* - SPP) und
- Teilnehmerdaten (*SUBscriber Profile* - SUP).

Die zwei letzten Klassen sind ohne größere Bedeutung für die Untersuchungen, die im Rahmen dieser Arbeit durchgeführt werden. Diese Daten haben einen statischen Charakter und eine wohldefinierte Speicherstelle in der betrachteten verteilten Datenbank. Aus diesem Grunde ist es auch nicht notwendig, eine Optimierung im Zusammenhang mit der Verteilung dieser Daten durchzuführen. Sie werden lediglich vollständigkeithalber kurz spezifiziert.

Auf die zwei ersten Klassen wird in dieser Arbeit vorrangig eingegangen.

3.1.1 Benutzerdaten

Der Benutzer - im Unterschied zum Teilnehmer - stellt die physikalische Person dar, die einen angebotenen Dienst verwendet. Der Teilnehmer dagegen ist eine juristische Person, die den Dienst kauft und den Preis für die Benutzung bezahlt.

Die Benutzerdaten stellen die wichtigste Klasse der Daten dar, auf die im Rahmen der Dienstangebotung regelmäßig zugegriffen wird. Sie sind in vier Unterklassen aufgeteilt:

- *subscription*-Unterklasse,
- *routing*-Unterklasse,
- *registration*-Unterklasse und
- *session*-Unterklasse.

Die *subscription*-Unterklasse speichert die Bedingungen, zu welchen die einzelnen Dienste vom Benutzer in Anspruch genommen werden können. Dies erlaubt dem Teilnehmer einfach zu spezifizieren, was der Benutzer machen darf und was nicht. Inhalte können z.B. sein: Listen der erlaubten Dienste, Maximum des Kredites oder zugelassene UMTS-Teilgebiete, in welchen der Benutzer unterstützt wird. Die *subscription*-Daten werden von dem Dienstanbieter und dem Teilnehmer spezifiziert. Der Benutzer hat nur das Leserecht auf die gespeicherten Daten.

Die *routing*-Unterklasse speichert für alle vom Benutzer beanspruchten Dienste die Leitwegnummern zum Festnetzknoten, unter der sich weitere Informationen, d.h. die *registration*-Unterklasse (und üblicherweise auch das TP) für den jeweiligen Dienst befinden. Zusätzlich wird hier auch der aktuelle Status des Dienstes abgespeichert. In Abhängigkeit der gewählten Datenverwaltungsstrategie speichert die *routing*-Information entweder direkt die Adresse des Knotens (MRN - *Mobile Routing Number*) oder lediglich die Adresse des UMTS-Teilgebietes (DI - *Domain Identifier*). Die *routing*-Daten werden ausschließlich vom DBMS verwaltet.

Die *registration*-Unterklasse speichert die Präferenzen, unter welchen der Benutzer jeden einzelnen Dienst für die ankommende Kommunikation in Anspruch nehmen will. Dies erlaubt dem Benutzer seine eigene, persönliche Kommunikationsumgebung zu kreieren. Ein Beispiel der Einträge stellt eine Liste mit den erwünschten und nicht erwünschten Anrufern oder eine andere Liste mit zeitlichen Restriktionen für weniger dringende Anrufe dar. Zusätzlich zu den Präferenzen wird hier die aktuelle Nummer des Endgerätes gespeichert, auf das die ankommende Kommunikation geleitet wird. Die *registration*-Daten werden vom Benutzer spezifiziert.

In der UMTS-Datenspezifikation (/Mone61/) ist ausdrücklich vorgesehen, daß die *registration*-Unterklasse nur für die ankommende Kommunikation zu benutzen ist. Dies ist auch der erste Grund, warum das UMTS-Konzept eine volle Personalisierung nicht unterstützt, wie im Kapitel 2.4.4 bereits erwähnt. Es wäre jedoch ohne Einschränkung der sonstigen UMTS-Funktionalität möglich, daß diese Unterklasse ebenfalls für die abgehende Kommunikation benutzt wird. Eine solche Erweiterung

verlangt eine Modifikation des Informationsflusses für den Anrufaufbau - die Abfrage des Inhaltes der *registration*-Daten wäre dann notwendig in Ergänzung zur Abfrage der *subscription*-Daten (vgl. /SSD94/ für die Grundschemas des Anrufaufbaus). Auf diese Art bekäme man eine einfache Möglichkeit, die Personalisierung der Benutzerumgebung im Sinne des PSCS-Konzepts (vgl. Kapitel 2.3.5) und der Mobilität im Sinne von UMTS miteinander zu verschmelzen. Auf diese Erweiterung wird jedoch in dieser Arbeit nicht näher eingegangen.

Die *session*-Unterklasse speichert Daten, die zur jeweiligen Sitzung gehören (z.B. den Status der Anrufgebühren, die in Anspruch genommenen Dienste und deren Anbieter, Statistikdaten des Anrufes usw.). Diese Unterklasse wird stets im lokalen Datenknoten gespeichert und existiert nur solange, wie die jeweilige Sitzung andauert. Nach dem Abschluß der Sitzung werden die relevanten *session*-Daten als ein Teil der Verwaltungsdaten bei den jeweiligen Netzbetreibern gespeichert. Die *session*-Daten werden ausschließlich vom DBMS verwaltet.

Der Aspekt der Datenorganisation ist mit der Spezifikation der einzelnen Datenklassen nur zum Teil gelöst; die Datenpartitionierung ist damit vorgeschrieben: die Daten, die zu einer Unterklasse gehören, werden immer zusammen gespeichert. Die Frage einer eventuellen Duplizierung ist dadurch nicht berührt worden. Eine eindeutige Platzierung der einzelnen Untersätze ist nur für die *session*-Daten vorgegeben. Die Verteilungsoptionen der sonstigen Benutzerdaten bleiben vorerst offen und werden erst im Kapitel 3.4.2.1 diskutiert.

3.1.2 Endgerätedaten

Im UMTS-Konzept bieten die Endgerätedaten (TP) eine begrenzte Beschreibungsmöglichkeit des Endgerätes. Folgende Daten werden hier gespeichert:

- die aktuelle Relation zwischen dem logischen und dem physikalischen Gerät (gemäß dem Konzept des virtuellen Endgerätes - vgl. /Mone61/),
- der aktuelle Standort des physikalischen Gerätes innerhalb des Lokationsgebietes (erlaubt die Funkverbindung innerhalb einer Zelle aufzubauen) und
- der aktuelle Status (informiert darüber, ob das Gerät ein- oder ausgeschaltet ist).

Die Trennung des logischen und des physikalischen Endgerätes erlaubt, daß mehrere Benutzer, jeder einem anderen logischen Gerät zugeordnet, dasselbe physikalische Endgerät gleichzeitig benutzen dürfen. Sollten mehrere Benutzer tatsächlich dasselbe Endgerät benutzen, hätte dies schwerwiegende Folgen. Würden manche Benutzer verschiedene, zum Teil sich deckende UMTS-Teilsysteme bevorzugen (z.B. wählen einige aus Kostengründen das private und die anderen das öffentliche System), entstünde das Problem, die Endgerätedaten konsistent zu halten, die jetzt gleichzeitig in verschiedenen Teildatenbanken zu speichern sind.

Das bisherige UMTS-Endgerätkonzept berücksichtigt nicht die Aspekte der verschiedenen Gerätefähigkeiten. Dies ist, neben den Einschränkungen der *registrati-*

on-Daten, der zweite Grund, warum das UMTS-Konzept eine volle Personalisierung nicht unterstützt (vgl. Kapitel 3.1.1). Eine Erweiterung des Endgerätedatenkonzeptes zum Zwecke der Personalisierung kann vorgeschlagen werden, indem

- die mobilitätsbezogene Unterklasse, die die herkömmlichen UMTS-Endgerätedaten speichert, und
- die fähigkeitsbezogene Unterklasse, die die Fähigkeitsmerkmale eines physikalischen Gerätes beschreibt,

separat spezifiziert werden.

Die mobilitätsbezogenen Daten werden durch die Verwaltungsfunktionen des UMTS-Systems kreiert, d.h. sie brauchen nicht im Netz permanent gespeichert zu werden. Die fähigkeitsbezogenen Daten können entweder in einem Datenknoten als Hauptkopie vorhanden sein oder über die Funkschnittstelle vom Gerät abrufbar sein.

Für die Leistungsbewertung wird zwecks Vereinfachung angenommen, daß

- die Benutzer, die das gleiche physikalische Gerät anwenden, stets in demselben UMTS-Teilsystem sind und
- die Problematik der Gerätefähigkeit, d.h. der technischen Charakteristik des Endgerätes, und die eventuell notwendige Modifikation der Informationsflüsse nicht betrachtet werden.

Die erste Annahme ist berechtigt, da die Mehrfachregistrierung aus verschiedenen Teilgebieten einen Ausnahmecharakter hat und deswegen keinen Einfluß auf die Leistungsfähigkeit des Systems ausübt. Die zweite Annahme ist ebenfalls berechtigt, da die Problematik der Gerätefähigkeiten Fragen aufwirft, die über die Grenzen des untersuchten UMTS-Systems hinausgehen.

3.1.3 Dienstanbieterdaten

Die Dienstanbieterdaten beschreiben die vom Dienstanbieter unterstützten Dienste und deren Merkmale, wie z.B. Preis oder technisch notwendige Voraussetzungen (Bandbreite oder Gerätefunktionalität). Da diese Daten nicht mit der Mobilität des Benutzers in Verbindung stehen, werden sie in permanenten Datenknoten gehalten und haben keinen Einfluß auf die gewählte Datenverwaltungsstrategie.

3.1.4 Teilnehmerdaten

Der Teilnehmer ist im Unterschied zum Benutzer die juristische Person (z.B. das Unternehmen), die Dienste kauft und dem Benutzer zur Verfügung stellt. Der Teilnehmer und der Benutzer können entweder unterschiedliche oder gleiche Personen sein.

Die Teilnehmerdaten beschreiben die von ihm gekauften Dienste und die Konditionen, zu denen die Dienste benutzt werden können. Da diese Daten ebenfalls nicht mit der Mobilität des Benutzers in Verbindung stehen, werden sie lediglich permanent in ausgewählten Datenknoten gehalten. Somit haben sie keinen Einfluß auf die endgültig gewählte Alternative und werden deshalb im weiteren nicht betrachtet.

3.2 Modell der UMTS-Datenbank

3.2.1 Struktur der Datenbank

Unter der Struktur (auch Architektur) einer verteilten Datenbank versteht man die Relationen der Knoten untereinander /Fort85/.

Die UMTS-Daten müssen den Benutzern und/oder den Netzfunktionen zur Verfügung gestellt werden. Dies soll möglichst effizient gestaltet werden. Welche Punkte im Zusammenhang mit der Aufwandminimierung zu betrachten sind, stellt Kapitel 3.5 dar. Viele Gründe sprechen für eine Verteilung der Daten auf mehrere Knoten:

- Multianbieterszenario,
- lokaler Charakter der meisten Funktionsaufrufe,
- konzeptionelle Skalierbarkeit,
- hohe Belastung eines zentralen Knotens,
- hohe Belastung der Verbindungsleitungen bei einer zentralen Lösung,
- Zuverlässigkeit bei einem Knotenausfall und
- Kostenvorteile durch mehrere kleine Knoten.

Aufgabe einer dezentralisierten Datenbank ist, die Daten entsprechend einer gewählten Strategie zu verteilen, zu duplizieren und zu erneuern (vgl. /SSD94/, /Mone14/ und /Mone61/). Dazu werden die Knoten der Datenbank in eine Relation zueinander gebracht; diese Relation wird nach /Fort85/ eine Struktur der Datenbank genannt.

Um der Komplexität und der Multianbieterkonfiguration Rechnung zu tragen, findet sich die vorgeschlagene zweistufige Hierarchie des UMTS-Systems (vgl. Kapitel 2.4.2.2) ebenfalls in der Struktur der Datenbank wieder (vgl. /SSD94/). Es wird hier die Datenbankstruktur innerhalb eines Teilgebietes und zwischen den Teilgebieten unterschieden.

Innerhalb eines UMTS-Teilgebietes wird die verteilte UMTS-Datenbank aus mehreren Datenknoten bestehen. Um die verteilten Daten zu verwalten, wurden hierarchisch angeordnete doppelverkettete Zeiger vorgeschlagen (/SSD94/, /Mone14/ und /Mone61/), die in den Verzeichnisknoten abgelegt werden. Dabei entsteht eine baumartige, verteilte Datenbank, die aus den Datenknoten als Blätter und den Verzeichnisknoten, auch DIR-Knoten genannt, als Zweige aufgebaut ist. Die Datenknoten speichern die UMTS-Daten und die DIR-Knoten die Zeiger auf diese Daten.

Die auf diese Art konstruierten Bäume stellen die verteilte Datenbank eines UMTS-Teilgebietes dar. Die Bäume der einzelnen Teilgebiete koexistieren gleichberechtigt nebeneinander (siehe Abbildung 3.1). Die gewählte Datenbankstruktur trägt ebenfalls dem vorgesehenen Nummernschema in UMTS Rechnung. Die UMTS-Benutzernummer soll lediglich eine Indikation des UMTS-Heimteilgebietes erlauben, d.h. des Teilgebietes, in dem der Benutzer zuhause ist. Sie erlaubt aber nicht, den Heimdatenknoten zu identifizieren: innerhalb eines Teilgebietes gibt es keine Zuordnung zwischen der Benutzernummer und dem Heimdatenknoten. Die geforderte Bindung der Nummer zum Heimteilgebiet kann unterschiedlich gehandhabt werden. Eine Möglichkeit ist, daß in einem solchen Fall eine neue Nummer zugewiesen wird, wobei natürlich der lebenslange Charakter verlorengeht. Eine andere Möglichkeit ergibt sich, wenn die Wurzelknoten Listen der Benutzer erhalten, die ihr ursprüngliches Heimteilgebiet gewechselt haben (vgl. /Syka94/).

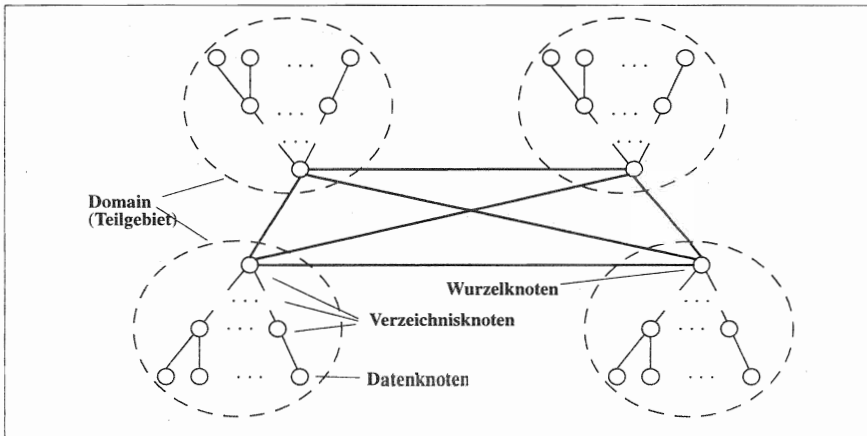


Abb. 3.1 Logische Datenbankarchitektur: Struktur innerhalb und zwischen Teilgebieten

Es soll hier betont werden, daß die beschriebene hierarchische Datenbankstruktur, wie in Abbildung 3.1 gezeigt, eine logische Eigenschaft und nicht eine physikalische Topologie des Systems darstellt. Dies bedeutet im speziellen, daß die Linien zwischen den Datenknoten und den Verzeichnisknoten nicht notwendig real existierenden Verbindungen entsprechen. Aus physikalischer Sicht ist zu erwarten, daß ein Trägernetz (z.B. B-ISDN, vgl. /Norp94a/) in Zukunft eine Transportmöglichkeit bieten wird, auf der eine Anwendung wie die des UMTS-Systems ausgeführt werden kann. Dieses Grundnetz wird aus LE (*Local Exchange*) und TX (*Transit eXchange*) bestehen, die mittels schneller und zuverlässiger Leitungen miteinander verbunden sind. In einer solchen Topologie werden die Daten- und die Verzeichnisknoten der UMTS-Datenbank, die ebenfalls eine Anwendung darstellt, an das Transportnetz angeschlossen und sind von dessen Kapazität und seiner gebotenen Bandbreite abhängig. Die Abbildung 3.2 macht diesen Ansatz deutlich.

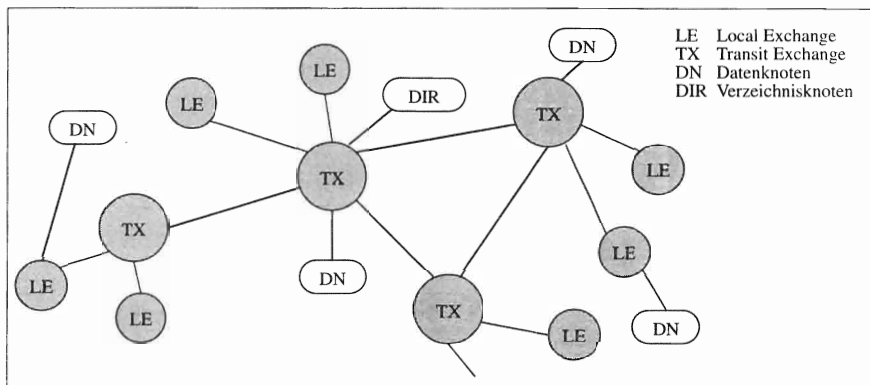


Abb. 3.2 Physikalische Datenbanktopologie (Beispiel)

3.2.2 Funktionale Architektur

Die funktionale Architektur der Daten- und Verzeichnisknoten ist unterschiedlich:

- Die Datenknoten verfügen jeweils über eine Speicherfunktion und eine lokale Verzeichnisfunktion, die die lokalen Daten verwaltet.
- Die Verzeichnisknoten verfügen jeweils über eine Verzeichnisfunktion, die in einer vordefinierten Art und Weise die Suche in der Datenbank ermöglicht (wie genau, das hängt von der gewählten Strategie ab).

Diese Funktionalität ist noch nicht ausreichend, um die Funktionalität der Datenbank zu gewährleisten. Folgende zwei Aspekte sind bis jetzt zu kurz gekommen:

1. Die Datenknoten speichern nicht nur die Daten, sondern verarbeiten auch die Anfragen des Vermittlungsteilnetzes zu einzelnen, datenbankgerechten Operationen.
2. Die Zweistufigkeit der Datenstruktur verlangt nach einer "intelligenten" Kommunikation zwischen den Datenbanken der einzelnen UMTS-Teilsysteme, d.h., die Wurzelknoten müssen die Steuerung der Kommunikation zwischen den Teilgebieten übernehmen.

Eine zusätzliche Funktion, der Anfragemanager (*query handler*), wird somit im Datenknoten und im Wurzelknoten eingeführt, die die Kommunikation mit dem Rest des UMTS-Systems ermöglicht (siehe Abbildung 3.3).

Zusammenfassend haben die einzelnen Funktionseinheiten folgende Aufgaben:

- DS-Funktion (Verzeichnisfunktion) - speichert und verwaltet die Zeiger, wie bereits beschrieben.
- DA-Funktion (Datenspeicherfunktion) - speichert die Benutzer- und Endgeräte-daten, wie bereits beschrieben.

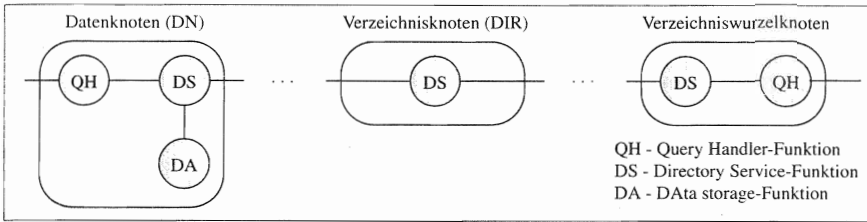


Abb. 3.3 Funktionale Architektur der Datenbank

- QH-Funktion (Anfragenverwalter)
 - im Datenknoten - bereitet die Einzelabfragen für die Datenbank aufgrund der externen Nachfragen aus dem Vermittlungsunternetz vor, und
 - im Wurzelknoten - koordiniert die Kommunikation zwischen den Teilsystemen der Datenbank.

3.3 UMTS-Benutzerprozeduren

Als Benutzerprozeduren sind spezifische UMTS-Prozeduren zu verstehen, die aufgrund der Benutzeraktionen im UMTS-Vermittlungsteilnetz initiiert werden.

Die Benutzerprozeduren können in zwei Gruppen aufgeteilt werden /SSD94/:

1. mobilitätsbezogene Prozeduren
2. nichtmobilitätsbezogene Prozeduren

Die erste Gruppe befaßt sich mit den Prozeduren, die aufgrund der Benutzer- und Endgerätemobilität notwendig sind, um jederzeit eine Verbindung mit dem Benutzer zu ermöglichen. Folgende Prozeduren gehören zu dieser Gruppe:

- *Location Update* (Aktualisierung des Aufenthaltsgebietes),
- *Domain Update* (Aktualisierung des Teilgebietes),
- *Terminal Attach / Detach* (Endgerät einschalten / ausschalten) und
- *Handover* (Anrufübergabe).

Die zweite Gruppe umfaßt die Prozeduren, die aufgrund der Dienstanforderung (z.B. Anrufaufbau) oder der Dienstspezifizierung (z.B. Personalisierung) notwendig sind. Folgende Prozeduren gehören zur zweiten Gruppe:

- *Session Set-Up / Session Release* (Sitzungsaufbau / -freigabe),
- *Call Set-Up / Call Release* (Anrufaufbau / -freigabe) und
- *User Registration / Deregistration* (Benutzerregistrierung / -deregistrierung).

Im folgenden werden Prozeduren der beiden Gruppen genauer definiert. Dabei werden die einzelnen Zugriffe des Vermittlungsteilnetzes auf das Datenteilnetz erläutert. Jeder Zugriff entspricht einer separaten sog. Datenanfrage, die vom

Anfragemanager (QH) weiterverarbeitet und in eine Nachrichtenfolge zerlegt wird. Die einzelnen Nachrichten aus einer Folge werden als Abfragen bezeichnet.

3.3.1 Mobilitätsbezogene Prozeduren

3.3.1.1 *Location Update*

Die *Location Update*-Prozedur (LU) wird jedesmal durchgeführt, wenn der Benutzer das bisherige Aufenthaltsgebiet (*Location Area* - LA) verlassen hat. Beide LA-Gebiete gehören dabei zum gleichen UMTS-Teilgebiet. Das LA-Gebiet kann entweder kleiner oder gleich dem DN-Gebiet sein, und somit kann die *Location Update*-Prozedur entweder innerhalb eines DN-Gebietes oder zwischen zwei DN-Gebieten bzw. innerhalb eines DN-Knotens oder zwischen zwei DN-Knoten erfolgen.

Es wird seitens des Vermittlungsteilnetzes einmal auf die Datenbank zugegriffen (LU_1 in Abbildung 3.4). Dabei werden folgende Aktionen durchgeführt:

- Aktualisierung der Information, die im Endgerätprofil (TP) gespeichert ist,
- Aktualisierung der Zeiger in den Verzeichnisknoten¹,
- Aktualisierung der Referenz auf TP in den *registration*-Daten¹,
- Transport des Benutzerprofils (UP) und TP in den neuen Datenknoten¹ und
- Aktualisierung der *routing*-Daten¹ und ².

3.3.1.2 *Domain Update*

Die *Domain Update*-Prozedur (DU) wird jedesmal durchgeführt, wenn der Benutzer das bisherige Aufenthaltsgebiet verlassen und dabei gleichzeitig das UMTS-Teilgebiet gewechselt hat.

Es wird seitens des Vermittlungsteilnetzes zweimal auf die Datenbank zugegriffen (DU_1 und DU_2 in Abbildung 3.4). Dabei werden folgende Aktionen durchgeführt:

1. Zugriff (DU_1):

- die TMTI-Nummer (*Temporary Mobile Terminal Identifier*) wird nachgefragt, um eine Netzverbindung aufbauen zu können.

2. Zugriff (DU_2):

- Abfrage der *subscription*-Daten,
- Aktualisierung der Information, die im TP gespeichert ist,
- Aktualisierung der Zeiger in den Verzeichnisknoten,
- Aktualisierung der Referenz auf TP in den *registration*-Daten,

1. nur wenn der Benutzer sich zwischen zwei Datenknoten bewegt hat und wenn die gewählte Methode der Datenbankorganisation einen Datentransport vorsieht

2. nur wenn die *routing*-Information auf den Datenknoten zeigt

- Transport des UP und TP in den neuen Datenknoten und
- Aktualisierung der *routing*-Daten.

3.3.1.3 Terminal Detach

Die *Terminal Detach*-Prozedur (TD) wird jedesmal durchgeführt (/SSD94/), wenn mindestens einer der folgenden Gründe zutrifft:

- der Benutzer hat das Endgerät ausgeschaltet oder
- das Endgerät wurde erfolglos gesucht (nach dem *paging*) oder
- das Endgerät hat sich nicht vorschriftsmäßig gemeldet.

Es gibt drei Möglichkeiten, die Ergebnisse der *Detach*-Prozedur zu speichern:

- kein Eintrag in der Datenbank,
- Aktualisierung nur im TP oder
- Aktualisierung im TP, in den *registration*- und den *routing*-Daten.

Es wird seitens des Vermittlungsteilnetzes einmal auf die Datenbank zugegriffen (TD_1 in Abbildung 3.4). Dabei werden folgende Aktionen durchgeführt:

- Aktualisierung der Einträge im TP und, wenn vorgesehen,
- Aktualisierung der Einträge im *registration*-Datensatz vom UP.

3.3.1.4 Terminal Attach

Die *Terminal Attach*-Prozedur (TA) wird jedesmal durchgeführt, wenn der Benutzer das Endgerät einschaltet. Darüber hinaus sind im UMTS-System verschiedene Arten der *Terminal Attach*-Prozedur vorgesehen, die entsprechend der *Terminal Detach*-Prozedur abgearbeitet werden.

Es wird seitens des Vermittlungsteilnetzes einmal auf die Datenbank zugegriffen (TA_1 in Abbildung 3.4). Dabei werden folgende Aktionen durchgeführt:

- TMTI-Nummer wird nachgefragt, um eine Netzverbindung herzustellen,
- Aktualisierung der Einträge im TP,
- wenn vorgesehen, dann Aktualisierung der Einträge in den *registration*- und *routing*-Daten, und
- gegebenenfalls, d.h., wenn der Benutzer in das alte Aufenthaltsgebiet gewechselt hat, werden die bereits vorgestellten Prozeduren LU_1 oder DU_2 durchgeführt.

3.3.1.5 Handover

Die *Handover*-Prozedur wird jedesmal durchgeführt, wenn der Funkkanal gewechselt wird, auf dem die Verbindung zwischen dem Endgerät und der Basisstation be-

reits existierte. Im UMTS-System ist in Erweiterung zu den bestehenden Konzepten vorgesehen, daß die *Handover*-Prozedur sowohl zwischen Teilgebieten als auch zwischen Zellen unterschiedlicher Hierarchiestufen (z.B. Mikro- und Makrozellen) durchgeführt werden kann. Die Prozedur ist aus Sicht des Vermittlungsteilnetzes sehr komplex und verlangt eine Synchronisation zwischen dem ursprünglichen und dem zukünftig benutzten Kanal. Das aus GSM bekannte Ankerprinzip wird aufgegeben /Chia91/. Eine genauere Spezifikation der *Handover*-Prozedur kann /Carr94/ entnommen werden.

Aus Sicht der verteilten Datenbank ist die *Handover*-Prozedur jedoch sehr einfach, weil in ihrem Verlauf höchstens einmal seitens des VTN auf das DTN zugegriffen wird, und zwar nur, wenn ein Teilgebietwechsel vorliegt. Dann wird der Inhalt der *subscription*-Daten dahingehend ausgewertet, ob der Benutzer im neuen Teilgebiet unterstützt werden kann/darf. Es ist jedoch bis jetzt nicht spezifiziert, ob dieser Zugriff während des Anrufes oder nach dem Anruf erfolgen soll. Hier wird angenommen, daß die *Handover*-Prozedur keine Anforderungen an die Datenbank stellt.

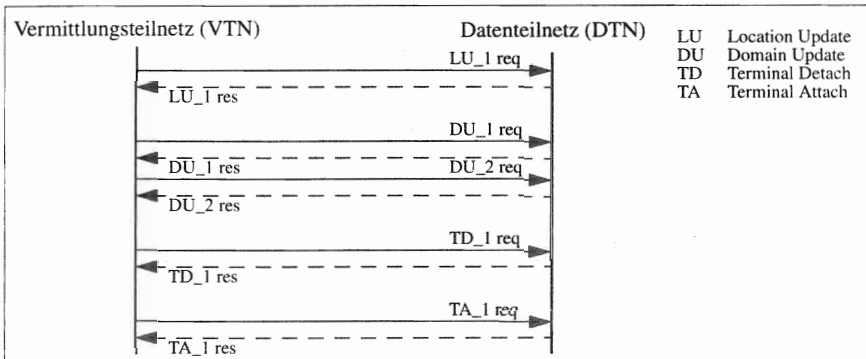


Abb. 3.4 Kommunikation zwischen VTN und DTN bei mobilitätsbezogenen Prozeduren

3.3.2 Nichtmobilitätsbezogene Prozeduren

3.3.2.1 Session Set-Up / Session Release

Die *Session Set-Up*-Prozedur (SSU) wird jedesmal durchgeführt, wenn der Benutzer eine Aktion vornimmt, die einen Netzzugang erfordert. Das sog. vertraute Verhältnis zwischen dem Benutzer und dem Netz wird dann aufgebaut /SSD94/.

Wenn es sich um die Anfangsprozedur (*initial Session Set-Up*) handelt, wird seitens des Vermittlungsteilnetzes dreimal auf die Datenbank zugegriffen. Dabei werden folgende Aktionen durchgeführt:

1. Zugriff (SSU_1):
 - Die TMTI-Nummer wird nachgefragt, um eine Netzverbindung herzustellen.

2. Zugriff (SSU_2):

- Der Dateninhalt der *subscription*-Daten wird überprüft, ob es dem Benutzer erlaubt ist, die Sitzung aufzubauen.

3. Zugriff (SSU_3):

- Die *session*-Daten werden angelegt, nachdem die Authentifizierung durchgeführt worden ist.

Wenn es sich um die einfache Prozedur (*simple Session Set-Up*) handelt, wird seitens des Vermittlungsteilnetzes zweimal auf die Datenbank zugegriffen. Dabei werden folgende Aktionen durchgeführt:

1. Zugriff (SSU_1):

- Die TMTI-Nummer wird nachgefragt, um eine Netzverbindung herzustellen.

2. Zugriff (SSU_3):

- Die *session*-Daten werden angelegt, nachdem die Authentifizierung durchgeführt worden ist.

Die Aktion SSU_1 kann ausgelassen werden, wenn bereits eine Netzverbindung besteht.

Die *Session Release*-Prozedur wird jedesmal durchgeführt, wenn der Benutzer eine Aktion, z.B. einen Anruf oder eine *user registration*-Aktion, abgeschlossen hat.

Es wird seitens des Vermittlungsteilnetzes einmal auf die Datenbank zugegriffen (SR_1). Dabei werden folgende Aktionen durchgeführt:

- Die *session*-Daten werden gelöscht und die erzeugten Daten nach Bedarf verarbeitet und im Netz gespeichert, z.B. in einem Dienstanbieter-Datensatz.

3.3.2.2 Call Set-Up/Call Release

Beim *Call Set-Up* (CSU) wird seitens des Vermittlungsteilnetzes mehrmals auf die Datenbank zugegriffen. Dabei werden folgende Aktionen durchgeführt:

- auf der anrufenden (*originating*) Seite:

1. Zugriff (o_CSU_1):

- Der Inhalt der *subscription*-Daten wird dahingehend überprüft, ob es dem Benutzer erlaubt ist, den Anruf durchzuführen.

2. Zugriff (o_CSU_2):

- Die *locating*-Anfrage wird durchgeführt, um den aktuellen Standort des angerufenen Benutzers herauszufinden.

- auf der angerufenen (*terminating*) Seite:

1. Zugriff (t_CSU_1):

- Der Inhalt der *subscription*- und *registration*-Daten wird überprüft und das aktuell verfügbare Endgerät ermittelt.

2. Zugriff (t_CSU_2):

- Der Inhalt des TP wird ausgelesen, um das Endgerät ausfindig zu machen.

3. Zugriff (t_CSU_3):

- Der Inhalt der *registration*- und *session*-Daten wird aktualisiert.

Beim *Call Release* (CR) wird ebenfalls seitens des Vermittlungsteilnetzes mehrmals auf die Datenbank zugegriffen. Dabei werden folgende Aktionen durchgeführt:

- auf der anrufenden Seite:
 1. Zugriff (o_CR_1):
 - Die Inhalte der *subscription*-, *registration*- und *session*-Daten des Anrufers werden aktualisiert.
- auf der angerufenen Seite:
 2. Zugriff (t_CR_1):
 - Die Inhalte der *subscription*-, *registration*- und *session*-Daten des Angerufenen werden aktualisiert.

3.3.2.3 User Registration/Deregistration

Die *User Registration*-Prozedur (UR) wird jedesmal durchgeführt, wenn der Benutzer eine Änderung der *registration*-Daten vornimmt.

Es wird seitens des Vermittlungsteilnetzes zweimal auf die Datenbank zugegriffen. Dabei werden folgende Aktionen durchgeführt:

1. Zugriff (UR_1):
 - Der Inhalt der *subscription*-Daten wird dahingehend überprüft, ob es dem Benutzer erlaubt ist, die Registrierung vorzunehmen.
2. Zugriff (UR_2):
 - Der Inhalt der *registration*- und *routing*-Daten wird aktualisiert, und
 - die Zeiger werden gegebenenfalls angelegt.

Die *User Deregistration*-Prozedur (UDR) wird jedesmal durchgeführt, wenn der Benutzer eine Änderung der *registration*-Daten abgeschlossen hat.

Es wird seitens des Vermittlungsteilnetzes einmal auf die Datenbank zugegriffen (UDR_1). Dabei werden folgende Aktionen durchgeführt:

- Der Inhalt der *registration*- und *routing*-Daten wird aktualisiert, und
- die Zeiger werden gegebenenfalls gelöscht.

3.3.3 Einfluß der Prozeduren auf die internen Transaktionen

Die Präsentation der UMTS-Benutzerprozeduren hat deutlich gemacht, daß nicht alle Prozeduren die gleiche Information aus der verteilten Datenbank benötigen.

Aufgrund der Datenverteilung ist es darüber hinaus möglich, daß die gesuchte Information auf die Datenbankknoten verteilt ist. Zwecks Übersichtlichkeit wird eine Systematik der Aktionen auf der Basis ihrer Komplexität eingeführt:

1. Typ:
Lokale Aktionen
2. Typ:
1. Typ oder Aktionen auf einem DN-Knoten mit bekannter Adresse
3. Typ:
2. Typ oder Aktionen auf einem DN-Knoten mit unbekannter Adresse
4. Typ:
3. Typ und Aktualisierungsaktionen auf den DIR-Knoten

Aus Sicht der Optimierungsaufgabe der verteilten Datenbank sind die Aktionen, die dem 1. und 2. Typ zuzuordnen sind, ohne Bedeutung, da sie keine Optimierung erlauben; sie sind in ihrem Ablauf fest vorgegeben.

Aktionen des 3. Typs werden während der Zugriffe TA_1, SSU_1, o_CSU_1 und o_CSU_2 durchgeführt. Dabei gilt, wenn Zugriffe TA_1, SSU_1 und o_CSU_1 auf einen DN-Knoten mit einer unbekannten Adresse stattfinden, daß eine Benutzerbewegung ohne Aktualisierung stattgefunden hat. Um die fehlende Aktualisierung nachzuholen, müssen in solchen Fällen Aktionen durchgeführt werden, die den Zugriffen LU_1 oder DU_2 entsprechen (siehe unten). Die während TA_1, SSU_1 und o_CSU_1 verbleibenden Restaktionen sind höchstens vom 2. Typ. Damit ist der Zugriff o_CSU_2 der einzige, bei dem Aktionen des 3. Typs durchgeführt werden [D14]. Im folgenden wird o_CSU_2 als der Suchzugriff (oder die Suchanfrage) bezeichnet.

Zu den Aktionen des 4. Typs gehören LU_1 und DU_2, die im folgenden als Aktualisierungszugriffe (oder Aktualisierungsanfragen) bezeichnet werden.

Die Zugriffe UR_2 und UDR_1 verlangen ebenfalls eine Aktualisierung des DIR-Baumes. Diese werden jedoch wegen ihrer geringen Häufigkeit nicht untersucht.

Alle anderen Zugriffe gehören zum 1. oder 2. Typ und brauchen im weiteren nicht betrachtet zu werden.

3.4 Alternativen der UMTS-Datenbank

Im vorliegenden Kapitel wird eine detaillierte Darstellung der Alternativen gegeben, die als entscheidend für die Leistungsfähigkeit der UMTS-Datenbank angesehen werden [Lawn95]. Diese Alternativen stellen die Grundlage der quantitativen Auswertung dar, die im Kapitel 6 durchgeführt wird.

Folgende zwei Gruppen von Alternativen können unterschieden werden:

- die strukturellen Alternativen und
- die Datenverteilungs- und Datenverwaltungsalternativen.

3.4.1 Strukturelle Alternativen

Wenn hier von strukturellen Alternativen die Rede ist, so sind ausschließlich die Relationen zwischen den Daten- und den Verzeichnisknoten eines Teilgebietes gemeint. Nur die Struktur innerhalb eines Teilgebietes ist beeinflussbar und wird hier untersucht. Es handelt sich dabei um eine logische Struktur, die auf die jeweilige physikalische Topologie nicht explizit Rücksicht nimmt (vgl. 3.2.1 und /Lawn95/).

Der Einfluß der Struktur auf die Leistungsfähigkeit der Datenbank wird offensichtlich, wenn man sich beispielsweise einen Vergleich zwischen Verzögerungszeiten einer Suche bei einer stark hierarchischen und einer flachen Struktur bei gleichbleibender Knotenausnutzung vorstellt. Aufgrund unterschiedlich langer Suchpfade ergeben sich hier unterschiedliche Suchzeiten.

Folgende strukturelle Alternativen werden hier als unabhängig betrachtet:

- vertikaler Aufbau der Datenbank, d.h. Anzahl der Verzeichnisebenen,
- Organisation der höchsten Verzeichnisebene, sog. Wurzelebene,
- Gleichmäßigkeit der Datenbankstruktur und
- Zeitvariabilität der Datenbankstruktur.

Es wird vorausgesetzt, daß die Datenknoten eine fest vorgegebene Topologie haben, die sich anhand der Dichte der Heimbenutzer und der verfügbaren Datenknotenkapazität spezifizieren läßt, z.B. soll ein Datenknoten 100 000 Benutzer unterstützen.

3.4.1.1 Vertikaler Aufbau der verteilten Datenbank

Die Struktur der baumähnlichen UMTS-Datenbank kann entweder stark hierarchisch oder flach aufgebaut sein.

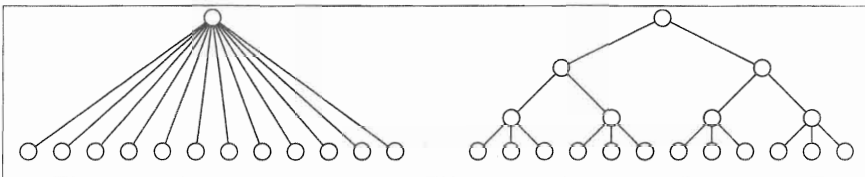


Abb. 3.5 Flache (links) vs. hohe (rechts) Baumstruktur

Beide Lösungen haben aus Sicht der Leistungsfähigkeit Vor- und Nachteile:

- Eine flache Struktur vereinfacht und beschleunigt den Prozeß der Verzeichnisaktualisierung. Darüber hinaus bietet sie kürzere Suchwege, da weniger Sprünge notwendig sind, wenn Daten mit Hilfe eines Verzeichnisses gesucht werden. Andererseits führt eine Abflachung des Baumes zu höheren Lasten in den wenigen Verzeichnisknoten. Die einzelnen Abfragen werden mangels eines regionalen Knotens direkt an den zentralen Knoten geleitet. Die Verbindungsbelastung steigt deswegen ebenfalls an.

- Eine stark hierarchische Struktur unterstützt eine gleichmäßige Verteilung der Abfragen auf mehrere Verzeichnisknoten und senkt damit die maximale Knotenbelastung. Für die Verbindungsbelastung wird erwartet, daß sie einen niedrigeren Wert annimmt, weil ein höherer Anteil der Anfragen regional beantwortet werden kann (die zentralen Verzeichnisknoten werden nicht involviert). Die Gesamtanzahl der Knoten steigt an, und alle Anfragen, die auf das Verzeichnis zugreifen, benötigen längere Zeit und sind komplizierter.

Diese Vor- und Nachteile sind schwierig gegeneinander aufzuwiegen, da das relative Verhältnis der einzelnen Charakteristika von vielen Parametern abhängig ist, wie z.B. Benutzerverhalten, Topologie des Netzes usw. (vgl. Kapitel 6.2). Eine numerische Auswertung ist hier notwendig, um eine fundierte Aussage zu liefern.

Im Zusammenhang mit der Baumhöhe soll hier noch auf die Zuverlässigkeit des Systems hingewiesen werden: Je höher die Hierarchie, desto höher ist die Anfälligkeit sowohl für Knoten- als auch für Nachrichtenfehler. Auf eine quantitative Analyse dieses Punktes wird in der vorliegenden Arbeit aus Platzgründen verzichtet.

3.4.1.2 Organisation der DIR-Wurzelebene

Unabhängig von der Frage nach der optimalen Baumhöhe muß das Problem der Organisation der höchsten DIR-Ebene, der Wurzelebene, gelöst werden. Zwei grundsätzliche Optionen sind hier möglich:

1. Der DIR-Baum ist mit einem gemeinsamen Wurzelknoten abgeschlossen (die traditionelle Lösung mit einer Ein-Baum-Struktur).
2. Auf der Wurzelebene gibt es mehrere koexistierende Knoten, von denen jeder nur für einen Unterbaum zuständig ist (Multi-Baum-Struktur).

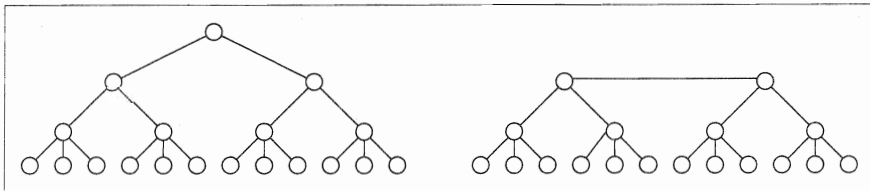


Abb. 3.6 Ein-Baum-Struktur (links) vs. Multi-Baum-Struktur (rechts) mit zwei Wurzelknoten

In der ersten Option enthält der Wurzelknoten die vollständige Information über alle im Teilgebiet gespeicherten Daten. Die Suche nach einem Eintrag wird spätestens im Wurzelknoten erfolgreich sein.

In der zweiten Option ist das betrachtete UMTS-Teilgebiet unterteilt und wird mittels mehrerer Verzeichnisteilbäume verwaltet. Ein gesuchter Eintrag kann sich in einem anderen Teilbaum befinden als der die Suche initiiierende Knoten. Dadurch ist eine einfache Zeigerverwaltung wie in der ersten Option nicht ausreichend. Hier ist es erforderlich, zusätzlich festzulegen, wie die Zeigerverwaltung auf der Wurzelebene gestaltet wird. Aus dieser Perspektive der im Rahmen dieser Arbeit untersuchten

zwei Anfragearten (Suche und Aktualisierung - vgl. Kapitel 3.3.3) lassen sich zwei Möglichkeiten unterscheiden, eine Multi-Baum-Struktur zu handhaben:

1. Während der Aktualisierung des Teilbaumes macht die Prozedur keinen Halt beim eigenen Wurzelknoten. Sie teilt die Änderung allen Wurzelknoten des eigenen Teilgebietes mit. Dann ist der Suchvorgang identisch zur Situation mit einem Wurzelknoten, da jeder Wurzelknoten die relevanten Zeiger zu den anderen Teilbäumen besitzt.

Wie die Aktualisierung auf der Wurzelebene geschieht, ist in Kapitel 3.4.2.2 dargestellt, wo die einzelnen Abfragetechniken erläutert werden.

2. Während der Aktualisierung wird die Prozedur spätestens im eigenen Wurzelknoten abgeschlossen. Dann muß die Suchprozedur den richtigen Teilbaum herausfinden, wo die Suche gegebenenfalls fortgesetzt werden soll.

Wie die Suche auf der Wurzelebene geschieht, ist im Kapitel 3.4.2.2 dargestellt.

Beide Wege der Handhabung der Multi-Baum-Struktur haben ihre Vor- und Nachteile, die sich aus dem unterschiedlichen Aufwand ergeben, der gegebenenfalls zur Durchführung der Aktualisierung oder der Suche notwendig ist. Eine quantitative, numerische Auswertung ist notwendig, um fundierte Aussagen zu liefern.

Um Mißverständnissen vorzubeugen, soll hier darauf hingewiesen werden, daß es ein grundsätzlicher Unterschied ist, ob man von mehreren Bäumen in unterschiedlichen Teilgebieten spricht, oder ob man ein Teilgebiet mittels mehrerer Teilbäume verwaltet. Folgendes ist zu beachten:

- Unterschiedliche Teilgebiete können verschiedenen Netzbetreibern gehören. Somit wird die Kommunikation zwischen den Teilgebieten an Sicherheitsvorschriften gebunden sein /Horn94/. Demgegenüber ist die Kommunikation innerhalb des Teilgebietes stets bei ein und demselben Netzbetreiber.
- Die personenbezogene Nummer zeigt auf das Heimteilgebiet des Benutzers, nicht aber auf einen eventuellen Teilbaum innerhalb dieses Gebietes.

3.4.1.3 Gleichmäßigkeit der Datenbankstruktur

Im folgenden wird der Einfluß der Gleichmäßigkeit der DIR-Struktur und deren zeitliche Variabilität auf die Leistungsfähigkeit der UMTS-Datenbank diskutiert.

Zwei Alternativen stehen hier zur Wahl:

1. Die DIR-Struktur ist unveränderbar und wird lediglich aufgrund der Benutzeranzahl festgelegt.
2. Die Struktur ist flexibel und zieht andere Einflußparameter zusätzlich in Betracht, z.B. das Nutzerverhalten. Die Flexibilität kann hier unterschiedlich gehandhabt werden:
 - Die DIR-Struktur ist für alle Benutzer gleich, d.h., sie ist für den statistischen Benutzer festgelegt, und ist damit eine Datenbankeigenschaft, die sich jedoch zeitlich ändern kann.

- Die DIR-Struktur ist benutzerabhängig und nimmt das Verhalten der einzelnen Benutzer als Grundlage für die Strukturfestlegung, d.h., sie ist eine Benutzereigenschaft.

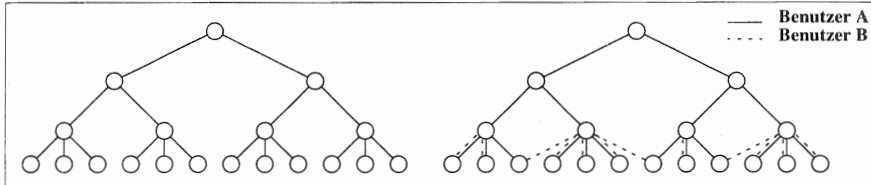


Abb. 3.7 Symmetrische (links) vs. unsymmetrische (rechts) Datenbankstruktur

Die erste Alternative ist sehr einfach zu verwalten und wird im wesentlichen durch die bereits diskutierte Hierarchiestruktur vollständig beschrieben. Der offensichtliche Nachteil ist die zu erwartende Ineffizienz in der Handhabung der Prozeduren, da diese durch das Benutzerverhalten beeinflusst werden.

Die zweite Alternative verlangt die Kenntnis des Benutzerverhaltens. Als einfaches, aber sehr gut geeignetes Merkmal erscheint hier z.B. die Verteilungswahrscheinlichkeit des Aufenthalts der Benutzer. Die Verwendung dieser Verteilung ist deshalb sinnvoll, weil das Anrufverhalten, d.h. die Verteilung der Anrufziele mit der Verteilung des Aufenthaltes im Regelfall korreliert: Die Benutzer bewegen sich häufiger zwischen den Orten, mit denen sie in irgendeiner persönlichen Beziehung stehen und auch öfters telefonisch kommunizieren. Diese Alternative kann verfeinert werden, indem die Verteilung der einzelnen Benutzer unabhängig betrachtet und damit persönliche DIR-Strukturen kreiert werden. Es handelt sich hier um keine dynamische Lösung, weil sie in der gewählten Konfiguration hinreichend lange Zeit fortbestehen wird, nachdem die Benutzerstruktur spezifiziert ist. Die Netzbelastung, die durch die Verwaltung der Strukturänderungen verursacht wird, ist gegenüber der Grundbelastung wesentlich geringer und kann deswegen unbeachtet bleiben.

Es erscheint naheliegend, daß die benutzerangepaßte Struktur eine Belastungsreduktion mit sich bringen wird. Wie groß die damit erzielte Verbesserung ist, wird im Rahmen dieser Arbeit im Kapitel 5 numerisch untersucht.

3.4.1.4 Zeitvariabilität der Datenbankstruktur

Die Anpassung der besprochenen benutzerangepaßten DIR-Struktur soll hinreichend selten notwendig sein. Dies bedeutet praktisch eine quasi-statische Struktur.

Eine berechnete Frage wäre die nach einer Erweiterung dieses Konzeptes auf eine dynamische, d.h. sich oft ändernde Struktur. Einerseits würde so ein Ansatz den Vergleich zwischen den Verwaltungs- und den Nutzprozeduren erforderlich machen. Das Ergebnis eines solchen Vergleiches ist zunächst unbekannt. Andererseits ist es notwendig, die Überlegung anzustellen, ob es überhaupt sinnvoll ist, eine solche Struktur anzustreben. Die Beobachtung des Benutzerverhaltens läßt den Schluß zu,

daß die Änderungen des Benutzerverhaltens im Vergleich zur Häufigkeit des Dienstzugriffs oder der Mobilität langsam, d.h. quasi-statisch verlaufen /Mone54/.

Aus obigen Gründen wird der Ansatz einer dynamischen Datenbankstruktur in der vorliegenden Arbeit nicht weiter verfolgt.

3.4.2 Alternativen der Datenverteilung und Datenverwaltung

Die zweite Gruppe der Alternativen befaßt sich mit den Konsequenzen für die Leistungsfähigkeit, die sich aus den Variationen der Datenverteilung - die Datenverwaltung resultiert daraus - ergeben. Es ist zum Teil unvermeidbar, daß die Datenverwaltungskonzepte in einer engen Relation zu den bereits besprochenen strukturellen Alternativen stehen. So geben z.B. die Zeigerinhalte in den DIR-Knoten das Spiegelbild der DIR-Struktur wieder. In den folgenden Ausführungen wird an Stellen, wo es zu einer Überschneidung kommt, gesondert darauf hingewiesen.

Hier werden folgende Verteilungs- und Verwaltungsalternativen als unabhängige Dimensionen betrachtet:

- die Verteilung der Benutzerdaten,
- die internen Abfragetechniken und
- die Optionen der Verzeichniszeiger.

3.4.2.1 Datenverteilungsalternativen (Datenszenarien)

Die Datenverteilung oder besser gesagt die Verteilung und Duplizierung der einzelnen Unterklassen der UMTS-Benutzerdaten (vgl. Kapitel 3.1) ist eine der wichtigsten Möglichkeiten zur Verbesserung der Leistungsfähigkeit der UMTS-Datenbank. Das ist die Konsequenz, die aus zwei entgegenwirkenden Bestrebungen resultiert:

- Für die meisten Aktionen in der UMTS-Datenbank ist es von Vorteil, wenn sich die Daten in einem der DN-Knoten befinden, der möglichst nahe an dem aktuellen Benutzerstandort liegt (vgl. Aktionen vom 1. und 2. Typ im Kapitel 3.3.3).
- Aufgrund der Mobilität verändert der Benutzer seinen Standort bezüglich des Netzes. In der Konsequenz muß die Datenverteilung alsbald aktualisiert werden, was zu einem hohen mobilitätsabhängigen Datenverkehr führt.

Bei der Darstellung der unterschiedlichen Datenverteilungsszenarien wird davon ausgegangen, daß die Anzahl der Datenknoten und deren topologische Anordnung durch das Teilgebiet sowie durch die Anzahl der Heimbenutzer und die DN-Knotenzapazität gegeben ist. Somit bleibt nun zu beantworten, wie die einzelnen UMTS-Datenunterklassen aufgrund der Mobilität verteilt und/oder dupliziert sein sollten.

Die Benutzer können sich aufgrund ihrer Mobilität in verschiedenen UMTS-Teilgebieten aufhalten. Das UMTS-Teilgebiet, in dem die Benutzer zu Hause sind und ihre Daten permanent gespeichert haben, wird das Heimteilgebiet genannt. Die sonstigen UMTS-Teilgebiete sind in bezug auf diese Benutzer die sog. Besuchsteilgebiete.

Aufgrund der Mobilität können sich die Benutzer in verschiedenen DN-Gebieten innerhalb eines UMTS-Teilgebietes aufhalten. Das DN-Gebiet, in dem die Benutzerdaten permanent gespeichert sind, wird das DN-Heimgebiet genannt (und der zugehörige DN-Knoten der Heimknoten). Die sonstigen DN-Gebiete sind in bezug auf diese Benutzer die sog. DN-Besuchsgebiete (und entsprechend die zugehörigen DN-Knoten die Besuchsknoten). Im Heimteilgebiet existiert immer ein Heimknoten. Die Existenz des Besuchsknotens im Heimteilgebiet und des Heimknotens in einem Besuchsteilgebiet ist vom gewählten Datenszenario abhängig.

Es soll folgendes stets beachtet werden:

- Die Dienstanbieter- und Teilnehmerdatensätze stehen in keiner Relation zur Benutzermobilität und werden deswegen von den UMTS-Benutzerprozeduren, wie im Kapitel 3.3 definiert, nicht benötigt. Dadurch haben sie auch keinen Einfluß auf die vorgenommene Optimierung und brauchen nicht betrachtet zu werden.
- Der Endgerätedatensatz (TP) wird üblicherweise im gleichen DN-Knoten gespeichert, wo die aktuellen Benutzerdaten gespeichert sind.
- Die zu untersuchenden Möglichkeiten der Verteilung von Benutzerdaten werden noch weiter eingeschränkt, indem die *session*-Unterklasse immer im lokalen DN-Knoten abgelegt wird.

Aus dem obigen folgt, daß die verbleibenden *subscription*-, *routing*- und *registration*-Unterklassen die einzigen sind, die tatsächlich untersucht werden müssen.

In Abbildung 3.8 sind Kombinationen zusammengestellt, die die Beziehung der drei relevanten Datenunterklassen in einem UMTS-Heimteilgebiet untereinander beschreiben. Die eingekreisten Nummern spezifizieren die möglichen Optionen unter den Datenunterklassen.

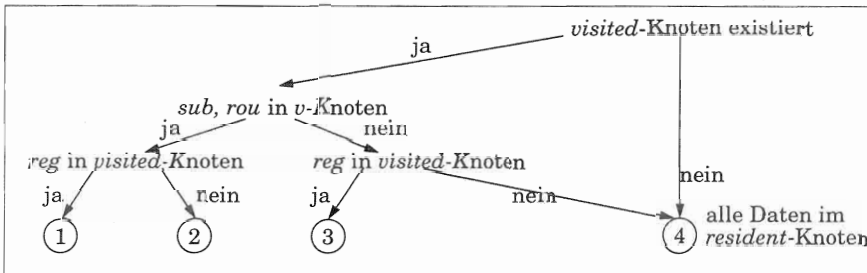


Abb. 3.8 Beziehung zwischen den Benutzerdaten innerhalb des Heimteilgebiets

Die obige Darstellung der Beziehung stellt nicht alle Aspekte der jeweiligen Datenverteilung dar. Zusätzlich müssen noch folgende Punkte festgelegt werden:

- Sind die *subscription*- und *routing*-Daten im V-Knoten die einzigen Hauptkopien oder bloß Zweitkopien der Hauptkopie, die im V-Knoten liegt?
- Enthält der *resident*-Knoten eine Referenz auf den *visited*-Knoten oder nicht?

Unabhängig von der Verteilung innerhalb des Heimteilgebietes kann auch die Beziehung unter den Datenunterklassen innerhalb eines UMTS-Besuchsteilgebietes spezifiziert werden. Es erscheint zweckmäßig, diese in Anlehnung an die Situation im UMTS-Heimteilgebiet zu spezifizieren - natürlich unter Beachtung der Vereinfachungen, die sich in einem UMTS-Besuchsteilgebiet ergeben (z.B. fehlender Heimknoten oder fehlende Daten einzelner Unterklassen). In Abbildung 3.9 ist eine schematische Darstellung der Beziehung in einem Besuchsteilgebiet gegeben.

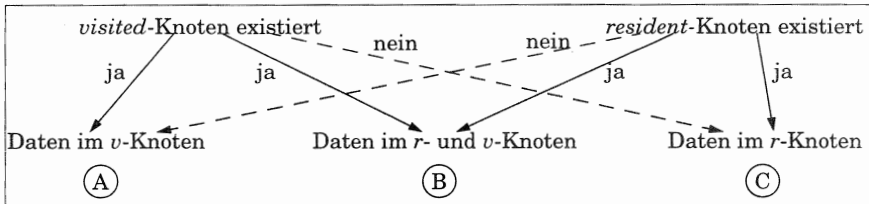


Abb. 3.9 Beziehung zwischen den Benutzerdaten innerhalb des Besuchsteilgebietes

Für weitere, genauere Untersuchungen wurden aus den möglichen Datenverteilungsszenarien vier Grundszenarien (/Mone61/) gewählt, die sich mittels der obigen Beziehungen spezifizieren lassen:

- *static*-Szenario (ST-Szenario) - basiert auf einer permanenten Speicherung der Benutzer- und Endgerätedaten im *resident*-DN-Knoten. Dies entspricht der Option 4 aus der Abbildung 3.8 und der Option C aus der Abbildung 3.9. Alle Such- und Aktualisierungsanfragen werden auf dem *resident*-Knoten durchgeführt. Alle Zeiger zeigen auf den *resident*-Knoten, und somit ist eine Aktualisierung der Zeiger in den DIR-Knoten nicht notwendig.
- *resident*-Szenario (R-Szenario) - basiert auf einer dynamischen Speicherung der Benutzer- und Endgerätedaten im *visited*-DN-Knoten. Dies entspricht der Option 1 aus der Abbildung 3.8 und der Option B aus der Abbildung 3.9. Im *resident*-DN-Knoten befindet sich eine Referenz zum *visited*-DN-Knoten. Alle Such- und Aktualisierungsanfragen werden auf dem *visited*-DN-Knoten durchgeführt. Während der Aktualisierung wird die Referenz zum *visited*-DN-Knoten erneuert. Alle Zeiger zeigen auf den *resident*-Knoten, und somit ist eine Aktualisierung der Zeiger in den DIR-Knoten nicht notwendig.
- *visited*-Szenario - basiert auf einer dynamischen Speicherung der Benutzerdaten im *visited*-DN-Knoten als Zweitkopien und Endgerätedaten als Hauptkopien. Dies entspricht der Option 1 aus der Abbildung 3.8 und der Option A aus der Abbildung 3.9. Alle Such- und Aktualisierungsanfragen werden auf dem *visited*-Knoten durchgeführt. Alle Zeiger zeigen auf den *visited*-Knoten und werden während der Aktualisierung erneuert.
- *resident/visited*-Szenario - basiert auf einer dynamischen Speicherung der Benutzerdaten im *visited*-DN-Knoten als Zweitkopien und Endgerätedaten als Hauptkopien. Dies entspricht der Option 1 aus der Abbildung 3.8 und der Option A aus der Abbildung 3.9. Im *resident*-DN-Knoten befindet sich eine Referenz zum *visited*-DN-Knoten. Alle Such- und Aktualisierungsanfragen werden auf

dem *visited*-Knoten durchgeführt. Während der Aktualisierung werden die Referenz zum *visited*-DN-Knoten und alle Zeiger zum *visited*-Knoten erneuert. Während der Suche werden beide Zeiger zum *resident*- und zum *visited*-Knoten beachtet.

Von den sonstigen Kombinationen sind einige sicherlich interessant und könnten im Hinblick auf die Reduktion des Verkehrsaufkommens im Kommunikationsnetz untersucht werden. Hier wurden jedoch nur die vier dargestellten Grundszenarien betrachtet, da sie die vorgeschlagenen UMTS-Optionen repräsentieren /Mone61/.

Andere Optionen können mit dem nachfolgend beschriebenen Modell ebenfalls untersucht werden - es gibt diesbezüglich keine prinzipiellen Beschränkungen. Die jeweiligen Protokolle müßten jedoch neu spezifiziert und zwecks Auswertung implementiert werden. Näherungsweise lassen sich diese Optionen mit den Ergebnissen der untersuchten Szenarien in ihrem Verhalten abschätzen.

3.4.2.2 Interne Abfragetechniken

Die internen Abfragetechniken gewährleisten eine geordnete, schrittweise Durchführung des Datenbankzugriffs, indem die einzelnen Protokollgrundelemente des Zugriffs in eine Nachrichtenfolge zerlegt werden. Dabei ist zu beachten, daß die Daten in den DN-Knoten verteilt und/oder dupliziert sein können und daß die Zeiger in den DIR-Knoten nach einem bestimmten Schema verteilt sind.

Im Kapitel 3.3.3 wurde erläutert, warum nur Such- und Aktualisierungsanfragen für die Optimierung der UMTS-Datenbank von Bedeutung sind. Dabei ist zu beachten, daß z.B. während der Aktualisierungsanfrage je nach dem Stand der Information in den einzelnen Datenknoten ebenfalls eine Suche nach einer bestimmten Information notwendig ist. Deshalb wird in der folgenden Diskussion von der Anfrageart abgesehen und lediglich allgemein unterschieden, ob es sich um die Suche nach einer Information oder um die Aktualisierung einer Information handelt:

- Während der Suche, die in mehreren Einzelschritten durchgeführt wird, führt nicht unbedingt der erste Schritt im eigenen Knoten zu einer positiven Antwort. Deshalb müssen die einzelnen Schritte miteinander logisch verkettet werden. Die Art der Verkettung gibt der jeweiligen Abfragetechnik den Namen.
- Während der Aktualisierung ist der Vorgang ebenfalls nicht immer nach einem Schritt beendet. Ähnlich wie bei der Suche können die Einzelschritte unterschiedlich verkettet werden. Zusätzlich ist zu beachten, daß die Reihenfolge der Schritte hier durch die Prinzipien der verteilten Transaktionen beeinflusst wird.

Im Prinzip können während der Suche drei Phasen unterschieden werden:

1. Abfragen innerhalb der hierarchisch angeordneten Knoten der DIR-Struktur eines UMTS-Teilgebietes,
2. Abfragen innerhalb der Knoten einer DIR-Ebene eines UMTS-Teilgebietes, z.B. unter den Knoten der Wurzelebene in einer Multi-Baum-Struktur, und

3. Abfragen unter den Übergangsknoten der UMTS-Teilgebiete.

Die 1. Phase wird immer ausgeführt; in bestimmten Situationen fallen die 2. und 3. Phase weg. Folgende Abfragetechniken können in der 1. Phase angewendet werden (siehe Abbildung 3.10):

- *passing*
 - sehr einfach in der Handhabung,
 - erfordert eine Zusatzinformation, die es dem Ziel direkt erlaubt, mit dem Ursprung zu kommunizieren,
 - sehr schlechte Fehlerkorrekturmöglichkeit, da die Fehlerquelle nicht lokalisiert werden kann,
- *chaining*
 - eine aufwendige Ausgestaltung in den mittelbaren Knoten, da die Verwaltung der abgeschickten und noch nicht zurückgekommenen Nachrichten notwendig ist,
 - hohe Belastung der mittelbaren Knoten, die in UMTS hauptsächlich aus den DIR-Knoten bestehen,
 - relativ einfache Handhabung,
- *referral*
 - geringe Intelligenz ist in den DIR-Knoten notwendig, da die gesamte Prozedur durch die DN-Knoten gesteuert wird,
 - sehr gute Fehlerkorrekturmöglichkeiten, da der Fehler und seine Quelle direkt entdeckt werden,
 - sehr hohe Belastung der Steuerungsknoten - hier die DN-Knoten,
 - eine ungünstige Verkehrssituation im Falle der Kommunikation zwischen UMTS-Teilgebieten, da zusätzlicher Nachrichtenaustausch zwischen den DN- und dem Wurzelknoten - der letzte funktioniert hier als Netzübergang - notwendig ist,
- *chaining / referral*
 - eine Kombination der beiden Techniken, die deren Vorteile abgeschwächt enthält und deren Nachteile teilweise beseitigt,
- *direct*
 - angewandt, wenn die Speicherstelle der gesuchten Daten aus früheren Zugriffen bekannt ist, z.B. ist der alte Standort des Benutzers während der Aktualisierungsprozedur im Endgerät gespeichert und wird über die Funkschnittstelle dem neuen DN-Besuchsknoten mitgeteilt.

Unabhängig davon können folgende Abfragetechniken in der 2. Phase angewendet werden:

- *broadcasting*
 - die schnellste Methode, um eine Information zu verteilen,

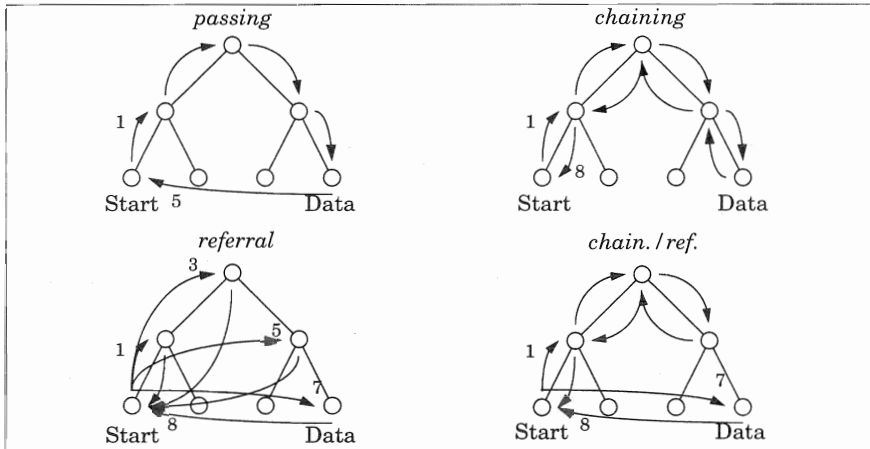


Abb. 3.10 Interne Abfragetechniken in der 1. Suchphase

- die leistungsintensivste Methode, die das Netz am stärksten belastet,
- sehr gut geeignet, wenn die Information auf alle Knoten verteilt werden soll, z.B. wenn die Aktualisierung in den Wurzelknoten einer Multi-Baum-Struktur durchgeführt wird,
- *chaining*
 - entspricht der *chaining* Technik aus der 1. Phase,
- *referral*
 - entspricht der *referral* Technik aus der 1. Phase,
- *direct*
 - wird angewendet, wenn die Speicherstelle der gesuchten Daten bereits bekannt ist, so daß die Suchprozedur diese Information ausnutzen und den entsprechenden Knoten direkt abfragen kann,
 - bringt zusätzliche Vorteile, wenn die gesuchten Daten nicht mehr im nachfragenden UMTS-Teilgebiet vorhanden sind, da das ansonsten notwendige Warten auf Absagen von allen Wurzelknoten lange dauern kann.

In der 3. Phase werden die Abfragen zwischen den Teilgebieten voraussichtlich unter der Ausnutzung der *referral*-Abfragetechnik durchgeführt. Dies erscheint zweckmäßig wegen der identischen Gesamtnetzbelastung im Vergleich zur *chaining*-Abfragetechnik und wegen der gleichzeitigen besseren Überwachung des Informationsflusses.

Die Aktualisierung wird in vier Phasen durchgeführt:

1. Kopieren der relevanter Daten
2. Aktualisieren des Verzeichnisbaumes

3. Löschen der alten Verzeichniseinträge
4. Aktualisieren der Referenzen im Heimdatenknoten

Die Aktionen, die das Verzeichnis aktualisieren (2. und 3. Phase), können mit den gleichen Techniken wie die Suche in der 1. und 2. Phase durchgeführt werden. Die sonstigen Phasen werden unter Zuhilfenahme der *referral*-Technik ausgeführt, um dadurch die maximale Überwachung des Informationsflusses zu gewährleisten.

3.4.2.3 Zeigeroptionen in den DIR-Knoten

Die dargestellten Abfragetechniken sind auf die Zeiger angewiesen, die sich in den DIR-Knoten befinden. Die Zeiger wiederum spiegeln die gewählte Datenverteilung in den DN-Knoten wider.

Die Möglichkeiten zur Organisation der Inhalte der DIR-Zeiger sind bis jetzt nicht erläutert worden. Hier sind zwei prinzipielle Aspekte zu untersuchen:

- die zeitliche Änderung der Information; man unterscheidet
 - statische Zeiger, die eine Eigenschaft des DIR-Baumes sind, und
 - dynamische Zeiger, die erst nach Bedarf angelegt werden,
- die Art der gegenseitigen Verknüpfung der Zeiger; man unterscheidet
 - direkte Zeiger auf den DN-Knoten, in dem sich die Daten befinden, und
 - indirekte Zeiger auf einen DIR-Knoten der Ebene direkt darunter.

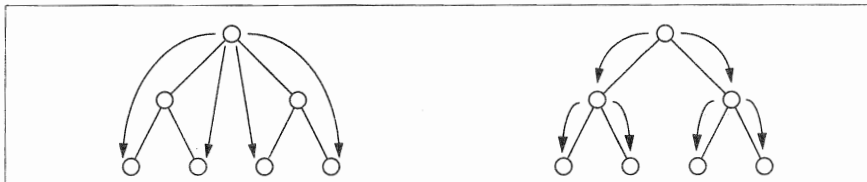


Abb. 3.11 Direkte (links) vs. indirekte (rechts) Zeiger

Die Optionen der zeitlichen Variabilität der Zeiger lassen sich wie folgt erklären:

- Statische Zeiger sind in den DIR-Knoten bereits angelegt, bevor die Daten in den DN-Knoten gespeichert werden. Nachdem die Daten in den DN-Knoten gespeichert worden sind, wird eine Relation zwischen den Zeigern und den Daten hergestellt. Diese Relation wird in einer bekannten Speicherstelle abgelegt, z.B. im Benutzerheimknoten.
- Dynamische Zeiger werden in den DIR-Knoten erst dann angelegt, wenn die Daten in den DN-Knoten gespeichert worden sind. Die dabei benutzte Kennung, die in den DIR-Knoten angelegt wird, verweist auf den gespeicherten Datensatz im DN-Knoten.

Die statische Methode ist praktisch nur zur Verwaltung der Endgerätedaten anwendbar. Die Endgeräte werden mit Hilfe von TMTI-Nummern verwaltet, die im

voraus jedem DN-Gebiet zugeordnet werden können. Jedes DN-Gebiet verfügt damit über eine Anzahl von TMTI-Nummern. In einem solchen Szenario erhält jedes Endgerät, nachdem es in ein neues Gebiet gekommen ist, eine neue TMTI-Nummer. Diese wird wiederum in den *registration*-Daten gespeichert. Die Anwendung der statischen Methode für Benutzerdaten setzt voraus, daß der neu zugewiesene Zeiger in einer bekannten Stelle gespeichert ist, z.B. im Heimknoten. Das würde jedoch eine zweimalige Suche notwendig machen: zuerst nach dem Heimknoten und dann nach dem Besuchsknoten. Das ist ein klarer Nachteil gegenüber dem RV-Szenario, wo eine im Heimknoten gespeicherte Referenz die zweite Suche überflüssig macht.

Die Unterscheidung, ob statische oder dynamische Endgerätezeiger verwendet werden, erweist sich jedoch aus folgenden Gründen als nicht relevant:

- Es wird angenommen, daß die Endgerätedaten im gleichen DN-Knoten gespeichert sind wie die jeweiligen *registration*-Benutzerdaten. Eine Ausnahme von diesem Ansatz wäre nur dann als sinnvoll zu betrachten, wenn ein Endgerät gleichzeitig von mehreren Benutzern in verschiedenen Teilgebieten verwendet werden würde (vgl. Kapitel 3.1.2). Da dieses jedoch selten zu erwarten ist, wird der Einfluß auf die Leistungsfähigkeit nicht von Bedeutung sein.
- Wenn die Benutzerdaten sich der dynamischen Zeiger bedienen, dann ist auch eine Aktualisierung des DIR-Baumes aufgrund von Benutzerbewegungen notwendig. Wenn aber die Aktualisierung in den DIR-Knoten aufgrund der Benutzerdatenzeiger vorgenommen werden muß, fällt auch eine eventuelle Aktualisierung der Endgerätezeiger nicht ins Gewicht.

Die Verwaltung der dynamischen Zeiger kann unterschiedlich organisiert werden:

- Die Heimknotenzeiger werden immer im gesamten Pfad des Heimknotens angelegt.
- Die Besuchsknotenzeiger können auf drei Arten verknüpft werden (siehe Abbildung 3.12 - die schattierten Knoten enthalten Zeiger auf den Besuchsknoten):
 1. Option: Die Zeiger werden im gesamten Pfad des Besuchsknotens angelegt.
 2. Option: Der Zeiger wird nur im Brückenknoten des Heim- und Besuchsknotenpfades aufgebaut.
 3. Option: Die Zeiger werden in allen DIR-Knoten auf dem Weg vom Heim- zum Besuchsknoten aufgebaut.

In der numerischen Auswertung im Kapitel 5 wird die 1. Option angewendet. Diese Wahl wurde aus folgenden Gründen getroffen:

- Diese Option wurde als die bevorzugte Lösung vorgeschlagen (vgl. /Monne61/).
- Diese Option stellt einen Kompromiß zwischen den beiden anderen Optionen in bezug auf den Such- bzw. Aktualisierungsaufwand dar.

Die Art der Verknüpfung der Zeiger hat einen entscheidenden Einfluß auf die Effizienz der verschiedenen Prozeduren:

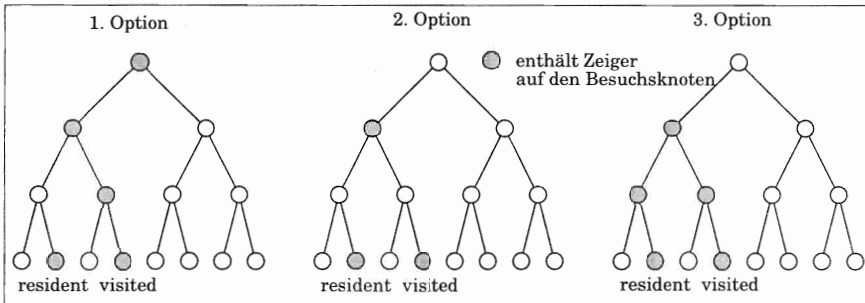


Abb. 3.12 Optionen der dynamischen Zeiger

- Einerseits verkürzen die direkten Zeiger die Suchpfade, so daß alle Prozeduren, die eine Suche notwendig machen, schneller werden und mit weniger Ressourcen auskommen. Andererseits aber verlangt die Aktualisierung immer eine konsequente Durchführung in allen Knoten des gesamten DIR-Pfades bis in die Wurzelebene. Jegliche Bewegung zwischen zwei DN-Knoten muß allen DIR-Knoten im Pfad darüber mitgeteilt werden.
- Die indirekten Zeiger verlängern die Suchpfade. Gleichzeitig aber vereinfachen sie die Aktualisierung, da diese nur bis zum sog. Brückenknoten, d.h. dem gemeinsamen DIR-Knoten im alten und neuen DIR-Pfad durchgeführt wird.

Aus der Charakteristik der direkten und indirekten Zeiger läßt sich schließen, daß die direkten Zeiger von Vorteil sind, wenn die Daten in den DN-Knoten nie oder nur sehr selten ihre Speicherstelle ändern. Dies trifft dann zu, wenn man z.B. das *residential*- und *static*-Szenario betrachtet.

Im Falle der Daten, die ihre Speicherstelle relativ häufig ändern, ist es notwendig, eine quantitative Analyse vorzunehmen. Eine solche Analyse ist auch deshalb notwendig, da Vor- und Nachteile der beiden Verknüpfungsarten von der Höhe des DIR-Baumes und dessen Relation zum Benutzermobilitätsmuster abhängig sind.

3.5 Bewertungskriterien der Leistungsfähigkeit

Die Auswertungsergebnisse der Leistungsfähigkeit der einzelnen Lösungen sind von den gewählten Kriterien abhängig. Hier werden zuerst die Kriterien genannt, die während der Auswertung im Kapitel 4 benutzt werden. Zu diesen gehören:

- die gesamte und maximale Prozessorkapazität,
- die gesamte und maximale Speicherkapazität,
- die Belastung der Verbindungsleitungen,
- die Verzögerungszeit der Prozedur, verursacht durch die Bearbeitungszeiten in den Knoten, durch die Bandbreitenbegrenzung auf den Verbindungsleitungen und eventuell durch die Ausbreitungsgeschwindigkeit der Signale und

- die Zuverlässigkeit und/oder Fehlertoleranz, d.h. Fehlervorbeugung und/oder Reaktion auf einen Knotenausfall oder auf den Protokollabbruch.

Den verschiedenen Kriterien könnte unterschiedliche Wichtung zugeteilt werden, und auf diese Weise wäre es möglich, einen Gesamtvergleich durchzuführen. Die Festlegung der einzelnen Gewichte hängt jedoch vom jeweiligen Ziel ab, das bei der Auswertung verfolgt werden soll. Aus diesem Grunde ist es nicht möglich, sich für eine allgemeine Festlegung zu entscheiden, besonders wenn verschiedene Merkmale unterschiedliche Tendenzen aufweisen. Im Rahmen dieser Arbeit wird ein allgemeiner Vergleich durchgeführt, der anhand gewählter Beispiele ein fundiertes Bild über Vor- und Nachteile der einzelnen Lösungen bietet. Dies erfordert, daß die einzelnen Kriterien separat ausgewertet werden.

3.5.1 Anforderungen an die Datenbankknoten

Die zwei ersten Kriterien, die Anforderungen an die Prozessor- und Speicherkapazität in den Datenbankknoten stellen, werden hier zusammen diskutiert.

3.5.1.1 Anforderung an die Prozessorkapazität

Die Ermittlung des absoluten Leistungsbedarfs der Prozessoren in den einzelnen Knoten ist eine schwierige Aufgabe, die erst nach der Implementierung wirklichkeitstreu gelöst werden kann. Erst dann kann man exakt den Aufwand ermitteln, der für einzelne Prozeduren notwendig ist. Darüber hinaus ist die eingesetzte Hardware maßgeblich für das Laufzeitverhalten verantwortlich.

Der hier gewählte Weg liefert lediglich eine vergleichende und vereinfachte Studie. Der Aufwand soll lediglich größenordnungsmäßig abgeschätzt werden. Zu diesem Zweck kann für verschiedene Operationen in den Knoten eine relative Komplexität definiert werden. Die damit ermittelte Gesamtknotenbelastung stellt dann für jeden Knoten ein Maß für die notwendige Prozessorleistungsfähigkeit dar. In der Arbeit werden drei Operationsarten unterschieden:

- Bearbeitung eines Schreibvorgangs,
- Bearbeitung eines Lesevorgangs und
- sonstiger Vorgang.

Um dem Problem einer numerischen Vorfestlegung der Komplexitätswerte aus dem Weg zu gehen, wurden die unterschiedlichen Operationen separat quantifiziert.

3.5.1.2 Anforderung an die Speicherkapazität

Wenn die notwendige Speicherkapazität betrachtet wird, sind zwei Aspekte von Bedeutung:

1. Speicherbedarf
2. Speicherdurchsatz

Der erste Aspekt, der Speicherbedarf, ist in der betrachteten UMTS-Datenbank von den folgenden Faktoren abhängig:

- Größe der einzelnen Datensätze,
- Größe des Systems und
- Datenduplikation der einzelnen Sätze.

Die zwei ersten Faktoren sind absolute Größen, d.h., sie sind durch äußere, nicht von der Datenbank abhängige Umstände vordefiniert. Aus diesem Grunde werden sie hier als gegeben und dadurch auch nicht optimierbar vorausgesetzt.

Der dritte Faktor ist von der gewählten Datenorganisationsmethode abhängig und wird als ein weiteres Kriterium angesehen. Es soll darauf hingewiesen werden, daß die Datenduplikation aufgrund der Sicherheitskopien in einzelnen Knoten nicht in die Auswertung einbezogen wird, da dies ein generelles Datenbankkonzept darstellt. Hier geht es um die Datenduplikation in mehreren Knoten, die durch das gewählte Datenverteilungsszenario verursacht wird.

In der Auswertung sind sowohl der Gesamtspeicherbedarf als auch der maximale Speicherbedarf von Bedeutung, da der Speicheraufwand bekanntlich nicht linear mit dem Speicherbedarf ansteigt.

Der zweite Aspekt, der Speicherdurchsatz, ist von folgenden Faktoren abhängig:

- Technologie,
- Größe der einzelnen Datensätze und
- Handhabung der Sicherheitskopien.

Der erste Faktor wird hier nicht betrachtet, da die Arbeit relative Ergebnisse liefert und keinen Anspruch auf absolute Aussagen erhebt. Der zweite Faktor soll hier ebenfalls unbeachtet bleiben, da die Größe des Datensatzes bei den erwarteten kleinen Datenmengen keinen nennenswerten Einfluß auf den Durchsatz haben wird /Bell92/. Der dritte Faktor kann für jeden einzelnen Knoten ebenfalls als gegeben angesehen werden und wird deswegen in der Arbeit nur insoweit beachtet, wie sein Einfluß auf die relative Komplexität (vgl. Kapitel 3.5.1.1) dies verlangt.

3.5.1.3 Knotenorganisation und dessen Speicherdurchsatz

Im Zusammenhang mit dem Speicherdurchsatz soll hier auf die Bedeutung der Knotenorganisation eingegangen werden (vgl. /Bell92/ und /SSD94/). Entscheidend für den Einfluß der Knotenorganisation auf die Leistungsfähigkeit ist die Relation zwischen den Lese- und Schreiboperationen. Zwei Aspekte, die im folgenden als unabhängig betrachtet werden, sind hier von entscheidender Bedeutung:

- die Art, wie man den Durchsatz des Knotens zu erhöhen versucht, und
- die Handhabung der Sicherheitskopien.

Zuerst sei angenommen, daß der Durchsatz eines Knotens nicht ausreichend ist. Zwei Methoden der Speichererweiterung wären dann grundsätzlich denkbar:

1. Die Datensätze werden vervielfältigt, und es gibt dabei keine gemeinsame Speichersteuerung, d.h. der Speicherknoten besteht aus quasi-unabhängigen Speicherblöcken.
2. Die Datensätze werden partitioniert und mit Hilfe einer gemeinsamen Tabelle gesteuert, d.h. der Speicherknoten besteht aus Speicherblöcken, die gemeinsam verwaltet werden.

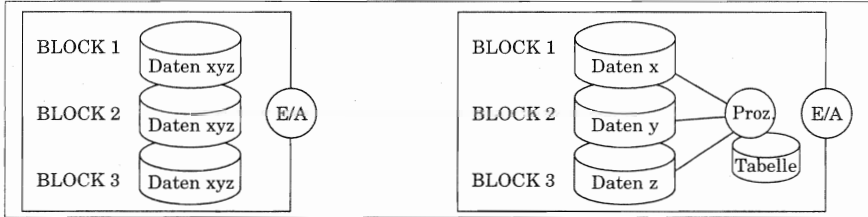


Abb. 3.13 Speichervervielfältigung (links) vs. Speicherpartitionierung (rechts)

Die erste Methode bringt eine Komplikation mit sich, die sich aufgrund der notwendigen Speicherkonsistenz ergibt. Wird eine Leseoperation durchgeführt, so ist diese Lösung von klarem Vorteil gegenüber einer Einblocklösung und erlaubt einen höheren Durchsatz. Wenn man jedoch die Schreiboperationen betrachtet, müssen sie wegen der Konsistenz auf allen Speicherblöcken durchgeführt werden. Als Schreiboperationen sind hier alle Modifizierungsoperationen zu verstehen.

Die Konsequenz ist, daß der Durchsatz eines Knotens nur unterproportional mit der Anzahl der Speichereinheiten ansteigt. Analytisch formuliert ergibt sich der folgende Zusammenhang zwischen der Auslastung der Speichereinheiten nach und vor der Speicherblockerweiterung:

$$\frac{\rho_{neu}}{\rho_{alt}} = \frac{N_{write} \cdot t_{write} + (N_{read} \cdot t_{read}) / Mem_No_{neu}}{N_{write} \cdot t_{write} + (N_{read} \cdot t_{read}) / Mem_No_{old}} \quad (3.1)$$

Hieraus ergibt sich:

$$\frac{\rho_{neu}}{\rho_{alt}} = \frac{1 + K / Mem_No_{neu}}{1 + K / Mem_No_{old}}, \text{ wobei } K = \frac{N_{read} \cdot t_{read}}{N_{write} \cdot t_{write}} \quad (3.2)$$

Es bedeuten:

N_{write}	Anzahl der Schreiboperationen
t_{write}	mittlere Zeit der Schreiboperationen
N_{read}	Anzahl der Leseoperationen
t_{read}	mittlere Zeit der Leseoperationen

Mem_No_{old} Anzahl der alten Speicherblöcke

Mem_No_{new} Anzahl der neuen Speicherblöcke

Das Maximum dieses Verhältnisses, das proportional zum maximalen Durchsatzgewinn ist, ergibt sich, wenn:

$$\begin{aligned} Mem_No_{old} &= 1 \text{ und} \\ Mem_No_{new} &\rightarrow \infty, \end{aligned}$$

d.h., wenn man die Speichererweiterung auf der Grundlage eines Einblockspeichers baut, der nun beliebig erweiterbar ist.

Dann erhält man als das bestmögliche Verhältnis:

$$\frac{\rho_{neu}}{\rho_{alt}} = \frac{1}{1 + K} \quad (3.3)$$

Der obige Quotient drückt die maximal erreichbare Verbesserung des Durchsatzes eines Knotens mit Hilfe der Speichervervielfältigung aus. Wenn die Anzahl der neuen Speicherblöcke nicht unendlich groß ist, dann erhält man eine geringere Verbesserung:

$$\frac{\rho_{neu}}{\rho_{alt}} = \frac{1 + K/Mem_No_{neu}}{1 + K} \quad (3.4)$$

Zur Anschaulichkeit ist diese Funktion in Abbildung 3.14 dargestellt.

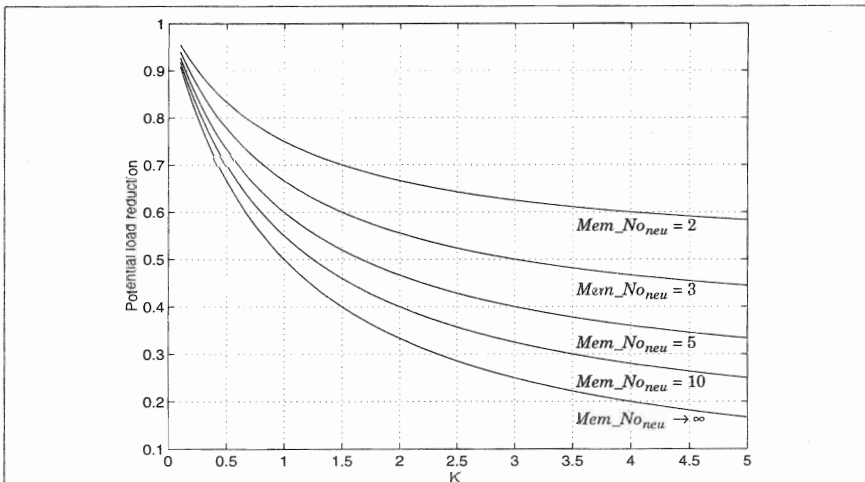


Abb. 3.14 Potentielle Auslastungsreduktion durch Speicherblockerweiterung

Obige Betrachtung macht deutlich, daß die Anwendbarkeit der Speichervervielfältigung vom Verhältnis K abhängig ist und die Grenzen dieses Verfahrens sehr schnell erreicht sind.

Als eine von K unabhängige Lösung bleibt die zweite Methode, in der die Datensätze partitioniert werden und jeweils als einzige Kopie vorhanden sind (die Duplizierung zwecks Datensicherung soll vorerst außer Betracht gelassen werden). Diese Lösung ist durch den Durchsatz des Speichers begrenzt, in dem die gemeinsame Tabelle angelegt ist. Der Durchsatz dieses Speichers muß nämlich in gleichem Maße erhöht werden wie der geforderte Gesamtknotendurchsatz. Diese Eigenschaft führt dazu, daß die zweite Möglichkeit ebenfalls nur begrenzt einsetzbar ist und die Skalierbarkeit durch die Technologie eingeschränkt ist.

Eine Kombination aus den beiden dargestellten Möglichkeiten kann insoweit eine Abhilfe schaffen, indem sie sowohl den Durchsatz mittels Technologie als auch mittels der Vervielfältigung erhöht.

Der 2. genannte Aspekt der Knotenorganisation ist die Handhabung der Sicherheitskopien in den einzelnen Knoten. Es sei vorausgesetzt, daß im betrachteten Knoten zusätzlich zur Hauptkopie eine Sicherheitskopie existiert.

Ein wesentlicher Einfluß auf den Durchsatz wird durch die Handhabung der Schreiboperationen ausgeübt. Zwei Methoden sind dabei denkbar:

1. Die Sicherheitskopie wird unabhängig von der Fertigstellungsantwort angelegt, die den Speicher des Knotens als bereit für die nächste Operation anmeldet.
2. Die Sicherheitskopie wird zuerst angelegt und danach die Fertigstellungsantwort an die zentrale Knotenverwaltung geschickt.

Bei Anwendung der ersten Methode ist der Verzögerungsunterschied zwischen den Lese- und Schreiboperationen von geringerer Bedeutung, da beide blockorientiert auf ein und denselben Speicherbereich erfolgen. Bei Anwendung der zweiten Methode ist dagegen die Verzögerung einer Schreiboperation wesentlich höher als die einer Leseoperation.

Die Wahl einer dieser Methoden hat deswegen auch einen großen Einfluß auf die relative Komplexität der Operationen und damit auf die Beurteilung der Qualität einzelner Datenverwaltungsszenarien.

3.5.2 Belastung der Verbindungen

Die Verteilung der Daten auf verschiedene Knoten der verteilten Datenbank und die Notwendigkeit, die Daten zu aktualisieren und zu verwalten, verursacht Datenverkehr auf den Verbindungen. Die Belastung der Verbindungen kann unterschiedlich gemessen werden.

Die maximale Belastung jeder einzelnen Verbindung entspricht der dort benötigten Bandbreite. Um diese zu ermitteln, ist es ausreichend, die Größe der gesendeten Nachrichten pro Zeiteinheit auf der jeweiligen Verbindung zu summieren.

Die summarische Belastung aller Verbindungen gibt einen Aufschluß über das Kommunikationsvolumen einer bestimmten Datenverwaltungsmethode.

Um eine Maßzahl für die Gesamtnutzung des Netzes zu erhalten, ist es zweckmäßig, die Größe der gesendeten Nachrichten mit der Entfernung zu wichten, auf der die einzelnen Nachrichten gesendet werden.

Daraus ergeben sich zwei Arten von Bewertungskriterien, die die Verbindungsbelastung charakterisieren:

- die Menge der Daten, die in einem bestimmten Datenbankszenario transportiert und entweder als
 - maximale Verbindungsmenge oder als
 - Gesamtmengein der Auswertung beachtet wird, und
- die entfernungsgewichtete Datenmenge, die zur Bewertung der Gesamtnutzung verwendbar ist.

Die obigen Ausführungen haben allgemeine Gültigkeit und sind von keinem speziellen Transportnetz abhängig. Die aufgrund dieser Kriterien ermittelten Zahlen stellen Mindestwerte an Verbindungskapazitäten dar, die in der verteilten Datenbank benötigt werden. Wenn man sich einer speziellen Netztopologie bedient, wie beispielsweise in Abbildung 3.2 präsentiert, dann ist die Verbindungsnutzung auch von der Topologie und vom Routingverfahren abhängig. In dieser Arbeit ist man jedoch bestrebt, die prinzipielle Situation zu bewerten, und deswegen werden lediglich die obigen idealen Mindestwerte miteinander verglichen.

3.5.3 Prozedurverzögerungszeit

Die Verzögerungszeit, die sich ergibt, wenn auf eine Antwort der Datenbank gewartet werden muß, kann ein sehr wichtiges Kriterium sein, das für oder gegen eine bestimmte Datenverwaltungsmethode spricht. Für die Anwendbarkeit dieses Kriteriums müssen jedoch zwei Bedingungen erfüllt sein:

1. Die ausgewertete Prozedur ist zeitkritisch, d.h., vom technischen Standpunkt aus oder aber aus Sicht des Benutzers darf eine bestimmte maximale Verzögerungszeit nicht überschritten werden. Die Wahrnehmung des Benutzers ist hier entscheidend, da es sich um die benutzerorientierte Dienstqualität handelt.
2. Die Verzögerungsunterschiede zwischen verschiedenen Methoden sind in ihren absoluten Werten relevant. Die Betonung liegt hier auf absolut. Wenn nämlich eine zeitkritische Prozedur vorliegt, ist die maximal zulässige Verzögerung in absoluten Einheiten spezifiziert und die absolute Verzögerung maßgebend.

Aufgrund der unterschiedlichen Anzahl der beteiligten Datenknoten während einer Anfrage werden die Verzögerungszeiten unterschiedliche Werte aufweisen. Die Verteilung der Werte wird von der Lage der Daten und des Ortes der Anfrageinitiierung abhängig sein. Es ist deswegen im allgemeinen notwendig, nicht nur mit der durch-

schnittlichen Verzögerungszeit t_{aver} zu arbeiten, sondern auch die minimale und maximale Verzögerungszeit t_{min} und t_{max} zu beachten.

In der Untersuchung im Kapitel 5 wird die Verzögerungszeit lediglich für die Suchanfrage numerisch ausgewertet. Die Verzögerungszeit der Aktualisierungsprozedur wird nicht quantitativ analysiert, da diese Prozedur ohnehin vom Benutzer nicht wahrgenommen wird und somit die erste der obigen Bedingungen nicht erfüllt.

3.5.4 Zuverlässigkeit und Fehlertoleranz

Eine Systemalternative wird im folgenden als zuverlässig bezeichnet, wenn sie nur gering fehleranfällig ist und wenn etwaige Fehler vom System automatisch erkannt und korrigiert werden. Dieses Kriterium entzieht sich in einem komplexen System wie der UMTS-Datenbank einem 'a priori'-meßbaren Vergleich. Zumindest ist das Ergebnis eines solchen Vergleichs sehr stark von der jeweiligen Technologie abhängig. Hier soll lediglich eine qualitative Diskussion wiedergegeben werden

Abgesehen von den Fehlern, die während der Kommunikation auftreten und durch die unteren OSI-Schichten korrigiert werden, können sich zwei grundsätzliche Fehlerarten ereignen, die von der Anwendungsschicht abgefangen werden müssen:

1. Die Nachrichtenfolge wird unterbrochen und die Durchführung der Aktion nicht fortgesetzt. In einem solchen Fall soll sich das Anwendungsprotokoll selber "reparieren", d.h., die Datenkonsistenz soll wieder hergestellt werden.
2. Ein Datenbankknoten, entweder ein Verzeichnis- oder ein Datenknoten, bricht zusammen - in einem solchen Fall muß es möglich sein, daß das Anwendungsprotokoll sowohl die Funktion dieses fehlerhaften Knotens durch einen anderen Knoten ersetzt als auch durch den Zugriff auf verteilte Sicherheitskopien (wenn die Kopie im Knoten selber nicht zugreifbar ist) die Funktionstüchtigkeit der Gesamtdatenbank aufrecht erhalten kann.

Üblicherweise wird die erste Fehlerart durch die Transaktionsprotokolle überwacht (vgl. Kapitel 2.1.2.4). Diese sind jedoch sehr aufwendig und sollten nur im Notfall angewendet werden - auch in einem fehlerfreien Fall wird dabei eine erhebliche Mehrbelastung des Netzes hervorrufen. Stattdessen kann durch einen geschickten Aufbau des Protokolls erzielt werden, daß die Konsistenz der Daten zu jeder Zeit der Anfrage erhalten ist. Dann ist auch möglich, an der Abbruchstelle mit der Anfrage fortzufahren, nachdem der Fehler erkannt und lokalisiert wurde (vgl. Kapitel 6.2).

Die zweite Fehlerart ist viel schwieriger zu reparieren. Es ist hier notwendig, daß sowohl die Protokoll- als auch die Datenbankarchitektur eine Fehlerkorrektur dieser Art vorsehen. Dies bedeutet im einzelnen, daß

- eine andere Zuordnung der DIR-Knoten zu den DN-Knoten, als ursprünglich vorgesehen, möglich sein muß, um den fehlerhaften Knoten zu übergehen, und
- die in den Datenknoten gespeicherte Information immer dupliziert wird.

Eine detaillierte Betrachtung der zweiten Fehlerart würde den Rahmen dieser Arbeit sprengen, da zusätzlich zur Systemanalyse eine exakte Spezifikation des verteilten und fehlertoleranten Datenbankprotokolls notwendig wäre.

Zusammenfassend kann festgehalten werden, daß die Aspekte der Zuverlässigkeit als wichtiges Kriterium in der Beurteilung angesehen werden. Aufgrund der mangelnden quantitativen Möglichkeiten, dieses Kriterium direkt anzuwenden, wird lediglich auf die Problematik hingewiesen, die mit den einzelnen vorgeschlagenen Lösungen verbunden ist.

3.5.5 Zusammenfassung der Bewertungskriterien

In den vergangenen Kapiteln wurden die Kriterien analysiert und deren Einfluß auf die Ergebnisse der Auswertungen geschildert. Einige Kriterien wurden dabei als nicht relevant, einige andere als nicht quantifizierbar eingestuft. Im Folgenden soll eine kurze Zusammenfassung dieser Kriterien gegeben werden, die bei den Auswertungen im Kapitel 5 benutzt werden:

- notwendige relative Prozessorkapazität, ausgedrückt durch die Anzahl der unterschiedlichen Operationen in den Knoten - dadurch wird die Möglichkeit gegeben, sowohl die Gesamtanzahl der Operationen als auch die relative Knotenauslastung (unter Zuhilfenahme der Komplexitäten) einfach zu ermitteln,
- notwendiger Speicherdurchsatz bei einer festgelegten Handhabung der Sicherheitskopien, was sich ebenfalls aus der Anzahl der unterschiedlichen Operationsarten an den Knoten ermitteln läßt,
- notwendiger Speicherbedarf aufgrund der Datenduplikation der einzelnen Datenunterklassen,
- verursachte Verbindungsnutzung, bewertet durch:
 - absolute Gesamtdatenmenge und
 - gewichtete Gesamtdatenmenge, d.h. Gesamtverbindungsnutzung unter Beachtung der Transportentfernung,
- minimale, durchschnittliche und maximale Verzögerungszeit - nur für die zeitkritische Suchanfrage.

Die Fehlertoleranz wird nur qualitativ ausgewertet und dient eher einer Orientierung als einer konkreten Empfehlung.

4 Analytisches Modell

Nach den allgemeinen Bemerkungen zu dem gewählten Modellierungsansatz werden die Benutzerverhaltenmodelle präsentiert. Die darauf folgende formale Beschreibung der einzelnen Anfragearten schließt die Auswertungsvorbereitung ab.

4.1 Allgemeine Bemerkungen

Das Benutzerverhalten wird durch zwei unterschiedliche Modelle beschrieben: ein Mobilitäts- und ein Anrufverteilungsmodell. Die beiden Modelle sind quasi-unabhängig, d.h., die Mobilität ist autonom, aber die Anrufverteilung ist zum Teil durch die Benutzerverteilung bestimmt. Die Modelle werden als Grundlage der Leistungsbewertung von der Aktualisierungs- und Suchanfrage im Kapitel 5 verwendet.

Die Beschreibung ist separat für die Ereignisse zwischen den Teilgebieten (*inter-domain*-Verhalten) und innerhalb eines Teilgebietes (*intra-domain*-Verhalten) gegeben. Die Modellierung des *inter-domain*-Verhaltens wird stark vereinfacht, da die Kommunikation zwischen den Teilgebieten kaum einen Spielraum für Verbesserungen übrig läßt - sie ist durch Standardisierung und Sicherheitsbestimmungen geprägt. Im Gegensatz dazu bietet die Kommunikation innerhalb jedes Teilgebietes großes Optimierungspotential (vgl. Kapitel 3.4), und deswegen wird sie relativ genau modelliert.

Bei der Auswertung soll folgendes beachtet werden: Aufgrund der Tatsache, daß sich in einem UMTS-Teilgebiet sowohl Heim- als auch Besuchsbewerber aufhalten können, die sowohl eine *intra-domain*- als auch eine *inter-domain*-Kommunikation betreiben, wird sich die insgesamt erzeugte Verkehrslast als eine Überlagerung aus der Kommunikation innerhalb des Teilgebietes und zwischen den Teilgebieten ergeben. Die letztere wird jedoch nur insofern betrachtet, als sie einen Beitrag zur Verkehrslast innerhalb des Teilgebietes liefert. Als Verkehrslast wird hier die Knoten- und Verbindungsbelastung bezeichnet.

4.2 Mobilitätsmodell

Das Mobilitätsmodell soll die Benutzerbewegungen möglichst realistisch und flexibel wiedergeben. Ein realistisches Modell vermag das komplexe Verhalten in seinen wesentlichen Eigenschaften korrekt zu beschreiben (/SSD94/ und /Jain91/). Ein flexibles Modell ist imstande, mittels einiger Parameter an verschiedene Verhaltensmuster des Benutzers angepaßt zu werden. Es erlaubt auf diese Art, Untersuchungen des Systems in verschiedenen Situationen vorzunehmen, z.B. sind die Benutzer mehr oder weniger mobil, oder sie bewegen sich geordnet statt stationär.

Das hergeleitete Mobilitätsmodell erfüllt diese Anforderungen, indem es die Benutzerverteilung und -bewegungen mit großer Genauigkeit beschreibt (/SSD94/, /Mone61/) und verschiedene Situationen mittels Parameter nachbilden kann.

4.2.1 Bewegungen innerhalb eines Teilgebietes

4.2.1.1 Allgemeine Annahmen

Die geographische Topologie eines Teilgebietes, d.h. die räumliche Aufteilung des Gebietes in einzelne Datenknotengebiete (DN-Gebiete), wird im künftigen UMTS-Netz durch die Dienstversorgungsanforderungen vorbestimmt. Mangels solcher Anforderungen in der jetzigen Entwicklungsphase des UMTS-Systems und wegen des hier erhobenen Anspruches auf eine allgemeine Lösung wird im Modell eine Grundtopologie verwendet, die sich aus hexagonalen DN-Gebieten zusammensetzt. Die Verwendung anders geformter, z.B. rechteckiger DN-Gebiete hat grundsätzlich keinen Einfluß auf den Modellierungsansatz, würde jedoch zur Verkomplizierung bei der Beschreibung von 'Eckenbewegungen' führen. Der hier vorliegende Ansatz ist ähnlich zu dem, der üblicherweise (/Fuhr90/ und /Pers93/) bei der Modellierung der Funkzellen in mobilen Netzen, z.B. in GSM, verwendet wird. Weil die räumliche Ausdehnung einzelner DN-Gebiete viel größer als die der Funkzellen ist, ist es auch hier berechtigt, lediglich den statistischen Benutzer zu betrachten.

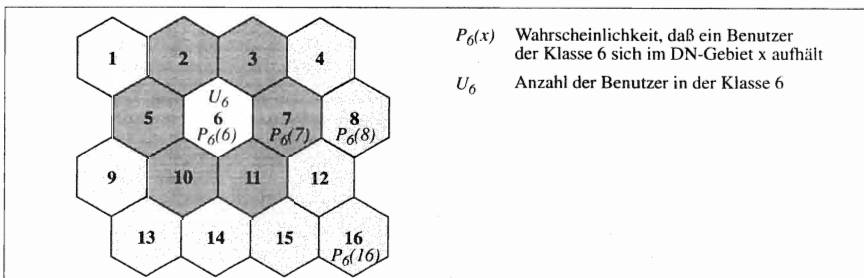


Abb. 4.1 Beispiel einer Topologie

Das in Abbildung 4.1 dargestellte Teilgebiet besteht aus 16 DN-Gebieten (allgemein: aus N DN-Gebieten). In diesem Beispiel ist der Datenknoten 6 der sog. Heimknoten (*resicent node*), die stark schattierten Datenknoten stellen die 6 Nachbargebiete dar (*neighbour nodes*), und die leicht schattierten Knoten markieren die externen DN-Gebiete (*external nodes*) zum Heimknoten 6.

Die jetzt folgende Herleitung basiert auf zwei Grundannahmen:

1. Die Anzahl der Heimbenutzer ist in jedem DN-Gebiet vorgegeben.
2. Die Verteilungswahrscheinlichkeit der Benutzer ist im jeweiligen Teilgebiet bekannt.

Beide Annahmen stellen wegen der Flexibilität des Modells keine ernsthaften Restriktionen dar.

Die erste Annahme ist sehr leicht zu modellieren. Wegen der Möglichkeit, daß die Anzahl von Heimbenutzern unterschiedliche Werte in verschiedenen Teilgebieten,

aufweisen kann, wird ein Vektorparameter U eingeführt, so daß jede Komponente U_r die Anzahl der Heimbewohner im Gebiet r beschreibt. Benutzer, die einem Heimknoten angehören, werden als eine Benutzerklasse bezeichnet.

Die zweite Annahme verlangt mehr Aufmerksamkeit. Um die Möglichkeit zu berücksichtigen, daß die Benutzer, die unterschiedlichen DN-Heimgebieten angehören, unterschiedliche Gewohnheiten haben können, wird die Wahrscheinlichkeitsverteilung unabhängig für jede dieser Benutzerklassen definiert. Dieser Zusammenhang wird mit Hilfe des Matrixparameters P erfaßt, dessen Vektorkomponente P_r die Verteilung der Benutzerklasse aus dem DN-Heimgebiet r beschreibt und zwar so, daß $P_r(i)$ die Wahrscheinlichkeit angibt, mit welcher die Benutzer aus dem DN-Heimgebiet r sich im DN-Gebiet i innerhalb des eigenen UMTS-Heimteilgebietes aufhalten. Mit anderen Worten beschreibt $P_r(i)$ die aktuelle Verteilung der Benutzeraufenthaltsorte. Die absolute Anzahl der Benutzer in einem DN-Gebiet i , die im DN-Gebiet r ihr Heim haben, läßt sich als Produkt $U_r \cdot P_r(i)$ ausdrücken.

Die Möglichkeit für die Anwendung der obigen Beschreibung bietet beispielsweise ein UMTS-Szenario, in dem sowohl die städtische, vorörtliche als auch dörfliche Umgebung gleichzeitig modelliert werden soll. Die Heimbewohneranzahl in den einzelnen DN-Gebieten definiert den Vektorparameter U . Die Werte von P_r können näherungsweise mit Hilfe einer Beschreibung der charakteristischen Aufenthaltsorte der Benutzer, z.B. der Arbeitsstätten, Einkaufs- oder Unterhaltungszentren, ermittelt werden. Wenn eine detaillierte Untersuchung eines speziellen UMTS-Teilgebietes durchgeführt werden soll, ist es selbstverständlich notwendig, empirische Messungen vorzunehmen, um die korrekten Werte der Vektoren P_r zu ermitteln. Die einzelnen DN-Gebiete stellen die Grundlage des Modells dar: Jedes einzelne DN-Gebiet wird durch einen separaten Zustand repräsentiert.

Die Werte der Verteilung des Benutzerstandortes sind wichtige Grunddaten sowohl für das Mobilitäts- als auch das Anrufmodell:

- Die Benutzerbewegungen werden aufgrund der Verteilung des Benutzeraufenthaltsortes ermittelt.
- Die Anrufverteilung ist von der Wahrscheinlichkeitsverteilung des Aufenthaltsortes abhängig.

Der im folgenden oft verwendete Begriff 'Bewegung' soll hier definiert werden: Von einer Bewegung des Benutzers im Sinne des folgenden Modells wird gesprochen, wenn der Benutzer sich zwischen zwei sog. Aufenthaltsgebieten (*Location Areas - LA*) bewegt hat. Eine solche Bewegung entspricht im Modell einem Zustandsübergang. Wenn ein DN-Gebiet aus mehreren Aufenthaltsgebieten besteht, ist es möglich, daß Bewegungen innerhalb eines DN-Gebietes erfolgen. Dies entspricht im Modell einem Übergang aus einem Zustand in sich selbst. Eine umgekehrte Situation, d.h. ein Aufenthaltsgebiet, das sich aus mehreren DN-Gebieten zusammensetzt, sei hier ausgeschlossen. Eine solche Situation würde erhebliche Schwierigkeiten bereiten: Ein Wechsel des DN-Gebietes würde nicht erkannt, obwohl sich der Benutzer de facto im Versorgungsbereich eines neuen DN-Knotens befinden würde.

Üblicherweise sind die Benutzerbewegungen durch die Aktualisierungsprozeduren (*Location Update* oder *Domain Update*) begleitet, vorausgesetzt, die Bewegungen werden vom Netz registriert (vgl. auch Erklärungen im Kapitel 3.3.1.1 und 3.3.1.2).

Ein einzelnes UMTS-Teilgebiet ist ein offenes System, und als solches kann es im Modell als endlicher Zustandsautomat mit externen Ankünften und Abgängen als Markovketten beschrieben werden. Die Ankunfts- und Abgangsraten in jedem DN-Gebiet j , die durch die Benutzer verursacht werden, dessen Heimknoten sich im DN-Gebiet r befindet, können nach /Bolc91/ durch folgende Gleichungen dargestellt werden:

$$A_r(j) = \lambda_r(0, j) + \sum_{i=1}^N \lambda_r(i) \cdot p_r(i, j) \quad j = 1, 2, \dots, N \quad (4.1)$$

$$D_r(j) = \lambda_r(j, 0) + \sum_{i=1}^N \lambda_r(j) \cdot p_r(j, i) \quad j = 1, 2, \dots, N \quad (4.2)$$

Dabei bedeuten:

$\lambda_r(k)$	die Gesamtrate der Benutzerbewegungen im DN-Gebiet k
$\lambda_r(0, j)$	die Ankunftsrate von außerhalb dieses UMTS-Teilgebietes in das DN-Gebiet j
$\lambda_r(j, 0)$	die Abgangsrate aus dem DN-Gebiet j in ein anderes UMTS-Teilgebiet
$p_r(i, j)$	die bedingte Übergangswahrscheinlichkeit aus dem DN-Gebiet i in das DN-Gebiet j
$A_r(j), D_r(j)$	die Ankunftsrate in das bzw. Abgangsrate aus dem Gebiet j
N	die Anzahl der Gebiete

Es wird angenommen, daß die Bewegungen in einem DN-Gebiet proportional der Benutzeranzahl in diesem Gebiet sind:

$$\lambda_r(i) = k_M \cdot U_r \cdot P_r(i) \quad (4.3)$$

Dabei ist k_M der Proportionalitätsfaktor der Mobilität.

Unter Verwendung von (4.3) können (4.1) und (4.2) umgeschrieben werden:

$$A_r(j) = \lambda_r(0, j) + k_M \cdot U_r \cdot \sum_{i=1}^N P_r(i) \cdot p_r(i, j) \quad j = 1, 2, \dots, N \quad (4.4)$$

$$D_r(j) = \lambda_r(j, 0) + k_M \cdot U_r \cdot P_r(j) \cdot \sum_{i=1}^N p_r(j, i) \quad j = 1, 2, \dots, N \quad (4.5)$$

Folgende zwei Ansätze sind bezüglich der Werte von $p_r(i, j)$ denkbar:

1. Wenn Bewegungen der Benutzer unabhängig vom Status des Terminals (*attached/detached*) immer durch das Netz erfaßt werden könnten, dann muß $p_r(i, j) = 0$ für alle DN-Gebiete i und j gelten, die nicht benachbart sind.

2. Wenn nur manche Bewegungen der Benutzer erfaßt werden, dann ist es möglich, daß $p_r(i, j) \neq 0$ für zwei nicht benachbarte DN-Gebiete i und j gilt.

In der Arbeit wird angenommen, daß die Aktualisierung der Datenbank infolge der Benutzermobilität (*Location Update* und *Domain Update*) nach dem ersten Ansatz zu modellieren ist. Es wird hier vereinfachend angenommen, daß jede Benutzerbewegung während des Aufenthaltes im neuen Aufenthaltsgebiet erkannt wird. Die *User Registration*- oder *Session Set-Up*-Prozeduren müßten aber nach dem zweiten Ansatz modelliert werden, da dort durchaus zu erwarten ist, daß der Benutzer sich zwischenzeitlich vom Netz trennt und erst später erneut verbindet.

Für die Bewegungen in den DN-Gebieten an der Teilgebietsgrenze wird in dem zuerst betrachteten stationären Fall angenommen, daß sich die Abgangs- und Ankunftsdaten separat in jedem DN-Gebiet im Gleichgewicht befinden. Drei Gründe sprechen für diese Vereinfachung:

1. Soll diese Annahme fallengelassen werden, so müßte ein zusätzlicher frei wählbarer Parameter eingeführt werden, der den Zusammenhang zwischen den externen Ab- und Zugängen beschreibt. Dies erhöht den Spezifizierungsaufwand erheblich, ohne exaktere Ergebnisse zu liefern, da die Kommunikation zwischen den Teilgebieten dadurch nur unwesentlich beeinflusst wird.
2. Die nichtstationären Lösungen, wie im Kapitel 4.2.1.3 und 4.2.1.4 präsentiert, können näherungsweise zur Beschreibung des Ungleichgewichtsfall herangezogen werden.
3. Das Ziel der vorliegenden Modellierung ist primär, das Verhalten innerhalb eines UMTS-Teilgebietes detailliert zu beschreiben, um eine Verbesserung der Leistungsfähigkeit zu erreichen. Die Kommunikation zwischen den einzelnen Teilgebieten soll demnach vereinfacht modelliert werden.

Mit der obigen Annahme erhält man ein neues System, das wie ein geschlossenes Wartenetz betrachtet werden kann:

$$A'_r(j) = k_M \cdot U_r \cdot \sum_{i=1}^N P_r(i) \cdot p_r(i, j), \text{ mit } A'_r(j) = A_r(j) - \lambda_r(0, j) \quad (4.4')$$

$j = 1, 2, \dots, N$

$$D'_r(j) = k_M \cdot U_r \cdot \sum_{i=1}^N P_r(j) \cdot p_r(j, i), \text{ mit } D'_r(j) = D_r(j) - \lambda_r(j, 0) \quad (4.5')$$

$j = 1, 2, \dots, N$

Die Charakteristik der Bewegungen zwischen den Teilgebieten wird separat berechnet:

- Die Rate (damit auch die Wahrscheinlichkeit) des Ereignisses, daß Benutzer das betrachtete UMTS-Teilgebiet im DN-Gebiet j betreten, wird wie in der Formel (4.6) modelliert:

$$\lambda_r(0, j) = k_M \cdot U_r \cdot \left(1 - \sum_j P_r(j) \right) \cdot P_r(j) \cdot p_r(0, j) \quad j = 1, 2, \dots, N \quad (4.6)$$

- Wegen der obigen Gleichgewichtsannahme gilt ähnliches für die Bewegungen aus dem UMTS-Teilgebiet:

$$\lambda_r(j, 0) = \lambda_r(0, j) \quad j = 1, 2, \dots, N \quad (4.7)$$

- Die in den Formeln (4.6) und (4.7) benötigte Übergangswahrscheinlichkeit wird folgendermaßen berechnet:

$$p_r(0, j) = p_r(j, 0) = cross \cdot \frac{len(j)}{\sum_{i=1}^N P_r(i) \cdot len(i)} \quad j = 1, 2, \dots, N \quad (4.8)$$

Dabei gelten folgende Abkürzungen:

$len(j)$	Anteil der Länge der Teilbegietsgrenze, die innerhalb des DN-Gebietes j verläuft
$cross$	Verhältnis der Bewegungen über die Teilbegietsgrenze zu allen Bewegungen in den Besuchsteilgebieten

Der Parameter $cross$ legt die absolute Anzahl der Bewegungen über die Teilbegietsgrenze fest:

- Bewegungen über die Heimteilbegietsgrenze - $cross$ wird als Quotient aus der Anzahl der Bewegungen über die Heimteilbegietsgrenze und aller Bewegungen in den Besuchsteilgebieten ausgedrückt (wie in 4.6).
- Bewegungen über die Besuchsteilbegietsgrenze - $cross$ wird als Quotient aus der Anzahl der Bewegungen über die Besuchsteilbegietsgrenze und aller Bewegungen in den Besuchsteilgebieten ausgedrückt .

Bemerkung:

In der Herleitung wurde stillschweigend angenommen, daß das Heimteilgebiet betrachtet wird. Ähnliche Formeln gelten für jedes Besuchsteilgebiet.

4.2.1.2 Stationäre Lösung

Zuerst wird explizit der stationäre Fall diskutiert, in dem die Anzahl der Benutzer in allen DN-Gebieten des UMTS-Teilgebietes im betrachteten Zeitintervall konstant bleibt, d.h., die Ankunftsrate ist überall gleich der Abgangsrate. Dann muß folgendes System von homogenen, linearen Gleichungen erfüllt werden:

$$A'_r(j) - D'_r(j) = 0 \quad j = 1, 2, \dots, N \quad (4.9)$$

Mit (4.4') und (4.5') erhält man (nach einer Division durch $k_M \cdot U(r)$):

$$\sum_{i=1}^N P_r(i) \cdot p_r(i, j) - \sum_{i=1}^N P_r(j) \cdot p_r(j, i) = 0 \quad j = 1, 2, \dots, N \quad (4.10)$$

Das System (4.10) beschreibt N Gleichungssysteme (für jede Benutzerklasse r), jedes aus N Gleichungen bestehend.

Die Werte der Wahrscheinlichkeitsverteilung P_r werden aus der Charakteristik des jeweiligen Teilgebietes hergeleitet. Für die vollständige Beschreibung des Benutzerverhaltens werden noch die Übergangswahrscheinlichkeiten $p_r(j, i)$ benötigt.

Wenn man das System (4.10) für die Benutzer einer Klasse betrachtet, so hat man N^2 Variablen $p_r(j, i)$ und maximal $(N-1)$ linear unabhängige Gleichungen (Klein76/). Um dieses Dilemma zu lösen, wird das Wissen genutzt, das man von den Benutzern hat, und ein Modell vorgeschlagen, das die Übergangs- mit den Zustandswahrscheinlichkeiten in eine einfache, plausible Relation setzt.

Der vorgeschlagene Lösungsweg beruht auf dem Ansatz, daß die bedingte Übergangswahrscheinlichkeit aus dem DN-Gebiet i in das DN-Gebiet j proportional zur Zustandswahrscheinlichkeit im Gebiet j ist. Die Proportionalitätsfaktoren existieren spezifisch für jedes einzelne DN-Zielgebiet und stellen die neuen Variablen des Systems dar (insgesamt N Variablen). Die bis jetzt fehlenden Übergangswahrscheinlichkeiten werden mittels der Variablen ausgedrückt. Die Normierung wird mit der Summe der Zustandswahrscheinlichkeiten im DN-Gebiet i und in allen DN-Gebieten j vorgenommen, wo $p_r(i, j) \neq 0$ gilt.

Die Proportionalitätsfaktoren können auch als Korrekturfaktoren bezeichnet werden: Von 1 unterschiedliche Faktorwerte sind nur dann notwendig, wenn nicht alle DN-Gebiete benachbart sind; die Faktoren beschreiben die notwendige Korrektur.

Zur Vereinfachung der numerischen Berechnung der Aktualisierungsprozedur wurde angenommen, daß alle Bewegungen durch das Netz erfaßt werden. Diese Annahme stützt sich auf drei Begründungen:

1. Der Fall, daß die Bewegung jedesmal durch das Netz erfaßt wird, ist der ungünstigste und verursacht den meisten Verkehr. Das Netz soll für diese Verkehrssituation ausgelegt werden.
2. Der Anteil der sonstigen Fälle, in denen die Bewegung nicht direkt registriert wird, wird voraussichtlich aus folgenden zwei Gründen niedrig sein:
 - Da die Benutzer im Regelfall erreichbar sein wollen, sind lediglich kurze Zeitintervalle zu erwarten, in denen das Endgerät keinen Kontakt zum Netz hat. Möchte der Benutzer sich vor unerwünschten Anrufen schützen, kann er das durch einen Eintrag in den *registration*-Datensatz erreichen und braucht nicht dafür das Endgerät abzuschalten.
 - Die Technologiefortschritte auf dem Gebiet des Batteriebaus machen das Ausschalten zwecks Batterieschonung überflüssig.
3. Ein Versuch, die Bewegungen zwischen nicht benachbarten DN-Gebieten quantitativ zu erfassen, ist auf spekulative Annahmen angewiesen, da jegliche Erfahrungswerte fehlen.

Formal kann der vorgeschlagene Ansatz zur Berechnung von Übergangswahrscheinlichkeiten wie folgt geschrieben werden:

$$p_r(i, j) = \frac{f_r(j) \cdot P_r(j)^{x_r}}{\sum_{k=1}^S P_r(k) + P_r(i) + P_r(i) \cdot p_r(i, 0) / \text{cross}} \quad (4.11)$$

Dabei sind:

- $f_r(j)$ der oben beschriebene Proportionalitätsfaktor
- x_r Parameter, dessen Wert zusätzlich die Abhängigkeit der Übergangswahrscheinlichkeiten beeinflusst:
wenn $x_r > 1$, dann werden die Unterschiede zwischen den Werten der Zustandswahrscheinlichkeiten verstärkt,
wenn $x_r < 1$, dann werden die Unterschiede zwischen den Werten der Zustandswahrscheinlichkeiten abgeschwächt
- S die DN-Gebiete, die die direkten Nachbarn des DN-Gebietes i sind (oder allgemeiner, wo $p_r(i, j) \neq 0$ gilt)

Basierend auf dem obigen Ansatz erhält man maximal $(N - 1)$ unabhängige Gleichungen mit $(N + 1)$ Variablen: N Proportionalitätsfaktoren $f_r(j)$ und einen Parameter x_r .

Die exakte Anzahl der unabhängigen Gleichungen ist durch die Anzahl der zyklischen Gruppen (Markovketten) im betrachteten Gleichungssystem definiert (/Gros85/). Das Auftreten dieses Phänomens ist auf die physikalische Tatsache zurückzuführen, daß innerhalb eines Teilgebietes einzelne Gruppen einer Benutzerklasse sich z.B. nur zwischen bestimmten DN-Gebieten bewegen. Dadurch entstehen Benutzerunterklassen, die ein eigenes Bewegungsmuster haben und die jeweils durch eine Teilmenge der ursprünglichen Gleichungen beschrieben werden. Jede Unterklasse, die aus n Zuständen besteht, wird durch $(n - 1)$ unabhängige Gleichungen beschrieben (/Klein76/). Die im folgenden hergeleitete Lösung ist für den Spezialfall anwendbar, daß alle DN-Gebiete des betrachteten Teilgebietes zu einer zyklischen Gruppe gehören, d.h., $(N - 1)$ Gleichungen sind voneinander unabhängig. Die Lösungen für mehrere zyklische Gruppen sind ähnlich zu konstruieren. Der notwendige zusätzliche Schritt ist, daß die einzelnen zyklischen Gruppen definiert und ihre Matrix der Übergangswahrscheinlichkeiten festgelegt werden müssen (vgl. BCMP-Netze in /Walk92a/). In der numerischen Lösung, die in dieser Arbeit zum Zwecke der quantitativen Analyse implementiert wurde, ist die Existenz mehrerer Gruppen vorgesehen, wobei angenommen wurde, daß die zyklischen Gruppen in sich geschlossen sind.

Da jetzt $(N - 1)$ Gleichungen unabhängig sind, dürfen zwei Variablen frei festgelegt werden. Es ist naheliegend, die Werte von x_r und von einem Proportionalitätsfaktor vorzuwählen. Die verbleibenden $(N - 1)$ Proportionalitätsfaktoren können dann mittels der $(N - 1)$ Gleichungen einfach numerisch berechnet werden.

Der Wert von x_r wird in Abhängigkeit der angestrebten Benutzermobilität festgelegt: Wenn eine höhere Mobilität modelliert wird, ist es zweckmäßig, den Wert von x_r klein zu wählen. Im umgekehrten Fall wird ein großer Wert von x_r gewählt.

Die Wahl des Wertes für den freien Proportionalitätsfaktor ist komplizierter. Nicht alle mathematischen Lösungen entsprechen einer möglichen physikalischen Wirklichkeit, z.B. negative Zahlenwerte für die Parameter widersprechen dem Lösungsansatz. Um physikalisch mögliche Lösungen zu erhalten, ist es notwendig, die Grenzbedingungen zu beachten. Zu diesem Zweck soll das Gleichungssystem (4.10) allgemein gelöst werden.

Aufgrund der Linearität aller Lösungen des Gleichungssystems, kann die Menge der möglichen Lösungen wie folgt dargestellt werden:

$$f_r = A_r + B_r \cdot \alpha \quad (4.12)$$

Es bedeuten:

α	reeller Parameter
f_r	Spaltenvektor, aus allen Proportionalitätsfaktoren $f_r(i)$ bestehend
A_r, B_r	Spaltenvektoren, die als Lösungen von (4.10) für zwei beliebige Werte des freien Proportionalitätsfaktors ermittelt werden

Wenn A_r gleich dem frei wählbaren Faktor $f_r(k)$ ist, dann können A_r und B_r als

$$\begin{aligned} A_r &= f_r, \alpha = 0 & \text{und} \\ B_r &= f_r, \alpha = 1 - A \end{aligned}$$

gerechnet werden.

Die obere Grenze der erlaubten Faktorwerte ist durch die Benutzeranzahl festgelegt, die sich bereits in jedem DN-Gebiet befinden. Die Gesamtanzahl der Benutzer, die jedes einzelne Gebiet während einer Beobachtungsperiode verlassen, kann nicht höher sein als die Anzahl der Benutzer, die sich am Anfang der Periode in jedem einzelnen DN-Gebiet befinden. Diese Bedingung ist an zwei Annahmen geknüpft:

- Die Benutzer bewegen sich während der Beobachtungsperiode maximal einmal zwischen zwei Aufenthaltsgebieten. Daraus folgt, daß
- Benutzer, die ein Aufenthaltsgebiet verlassen, sich bereits am Anfang der Beobachtungsperiode in diesem Aufenthaltsgebiet befinden, d.h., es gibt keine Bewegung über mehrere Aufenthaltsgebiete.

Die untere Grenze der erlaubten Faktorwerte ergibt sich aus der Bedingung, daß die Anzahl der Benutzer, die jedes einzelne DN-Gebiet verlassen, positiv sein muß.

Formal können beide Grenzbedingungen wie folgt geschrieben werden:

$$0 < d_r \cdot f_r < U_r \cdot P_r - \lambda_r(0) \quad (4.13)$$

Es gelten folgende Abkürzungen:

d_r	Abgangsmatrix, die die Abgänge aus jedem DN-Gebiet beschreibt: $d_r(j, i) = k_M \cdot U_r \cdot P_r(j) \cdot p_r(j, i)$ drückt die Abgänge aus dem DN-Gebiet j in das Gebiet i aus
-------	--

$\lambda_r(0)$ Spaltenvektor, der die Abgänge aus jedem DN-Gebiet in ein externes Teilgebietes beschreibt (vgl. Relation 4.7)

Die Bedingung (4.13) kann mit Hilfe von (4.12) umgeschrieben werden:

$$-d_r \cdot A_r < (d_r \cdot B_r) \cdot \alpha < U_r \cdot P_r - \lambda_r(0) - d_r \cdot A_r \quad (4.13')$$

Die obige Bedingung stellt ein System von N Einzelbedingungen für jedes DN-Gebiet dar. Das bedeutet, daß der Wert des Parameters α so festgelegt werden muß, daß jede einzelne Bedingung erfüllt ist. Daraus folgt, daß diese Bedingung umgeschrieben werden kann:

$$\max \left(\frac{d_r \cdot A_r}{d_r \cdot B_r} \right) < \alpha < \min \left(\frac{U_r \cdot P_r - \lambda_r(0) - d_r \cdot A_r}{d_r \cdot B_r} \right) \quad (4.14)$$

Unter Verwendung bekannter Relationen für d_r und für $\lambda(0)$ erhält man endgültig

$$\max \left(\frac{p_r \cdot A_r}{p_r \cdot B_r} \right) < \alpha < \min \left(\frac{1 - k_M \cdot \left(1 - \sum_j P(j) \right) \cdot p_r(0) - k_M \cdot p_r \cdot A_r}{k_M \cdot p_r \cdot B_r} \right), \quad (4.15)$$

wobei p_r die Matrix der Übergangswahrscheinlichkeiten darstellt, die alle einzelnen Wahrscheinlichkeiten $p_r(i, j)$ enthält und $p_r(0)$ die Summe der Abgangswahrscheinlichkeiten in ein externes Teilgebiet beschreibt.

Im folgenden soll eine spezielle Situation betrachtet werden, in der angenommen wird, daß die stationäre Lösung ausschließlich für alle Benutzer zusammen erfüllt ist und nicht wie bis jetzt für jede Benutzerklasse separat.

Der folgende Ansatz ist grundsätzlich äquivalent zu dem, der bereits hergeleitet worden ist, wobei die Beschreibung umfangreicher wird.

Die Bedingung für das lokale Gleichgewicht kann jetzt wie folgt formuliert werden:

$$\sum_{r=1}^N \sum_{i=1}^N U_r \cdot P_r(i) \cdot p_r(i, j) - \sum_{r=1}^N \sum_{i=1}^N U_r \cdot P_r(j) \cdot p_r(j, i) = 0 \quad j = 1, 2, \dots, N \quad (4.16)$$

Die Gleichungen können mittels durchschnittlicher Zustandswahrscheinlichkeiten umgeschrieben werden:

$$\sum_{i=1}^N P(i) \cdot p(i, j) - \sum_{i=1}^N P(j) \cdot p(j, i) = 0, \quad j = 1, 2, \dots, N \quad (4.17)$$

wobei die durchschnittliche Zustands- und Übergangswahrscheinlichkeit wie folgt definiert sind:

$$P(i) = \frac{\sum_{r=1}^N U_r \cdot P_r(i)}{\sum_{r=1}^N U_r} \quad (4.18)$$

$$p(i, j) = \frac{f(j) \cdot P(j)^x}{\sum_{k=1}^S P(k) + P(i) + P(i) \cdot p(i, 0)} \quad (4.19)$$

Die Bedingung (4.13) kann jetzt umgeschrieben werden:

$$0 < d \cdot f < U \cdot P - \lambda(0), \quad (4.20)$$

wobei folgende Abkürzungen gelten:

- d Abgangsmatrix, die die Abgänge aus jedem DN-Gebiet beschreibt, so daß
- $$d(j, i) = k_M \cdot \sum_{r=1}^N U_r \cdot P(j) \cdot p(j, i)$$
- die Abgänge aus dem DN-Gebiet j in das Gebiet i ausdrückt
- $\lambda(0)$ Spaltenvektor, der die Abgänge aus jedem DN-Gebiet in ein externes Teilgebiet beschreibt

Die obige Darstellung macht deutlich, daß die r -Abhängigkeit, die in der vorherigen Ableitung immer präsent war, nun nicht mehr besteht. Analog zu (4.14) erhält man:

$$\max\left(-\frac{d \cdot A}{d \cdot B}\right) < a < \min\left(\frac{\sum_{r=1}^N U_r \cdot P_r(j) - \lambda(0) - d \cdot A}{d \cdot B}\right) \quad (4.21)$$

Eine weitere Vereinfachung analog zur Bedingung (4.15) ist hier nicht möglich, da die Ausdrücke sich nicht entsprechend kürzen lassen.

Bei der Berechnung von unterschiedlichen Szenarien der Aktualisierungsanfrage ist es notwendig zu wissen, welchen Wert die Übergangswahrscheinlichkeit $p_r(i, j)$ hat. Im obigen r -unabhängigen Ansatz sind diese Werte nicht berechnet worden, da die Gleichungen die Benutzerklassen nicht explizit beachten. Der einzige Weg, dieses Problem zu lösen, ist eine Schätzung der Werte von $p_r(i, j)$. Ein einfacher Ansatz wäre hier die Annahme, daß $p_r(i, j)$ proportional zum Wert des Produktes von $U_r \cdot P_r(j)$ ist. Formal dargestellt erhält man folgendes:

$$p_r(i, j) = p(i, j) \cdot \frac{P_r(i) \cdot U_r}{\sum_{r=1}^N (P_r(i) \cdot U_r)} \quad (4.22)$$

4.2.1.3 Nichtstationäre Lösung innerhalb eines Teilgebietes

Der stationäre Fall beschreibt die Situation, in der sich ein statistischer Benutzer 'gedächtnislos' hin und her bewegt, wobei die Verteilung unverändert bleibt. Im Alltag tritt jedoch häufig die Situation ein, daß ein Teil einer Benutzerklasse sich geordnet von einem in ein anderes DN-Gebiet bewegt (sog. Benutzermigration), wenn z.B. viele Benutzer während der Morgenstunden von ihrem Zuhause zu entfernten Arbeitsstätten fahren oder wenn sie wegen einer Kultur- oder Sportveranstaltung sich zeitweise von Zuhause entfernen. Um den Prozeß der geordneten Bewegung zu beschreiben, die eine Verteilungsänderung begleitet, muß eine sog. nichtstationäre Lösung hergeleitet werden.

Eine solche geordnete Bewegung kann sowohl innerhalb als auch zwischen den Teilgebieten stattfinden. In diesem Kapitel wird zuerst angenommen, daß die Benutzermigration nur innerhalb des UMTS-Teilgebietes stattfindet. Der Fall einer Migration zwischen Teilgebieten ist im folgenden Kapitel dargestellt.

Die Vektoren der Zustandswahrscheinlichkeiten $P_{old, r}$ und $P_{new, r}$ beschreiben die Verteilung der Benutzer der Klasse r im Teilgebiet entsprechend vor und nach der Migration. Der Migrationsvektor $P_{mig, r}$ wird wie folgt definiert:

$$P_{mig, r} = P_{old, r} - P_{new, r} \quad (4.23)$$

Weil die Migration innerhalb eines Teilgebietes erfolgt, gilt:

$$\sum_{i=1}^N P_{new, r}(i) = \sum_{i=1}^N P_{old, r}(i) \quad (4.24)$$

$P_{mig, r}$ stellt die DN-Gebiete mit überschüssigen Benutzern als positive Werte und die DN-Gebiete, wohin die Benutzer migrieren, als negative Werte dar. Da die Benutzer nur innerhalb eines UMTS-Teilgebietes migrieren, gilt:

$$\sum_{i=1}^N P_{mig, r}(i) = 0 \quad (4.25)$$

Der Migrationsverlauf wird mit Hilfe des sog. quasi-Zufallsweges beschrieben. Ein ähnliches, jedoch einfacheres Verfahren, *semi-random walk* genannt, wurde in /SSD94/ und /Mone54/ beschrieben. Der in dieser Arbeit benutzte Algorithmus, der die Bestimmung des Zufallsweges beschreibt, kann wie folgt spezifiziert werden:

- Der quasi-Zufallsweg ist rekursiv, wobei jeder Rekursionsschritt k durch den Anfangsüberschußverteilsvektor $P_{inter, r, k-1}$ bestimmt ist:
 - $P_{inter, r, 0}$ ist gleich dem ersten Anfangsüberschußverteilsvektor, d.h., er entspricht den positiven Werten des Vektors $P_{mig, r}$.
 - $P_{inter, r, k}$ entspricht der im Schritt k erreichten Überschußverteilung.
- In jedem Schritt bewegen sich die 'überschüssigen' Benutzer aus den DN-Gebieten, die durch die Verteilung $P_{inter, r, k}$ beschrieben werden (sog. Quellen), in

Richtung solcher DN-Gebiete, die einen Mangel an Benutzern aufweisen, d.h. dorthin, wo $P_{mig, r} < 0$ gilt (sog. Senken).

- In jedem Schritt bewegen sich maximal nur so viele Benutzer, wie sich am Anfang des Schrittes in jedem Überschußgebiet befinden.
- Jeder einzelne Migrationsschritt wird so durchgeführt, daß die Anzahl der migrierenden Benutzer und deren Weg für jede Kombination aus Quelle und Senke separat nach folgendem Schema bestimmt wird:
 - Benutzer aus einer Quelle migrieren in Richtung aller Senken, wobei die Anzahl der Benutzer, die sich in Richtung einer bestimmten Senke bewegen soll, dem relativen Benutzermangel in der Senke entspricht.
 - Die Wahl des Weges zwischen einer Quelle und einer Senke wird für die bereits spezifizierte Anzahl der Benutzer nach folgenden Prinzipien festgelegt:
 - wenn Quelle und Senke benachbart sind, so migriert die berechnete Anzahl der Benutzer direkt;
 - wenn Quelle und Senke nicht benachbart sind, so migrieren die Benutzer in jedes Zwischengebiet, das sich physikalisch näher an der Senke befindet; wenn dabei mehrere Zwischengebiete in Frage kommen, so wählen die Benutzer den Weg zufällig, d.h., sie teilen sich gleichmäßig auf alle in Frage kommenden Zwischengebiete auf, ohne daß die Zwischengebiete nach verbleibender Entfernung gewichtet werden.

Die obige Spezifikation - obwohl sie angemessen erscheint - ist jedoch nicht hinreichend, um eine korrekte Lösung zu erhalten. Es soll hier auf zwei Punkte hingewiesen werden, in denen entweder Oszillation oder falsche Endverteilung der Benutzer entstehen kann:

1. Die Bestimmung der Anteile der Benutzer, die aus einer Quelle in Richtung der verschiedenen Senken migrieren sollen, erfolgt in jedem Schritt entweder mittels der Anfangs- oder der aktuellen Verteilung:
 - Bestimmung über die aktuelle Verteilung
Man betrachte zwei Senken i und j . Die Anzahl der fehlenden Benutzer in i und j ändert sich, wenn die ersten Benutzer eintreffen. Da die Migration auf verschiedenen Wegen erfolgt, kann sich auch die Anzahl der Benutzer in den verschiedenen Senken unterschiedlich schnell verändern. In der Konsequenz kann eine Situation auftreten, in der dieser Teil der Benutzer, der im vorherigen Schritt z.B. in Richtung i migrierte, im nächsten Schritt in Richtung j migrieren wird - Benutzer oszillieren.
 - Bestimmung über die Anfangsverteilung
Infolge der verschiedenen Migrationsgeschwindigkeiten ist eine falsche Endverteilung der Benutzer zu erwarten.
2. Die Festlegung des Weges in einzelnen Migrationsschritten erfolgt in jedem Schritt neu für alle Benutzer (vgl. Abbildung 4.2).
Das DN-Gebiet 14 ist die Quelle im Schritt k und die DN-Gebiete 1 und 4 sind die Senken. Im Schritt k werden die Benutzer aus dem Gebiet 14 in zwei Gruppen aufgeteilt: Die erste Gruppe migriert in Richtung des Gebietes 1 und be-

wegt sich deswegen in die Gebiete 10, 11 und 13. Die zweite Gruppe migriert in Richtung des Gebietes 4 und bewegt sich deshalb in die Gebiete 10, 11 und 15. Im nächsten Schritt ($k+1$) entsteht eine neue Situation, in der sich die Quellen in den Gebieten 10, 11, 13 und 15 befinden. Wenn jetzt die Wegfestlegung unabhängig vom Schritt k durchgeführt wird, so kann folgendes passieren: Die Benutzer z.B. aus dem DN-Gebiet 15 werden jetzt in zwei Gruppen aufgeteilt, die erste mit dem Migrationsziel DN-Gebiet 1 und die zweite mit dem Migrationsziel DN-Gebiet 4. Dadurch aber werden die Benutzer oszillieren, da aus der Beschreibung des Schrittes k folgt, daß alle Benutzer, die sich in das Gebiet 15 bewegten, als Migrationsziel das Gebiet 4 hatten.

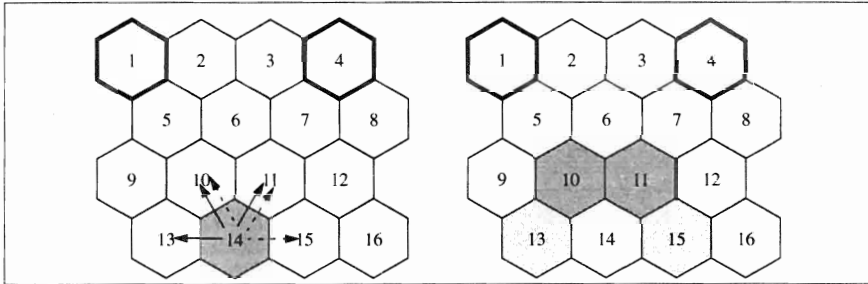


Abb. 4.2 Benutzeroszillation aufgrund der Wegfestlegung

Um Abhilfe für die dargestellten Probleme zu schaffen, wird eine Ergänzung zum Migrationsalgorithmus eingeführt: Den Benutzern wird vom ersten bis zum letzten Schritt gefolgt, d.h., im ersten Schritt wird festgelegt, wie die Benutzer aus allen Quellen sich auf alle Senken zu verteilen haben. Diese Zuordnung wird bis zum Ende der Migration beibehalten.

Unter Benutzung des beschriebenen Ansatzes ist es möglich, für jeden Schritt k den Vektor der Zustandswahrscheinlichkeit $P_{inter, r, k}$ und die Matrix der Übergangswahrscheinlichkeit $p_{inter, r, k}$ zu bestimmen. Da die Migration nicht synchron verläuft, ist es hilfreich, die durchschnittlichen Werte der Wahrscheinlichkeiten $P_{inter, r}$ und $p_{inter, r}$ zu ermitteln, um mit deren Hilfe weitere Berechnungen durchzuführen. Wenn die migrierenden Benutzer die einzigen sind, die betrachtet werden, so können die Durchschnittswerte nach den folgenden Formeln ermittelt werden (K ist die Gesamtanzahl der Schritte):

$$P_{inter, r}(i) = \frac{\sum_{k=1}^K P_{inter, r, k}(i)}{K} \quad i = 1, 2, \dots, N \quad (4.26)$$

$$p_{inter, r}(i, j) = \frac{\sum_{k=1}^K p_{inter, r, k}(i, j)}{\sum_{j=1}^N \sum_{k=1}^K p_{inter, r, k}(i, j)} \quad i, j = 1, 2, \dots, N \quad (4.27)$$

Es soll darauf hingewiesen werden, daß in Wirklichkeit immer eine Überlagerung von migrierenden und stationären Benutzern vorliegt, auch dann, wenn alle betrachteten Benutzer der systematischen Bewegung unterworfen sind. Aufgrund der unterschiedlichen Weglängen während der Migration erreichen manche Benutzer ihr Zielgebiet schneller als die anderen, und sie verhalten sich von diesem Zeitpunkt an stationär. Abgesehen von Sonderuntersuchungen ist deswegen bei einer Auswertung immer notwendig, beide Lösungen zu überlagern (vgl. Kapitel 4.2.1.5). Aus diesem Grunde ist es auch nicht erlaubt, die geordnete Bewegung in unabhängige Schritte zu zerlegen: Für das Erfassen der Gesamtbewegung ist immer eine Überlagerung der sich geordnet bewegendenden und der stationären Benutzer notwendig.

Bemerkung:

In der dargestellten Herleitung ist die Zugehörigkeit zu einer Benutzerklasse explizit beachtet worden. Eine Darstellung, die für alle Benutzer gilt, könnte analog zum stationären Fall gegeben werden. Die Migration wird dann für alle Benutzer statt für einzelne Benutzerklassen berechnet. In den Formeln (4.26) und (4.27) verschwindet dann die Abhängigkeit von der Benutzerklasse r . Auf eine explizite Darstellung wurde hier verzichtet, weil im Gegensatz zur stationären Lösung die systematische Bewegung immer für jede Benutzerklasse separat mit $P_{old, r}$ und $P_{new, r}$ spezifiziert werden kann. Damit ist eine globale Beschreibung überflüssig.

4.2.1.4 Nichtstationäre Lösung zwischen den Teilgebieten

Der im Kapitel 4.2.1.3 hergeleitete Ansatz zur Beschreibung der systematischen Benutzerbewegung wird jetzt für den Fall erweitert, daß die Bewegung zwischen Teilgebieten stattfindet.

Statt die Migration mittels zweier Vektoren $P_{old, r}$ und $P_{new, r}$ zu beschreiben, wird hier nur $P_{old, r}$ verwendet. Zusätzlich wird ein Verteilungsvektor $P_{ext, r, 0}$ eingeführt, um die Anfangsverteilung der überschüssigen Benutzer zu beschreiben, die aus dem Teilgebiet migrieren sollen. $P_{ext, r, 0}$ ist analog zu dem aus dem vorherigen Kapitel bekannten $P_{inter, r, 0}$. Die DN-Gebiete, in denen sich überschüssige Benutzer befinden, sind durch $P_{ext, r, 0} > 0$ charakterisiert. Diese Gebiete stellen die Quellen der Benutzer für den ersten Schritt dar. Folgende Bedingungen müssen erfüllt werden:

$$\sum_{i=1}^N P_{ext, r, 0}(i) < \sum_{i=1}^N P_{old, r}(i) \quad (4.28)$$

$$\forall_{i \in \{1 \dots N\}} (P_{ext, r, 0}(i) < P_{old, r}(i)) \quad (4.29)$$

Da der Lösungsansatz dem ähnlich ist, der innerhalb eines Teilgebietes angewendet wurde, muß noch die Verteilung der Migrationssinken im Teilgebiet beschrieben werden. Im betrachteten Fall findet die Migration in ein externes UMTS-Teilgebiet statt. Aus diesem Grund ist es zweckmäßig anzunehmen, daß die Senken am Teilgebietsrand platziert werden, was mit dem Verteilungsvektor $P_{bord, r}$ beschrieben wird.

Auf diese Art erhält man letztendlich eine äquivalente Situation wie bei der Beschreibung der Migration innerhalb des Teilgebietes. Damit ist der bereits einge-

führte Algorithmus anwendbar. Lediglich eine kleine Erweiterung ist notwendig, um den Sachverhalt wiederzugeben, daß die Benutzer nach dem Erreichen des Randgebietes im darauffolgenden Schritt das Teilgebiet verlassen.

Ähnlich wie im Kapitel 4.2.1.3 können hier die durchschnittlichen Werte für den Vektor der Zustandswahrscheinlichkeit $P_{ext,r}$ und die Matrix der Übergangswahrscheinlichkeit $p_{ext,r}$ bestimmt werden (K ist die Gesamtanzahl der Schritte):

$$p_{ext,r}(i,j) = \frac{\sum_{k=1}^K p_{ext,r,k}(i,j)}{\sum_{j=1}^N \sum_{k=1}^K p_{ext,r,k}(i,j) + \sum_{k=1}^K p_{ext,r,k}(i,0)} \quad (4.30)$$

$$P_{ext,r}(i) = \frac{\sum_{k=1}^K P_{ext,r,k}(i)}{K} \quad (4.31)$$

1. Bemerkung:

Durch die externe Migration wird nicht nur die Übergangswahrscheinlichkeit zwischen den einzelnen DN-Gebieten, sondern auch die Übergangswahrscheinlichkeit zwischen den Teilgebieten beeinflusst. Dies wird bei der implementierten numerischen Lösung selbstverständlich beachtet.

2. Bemerkung:

Sollte eine Lösung für alle Benutzerklassen gefunden werden, so gilt hier gleiches wie im vorherigen Kapitel.

4.2.1.5 Überlagerung unterschiedlicher Lösungen

Im folgenden wird die allgemeine Lösung hergeleitet, die sich durch eine Überlagerung der Speziallösungen ergibt. Dies entspricht der normalen Situation, die durch die gleichzeitige Existenz von drei Benutzerarten gekennzeichnet ist:

- gedächtnislose Benutzer, die sich zufällig in alle Richtungen bewegen,
- Benutzer, die innerhalb des Teilgebietes migrieren, und
- Benutzer, die zwischen den Teilgebieten migrieren.

Der allgemeine Lösungsansatz ist auf einer schrittweisen linearen Überlagerung der einzelnen Lösungen aufgebaut, d.h., jeder Schritt der nichtstationären Lösungen (innerhalb des Teilgebietes und zwischen den Teilgebieten) wird mit dem dazugehörenden stationären Anteil überlagert. Folgendes muß dabei beachtet werden:

- Die stationäre Lösung wird auf die Benutzer angewendet, die ihre Verteilung nicht verändern, und auf solche intern migrierenden Benutzer, die bereits das Zielgebiet erreicht haben.

- Die interne nichtstationäre Lösung wird auf die intern migrierenden Benutzer angewendet, die das Zielgebiet noch nicht erreicht haben.
- Die externe nichtstationäre Lösung wird auf die extern migrierenden Benutzer angewendet, die das Teilgebiet noch nicht verlassen haben.

Alle Lösungen müssen nach jedem Schritt neu berechnet werden. Das gilt auch für die stationäre Lösung, da sich der Anteil der Benutzer, die sich in jedem Schritt stationär verhalten, ständig ändert. Die Wahrscheinlichkeit, daß die Benutzer sich außerhalb des Teilgebietes befinden, ändert sich ebenfalls ständig.

Die durchschnittlichen Werte der Zustands- und der Übergangswahrscheinlichkeit im kombinierten Ansatz können wie folgt ermittelt werden (K ist die Gesamtanzahl der Schritte und K_{int}/K_{ext} ist die Gesamtanzahl der internen/externen Schritte):

$$P_r(i) = \frac{\sum_{k=1}^K (P_{stat,r,k}(i) + P_{int,r,k}(i) + P_{ext,r,k}(i))}{K} \quad (4.32)$$

$$p_r(i,j) = \frac{p_Z(i,j)}{p_N(i,j)} \quad (4.33)$$

Dabei gilt folgendes:

$$K = \max(K_{int}, K_{ext})$$

$$p_Z(i,j) = \sum_{k=1}^K (P_{stat,r,k}(i) \cdot p_{stat,r,k}(i,j) + \dots \quad (4.33')$$

$$P_{inter,r,k}(i) \cdot p_{inter,r,k}(i,j) + P_{ext,r,k}(i) \cdot p_{ext,r,k}(i,j))$$

$$p_N(i,j) = \sum_{j=1}^N p_Z(i,j) + \dots \quad (4.33'')$$

$$\sum_{k=1}^K P_{stat,r,k}(i) \cdot p_{stat,r,k}(i,0) + \sum_{k=1}^K P_{ext,r,k}(i) \cdot p_{ext,r,k}(i,0)$$

4.2.1.6 Ergänzung des Modells

Das hergeleitete Modell erlaubt eine Berechnung der Übergangswahrscheinlichkeiten auf der Basis der Zustandswahrscheinlichkeiten. Zusätzlich erlaubt es eine einfache, wahlweise separate oder überlagerte Behandlung der drei wichtigsten Situationen, die das Bewegungsverhalten der Benutzer beschreiben. Das Modell ist in der bisherigen Form jedoch nicht hinreichend spezifiziert, um die Situationen verständlich zu beschreiben. Zusätzliche Parameter müssen eingeführt werden.

Der wichtigste Parameter, der die Benutzerbewegungen innerhalb eines Teilgebietes im stationären Fall beschreibt, ist die sog. Beweglichkeit. Sie bringt zum Ausdruck, ob der Benutzer sich, unter Beibehaltung der gleichen Zustandsverteilung, sehr oft oder eher seltener hin und her bewegt. Die Beweglichkeit wird definiert als

$$P_{Bew} = (1 - \sum_{i=1}^N P(i) \cdot p(i, i)) \quad (4.34)$$

Der Wert des Parameters P_{Bew} wird bei der Lösung des stationären Falles als eine Vorgabegröße aufgefaßt, die den Charakter der Benutzerbewegungen beschreibt. Die freien Parameter werden dann so gewählt, daß die Lösung den vorgegebenen Wert von P_{Bew} erfüllt.

Um die Benutzerbewegungen zwischen den Besuchsteilgebieten zu beschreiben, wird der Parameter *cross_vis* (zusätzlich zum bereits eingeführten Parameter *cross* für die Bewegungen aus dem Heimteilgebiet) eingeführt. Er wird durch den Quotienten aus der Anzahl der Bewegungen in einem Besuchsteilgebiet und der Anzahl der Bewegungen über die Grenze des Besuchsteilgebietes ausgedrückt.

Mit Hilfe dieser zusätzlichen Parameter ist es nun möglich, die Bewegungen der Benutzer innerhalb eines Teilgebietes einfach und exakt zu beschreiben.

4.2.2 Bewegungen zwischen Teilgebieten

Alle bis jetzt hergeleiteten Modelle hatten das primäre Ziel, die Bewegungen innerhalb eines Teilgebietes formal zu beschreiben. Diese Aufgabe wurde mit großer Sorgfalt und Genauigkeit durchgeführt, da die Abläufe innerhalb des Teilgebietes den größten Einfluß auf die Optimierung der Datenbank haben. Im folgenden soll ein vereinfachtes Modell für die Bewegungen zwischen den Teilgebieten hergeleitet werden. Die eingeführten Vereinfachungen sind auf folgende Gründe zurückzuführen:

1. Mangel an zuverlässiger Information, die die Bewegung zwischen den Teilgebieten beschreibt.
2. Festlegung des Schwerpunktes der Untersuchungen auf die Prozesse innerhalb eines Teilgebietes.
3. Bewegungen zwischen den Teilgebieten werden eher einen geringen Teil der Gesamtbewegungen ausmachen, da die Teilgebiete mit geographischen Regionen der Größenordnung eines Landes zu assoziieren sind. Ein Teilgebiet entspricht der Deckung eines öffentlichen Netzes eines Netzbetreibers.
4. Leistungsfähigkeit einer Teilgebietsdatenbank, die von dem tatsächlichen Verhalten des Benutzers zwischen den Teilgebieten unabhängig ist.

Die ersten drei Punkte benötigen keine Erklärung - sie sind offensichtlich. Zu dem letzten Punkt soll folgendes erläutert werden: Die Kommunikation zwischen zwei Teilgebieten besteht aus drei Phasen - zwei davon werden innerhalb der Teilgebiete ausgeführt und die dritte zwischen den Teilgebieten. Es ist für jede einzelne Teilgebietsdatenbank relevant, ob die Kommunikation innerhalb des Teilgebietes oder zwi-

schen den Teilgebieten stattfindet. Es ist ohne Bedeutung zu wissen, zwischen welchen Teilgebieten die Kommunikation durchgeführt wird. In der Auswertung wird nur ein Teilgebiet betrachtet, in dem die internen Phasen überlagert werden.

Die Bewegungen zwischen den Teilgebieten werden mittels zweier Zustände modelliert, wie das in der Abbildung 4.3 dargestellt ist. In diesem Ansatz sind die Aspekte der räumlichen Entfernung zwischen den einzelnen Teilgebieten nicht beachtet. Ebenfalls ist nicht modelliert, welche zwei konkreten Teilgebiete in die Kommunikation involviert sind. Das Wesentliche ist, daß die Bewegung stattfindet und zur Datenkommunikation innerhalb der beiden Teilgebiete beiträgt.

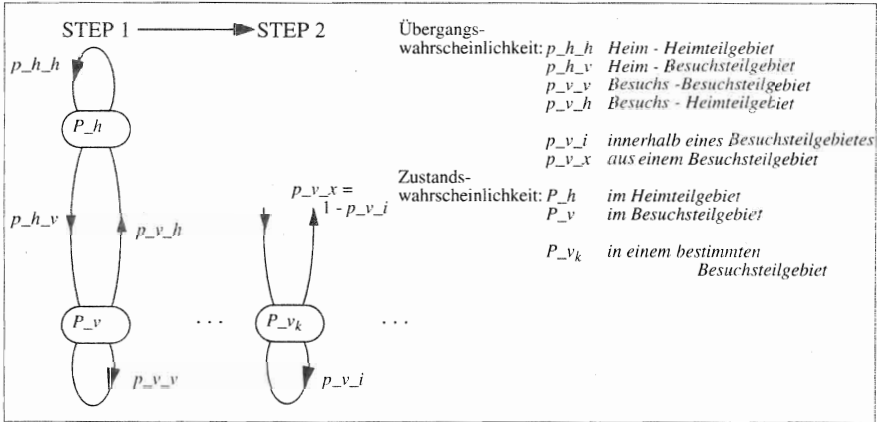


Abb. 4.3 Zustandsbeschreibung für die Bewegungen zwischen den Teilgebieten

Die Herleitung der gesuchten Wahrscheinlichkeiten (Raten) wird in zwei Schritten durchgeführt:

1. Nur die Bewegungen zwischen dem Heimteilgebiet und den Besuchsteilgebieten werden betrachtet, ohne auf die Bewegungen zwischen den Besuchsteilgebieten einzugehen. Die stationäre Lösung kann wie folgt beschrieben werden:

$$P_h \cdot p_{h_v} = P_v \cdot p_{v_h}; \quad P_h = 1 - P_v \quad (4.35)$$

2. Die Bewegungen zwischen den Besuchsteilgebieten werden modelliert.

Aus dem bereits hergeleiteten Modell innerhalb eines Teilgebietes kann der Wert der Zustandswahrscheinlichkeit P_h ausgedrückt werden als:

$$P_h = \frac{\sum_{r=1}^{N_{res}} \sum_{j=1}^N P_r(j) \cdot U_r}{\sum_{r=1}^{N_{res}} U_r} \quad (4.36)$$

$$P_v = 1 - P_h \quad (4.37)$$

Im folgenden werden die Raten und Übergangswahrscheinlichkeiten berechnet, die in Abbildung 4.3 präsentiert sind. Ihre Werte sind aus dem Mobilitätsmodell innerhalb jedes einzelnen Teilgebietes herzuleiten.

Die Abgangs- und Ankunftsdaten aus dem und in das Heimteilgebiet können folgendermaßen berechnet werden (unter der Verwendung des Parameters *cross*):

$$\lambda_{h_v} = \lambda_{v_h} = \sum_{r=1}^{N_{res}} \sum_{j=1}^N P_r(j) \cdot \lambda_r(j, 0) = cross \cdot k_M \cdot P_v \cdot \sum_{r=1}^{N_{res}} U_r \quad (4.38)$$

Die Rate der Bewegungen innerhalb des Heimteilgebietes ergibt sich zu:

$$\lambda_{h_h} = k_M \cdot P_h \cdot \sum_{r=1}^{N_{res}} U_r \quad (4.39)$$

Die Wahrscheinlichkeit, daß das Heimteilgebiet verlassen wird, ist dann:

$$p_{h_v} = \frac{\lambda_{h_v}}{\lambda_{h_v} + \lambda_{h_h}} \quad (4.40)$$

Mit der Gleichung (4.35) kann p_{v_h} berechnet werden:

$$p_{v_h} = p_{h_v} \cdot \frac{P_h}{1 - P_h}$$

Die Gesamtrate der Bewegungen innerhalb des Besuchsteilgebietes und aus diesem heraus ergibt:

$$\sum_{B_{TN}} (\lambda_{v_x} + \lambda_{v_i}) = \lambda_{v_h} \cdot \frac{1 - p_{v_h}}{p_{v_h}} = k_M \cdot P_v \cdot \sum_{r=1}^{N_{res}} U_r \quad (4.41)$$

Dabei bedeuten:

λ_{v_x}	alle Bewegungen aus einem Besuchsteilgebiet
λ_{v_i}	alle Bewegungen innerhalb eines Besuchsteilgebietes
B_{TN}	alle Besuchsteilgebiete

Die Rate λ_{v_x} kann nochmals in die

- Rate aus dem Besuchsteilgebiet ins Heimteilgebiet (λ_{v_h}) und die
- Rate aus dem Besuchsteilgebiet ins Besuchsteilgebiet (λ_{v_n})

zerlegt werden.

Der Wert von λ_{v_n} kann mittels des Parameters *cross* ausgedrückt werden:

$$\lambda_{v_n} = cross \cdot k_M \cdot P_v \cdot \sum_{r=1}^{N_{res}} U_r \quad (4.42)$$

4.3 Anrufverteilungsmodell

Die Beschreibung der Anrufverteilung stellt das zweite Grundgerüst neben dem Mobilitätsmodell dar, das notwendig ist, um die als wesentlich angesehenen Datenbankabfragen formal zu spezifizieren. Die Anrufverteilung bedeutet im folgenden die räumliche Verteilung der Anrufziele aus der Perspektive des DN-Gebietes, wo der Anruf gestartet wird. Das entwickelte Modell der Anrufverteilung gibt die Möglichkeit, die Verteilung der Suchziele wahrscheinlichkeitsbasiert zu beschreiben und mittels einiger Parameter an verschiedene Verhaltensmuster anzupassen.

Das Modell bildet ein realistisches Szenario nach, in dem die Abhängigkeit des Benutzerverhaltens vom aktuellen Standort in das Modell einbezogen wird.

4.3.1 Allgemeine Bemerkungen

Das Modell der Anrufverteilung beschreibt nicht nur flexibel das Verhalten im Heimknoten, sondern auch realistisch die Relation zwischen dem aktuellen Aufenthaltsort des Benutzers und der Verteilung der Anrufe.

Das hier hergeleitete Modell basiert auf der gleichen Topologie und der gleichen Zustandswahrscheinlichkeit im Teilgebiet, auf der das Mobilitätsmodell hergeleitet wurde. Es gibt zwei wichtige Unterschiede im Vergleich zum Mobilitätsmodell:

1. Die Beschreibung ist einerseits einfacher, da es nicht notwendig ist, die Übergangswahrscheinlichkeiten zu berechnen - nur die Aufenthaltswahrscheinlichkeiten sind für das Anrufverhalten maßgeblich.
2. Das Verhalten ist andererseits durch mehrere unabhängige Parameter beschrieben, im Unterschied zum Mobilitätsmodell, wo die Anzahl der freien Parameter eher beschränkt war - dies resultiert aus der Abhängigkeit des Anrufverhaltens vom aktuellen Aufenthaltsort.

Da die Beschreibung einfacher ist, sind auch explizit keine zwei separaten, unterschiedlich detaillierten Modelle für das Verhalten innerhalb eines Teilgebietes und zwischen Teilgebieten notwendig.

4.3.2 Das Modell

Die Verteilung der Anrufe zwischen den UMTS-Teilgebieten wird mittels zweier Parameter beschrieben:

1. Für alle Benutzer innerhalb des Heimteilgebietes wird angenommen, daß die Anrufe innerhalb des eigenen Heimteilgebietes mit $p_{c_{HD}}$ ausgeführt werden und mit $(1 - p_{c_{HD}})$ auf die sonstigen Teilgebiete verteilt sind.
2. Für alle Benutzer innerhalb des Besuchsteilgebietes wird angenommen, daß die Anrufe innerhalb des eigenen Besuchsteilgebietes mit $p_{c_{VD}}$ ausgeführt werden und mit $(1 - p_{c_{VD}})$ auf die sonstigen Teilgebiete verteilt sind.

Für alle Anrufe, die von außen in ein Teilgebiet kommen, wird angenommen, daß sie sich proportional zur Anzahl der Heimbenutzer (für Heimteilgebietsanrufe) oder der Besuchsbewohner (für Besuchsteilgebietsanrufe) auf die DN-Gebiete verteilen.

Diese sehr vereinfachte Beschreibung der Situation zwischen den Teilgebieten reflektiert den Schwerpunkt dieser Arbeit, der auf die Kommunikation innerhalb eines Teilgebietes gerichtet ist. Wohin der Anruf geleitet wird, wenn er das Gebiet verläßt, ist ohne Einfluß auf die interne Kommunikation. Für die ankommenden Anrufe ist andererseits ohne Bedeutung, aus welchem Teilgebiet sie eintreffen.

Das Anrufverhalten eines Benutzers innerhalb eines Teilgebietes, d.h., wenn der Benutzer Anrufe innerhalb dieses Gebietes durchführt, ist vom aktuellen Standort abhängig /SSD94/. Folgende drei Situationen werden dabei unterschieden:

1. Benutzer, die sich im eigenen DN-Heimgebiet innerhalb des Heimteilgebietes befinden, werden wahrscheinlich öfters lokale Anrufe durchführen. Das Anrufverhalten kann hier mittels einer Matrix $p_{h_r}(r, i)$ ausgedrückt werden, die die Wahrscheinlichkeit beschreibt, daß ein Benutzer aus dem DN-Heimgebiet r einen Anruf in das DN-Gebiet i plziert.
2. Benutzer, die sich in einem DN-Besuchsgebiet innerhalb des Heimteilgebietes befinden, werden sowohl lokale Anrufe als auch Anrufe nach Hause durchführen. Das Anrufverhalten wird somit durch folgende Parameter beschrieben:
 - Das lokale Verhalten wird mittels einer Matrix $p_{h_v}(j, i)$ ausgedrückt, die die Wahrscheinlichkeit beschreibt, daß ein Benutzer, der sich aktuell im DN-Besuchsgebiet j befindet, einen Anruf in das DN-Gebiet i plziert.
 - Das Heimverhalten wird mittels einer Matrix $p_{h_r}(r, i)$ ausgedrückt, die die Wahrscheinlichkeit beschreibt, daß ein Benutzer, der sich aktuell in irgendeinem DN-Besuchsgebiet befindet und dessen DN-Heimgebiet im DN-Gebiet r ist, einen Anruf in das DN-Gebiet i plziert.
 - Die Anteile des lokalen und des Heimverhaltens können mittels eines Parameters loc_b (*local behaviour*) ausgedrückt werden.
3. Das Anrufverhalten eines Benutzers im Besuchsteilgebiet wird mittels einer Matrix $p_{v_r/v}(j, i)$ beschrieben, die die Wahrscheinlichkeit beschreibt, daß ein Benutzer aus dem DN-Gebiet j einen Anruf in das DN-Gebiet i plziert.

Die Beschreibung des Verhaltens in der obigen 2. Situation ist eine lineare Überlagerung des lokalen und des Heimverhaltens. Das Gesamtverhalten eines Benutzers im DN-Besuchsgebiet j , der sein Heim im DN-Gebiet r hat, kann beschrieben werden:

$$p_{h_v_tot}(r, j, i) = loc_b \cdot p_{h_v}(j, i) + (1 - loc_b) \cdot p_{h_r}(r, i) \quad (4.43)$$

Bei der Beschreibung des Verhaltens in der 3. Situation braucht man nicht das lokale und das Heimverhalten zu unterscheiden, da im Besuchsteilgebiet alle DN-Gebiete für einen Benutzer 'fremd' sind. Damit wird das Anrufverhalten immer ausschließlich einen lokalen Charakter aufweisen. Das Heimverhalten könnte sich eventuell in der Verteilung der Anrufe auf die Teilgebiete zeigen: Der Benutzer plziert einen Sonderteil der Anrufe in seinem Heimteilgebiet (entgegen der Annahme,

daß die Anrufe gleich verteilt sind). In einem solchen Fall würden die Anrufe nicht gleichmäßig über alle Teilgebiete verteilt. Das Wesentliche für die Kommunikation innerhalb eines betrachteten Zielteilgebietes wäre dann zu wissen, für welchen Teil der Anrufe es als Heim- und für welchen als Besuchsteilgebiet zu modellieren ist. Auf diese Verfeinerung wurde hier verzichtet.

Bemerkung:

Die beschriebenen Parameter gelten für die Verteilung der Anrufe auf die Heimgebiete der angerufenen Benutzer. Das aktuelle Zielgebiet, wo sich der angerufene Benutzer befindet, muß gemäß der Wahrscheinlichkeitsverteilung der Aufenthaltsorte ermittelt werden. Dieser Ansatz ist naheliegend unter der Annahme, daß das Anrufverhalten des Anrufers von der Mobilität des Angerufenen unabhängig ist.

4.3.3 Einfluß der Mobilität auf das Anrufverhalten

Das eingeführte Modell behandelt die Anrufer und die Angerufenen unterschiedlich:

- für den Anrufer (Anrufenden) gibt es unterschiedliche Parameter in Abhängigkeit davon, wo sich der Anrufende befindet;
- für den Angerufenen spezifiziert das Modell nur die Wahrscheinlichkeit, daß ein Anruf zu seinem Heimknoten geleitet werden soll.

Um eine vollständige Beschreibung zu erhalten, wird somit wie folgt vorgegangen:

- Wenn der Anrufende betrachtet wird, so werden die drei im Kapitel 4.3.2 definierten Situationen gemäß der Wahrscheinlichkeit überlagert, mit der sich der Anrufende in jeder der Situationen befindet.
- Wenn der Angerufene betrachtet wird, so muß das Anrufverhalten mit dem Bewegungsmuster des Angerufenen verknüpft werden, um seinen aktuellen Standort zu bestimmen. Dabei sind drei Fälle möglich:
 1. Der DN-Heimknoten soll gefunden werden - nur das Anrufverhalten ist maßgebend.
 2. Der DN-Besuchsknoten soll gefunden werden - die Wahrscheinlichkeit, daß ein einzelner Benutzer in seinem Heimknoten den Anruf erhalten soll, und die Verteilungswahrscheinlichkeit der Bewegungen dieses Benutzers müssen miteinander multipliziert werden (da die Ereignisse als unabhängig angenommen werden, kann die bedingte Wahrscheinlichkeit als Produkt der einzelnen Wahrscheinlichkeiten ermittelt werden /Ash93/).
 3. Der DN-Heimknoten oder der DN-Besuchsknoten sollen gefunden werden - die Anrufverteilung muß in diesem Fall unter Beachtung des Bewegungsmusters neu berechnet werden, um der Bedingung gerecht zu werden, daß die Suche immer auf dem kürzesten Pfad zu erfolgen hat.

Formal wird die Suchwahrscheinlichkeit in einem Besuchsknoten berechnet:

$$p_{call}(i, j) = \sum_{l=1}^{N_R} P_l(i) \cdot \left(\sum_{k=1}^{N_R} p_{res\ call}(i, k) \cdot P_k(j) \right) \quad (4.44)$$

Dabei ist $p_{res\ call}(i, k)$ die Wahrscheinlichkeit, daß der Anruf zu einem Benutzer geht, der den DN-Heimknoten k hat.

Bemerkung:

Es wurde mehrmals betont, daß beide Seiten, der Anrufende und der Angerufene, dem Mobilitätsmodell ausgesetzt sind. Offensichtlich ist es das gleiche Modell. Zum Zwecke der Auswertung ist jedoch wünschenswert, die beiden Beschreibungen unabhängig spezifizieren zu können. Das folgende Beispiel beschreibt einen solchen Fall:

Es soll eine Situation untersucht werden, in der die Anrufenden ausschließlich einer Benutzerklasse angehören. Eine solche Auswertung ist sinnvoll, wenn man spezielle Verzeichnisstrukturen untersuchen will oder wenn man die nicht relevanten Auswirkungen auf das Netz unterdrücken will. Wird jetzt die gleiche Beschreibung für beide Seiten des Anrufes angewendet, so sind die Anrufe nur zum eigenen DN-Gebiet erlaubt (das Anrufmodell spezifiziert die Verteilung auf die Heimgebiete). Dies ist natürlich nicht das gewünschte Szenario. Ein Ausweg aus diesem Dilemma bietet eine separate Beschreibungsmöglichkeit für die Anrufenden und für die Angerufenen.

In der Auswertung wird somit eine separate Beschreibung verwendet, die durch die Verteilungsvektoren $P_{orig, r}(i)$ und $P_{term, r}(i)$ für den anrufenden und den angerufenen Benutzer gegeben wird. Diese Vektoren sind allerdings im Regelfall in allen DN-Gebieten identisch, in denen Benutzer existieren.

Es soll beachtet werden, daß die eingeführten Anrufverteilungsparameter $p_{h_r}(r, i)$, $p_{h_v}(j, i)$, $p_{v_r/v}(j, i)$ und p_{cHD} , p_{cVD} sowohl das Verhalten des Anrufers als auch die Zahl der Angerufenen in der jeweiligen Benutzerklasse widerspiegeln sollen.

4.4 Modellierung der ausgewählten Anfragen

Die Auswahl der auszuwertenden Anfragen wurde im Kapitel 3.3 aufgrund zweier Hauptmerkmale vorgenommen:

1. Die untersuchte Anfrage soll einen wesentlichen Beitrag zum Datenverkehr im Teilgebiet, insbesondere im DIR-Baum, leisten.
2. Die Anfrage soll von der Wahl des DIR-Baumes und der Aktionen innerhalb eines Teilgebietes abhängig sein, so daß ein Optimierungspotential gegeben ist.

Als Konsequenz aus dieser Überlegung und in Verbindung mit der Darstellung der Anfragen wurden drei Anfragen gewählt:

1. Suchanfrage als ein Teil des Anrufaufbaus
2. Aktualisierung der Datenbankeinträge innerhalb eines Teilgebietes
3. Aktualisierung der Datenbankeinträge zwischen den Teilgebieten

Im folgenden werden diese Anfragen formal spezifiziert und deren Einfluß auf die Leistungsfähigkeit der UMTS-Datenbank vorgestellt. Die Spezifikation stellt die Basis für die numerische Auswertung im Kapitel 5 dar.

4.4.1 Suchanfragen

Die Suchanfrage ist ein sehr komplexer Prozeß mit einmaligen Eigenschaften:

- Die Häufigkeit, mit der die Anrufprozeduren gestartet werden, ist erwartungsgemäß die höchste im Vergleich mit allen anderen UMTS-Prozeduren /Bell92/.
- Der Beitrag zum Datenverkehr im Verzeichnisbaum ist sehr hoch.
- Bedingt durch die mangelnde Relation zwischen der personenbezogenen Nummer und der Speicherstelle der Benutzerdaten ist die Suchanfrage auf die Existenz des Verzeichnisbaumes angewiesen.

Nach einem kurzen Überblick wird die wahrscheinlichkeitsfundierte Beschreibung der Suchanfrage präsentiert. Die Beschreibung wird mit einer Diskussion der Einflüsse von Datenverteilungs- und Architekturalternativen auf die Anfrage abgeschlossen, die als Vorbereitung auf die darauf folgende Auswertung gedacht ist.

4.4.1.1 Überblick

Das Ziel der Suchanfrage ist, den aktuellen Aufenthaltsort des angerufenen Teilnehmers zu ermitteln. Sie besteht aus einer Nachrichtenfolge, die zum größten Teil dem Zweck dient, eine Suche gemäß den UMTS-Prinzipien durchzuführen (es gibt keine Relation zwischen der personenbezogenen Nummer und dem Heim- oder Besuchsknoten /SSD94/). Im Laufe der Suche werden die *routing*-, *registration*- und *terminal*-Daten des angerufenen Benutzers gefunden. Aus ihnen werden die Voreinstellungen, die Präferenzen des Benutzers und schließlich die Leitwegnummer des Endgerätes herausgelesen. Diese Informationen werden teilweise direkt ausgewertet. Beispielsweise wird das Endgerät gar nicht weitergesucht, wenn momentan der Zugang zum Benutzer nicht erlaubt ist. Zum Teil werden die Informationen an den Anrufer zwecks Weiterbearbeitung geschickt.

Die Suche wird vom Verzeichnisbaum mittels der dort gespeicherten Information unterstützt entweder durch doppeltverkettete Zeiger (innerhalb des Teilgebietes) oder unter Benutzung des Wurzelknotens als Übergangsknoten zu anderen Teilgebieten (/SSD94/ und /Mone61/).

Wie bereits im Kapitel 3.4.2.1 erwähnt, sind unterschiedliche Möglichkeiten erfolgversprechend, die Benutzerdaten in einem Teilgebiet auf die Besuchs- und Heimknoten zu verteilen und dadurch die Mobilität und Suche zu unterstützen; vier von diesen Szenarien wurden explizit für genauere Untersuchungen spezifiziert.

Bei der Suche wird unabhängig vom jeweiligen Szenario unterschieden, welche Knotentypen (Heim- oder Besuchsknoten) während der Anfrage gesucht werden. Drei unterschiedliche Fälle werden dabei betrachtet: Die Suche kann nach dem

- Heimknoten - *resident* (R),
- Besuchsknoten - *visited* (V) oder
- Heim- oder Besuchsknoten - (RV)

durchgeführt werden. Zusammen mit der Information über die Anrufverteilung und die Benutzerbewegungen können dann die aktuell geltenden Wahrscheinlichkeiten für die Verteilung der einzelnen Nachrichten ermittelt werden. Dadurch wird die Auswertung der speziellen Benutzerdatenverteilungsmuster und der entsprechenden Nachrichtenfolgen als eine Aneinanderreihung von definierten Sequenzen mit bekannten Häufigkeiten dargestellt.

Bemerkung:

Bei der Auswertung einzelner Szenarien ist es sowohl möglich, daß während der Anfrage mehrere der oben genannten Knotentypen gefunden werden müssen, als auch, daß in einem bestimmten Datenverteilungsszenario entweder der R-Knoten oder der V-Knoten nicht vorhanden sind.

4.4.1.2 Formale Beschreibung

Bevor die Formeln für die Ereignisraten (Wahrscheinlichkeiten) gemäß den drei genannten Fällen präsentiert werden, sind die Möglichkeiten der gegenseitigen Lage der Aufenthaltsorte des Anrufers und des Angerufenen zu systematisieren. Es können insgesamt fünf Hauptfälle unterschieden werden, wie die Aufenthaltsorte des Anrufenden und des Angerufenen auf die Teilgebiete verteilt sein können:

- Der angerufene Benutzer befindet sich auf dem Gebiet seines *home domain*.
 1. Hauptfall: Der anrufende Benutzer befindet sich auf dem Gebiet des *home domain* des Angerufenen.
 2. Hauptfall: Der anrufende Benutzer befindet sich außerhalb des Gebietes des *home domain* des Angerufenen.
- Der angerufene Benutzer befindet sich außerhalb des Gebietes seines *home domain*, d.h. in einer *visited domain*.
 3. Hauptfall: Der anrufende Benutzer befindet sich auf dem Gebiet des *visited domain* des Angerufenen.
 4. Hauptfall: Der anrufende Benutzer befindet sich auf dem Gebiet des *home domain* des Angerufenen.
 5. Hauptfall: Der anrufende Benutzer befindet sich außerhalb des Gebietes des *home domain* und des *visited domain* des Angerufenen.

Im Sinne der früheren Ausführungen wurden folgende Annahmen getroffen:

- Die Gesamtheit der Anrufe, die aus den Teilgebieten abgehen, sei gleichmäßig auf alle Teilgebiete verteilt.
- Die Anzahl der Anrufe, die das Teilgebiet verlassen, ist gleich der Anzahl, die in dieses Teilgebiet von außen eintreffen.
- Wenn die Verteilung der von außen kommenden Anrufe auf die einzelnen DN-Gebiete innerhalb eines Teilgebietes zu betrachten ist, dann gilt:
 - Handelt es sich um das Heimteilgebiet des Angerufenen, so wird eine Verteilung entsprechend der bestehenden Verteilungsdichte verwendet.

- Handelt es sich um das Besuchsteilgebiet des Angerufenen, so wird eine gleichmäßige Verteilung verwendet.
- Gemäß den allgemeinen Bemerkungen im Kapitel 4.1 gilt für die Kommunikation innerhalb eines Teilgebietes, daß sie eine Überlagerung von internen und externen, d.h. abgehenden und ankommenden Anrufen ist. Die Wahrscheinlichkeiten von abgehenden und ankommenden Anrufen werden separat berechnet.

Bezüglich der Indizierung sollen folgende Konventionen beachtet werden:

- Die Nummer i steht stets für das DN-Gebiet, aus dem der Anruf initiiert wird (DN-Gebiet des Anrufenden).
- Die Nummer j steht für das DN-Gebiet, in dem der Anruf terminiert (DN-Gebiet des Angerufenen).

Gemäß dieser Konvention bedeutet $\lambda(i, j)$ die Häufigkeit, mit der die Suche im Knoten i beginnt und im Knoten j endet. Wenn entweder i oder j gleich 0 ist, dann symbolisiert das eine externe Kommunikation über die Teilgebietsgrenze.

Es soll beachtet werden, daß die Existenz der *resident*- und der *visited*-Knoten im Heimteilgebiet bzw. im Besuchsteilgebiet von der Wahl des jeweiligen Datenverteilungsszenarios abhängig sind.

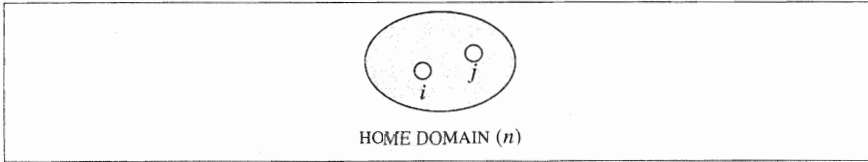


Abb. 4.4 Suchanfrage - Hauptfall 1

Im **Hauptfall 1** können folgende Ereignisraten spezifiziert werden:

- Suche nach dem *resident*-Knoten (**Fall_1R**):

$$\begin{aligned} \lambda_r(i, j)/k_{CS} = & \left(U_i \cdot P_i(i) \cdot p_{cHD} \cdot p_{h_r}(i, j) + \right. \\ & (1 - loc_b) \cdot \sum_{k \neq i} U_k \cdot P_k(i) \cdot p_{cHD} \cdot p_{h_r}(k, j) + \\ & loc_b \cdot \left(\sum_{k \neq i} U_k \cdot P_k(i) \right) \cdot p_{cHD} \cdot p_{h_v}(i, j) + \\ & \left. \sum_k U_k \cdot (1 - P_{in}) \cdot \sum_k P_{vis_k}(i) \cdot p_{cVD} \cdot p_{v_r/v}(i, j) \right) \cdot \sum_k P_j(k) \end{aligned} \quad (4.45)$$

- Suche nach dem *visited*-Knoten (**Fall_1V**):

$$\begin{aligned} \lambda_v(i, j)/k_{CS} = & \sum_l \left(U_i \cdot P_i(i) \cdot p_{cHD} \cdot p_{h_r}(i, l) + \right. \\ & (1 - loc_b) \cdot \sum_{k \neq i} U_k \cdot P_k(i) \cdot p_{cHD} \cdot p_{h_r}(k, l) + \end{aligned}$$

$$loc_b \cdot \left(\sum_{k \neq i} U_k \cdot P_k(i) \right) \cdot p_{cHD} \cdot p_{h_v}(i, l) + \sum_k U_k \cdot (1 - P_{in}) \cdot \sum_k P_{vis_k}(i) \cdot p_{cVD} \cdot p_{v_r/v}(i, l) \cdot P_l(j) \quad (4.46)$$

- Suche nach dem näheren der beiden Knoten (**Fall_1RV**):

$$\lambda_{rv}(i, j) = \lambda_r(i, j) + \lambda_v(i, j \mid d(i, k) > d(i, j)) - \lambda_v(i, j \mid d(i, k) < d(i, j)) \quad (4.47)$$

Die Symbole $d(i, k)$ und $d(i, j)$ stellen die Suchentfernungen im Verzeichnisbaum entsprechend zwischen den Knoten i und k sowie zwischen i und j dar.

Die Variable P_{in} drückt die Wahrscheinlichkeit aus, daß die angerufenen Teilnehmer sich innerhalb des eigenen Heimteilgebietes befinden:

$$P_{in} = \frac{\sum_k U_k \cdot \left(\sum_l P_k(l) \right)}{\sum_k U_k} \quad (4.48)$$

Die Wahrscheinlichkeit P_{vis} beschreibt die Verteilung der Benutzer innerhalb eines Besuchsteilgebietes (analog wie P die Verteilung der Benutzer innerhalb des Heimteilgebietes beschreibt, wobei P_{vis} die Verteilung ohne Unterscheidung der Benutzerklassen beschreibt, d.h. $\sum_k \sum_l P_{vis_k}(i) = 1$). Im Falle des fehlenden Heimknotens

im Besuchsgebiet (z.B. beim V-Szenario) gilt in allen Formeln:

$$\sum_k P_{vis_k}(i) = P_{vis}(i).$$

Der Parameter k_{CS} stellt die Relation zwischen der Anzahl der Benutzer und der Gesamtanzahl der Anrufe dar (wie analog dazu k_M die Relation zwischen der Anzahl der Benutzer und den Bewegungen darstellt).

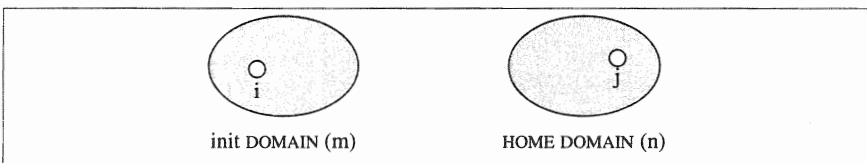


Abb. 4.5 Suchanfrage - Hauptfall 2

Im **Hauptfall 2** können folgende Ereignisraten spezifiziert werden:

- Innerhalb des anrufinitiiierenden Gebietes:

$$\lambda(i, 0)/k_{CS} = \left(\sum_k U_k \cdot P_k(i) \cdot (1 - p_{cHD}) + \sum_k U_k \cdot (1 - P_{in}) \cdot \sum_k P_{vis_k}(i) \cdot (1 - p_{cVD}) \right) \cdot P_{in} \quad (4.49)$$

- Im Heimteilgebiet des Angerufenen:

- Suche nach dem *resident*-Knoten (**Fall_2R**):

$$\lambda_r(0, j) = \sum_i \lambda(i, 0) \cdot \frac{U_j}{\sum_k U_k} \quad (4.50)$$

- Suche nach dem *visited*-Knoten (**Fall_2V**):

$$\lambda_v(0, j) = \sum_i \lambda(i, 0) \cdot \frac{U_l \cdot P_l(j)}{\sum_k U_l \cdot P_l(k)} \quad (4.51)$$

- Suche nach dem näheren der beiden Knoten (**Fall_2RV**):

$$\begin{aligned} \lambda_{rv}(0, j) &= \lambda_v(0, j), \text{ wenn der } \textit{visited}\text{-Knoten existiert} \\ &= \lambda_r(0, j), \text{ wenn nur der } \textit{resident}\text{-Knoten existiert} \end{aligned}$$

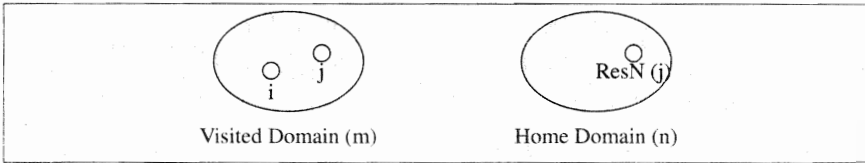


Abb. 4.6 Suchanfrage - Hauptfall 3

Im **Hauptfall 3** können folgende Ereignisraten spezifiziert werden:

- Suche nach dem *resident*-Knoten (**Fall_3R**):

$$\begin{aligned} \lambda_r(i, j)/k_{CS} = & \left(\sum_k U_k \cdot P_k(i) \cdot (1 - p_{cHD}) + \sum_k U_k \cdot (1 - P_{in}) \cdot \right. \\ & \left. \sum_k P_{vis_k}(i) \cdot (1 - p_{cVD}) \right) \cdot \frac{(1 - P_{in})}{N_D - 1} \cdot \sum_l P_{vis_f}(l) \end{aligned} \quad (4.52)$$

- Suche nach dem *visited*-Knoten (**Fall_3V**):

$$\begin{aligned} \lambda_v(i, j)/k_{CS} = & \left(\sum_k U_k \cdot P_k(i) \cdot (1 - p_{cHD}) + \sum_k U_k \cdot (1 - P_{in}) \cdot \right. \\ & \left. \sum_k P_{vis_k}(i) \cdot (1 - p_{cVD}) \right) \cdot \frac{(1 - P_{in})}{N_D - 1} \cdot \sum_l P_{vis_f}(l) \end{aligned} \quad (4.53)$$

- Suche nach dem näheren der beiden Knoten (**Fall_3RV**): wie (4.47).

Im **Hauptfall 4** können folgende Ereignisraten spezifiziert werden:

- Im Heimteilgebiet des Angerufenen:
 - Suche nach dem *resident*-Knoten (**Fall_4_H**):

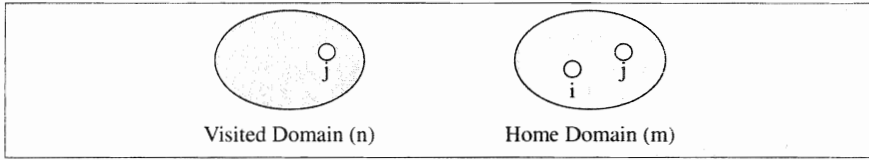


Abb. 4.7 Suchanfrage - Hauptfall 4

$$\lambda_{H,r}(i,j)/k_{CS} = \left(U_i \cdot P_i(i) \cdot p_{cHD} \cdot p_{h,r}(i,j) + \right. \\ (1 - loc_b) \cdot \sum_{k \neq i} U_k \cdot P_k(i) \cdot p_{cHD} \cdot p_{h,r}(k,j) + \\ loc_b \cdot \left(\sum_{k \neq i} U_k \cdot P_k(i) \right) \cdot p_{cHD} \cdot p_{h,v}(i,j) + \\ \left. \sum_k U_k \cdot (1 - P_in) \cdot \sum_k P_vis_k(i) \cdot p_{cVD} \cdot p_{v,r/v}(i,j) \right) \cdot (1 - \sum_k P_j(k)) \quad (4.54)$$

- Im Besuchsteilgebiet des Angerufenen:

- Suche nach dem *resident*-Knoten (**Fall_4_BR**):

$$\lambda_{V,r}(0,l) = \sum_i \sum_j \lambda_{H,r}(i,j) \cdot \sum_k P_vis_l(k) \quad (4.55)$$

- Suche nach dem *visited*-Knoten (**Fall_4_BV**):

$$\lambda_{V,v}(0,l) = \sum_i \sum_j \lambda_{H,r}(i,j) \cdot \sum_k P_vis_k(l) \quad (4.56)$$

- Suche nach dem näheren der beiden Knoten (**Fall_4_BRV**):

$$\lambda_{V,rv}(0,l) = \lambda_{V,v}(0,l), \text{ wenn der } visited\text{-Knoten existiert} \\ = \lambda_{V,r}(0,l), \text{ wenn nur der } resident\text{-Knoten existiert} \quad (4.57)$$

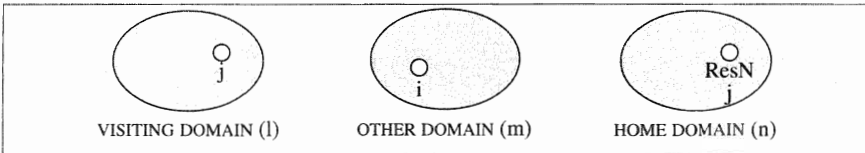


Abb. 4.8 Suchanfrage - Hauptfall 5

Im **Hauptfall 5** können folgende Ereignisraten spezifiziert werden: $\sum_l P_vis_j(l)$

- Im initiierten Teilgebiet (**Fall_5_ini**):

$$\lambda_I(i,0)/k_{CS} = \left(\sum_k U_k \cdot P_k(i) \cdot (1 - p_{cHD}) + \right. \\ \left. \sum_k U_k \cdot (1 - P_in) \cdot \sum_k P_vis_k(i) \cdot (1 - p_{cVD}) \right) \cdot (1 - P_in) \cdot \frac{N_D - 2}{N_D - 1} \quad (4.58)$$

- Im Heimatgebiet des Angerufenen:

$$\lambda_H(0, j) = \sum_i \lambda_I(i, 0) \cdot \frac{U_j}{\sum_k U_k} \quad (4.59)$$

- Im Besuchsteilgebiet des Angerufenen:
 - Suche nach dem *resident*-Knoten (**Fall_5R**):

$$\lambda_{V,r}(0, j) = \sum_i \lambda_I(i, 0) \cdot \sum_l P_{vis_j}(l) \quad (4.60)$$

- Suche nach dem *visited*-Knoten (**Fall_5V**):

$$\lambda_{V,v}(0, j) = \sum_i \lambda_I(i, 0) \cdot \sum_l P_{vis_l}(j) \quad (4.61)$$

- Suche nach dem näheren der beiden Knoten, entweder nach dem *resident*- oder dem *visited*-Knoten (**Fall_5RV**):

$$\begin{aligned} \lambda_{V,rv}(0, j) &= \lambda_{V,v}(0, j), \text{ wenn der } \textit{visited}\text{-Knoten existiert} \\ &= \lambda_{V,r}(0, j), \text{ wenn nur der } \textit{resident}\text{-Knoten existiert} \end{aligned}$$

4.4.1.3 Leistungsfähigkeit der Suchanfragen

Nach einer formalen Beschreibung der Relation zwischen dem Anrufenden und dem Angerufenen wird hier der Einfluß der Datenverteilungs- und Architekturalternativen auf die Leistungsfähigkeit der Suchanfrage qualitativ diskutiert.

Folgende Abhängigkeit zwischen den Datenbankalternativen und der Leistungsfähigkeit der Suchanfrage läßt sich vermuten:

1. In bezug auf die Anzahl der DIR-Ebenen läßt sich die Vermutung äußern, daß ein hoher Baum die Suchanfrage verlängern wird, wenn man die gleiche Knotenauslastung voraussetzt. Eine unveränderte Knotenauslastung verlangt eine Anpassung der Leistungsfähigkeit in den einzelnen Knoten; die maximal notwendige Leistung in den einzelnen Knoten steigt mit der Abflachung der Baumhierarchie.
2. Eine Benutzeranpassung der Struktur des DIR-Baumes wird geringere Bedeutung haben, da die Anpassung der Mobilität des Angerufenen durchgeführt wird. Soll jedoch die Verteilung der Anrufe und der Mobilität stark korrelieren, dann ist eine Reduktion der Suchzeiten und der Netzbelastung zu erwarten.
3. In bezug auf die Verteilung der Benutzerdaten ist eine eindeutige Präferenz schwierig zu bestimmen, da die Netzbelastung stark davon abhängig ist, wie die Relation zwischen der Mobilität und der Anrufverteilung aussieht. Im Falle einer niedrigen Mobilität des Angerufenen kann erwartet werden, daß solche Anfragen vorteilhaft sein werden, die den *resident*-Knoten suchen. Wenn eine hohe Mobilität die angerufenen Benutzer charakterisiert, kann dagegen erwartet werden, daß Anfragen, die den *visited*-Knoten suchen, Vorteile bieten werden.

4. In bezug auf die Wahl der Abfragetechnik ist es naheliegend anzunehmen, daß die *passing*-Technik zu bevorzugen ist, besonders wenn die Aspekte der Zuverlässigkeit ausgelassen werden.
5. In bezug auf die Zeigerverwaltung im DIR-Baum kann aus Sicht der Suchanfrage eine eindeutige Präferenz für die direkten Zeiger erwartet werden.

4.4.2 Aktualisierungsanfragen

Die Beschreibung der beiden Aktualisierungsanfragen, d.h. während der *Location* und *Domain Update*-Prozedur, wird hier wegen ihrer grundsätzlichen Ähnlichkeit zusammen durchgeführt. Der Unterschied zwischen den beiden Prozeduren liegt lediglich in deren Komplexität: Bei der *Domain Update*-Prozedur wird auch global in den Datenbanken der beteiligten Teilgebiete für Konsistenz gesorgt. Die während der Prozeduren initiierten Datenbankabfragen sind grundsätzlich identisch.

Die Aktualisierungsanfragen zeichnen sich durch folgende Eigenschaften aus:

- Es sind die wesentlichen Anfragen, die eine Modifizierung der Einträge im Verzeichnisbaum vornehmen; ansonsten wird die Modifizierung während der *User Registration / User Deregistration*-Prozedur durchgeführt.
- Im Gegensatz zur Suchanfrage sind diese Anfragen von der Beweglichkeit des Benutzers stark abhängig.

In folgenden Unterkapiteln wird ein kurzer Überblick der beiden Aktualisierungsanfragen gegeben, dem eine wahrscheinlichkeitbasierte Beschreibung folgen wird. Eine Diskussion der Einflüsse von Datenverteilungs- und Architekturalternativen auf die Anfragen schließt das Kapitel ab.

4.4.2.1 Überblick

Ziel der Aktualisierungsanfrage ist, die Bewegungen des Teilnehmers innerhalb des Teilgebietes zu erfassen und notwendige Aktionen vorzunehmen, um die Konsistenz der Daten in der verteilten UMTS-Datenbank zu gewährleisten. Die Anfrage besteht somit aus einer Nachrichtenfolge, die folgende Hauptaufgaben zu erfüllen hat:

1. gegebenenfalls das Auffinden der Daten im 'alten' DN-Knoten,
2. Modifizierung der durch die Bewegungen beeinflussten Benutzer- und Endgerätedaten in den DN-Knoten, d.h. *routing*-, *registration*- und *terminal*-Daten, sowie
3. Modifizierung der durch die Bewegung beeinflussten Zeiger in den *DIR*-Knoten.

Bei der Aktualisierung ist wesentlich zu unterscheiden, in welcher Relation die beteiligten Teil- und DN-Gebiete mit dem Heimteilgebiet und dem eventuell vorhandenen *resident*-DN-Knoten stehen. Dies wird auch als die Grundlage der formalen Beschreibung verwendet. Ähnlich wie bei der Suchanfrage wird dadurch die Auswertung der speziellen Benutzerdatenverteilungsmuster und der entsprechenden

Nachrichtenfolgen zu einer Aneinanderreihung von definierten Sequenzen mit bekannten Häufigkeiten reduziert.

Folgendes sei bei der Auswertung von Aktualisierungsanfragen zu beachten:

- Die wahrgenommenen Bewegungen des Benutzers werden nicht nur durch die absolute Beweglichkeit, sondern auch durch die Größe der DN-Gebiete und durch die Relation zwischen den Aufenthalts- und den DN-Gebieten beeinflusst:
 - Große Aufenthaltsgebiete reduzieren die Anzahl der Bewegungen innerhalb eines DN-Gebietes, und
 - große DN-Gebiete reduzieren die Anzahl der Bewegungen zwischen den DN-Gebieten.
- Die Relation zwischen der zeitlichen Dauer und der Häufigkeit der Ein- oder Ausschaltphasen des Endgerätes beeinflusst die Erfassungsrate der Bewegungen: Lange Ausschaltphasen bei gleicher Beweglichkeit reduzieren die Anzahl der erfaßten Bewegungen und damit auch die gesamte Netzbelastung.

In den folgenden Ausführungen wird angenommen, daß alle Benutzerbewegungen von der Datenbank erfaßt werden. Dies ist keine sehr strenge Forderung, wenn man die erwarteten Ausdehnungen der DN- und erst recht der Teilgebiete (/SSD94/) und somit die innerhalb eines Gebietes verbrachte Zeit beachtet. Die Aktualisierungsanfrage wird demnach immer durch die Bewegung ausgelöst (nicht unbedingt direkt an der Gebietsgrenze). Diese Annahme stellt den schlimmsten Fall der Netzbelastung dar und soll als Grenzwert angesehen werden.

4.4.2.2 Formale Beschreibung

Es gibt insgesamt fünf Hauptfälle, wie der neue und der alte Standort des Benutzers in Relation zum Heimteilgebiet stehen. Im folgenden stehen HD für das Heimteilgebiet (*Home Domain*), VD für das Besuchteilgebiet (*Visited Domain*), o für das alte (*old*), gerade verlassene Teilgebiet und n (*new*) für das neue Gebiet:

- Der Benutzer bewegt sich innerhalb eines Teilgebietes (*Location Update*):
 1. Hauptfall: innerhalb des Gebietes des *Home Domain* ($VD_n = VD_o = HD$)
 2. Hauptfall: innerhalb des Gebietes des *Visited Domain* ($VD_n = VD_o \neq HD$)
- Der Benutzer bewegt sich zwischen den Teilgebieten (*Domain Update*):
 3. Hauptfall: aus dem Gebiet des *Home Domain* in das eines *Visited Domain* ($VD_n \neq VD_o$) & ($HD = VD_o$)
 4. Hauptfall: aus dem Gebiet eines *Visited Domain* in das des *Home Domain* ($VD_n \neq VD_o$) & ($HD = VD_n$)
 5. Hauptfall: aus dem Gebiet eines *Visited domain* in das eines anderen ($VD_n \neq VD_o$) & ($HD \neq VD_o$) & ($HD \neq VD_n$).

Unabhängig von obigem gibt es sechs Unterfälle, wie der neue und der alte Standort des Benutzers in Relation zum DN-Heimgebiet stehen. Im folgenden stehen RN für das DN-Heimgebiet (*resident*-DN-Gebiet), VN für das DN-Besuchsgebiet (*visited*-DN-Gebiet), o für das alte, gerade verlassene DN-Gebiet und n für das neue Gebiet:

- Der Benutzer bewegt sich innerhalb eines DN-Gebietes:
 1. Unterfall: innerhalb des Gebietes des *resident*-DN-Knotens
($VN_n = VN_o = RN$)
 2. Unterfall: innerhalb des Gebietes des *visited* DN-Knotens
($VN_n = VN_o \neq RN$)
- Der Benutzer bewegt sich zwischen den DN-Gebieten:
 3. Unterfall: aus dem Gebiet des *resident*- in das eines *visited*-DN-Knotens
($VN_n \neq VN_o$) & ($RN = VN_o$)
 4. Unterfall: aus dem Gebiet eines *visited*- in das des *resident*-DN-Knotens
($VN_n \neq VN_o$) & ($RN = VN_n$)
 5. Unterfall: aus dem Gebiet eines *visited*-DN-Knotens in das eines anderen
($VN_n \neq VN_o$) & ($RN \neq VN_o$) & ($RN \neq VN_n$)
 6. Unterfall: aus dem Gebiet eines *resident*-DN-Knotens in das eines anderen
($VN_n \neq VN_o$) & ($RN = VN_o$) & ($RN = VN_n$).

Folgendes ist zu beachten, wenn die Unterfälle in die Hauptfälle eingebettet werden:

- Der 6. Unterfall kann nur eintreten, wenn das Teilgebiet gewechselt wird.
- Der 1. und 2. Unterfall können nur eintreten, wenn die Bewegung innerhalb eines Teilgebietes stattfindet.
- Die Unterfälle, in denen der *resident*-DN-Knoten beteiligt ist, können nur dann eintreten, wenn die Existenz dieses Knotens in einem betrachteten Datenverteilungsszenario (vgl. Kapitel 3.4.2.1) vorgesehen ist.

Folgende Konventionen bezüglich der Indizierung sollen beachtet werden:

- Die Nummer i steht stets für das DN-Gebiet, das gerade vom Benutzer verlassen worden ist, und
- die Nummer j steht für das DN-Gebiet, das vom Benutzer neu betreten worden ist.

Gemäß dieser Bezeichnung bedeutet $\lambda(i, j)$ die Aktualisierungsrate (und $prob(i, j)$ die entsprechende Wahrscheinlichkeit) aus dem DN-Gebiet i in das DN-Gebiet j . Die Werte $i = 0$ und $j = 0$ symbolisieren entsprechend die in ein Teilgebiet ankommenden und die aus einem Teilgebiet abgehenden Bewegungen.

Es sei im weiteren zu beachten, daß die Auswertung die Kommunikation innerhalb eines Teilgebietes erfaßt. Die Aktionen, die sich zwischen Teilgebieten ereignen, werden in zwei Komponenten zerlegt: innerhalb der Teilgebiete und zwischen den Teilgebieten. Die Aktionen innerhalb der Teilgebiete werden so verrechnet, als ob sie

sich innerhalb des gleichen Teilgebietes ereignen würden. Das betrachtete Teilgebiet wird somit gleichzeitig als das alte und das neue Teilgebiet angesehen.

Im **Hauptfall 1** ($VD_n = VD_o = HD$) können folgende Aktualisierungsraten spezifiziert werden:

- Unterfall 1 ($VN_n = VN_o = RN$):

$$\lambda(i, i)/k_M = U_i \cdot P_i(i) \cdot p_i(i, i) \quad (4.62)$$

- Unterfall 2 ($VN_n = VN_o \neq RN$):

$$\lambda(i, i)/k_M = \sum_{k, k \neq i} U_k \cdot P_k(i) \cdot p_k(i, i) \quad (4.63)$$

- Unterfall 3 ($VN_n \neq VN_o$) & ($RN = VN_o$):

$$\lambda(i, j)/k_M = U_i \cdot P_i(i) \cdot p_i(i, j) \quad (4.64)$$

- Unterfall 4 ($VN_n \neq VN_o$) & ($RN = VN_n$):

$$\lambda(i, j)/k_M = U_j \cdot P_j(i) \cdot p_j(i, j) \quad (4.65)$$

- Unterfall 5 ($VN_n \neq VN_o$) & ($RN \neq VN_o$) & ($RN \neq VN_n$):

$$\lambda(i, j)/k_M = \sum_{k, k \neq i \neq j} U_k \cdot P_k(i) \cdot p_k(i, j) \quad (4.66)$$

Im **Hauptfall 2** ($VD_n = VD_o \neq HD$) können folgende Aktualisierungsraten spezifiziert werden:

- Unterfall 1 ($VN_n = VN_o = RN$):

$$\lambda(i, i)/k_M = \sum_k U_k \cdot (1 - P_{in}) \cdot P_{vis_i}(i) \cdot p_{vis_i}(i, i) \quad (4.67)$$

- Unterfall 2 ($VN_n = VN_o \neq RN$):

- *resident-DN-Knoten existiert:*

$$\lambda(i, i)/k_M = \sum_n U_n \cdot (1 - P_{in}) \cdot \sum_{k, k \neq i} P_{vis_k}(i) \cdot p_{vis_k}(i, i) \quad (4.68)$$

- *resident-DN-Knoten existiert nicht:*

$$\lambda(i, i)/k_M = \sum_n U_n \cdot (1 - P_{in}) \cdot P_{vis}(i) \cdot p_{vis}(i, i) \quad (4.69)$$

- Unterfall 3 ($VN_n \neq VN_o$) & ($RN = VN_o$):

$$\lambda(i, j)/k_M = \sum_n U_n \cdot (1 - P_{in}) \cdot P_{vis_i(i)} \cdot p_{vis_i(i, j)} \quad (4.70)$$

- Unterfall 4 ($VN_n \neq VN_o$) & ($RN = VN_n$):

$$\lambda(i, j)/k_M = \sum_n U_n \cdot (1 - P_{in}) \cdot P_{vis_j(i)} \cdot p_{vis_j(i, j)} \quad (4.71)$$

- Unterfall 5 ($VN_n \neq VN_o$) & ($RN \neq VN_o$) & ($RN \neq VN_n$):

- *resident-DN-Knoten* existiert:

$$\lambda(i, j)/k_M = \sum_n U_n \cdot (1 - P_{in}) \cdot \sum_{k, k \neq i \neq j} P_{vis_k(i)} \cdot p_{vis_k(i, j)} \quad (4.72)$$

- *resident-DN-Knoten* existiert nicht:

$$\lambda(i, j)/k_M = \sum_n U_n \cdot (1 - P_{in}) \cdot P_{vis(i)} \cdot p_{vis(i, j)} \quad (4.73)$$

Die Parameter P_{in} und P_{vis} wurden bereits im Kapitel 4.4.1.2 definiert.

Die Wahrscheinlichkeit p_{vis} entspricht der Wahrscheinlichkeit p (vgl. Kapitel 4.2), wobei p_{vis} die Werte für die Benutzer innerhalb des Besuchsteilgebietes beschreibt.

Wenn ein betrachtetes Szenario keinen *resident-DN-Knoten* im Besuchsteilgebiet vorsieht, sind nur die Unterfälle 2 und 5 im Besuchsteilgebiet zulässig.

Die Beschreibungen von $P_{vis_k(i)}$ und $p_{vis_k(i, j)}$ vereinfachen sich entsprechend zu $P_{vis(i)}$ und $p_{vis(i, j)}$. Dies gilt für alle folgenden Betrachtungen.

Im **Hauptfall 3** ($VD_n \neq VD_o$) & ($HD = VD_o$) können folgende Aktualisierungsraten spezifiziert werden:

- aus der Perspektive des alten Teilgebietes:

- Unterfall 3 ($VN_n \neq VN_o$) & ($RN = VN_o$) & ($RN \neq VN_n$) zusammen mit Unterfall 6 ($VN_n \neq VN_o$) & ($RN = VN_o$) & ($RN = VN_n$):

$$\lambda_{3/6}(i, 0)/k_M = U_i \cdot (1 - \sum_l P_i(l)) \cdot P_i(i) \cdot p_i(i, 0) \quad (4.74)$$

- Unterfall 4 ($VN_n \neq VN_o$) & ($VN_o \neq RN$) & ($VN_n = RN$) zusammen mit Unterfall 5 ($VN_n \neq VN_o$) & ($VN_o \neq RN$) & ($VN_n \neq RN$):

$$\lambda_{4/5}(i, 0)/k_M = \sum_{k, k \neq i} U_k \cdot (1 - \sum_l P_k(l)) \cdot P_k(i) \cdot p_k(i, 0) \quad (4.75)$$

- aus der Perspektive des neuen Teilgebietes:

- Unterfall 3 ($VN_n \neq VN_o$) & ($VN_o = RN$) & ($VN_n \neq RN$) zusammen mit Unterfall 5 ($VN_n \neq VN_o$) & ($VN_o \neq RN$) & ($VN_n \neq RN$)

- *resident-DN-Knoten* existiert:

$$\lambda(0, j) = \sum_i \left(\lambda_{3/6}(i, 0) + \lambda_{4/5}(i, 0) \right) \cdot \sum_{k, k \neq j} P_{vis_k(j)} \cdot p_{vis_k(0, j)} / norm_{vis} \quad (4.76)$$

- *resident-DN-Knoten* existiert nicht:

$$\lambda(0, j) = \sum_i \left(\lambda_{3/6}(i, 0) + \lambda_{4/5}(i, 0) \right) \cdot P_{vis}(j) \cdot p_{vis}(0, j) / norm_{vis} \quad (4.77)$$

- Unterfall 4 ($VN_n \neq VN_o$) & ($VN_o \neq RN$) & ($VN_n = RN$) zusammen mit Unterfall 6 ($VN_n \neq VN_o$) & ($VN_o = RN$) & ($VN_n = RN$):

$$\lambda(0, j) = \sum_i \left(\lambda_{3/6}(i, 0) + \lambda_{4/5}(i, 0) \right) \cdot P_{vis_j}(j) \cdot p_{vis_j}(0, j) / norm_{vis} \quad (4.78)$$

Dabei gilt:

$$norm_{vis} = \sum_k \sum_j P_{vis_k}(j) \cdot p_{vis_k}(0, j)$$

Wenn ein betrachtetes Szenario keinen *resident-DN-Knoten* im Besuchsteilgebiet vorsieht, dann sind nur die Unterfälle 3 und 5 im neuen Teilgebiet zulässig. Es soll beachtet werden, daß für $p_{vis_k}(0, i)$ ähnliche Formel wie (4.8) gilt, wobei für den Wert des Parameters *cross* die Bemerkung auf der Seite 73 gilt.

Im **Hauptfall 4** ($VD_n \neq VD_o$) & ($HD = VD_n$) können folgende Aktualisierungsraten spezifiziert werden:

- aus der Perspektive des alten Teilgebietes:
 - Unterfall 3 ($VN_n \neq VN_o$) & ($VN_o = RN$) & ($VN_n \neq RN$) zusammen mit Unterfall 6 ($VN_o \neq VN_n$) & ($VN_o = RN$) & ($VN_n = RN$):
 - *resident-DN-Knoten* existiert:

$$\lambda_{3/6}(i, 0)/k_M = \sum_n U_n \cdot (1 - P_{in}) \cdot P_{vis_i}(i) \cdot p_{vis_i}(i, 0) \cdot P_{vis_home} \quad (4.79)$$

- *resident-DN-Knoten* existiert nicht:

$$\lambda_{3/6}(i, 0)/k_M = \sum_n U_n \cdot (1 - P_{in}) \cdot P_{vis}(i) \cdot p_{vis}(i, 0) \cdot P_{vis_home} \quad (4.80)$$

- Unterfall 4 ($VN_n \neq VN_o$) & ($VN_o \neq RN$) & ($VN_n = RN$) zusammen mit Unterfall 5 ($VN_o \neq VN_n$) & ($VN_o \neq RN$) & ($VN_n \neq RN$):

$$\lambda_{4/5}(i, 0)/k_M = \sum_n U_n \cdot (1 - P_{in}) \cdot \sum_{k, k \neq i} P_{vis_k}(i) \cdot p_{vis_k}(i, 0) \cdot P_{vis_home} \quad (4.81)$$

- aus der Perspektive des neuen Teilgebietes:
 - Unterfall 3 ($VN_n \neq VN_o$) & ($VN_o = RN$) & ($VN_n \neq RN$) zusammen mit Unterfall 5 ($VN_o \neq VN_n$) & ($VN_o \neq RN$) & ($VN_n \neq RN$):

$$\lambda(0, j) = \sum_i \left(\lambda_{3/6}(i, 0) + \lambda_{4/5}(i, 0) \right) \cdot \sum_{k, k \neq j} P_k(j) \cdot p_k(0, j) / norm \quad (4.82)$$

- Unterfall 4 ($VN_n \neq VN_o$) & ($VN_o \neq RN$) & ($VN_n = RN$) zusammen mit Unterfall 6 ($VN_o \neq VN_n$) & ($VN_o = RN$) & ($VN_n = RN$):

$$\lambda(0, j) = \sum_i \left(\lambda_{3/6}(i, 0) + \lambda_{4/5}(i, 0) \right) \cdot P_j(j) \cdot p_j(0, j) / norm \quad (4.83)$$

Wenn ein betrachtetes Szenario keinen *resident-DN*-Knoten im Besuchsteilgebiet vorsieht, dann sind nur die Unterfälle 3 und 6 im alten Teilgebiet zulässig. Die Wahrscheinlichkeit P_{vis_home} beschreibt die bedingte Wahrscheinlichkeit für den Übergang aus dem Besuchsteilgebiet ins Heimteilgebiet, wenn ein Benutzer das Besuchsteilgebiet verläßt ($P_{vis_home} = \lambda_{v_h} / \lambda_{v_x}$).

Im **Hauptfall 5** ($VD_n \neq VD_o$) & ($HD \neq VD_o$) & ($HD \neq VD_n$) können folgende Aktualisierungsraten spezifiziert werden:

- aus der Perspektive des alten Teilgebietes:
 - Unterfall 3 ($VN_n \neq VN_o$) & ($VN_o = RN$) & ($VN_n \neq RN$) zusammen mit Unterfall 6 ($VN_o \neq VN_n$) & ($VN_o = RN$) & ($VN_n = RN$):
 - *resident-DN*-Knoten existiert:

$$\lambda_{3/6}(i, 0) / k_M = \sum_n U_n \cdot (1 - P_{in}) \cdot P_{vis_i}(i) \cdot p_{vis_i}(i, 0) \cdot (1 - P_{vis_home}) \quad (4.84)$$

- *resident-DN*-Knoten existiert nicht:

$$\lambda_{3/6}(i, 0) / k_M = \sum_n U_n \cdot (1 - P_{in}) \cdot P_{vis}(i) \cdot p_{vis}(i, 0) \cdot (1 - P_{vis_home}) \quad (4.85)$$

- Unterfall 4 ($VN_n \neq VN_o$) & ($VN_o \neq RN$) & ($VN_n = RN$) zusammen mit Unterfall 5 ($VN_o \neq VN_n$) & ($VN_o \neq RN$) & ($VN_n \neq RN$):

$$\lambda_{4/5}(i, 0) / k_M = \sum_n U_n \cdot (1 - P_{in}) \cdot \sum_{k, k \neq i} P_{vis_k}(i) \cdot p_{vis_k}(i, 0) \cdot (1 - P_{vis_home}) \quad (4.86)$$

- aus der Perspektive des neuen Teilgebietes:
 - Unterfall 3 ($VN_n \neq VN_o$) & ($VN_o = RN$) & ($VN_n \neq RN$) zusammen mit Unterfall 5 ($VN_o \neq VN_n$) & ($VN_o \neq RN$) & ($VN_n \neq RN$)
 - *resident-DN*-Knoten existiert:

$$\lambda(0, j) = \sum_i \left(\lambda_{3/6}(i, 0) + \lambda_{4/5}(i, 0) \right) \cdot \sum_{k, k \neq j} P_{vis_k}(j) \cdot p_{vis_k}(0, j) / norm_{vis} \quad (4.87)$$

- *resident-DN*-Knoten existiert nicht:

$$\lambda(0, j) = \sum_i \left(\lambda_{3/6}(i, 0) + \lambda_{4/5}(i, 0) \right) \cdot P_{vis}(j) \cdot p_{vis}(0, j) / norm_{vis} \quad (4.88)$$

- Unterfall 4 ($VN_n \neq VN_o$) & ($VN_o \neq RN$) & ($VN_n = RN$) zusammen mit Unterfall 6 ($VN_o \neq VN_n$) & ($VN_o = RN$) & ($VN_n = RN$):

$$\lambda(0, j) = \sum_i \left(\lambda_{3/6}(i, 0) + \lambda_{4/5}(i, 0) \right) \cdot P_{vis_j}(j) \cdot p_{vis_j}(0, j) / norm_{vis} \quad (4.89)$$

Wenn ein betrachtetes Szenario keinen *resident*-DN-Knoten im *visited domain*-Gebiet vorsieht, dann sind nur die Unterfälle 3 und 6 im alten Teilgebiet sowie nur die Unterfälle 3 und 5 im neuen Teilgebiet zulässig.

4.4.2.3 Leistungsfähigkeit der Aktualisierungsanfragen

Nach der formalen Beschreibung der Bewegungsraten in den einzelnen Fällen wird hier der Einfluß der Datenverteilungs- und Architekturalternativen auf die Leistungsfähigkeit der Aktualisierungsanfragen qualitativ diskutiert.

Folgende Abhängigkeit zwischen den Datenbankalternativen und der Leistungsfähigkeit der Aktualisierungsanfragen läßt sich vermuten:

1. In bezug auf die Anzahl der DIR-Ebenen läßt sich die Vermutung äußern, daß ein hoher Baum die Aktualisierungsanfrage verlängern wird, wenn man die gleiche Knotenauslastung voraussetzt. Eine unveränderte Knotenauslastung verlangt eine Anpassung der Leistungsfähigkeit in den einzelnen Knoten; die maximal benötigte Leistung steigt mit einer Abflachung der Hierarchie an.
2. In bezug auf die Gleichmäßigkeit der Struktur des DIR-Baumes kann man erwarten, daß ein starker Einfluß zu beobachten sein wird, da die Anpassung der Struktur aufgrund der Mobilität des Benutzers durchgeführt wird. Eine dem Benutzer angepaßte Struktur wird den Aktualisierungsaufwand minimieren.
3. In bezug auf die Verteilung der Benutzerdaten ist eine eindeutige Präferenz schwierig festzustellen, da die Ergebnisse von der jeweiligen Relation zwischen den Inputdaten des Mobilitätsmodells sehr stark abhängig sind. Wenn der betrachtete Benutzer relativ unbeweglich ist und sich hauptsächlich im Gebiet des DN-Heimknotens aufhält, ist zu erwarten, daß eine auf dem *resident*-Knoten basierende Datenverteilung vorteilhaft sein wird. Ansonsten ist zu erwarten, daß eine auf dem *visited*-DN-Knoten basierende Datenverteilung besser ist.
4. In bezug auf die Wahl der Abfragetechnik ist es offensichtlich, daß die *passing*-Technik zu bevorzugen ist, besonders wenn die Aspekte der Zuverlässigkeit ausgelassen werden.
5. In bezug auf die Zeigerverwaltung sind zwei Fälle zu unterscheiden:
 - Bei Anwendung der Alternative, in der das Endgerät die Information über den alten Aufenthaltsort übermittelt, ist keine Suche notwendig. Hier ist eine indirekte Zeigeroption für die Aktualisierung vorteilhaft.
 - Betrachtet man die Alternative, die eine Suche nach dem alten Aufenthaltsort vorsieht, dann kann keine eindeutige Präferenz beobachtet werden. Hier ist die Relation zwischen der DIR-Ebene, bis zu der gesucht wird, und der Anzahl der DIR-Ebenen, welche noch darüber liegen, von entscheidender Bedeutung.
6. Im Zusammenhang mit den obigen zwei Alternativen ist folgender Einfluß zu erwarten: Wenn die Kennung des vorherigen Aufenthaltsgebietes im Endgerät gespeichert und diese dem neuen DN-Knoten mitgeteilt wird, reduziert sich der Gesamtaufwand erheblich, weil der 'alte' DN-Knoten nicht gesucht werden muß.

5 Leistungsbewertung

Im vorliegenden Kapitel werden die gewählten Anfragen, d.h. das Suchen nach dem Aufenthaltsort des Angerufenen und das Aktualisieren der Datenbank aufgrund der Benutzerbewegungen, numerisch ausgewertet.

Mit Hilfe der hergeleiteten Modelle ist es möglich, die Beschreibung der Charakteristika der einzelnen Anfragen zu formalisieren. Trotz der getroffenen Vereinfachungen ist die Anzahl der wechselwirkenden Parameter und Alternativen recht groß. Jede spezielle Kombination von Parametern und Alternativen wird im folgenden als Konfiguration bezeichnet. Die Hauptschwierigkeit der Bewertung liegt darin, daß es nicht immer zulässig ist, aus der untersuchten Teilmenge aller Konfigurationen auf das globale Optimum zu schließen /Jain91/. Dies ist um so schwieriger, je unterschiedlicher die einzelnen Tendenzen ausfallen. Hier wird wie folgt vorgegangen:

- Das Suchen und das Aktualisieren werden getrennt untersucht.
- Die Auswirkungen der Parameteränderungen werden separat präsentiert.
- Alle Abfragetechniken und Datenverteilungsszenarien werden, so weit es angemessen erscheint, in jeder Konfiguration komplett ausgewertet.
- Die *inter-domain*- und die *intra-domain*-Prozesse werden separat untersucht.

Die benutzten Parameter, Benutzerzahlen und Topologien werden zusammen mit den jeweiligen Ergebnissen präsentiert, um die Relation zwischen den Ergebnissen und den Annahmen zu erkennen. Aus den untersuchten Konfigurationen werden dann Schlußfolgerungen abgeleitet. Einige Sonderfälle werden separat diskutiert, z.B. das GSM-ähnliche Szenario.

Auf eine Wiedergabe der Zahlenwerte, mit Ausnahme der exemplarischen Grundkonfiguration, wird hier verzichtet, da dies den Lesefluß unnötig stören würde. Zwecks Vollständigkeit wurden die Ergebnisse im Anhang B zusammengestellt.

5.1 Programmierungsumgebung

Die durchgeführte Auswertung hat einen quantitativen Charakter: Auf der Grundlage der statistischen Modelle und der entwickelten Informationsflüsse für die untersuchten Anfragen wurde mittels Matlab®, Version 4.2a für Unix®, ein Softwarepaket implementiert. Mit dessen Hilfe wurden die numerischen Ergebnisse berechnet und aufgearbeitet.

5.1.1 Matlab

Matlab ist ein mathematisches Tool, dessen Funktionalität erlaubt, in einer C-ähnlichen Sprache Skripte und Funktionen zu erstellen (/MatU92/ und /MatR92/). Die auf diese Art vorbereiteten Matlab-Programme werden in der Ausführungsphase von der sog. Matlab-Maschine interpretiert und liefern dann die geforderten Ergebnisse.

Die Vorteile von Matlab gegenüber anderen mathematischen Werkzeugen sind:

- die Leistungsfähigkeit - Matlab arbeitet matrixorientiert, d.h. jede Variable ist als eine Matrix dargestellt, und
- die Verfügbarkeit einer umfangreichen Bibliothek.

Die Vorteile der verwendeten Programmierumgebung gegenüber einer der traditionellen Programmiersprachen liegen zum einen in einer geringeren Fehlertoleranz, wodurch ein fehlerarmes Skript leichter zu erstellen ist, und zum anderen in der direkten Verwendbarkeit, ohne zu übersetzen und zu binden.

5.1.2 Softwarestruktur

Der Gesamtumfang der erstellten Skripte liegt in der Größenordnung von über 8000 NLOC (*Non commented Lines Of Code*). Die Struktur der erstellten Software ist modular aufgebaut:

1. Spezifikation der notwendigen Parameter
2. Auflösung des Mobilitätsmodells
3. Analyse der gewählten Prozedur
4. Aufarbeitung der Ergebnisse

Im ersten Schritt werden die Parameter gewählt, die die Größe, Struktur und Topologie der Datenbank sowie die Anzahl der Benutzer im Heimteilgebiet und außerhalb dessen spezifizieren. Des weiteren werden die Parameter gewählt, die die Mobilität und das Anrufverhalten beschreiben. Zuletzt werden die Hilfsparameter festgelegt sowie die jeweilige Anfrageart und der Hauptfall (z.B. Aktualisierung innerhalb des Heimteilgebietes), das Datenverteilungsszenario, die interne Abfragetechnik und die Zeigerverwaltungsoption angegeben.

Im zweiten Schritt wird das Mobilitätsmodell gemäß der Voreinstellung (stationärer Fall, nichtstationärer Fall oder Überlagerung von beiden) gelöst.

Im dritten Schritt, der den größten Umfang des Matlab-Skriptes einnimmt, werden die Anfragen ausgewertet. Dies geschieht folgendermaßen:

1. Für jedes Datenknotenpaar, das in der Anfrage als das Start-Ende-Knotenpaar eventuell beteiligt werden könnte, wird die Wahrscheinlichkeit berechnet, daß dieses Paar tatsächlich in die Anfrage involviert ist.
Beispielsweise werden für die Suche die Möglichkeiten zwischen jedem Knotenpaar und für die Aktualisierung nur zwischen benachbarten Knoten betrachtet.
2. Die ermittelte Wahrscheinlichkeit wird mit der tatsächlichen Benutzeranzahl multipliziert und mittels der implementierten Informationsflüsse auf die Knoten der Datenbank verrechnet. Daraus ergibt sich für die betrachtete Datenbankalternative eine quantitative Beschreibung der Netzsituation wie
 - Belastung der einzelnen Knoten,
 - Belastung der Verbindungen und

- Verzögerungszeit.

Im vierten Schritt werden die erzielten Ergebnisse nachbearbeitet, d.h. zwecks einer vergleichenden Analyse komprimiert und dargestellt.

Die recht knappe Beschreibung der Softwareumgebung, die hier gegeben wurde, soll nur als Hinweis auf den Umfang und die Grundstruktur der programmtechnischen Realisierung des in der Arbeit verfolgten Ansatzes dienen. Aus Sicht der Arbeit stellt das Programm lediglich ein Mittel zur Lösung der entwickelten Modelle dar.

5.2 Grundkonfiguration

Um ein Vergleichsmaß für die Auswertung unterschiedlicher Alternativen festzulegen, wird eine sog. Grundkonfiguration gewählt, die dann in den untersuchten Alternativen entsprechend modifiziert wird. Der Begriff Grundkonfiguration soll hier keineswegs eine Wertung dieser Konfiguration geben. Sie ist ausschließlich als Bezugsmöglichkeit eingeführt worden.

5.2.1 Topologie und Parameter

Für die Grundtopologie wurde ein Teilgebiet, bestehend aus (6 mal 6)-DN-Gebieten, gewählt (Abbildung 5.1). Eine (4 mal 4)-Topologie wurde bereits in geringerem Umfang in /SSD94/ untersucht. Die dort vorliegenden Ergebnisse stimmen mit den hier präsentierten überein.

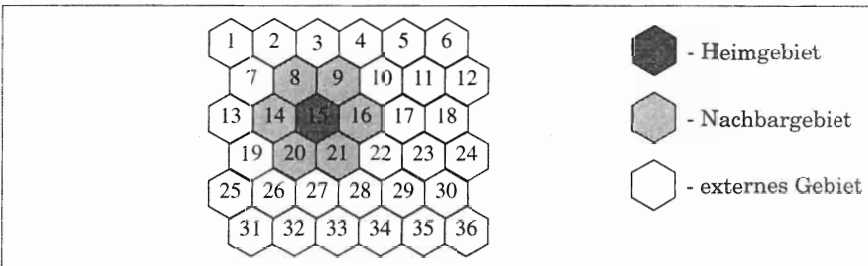


Abb. 5.1 Die Grundtopologie des UMTS-Teilgebietes

In der Grundkonfiguration wird nur eine Benutzergruppe betrachtet, die ihren Heimknoten im DN-Gebiet 15 hat. Die Gebiete 8, 9, 14, 16, 20 und 21 sind Nachbargebiete. Alle anderen Gebiete stellen externe Datenknotengebiete innerhalb des UMTS-Heimteilgebietes dar. Zwecks Vereinfachung werden in folgender Betrachtung lediglich diese drei Gebietstypen mit den assoziierten Aufenthaltswahrscheinlichkeiten unterschieden.

Die Größe der hexagonalen DN-Gebiete basiert auf einer Kantenlänge von 0,5 Einheiten - der diagonale Durchmesser entspricht dann einer Einheit.

Das Verzeichnis der Grundkonfiguration wird auf eine symmetrische 36-4-1-Struktur festgelegt, die in Abbildung 5.2 präsentiert ist. Die Zahlen in der Strukturbezeichnung (hier 36-4-1) beschreiben, wieviele Knoten sich in den einzelnen Ebenen befinden, wobei die erste Zahl die Anzahl der Datenknoten angibt und die sonstigen Zahlen sich auf die Anzahl der Verzeichnisknoten beziehen.

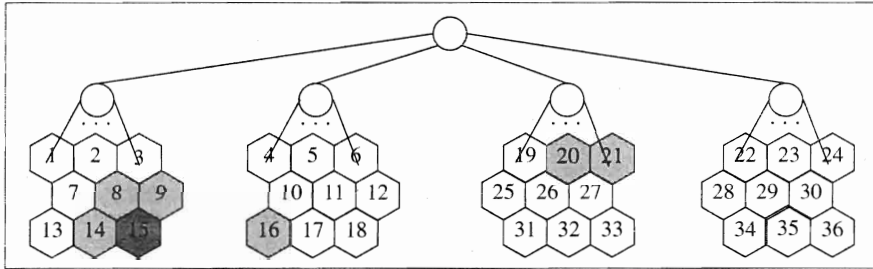


Abb. 5.2 Die Grundstruktur des Verzeichnisses

Der geographische Standort jedes einzelnen DIR-Knotens innerhalb des Gebietes ergibt sich idealisiert aus dem arithmetischen Mittelwert der Koordinaten der Standorte der DN-Knoten, die mit Hilfe dieses DIR-Knotens verbunden werden.

Die Verwaltung der DIR-Zeiger wird in der Grundkonfiguration in folgender Art organisiert:

- Die Datenverteilungsszenarien V und RV verwenden indirekte Zeiger.
- Die Datenverteilungsszenarien ST und R verwenden direkte Zeiger.

Folgende Mobilitätsparameter wurden für die Grundkonfiguration gewählt:

- Die Wahrscheinlichkeitsverteilung P_r im Heimteilgebiet entspricht einer **mittleren** Konzentration der Benutzer um den Heimknoten. Dabei werden Aufenthaltswahrscheinlichkeiten für die drei genannten Gebietstypen festgelegt:

$$P(\text{Heimgebiet}) = 0,4,$$

$$P(\text{Nachbargebiet}) = 0,45$$

$$P(\text{externes Gebiet}) = 0,05$$

- Die Wahrscheinlichkeit, daß der Benutzer sich außerhalb des Heimteilgebietes aufhält, ergibt sich zu:

$$P_v = 1 - P_h = 1 - (0,4 + 0,45 + 0,05) = 0,1$$

- Die Wahrscheinlichkeit, daß ein Benutzer während seiner Bewegung das DN-Gebiet wechselt und dabei innerhalb des Heimteilgebietes verbleibt, beträgt:

$$P_{out} = 0,1$$

- Die Wahrscheinlichkeitsverteilung P_v im Besuchsteilgebiet wird angenommen:

$$P_v(\text{beliebiges DN-Gebiet}) = 0,1/36 \text{ (gleichmäßige Verteilung)}$$

- Die Wahrscheinlichkeit, daß ein Benutzer während seiner Bewegung das DN-Gebiet wechselt und dabei innerhalb des Heimteilgebietes verbleibt, beträgt:

$$P_{out} = 0,2$$

Folgende Anrufparameter wurden für die Grundkonfiguration gewählt:

- Die Wahrscheinlichkeit, daß die Suche innerhalb des gleichen Teilgebietes (sowohl für die Benutzer im Heimteilgebiet als auch in einem Besuchsteilgebiet) durchgeführt wird, beträgt:

$$p_{cHD} = p_{cVD} = 0,7$$

- Die Wahrscheinlichkeit, daß sich ein Benutzer außerhalb des Heimknotens lokal verhält, beträgt:

$$loc_b = 0,5$$

- Die Wahrscheinlichkeit, daß die Suche im lokalen DN-Gebiet durchgeführt wird (für Benutzer im Heim-DN-Gebiet, Besuchs-DN-Gebiet und in einem DN-Gebiet außerhalb des Heimteilgebietes), beträgt:

$$p_{h_r}(i, i) = p_{h_v}(i, i) = p_{v_r/v}(i, i) = 0,7$$

Die sonstigen Anrufe sind gleichmäßig auf das UMTS-Teilgebiet verteilt.

In der obigen Grundkonfiguration, im folgenden auch die Grundkonfiguration_1 genannt, liegt das DN-Heimgebiet an der Grenze mehrerer DIR-Gebiete, was eventuell zu einem Sonderfall bezüglich der Ergebnisse führen könnte. Aus diesem Grunde wird eine zweite Möglichkeit untersucht, auch als Grundkonfiguration_2 bezeichnet, in der der Knoten 8 die Funktion des DN-Heimknotens übernimmt.

5.2.2 Präsentation der Ergebnisse

Die numerischen Ergebnisse der Grundkonfiguration_1 werden exemplarisch im folgenden präsentiert. Alle anderen Ergebnisse (einschließlich Grundkonfiguration_2) wurden in den Anhang B ausgelagert, um eine gute Übersichtlichkeit dieses Kapitels zu bewahren.

Alle Ergebnisse werden graphisch als Balkendiagramme präsentiert und zwar separat für Aktionen innerhalb des Heimteilgebietes sowie gegebenenfalls innerhalb des Besuchsteilgebietes und zwischen den Teilgebieten.

Die in diesem Kapitel enthaltenen Diagramme werden folgendermaßen aufgebaut:

- Die Knotenbelastung und Verbindungsbelastung werden getrennt dargestellt.
- Es werden immer vier unterschiedliche Szenarien verglichen (ST, R, V, RV).
- Die Knotenbelastung wird nach folgendem Schema dargestellt:
 - Die Knotenoperationen werden separat für jede Datenbankebene dargestellt (DN-Ebene, DIR1-Ebene, DIR2-Ebene,...). Dabei wird immer die Gesamtanzahl der Operationen pro Ebene angegeben.
 - Die Operationen werden nach Lesen (*read*), Schreiben (*write*) und sonstigen Operationen (*other*) unterschieden. Weil sowohl beim Lesen als auch beim Schreiben die Anzahl der Operationen von der Wahl der Abfragetechnik un-

abhängig ist, wird lediglich bei den sonstigen Operationen nach der Abfragetechnik unterschieden.

- Die Leitungsbelastung wird nach folgendem Schema dargestellt:
 - Es wird sowohl die gesamte Datenmenge als auch die gewichtete Datenmenge dargestellt.
 - Es wird nach der Abfragetechnik unterschieden.

Die im Anhang enthaltenen Ergebnisse werden in einer reduzierten Form wiedergegeben, um den Gesamtumfang der Darstellung sinnvoll zu begrenzen. Dabei wurden folgende Vereinfachungen vorgenommen:

- Es werden lediglich Ergebnisse für die *chaining/referral*-Abfragetechnik präsentiert, da die Relationen zwischen den einzelnen Techniken prinzipiell gleich sind und eine vollständige Wiedergabe aller Techniken keine zusätzliche Erkenntnis mit sich bringt.
- Es wird lediglich die gewichtete Leitungsbelastung dargestellt, da die Tendenzen für beide Belastungsarten (Gesamt- und gewichtete Belastung) gleich sind.
- Die zusammengehörenden Ergebnisse für die Knoten- und die Leitungsbelastung werden platzsparend im gleichen Diagramm untergebracht.

Wenn bei der Auswertung von V- und RV-Szenarien beide Optionen der Zeigerverwaltung untersucht werden (direkte und indirekte Zeiger), so werden sie in einem Diagramm als überlagerte Balken dargestellt (vgl. nächste Seite). Dabei sind die verwendeten Symbole wie in Abbildung 5.3 zu verstehen.

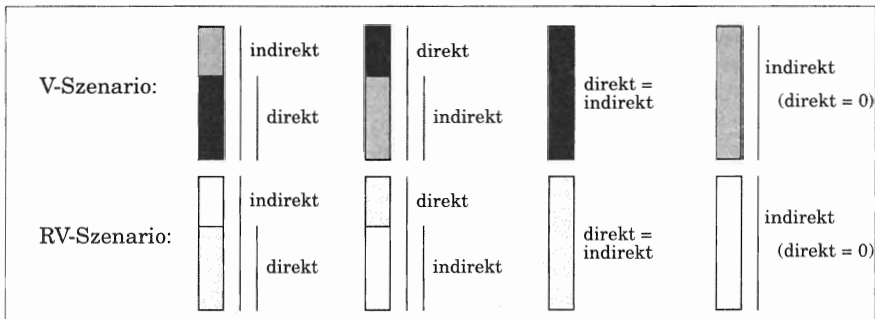
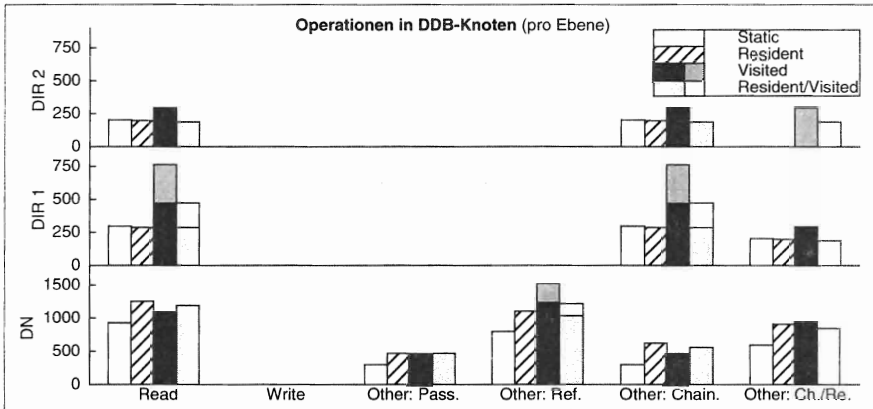


Abb. 5.3 Verwendete Symbole für V- und RV-Szenarien (Höhe des Balkens entspricht der Belastung).

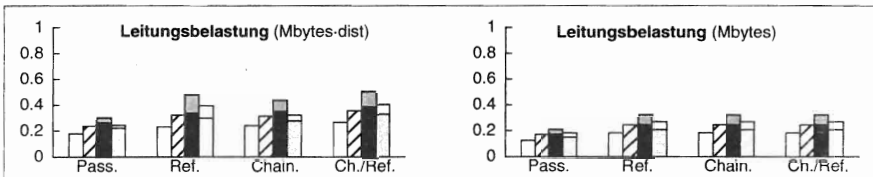
5.2.3 Suchanfragen in der Grundkonfiguration

Im folgenden sind die numerischen Ergebnisse für die Grundkonfiguration_I zusammengestellt und werden eingehend diskutiert. Anhand dieser Ergebnisse ist es dem Leser möglich, sich ein Bild über die Art der Ergebnisse zu verschaffen. Die Ergebnisse für die Kommunikation innerhalb des Besuchsteilgebietes wurden nicht wie-

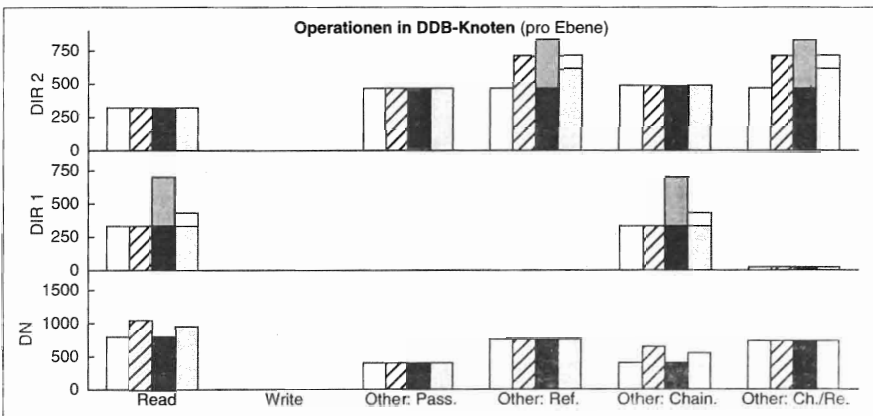
dergegeben, da in diesem Fall sowohl die Knoten- als auch die Leitungsbelastung einen sehr geringen Beitrag zur Gesamtbelastung liefern.



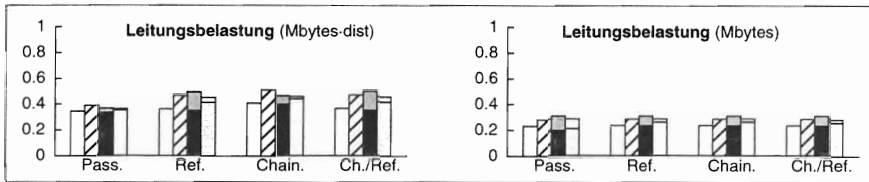
Diag. 5.1 Grundkonfiguration_1 - Suchanfrage innerhalb des Heimteilgebietes (Knotenoperationen)



Diag. 5.2 Grundkonfiguration_1 - Suchanfrage innerhalb des Heimteilgebietes (Leitungsbelastung)



Diag. 5.3 Grundkonfiguration_1 - Suchanfrage zwischen den Teilgebieten (Knotenoperationen)



Diag. 5.4 Grundkonfiguration_1 - Suchanfrage zwischen den Teilgebieten (Leitungsbelastung)

Die Ergebnisse der Suchanfrage in der Grundkonfiguration_1 lassen folgende Schlussfolgerungen zu:

- Die Anzahl der Zugriffe auf den Speicher, um die Benutzerinformation zu lesen (*read*), ist bei allen Abfragetechniken identisch - dieser Sachverhalt wird bei jeder Alternative zu beobachten sein. Die Unterschiede zwischen den einzelnen internen Techniken sind ausschließlich unter den sonstigen Operationen (*other*) zu finden. Unter Beachtung dieser Tatsache ist im Vergleich der einzelnen Abfragetechniken die *passing*-Technik mit dem geringsten sonstigen Aufwand und somit auch mit dem geringsten Gesamtaufwand verbunden. Es soll jedoch an dieser Stelle angemerkt werden, daß die *passing*-Technik am empfindlichsten auf Informationsflußstörungen reagiert, weil wegen der mangelnden Überwachung des aktuellen Flusses die etwaigen Fehler nur mit einem großen Aufwand zu lokalisieren sind. Die übrigen Abfragetechniken unterscheiden sich grundsätzlich in der Verteilung der Last auf die einzelnen Knoten. Dabei erweist sich die vierte Option, die *chaining/referral*-Technik, als eine gute Kompromißlösung, die im Vergleich mit anderen Techniken die gesamte Last gleichmäßiger auf Daten- und Verzeichnisknoten aufteilt.
- Der Anteil des Datenverkehrs innerhalb des Besuchsteilgebietes ist in der betrachteten Konfiguration nicht relevant, was sich grundsätzlich aus den gewählten Parametern ergibt. Speziell ist die Annahme über die Gesamtgröße des UMTS-Gebietes von Bedeutung - es besteht hier aus 25 Teilgebieten (siehe die Formel für den 3. Fall im Kapitel 4.4.1.2, die die Suchanfrage innerhalb eines Besuchsteilgebietes bestimmt). Aus der formalen Beschreibung ist offensichtlich, daß der Anteil innerhalb des Besuchsteilgebietes in allen ausgewogenen Situationen niedrig sein wird. Aus diesem Grund wird auch in den anderen Alternativen dieser Fall nicht weiter untersucht.
- Aus dem Vergleich der unterschiedlichen Datenverteilungsszenarien ergibt sich die Folgerung, daß bei den gegebenen Parametern sich das ST-Szenario als die beste Alternative präsentiert. Die zweitbeste Alternative ist das R-Szenario, gefolgt von den RV- und V-Szenarien.
- Die obigen Ergebnisse gelten unter dem Vorbehalt der gewählten Zeigeroption, d.h., die ST- und R-Szenarien verwenden die direkten Zeiger und die V- und RV-Szenarien die indirekten Zeiger. Die letzteren führen zu einer höheren lesebedingten Knotenlast in den mittleren Verzeichnisebenen. Ein Vergleich mit den V- und RV-Szenarien unter Verwendung der direkten Zeiger zeigt dies deutlich -

hier ist das RV-Szenario durchaus mit dem R-Szenario vergleichbar. Das V-Szenario bleibt weiterhin ungünstig wegen der hohen Anzahl der Operationen in den Verzeichnisknoten.

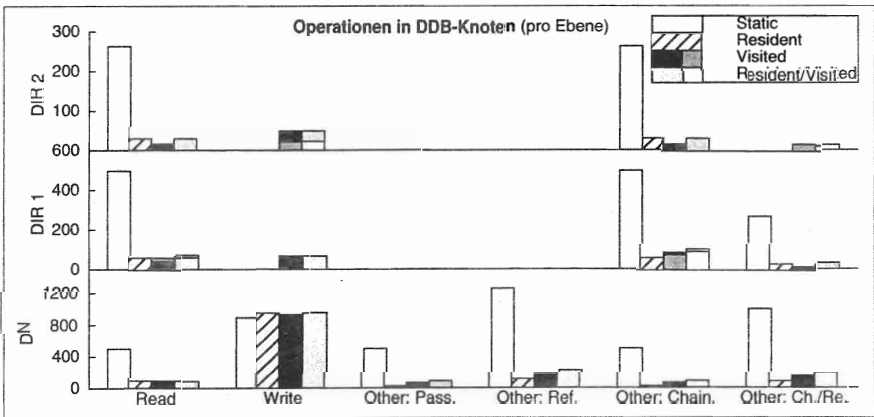
Für die Grundkonfiguration_2 (vgl. Anhang B, Diagramme B.1 und B.2) lassen sich nachstehende Folgerungen im Vergleich zur Grundkonfiguration_1 herleiten:

- Die Belastung der DN-Knoten ist naturgemäß unverändert, da die Benutzerverteilung gleich geblieben ist.
- Die Belastung der DIR-Knoten ist geringer geworden, was sich auf die Tatsache zurückführen läßt, daß die Benutzerbewegungen sich jetzt hauptsächlich innerhalb einzelner DIR-Gebiete abspielen, d.h. innerhalb solcher DN-Gebiete, die von einem gemeinsamen DIR-Knoten der ersten Verzeichnisebene überwacht werden.
- Ähnliches wie oben gilt für die Verbindungsbelastung.
- Die Folgerungen bezüglich der bevorzugten Abfragetechniken und des besten Szenarios gelten unverändert.

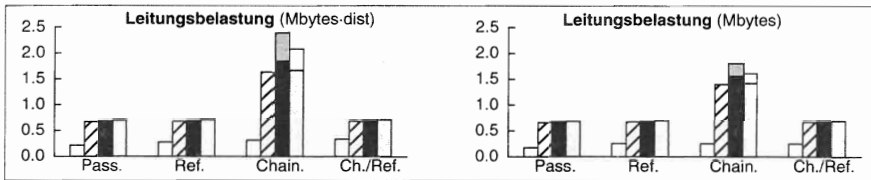
Wegen der prinzipiellen Äquivalenz der beiden Grundkonfigurationen wird einfachheitshalber die erste als Ausgangspunkt für die folgenden Untersuchungen aller Alternativen der Suchanfrage genommen.

5.2.4 Aktualisierungsanfragen in der Grundkonfiguration

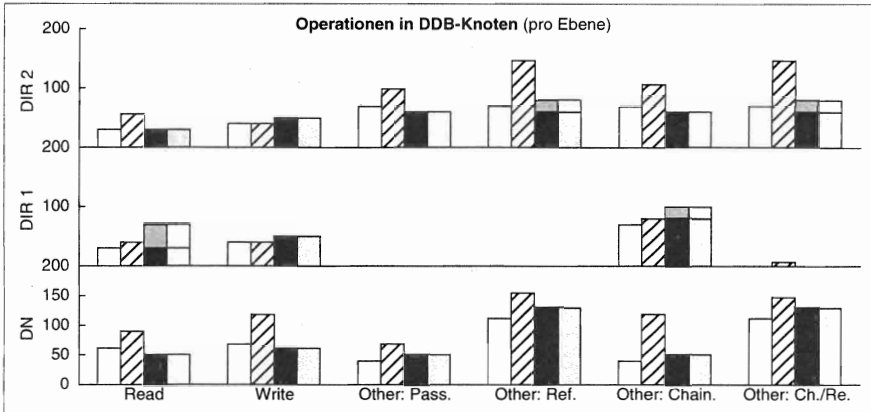
Im folgenden sind die numerischen Ergebnisse für die Grundkonfiguration_1 zusammengestellt und werden eingehend diskutiert. Die Ergebnisse für die Kommunikation innerhalb des Besuchsteilgebietes wurden nicht wiedergegeben, da sowohl die Knoten- als auch die Leitungsbelastung (mit Ausnahme des ST-Szenarios) einen vernachlässigbaren Beitrag zur Gesamtbelastung liefern.



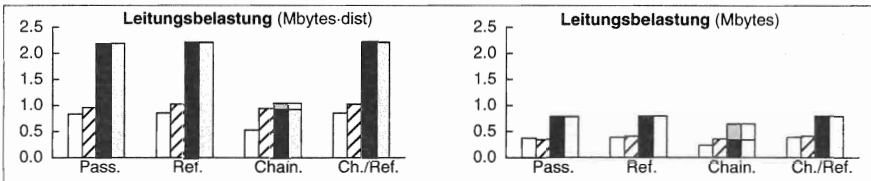
Diag. 5.5 Grundkonfiguration_1 - Aktualisierung innerhalb des Heimteilgebietes (Knotenoperationen)



Diag. 5.6 Grundkonfiguration_1 - Aktualisierung innerhalb des Heimteilgebietes (Leitungsbelastung)



Diag. 5.7 Grundkonfiguration_1 - Aktualisierung zwischen Teilgebieten (Knotenoperationen)



Diag. 5.8 Grundkonfiguration_1 - Aktualisierung zwischen Teilgebieten (Leitungsbelastung)

Die Ergebnisse der Aktualisierungsanfrage in der Grundkonfiguration_1 lassen folgende Schlußfolgerungen zu:

- Die Anzahl der Zugriffe auf den Speicher, um die Benutzerinformation zu lesen oder zu schreiben (*read* oder *write*), ist bei allen Techniken identisch. Dieser Sachverhalt wird in jeder Alternative zu beobachten sein, ähnlich wie dies für die Suchanfrage galt. Die Unterschiede zwischen den einzelnen internen Abfragetechniken sind ausschließlich unter sonstigen Operationen (*other*) zu finden. Die Folgerungen bezüglich der Verwendbarkeit der jeweiligen Abfragetechnik gelten äquivalent zu denen, die für die Suchanfrage dargelegt wurden.

- Der Anteil der Netzbelastung, der sich aufgrund der Bewegungen innerhalb des Besuchsgebietes ergibt, ist, mit Ausnahme des ST-Szenarios, in der betrachteten Konfiguration nicht relevant. Der Grund für den unproportional hohen Beitrag im ST-Szenario liegt darin, daß in allen anderen Szenarien die Aktualisierung meistens im lokalen Besuchsknoten durchgeführt wird. Im ST-Szenario sind dagegen alle Änderungen im abgelegenen Heimknoten durchzuführen. Der Heimknoten ist im Besuchsteilgebiet deswegen fast immer abgelegenen, weil er quasi zufällig deklariert wird, z.B. dort, wo der Benutzer das neue Teilgebiet betreten hat. Deswegen wird das ST-Szenario immer sehr ungünstig sein.
- Aus dem Vergleich der einzelnen Szenarien für die Benutzerbewegungen innerhalb des Heimteilgebietes läßt sich die Schlußfolgerung herleiten, daß das ST-Szenario am schlechtesten beim Lesen und bei sonstigen Operationen, am besten aber beim Schreiben abschneidet. Die Schreibvorteile werden jedoch bei weitem durch die Nachteile beim Lesen und bei sonstigen Operationen überwogen. Nur ein extremer Schreibaufwand würde die Lesenachteile aufwiegen können. Aufgrund dieses Sachverhaltes ist das R-Szenario bezüglich der Knotenbelastung am besten, dicht gefolgt vom V-Szenario. Bezüglich der Leitungsbelastung ist das ST-Szenario am besten.
- Obwohl bei den V- und RV-Szenarien eine indirekte Zeigerverwaltung angewendet wurde, die gegenüber der direkten Verwaltung beim Aktualisieren vorteilhaft ist, sind keine offensichtlichen Vorzüge von V- und RV- gegenüber dem R-Szenario sichtbar. Dieser Sachverhalt resultiert aus der Tatsache, daß in den statischen Szenarien (ST und R) keine Aktualisierung des DIR-Baumes notwendig ist, solange sich die Benutzer innerhalb des Teilgebietes befinden.
- Die Anwendung der direkten Zeiger in den dynamischen Szenarien führt zu einer leichten Senkung der Belastung durch die Leseoperationen und zu einer Steigerung der Belastung bei den Schreiboperationen. Die Unterschiede sind jedoch nicht so gravierend, daß eine eindeutige Präferenz festzustellen wäre. Es können sowohl die direkten als auch die indirekten Zeiger angewendet werden.

5.3 Auswertung der Suchanfragen

Im vorliegenden Kapitel werden die Ergebnisse der Auswertung unterschiedlicher Alternativen dargestellt. Um dabei den Überblick zu bewahren und den Einfluß der einzelnen Alternativen auf die Änderung der Ergebnisse verfolgen zu können, werden die jeweiligen Alternativen immer in bezug zur 1. Grundkonfiguration gebracht. Eine Zusammenfassung der Schlußfolgerungen wird in jedem Unterkapitel gegeben. Alle Diagramme sind im Anhang B zu finden.

5.3.1 Einfluß der Verteilungsparameter

5.3.1.1 Untersuchte Alternativen

Gemäß dem hergeleiteten Modell ist das Anrufverhalten nicht von der Benutzerbeweglichkeit abhängig. Als einziger Mobilitätsparameter bleibt somit die Verteilungs-

wahrscheinlichkeit der Benutzer. In der Grundkonfiguration wird eine **mittlere** Konzentration der Benutzer um den Wohnort betrachtet. Es werden folgende Varianten der Benutzerverteilung untersucht:

- Alle Benutzer sind um ihr Heimgebiet verteilt, d.h., sie bewegen sich um ihr Heimgebiet herum:
 1. Variante: **starke** Benutzerkonzentration um den Heimknoten (Diagramme B.5 und B.6):
 $P(\text{Heimgebiet}) = 0,75$
 $P(\text{Nachbargebiet}) = 0,10$
 $P(\text{externes Gebiet}) = 0,05$
 2. Variante: **schwache** Benutzerkonzentration (starke Benutzerverteilung) um das Heimgebiet (Diagramme B.7 und B.8):
 $P(\text{Heimgebiet}) = 0,10$
 $P(\text{Nachbargebiet}) = 0,10$
 $P(\text{externes Gebiet}) = 0,70$
- Benutzer sind um ein Besuchsgebiet (hier beispielsweise das Besuchsgebiet 23) statt um ihr Heimgebiet (DN-Gebiet 15) verteilt. Diese Art der Verteilung wird im weiteren die verschobene Verteilung (Konzentration) genannt. Sie beschreibt eine Situation, in der viele Bewohner zeitweise ihr Heimgebiet verlassen, z.B. weil die Arbeitsstätte entfernt liegt oder eine große Kultur- oder Sportveranstaltung außerhalb des Wohngebietes organisiert wird; alle anderen Benutzer des UMTS-Teilgebietes verhalten sich wie in der Grundkonfiguration:
 3. Variante: **mittlere** Benutzerkonzentration um einen Besuchsknoten (Diagramme B.9 und B.10):
 $P(\text{Heimgebiet}) = 0,40$
 $P(\text{Nachbargebiet}) = 0,45$
 $P(\text{externes Gebiet}) = 0,05$
 4. Variante: **starke** Benutzerkonzentration um einen Besuchsknoten (Diagramme B.11 und B.12):
 $P(\text{Heimgebiet}) = 0,75$
 $P(\text{Nachbargebiet}) = 0,10$
 $P(\text{externes Gebiet}) = 0,05$
 5. Variante: **schwache** Benutzerkonzentration wird nicht mehr präsentiert, da gemäß der Erwartungen die erhaltenen Ergebnisse mit der Variante 2 prinzipiell übereinstimmen.

5.3.1.2 Ergebnisse der Auswertung

Nachstehende Folgerungen lassen sich aus den Ergebnissen herleiten:

- Aus dem Vergleich der Ergebnisse für die Varianten 1 und 2 ist ersichtlich, daß die steigende Konzentration um den Heimknoten wie in der 1. Variante zu einer Reduktion des Datenverkehrs und damit zu einer geringeren Belastung in allen Knoten und Verbindungen führt. Umgekehrt führt die zunehmende Benutzer-

verteilung auf das gesamte Teilgebiet wie in der 2. Variante zu einer Steigerung der Belastung.

- Die aus der Grundkonfiguration bekannten Folgerungen bezüglich der Abfrage-techniken und der jeweiligen Szenarien bleiben im untersuchten Parameterbereich erhalten. Es ist jedoch aus den Ergebnissen im Diagramm B.5 ersichtlich, daß bei einer gestiegenen Benutzerkonzentration die Vorteile des ST-Szenario gegenüber dem R-Szenario sinken.
- Der Einfluß des Verteilungsparameters auf verschiedene Szenarien ist unterschiedlich stark ausgeprägt. Die relativen Änderungen der Netzbelastung sind unter Verwendung des V-Szenarios am stärksten. Das macht das V-Szenario zu einer schwierig vorausberechenbaren Alternative.
- Die Ergebnisse in den Diagrammen B.5, B.6, B.7 und B.8 zeigen Änderungen, die sich aus der Anwendung der direkten Zeiger in den V- und RV-Szenarien ergeben. Daraus ist erkennbar, daß die ursprünglich starke Belastung der Knoten auf der ersten Verzeichnisebene reduziert wurde. Im Falle des RV-Szenarios ist die Reduktion auf das Niveau des R-Szenario erreicht worden. Dies gilt sowohl innerhalb als auch zwischen den Teilgebieten. Ebenfalls wurde die Leitungshlastung entsprechend reduziert.

Aufgrund der Tatsache, daß das RV-Szenario eine geringe Lesebelastung des Wurzelknotens verursacht, stellt dieses Szenario aus Sicht der Belastung der DIR-Knoten eine optionale Ausweichmöglichkeit zum R-Szenario dar.

Der direkte Vergleich mit dem ST-Szenario hebt die unterschiedliche Charakteristika hervor: ST ist besser bei den Datenknoten und RV besser bei den Verzeichnisknoten (besonders beim Lesen). Die Ausprägung dieser Unterschiede verstärkt sich mit zunehmender Verteilung der Benutzer.

- Die Vorteile der V- und RV-Szenarien werden erst in den Varianten 3 und 4 sichtbar. Bei der Verteilung der Benutzer um einen Besuchsknoten erscheint die Suche nach dem Besuchsknoten durchaus sinnvoll. Obwohl hier die indirekten DIR-Zeiger in den V- und RV-Szenarien verwendet werden, sind die Ergebnisse dieser Szenarien gegenüber den statischen Szenarien konkurrenzfähig. Die Verwendung der direkten Zeiger führt zu einer nochmaligen Verbesserung der Qualität der V- und RV-Szenarien, wie bereits in den Varianten 1 und 2 exemplarisch gezeigt worden ist.

Das RV-Szenario ist das beste, V ist das zweitbeste und R ist das drittbeste Szenario.

Dabei ist zu beachten, daß hier ausschließlich die Heimbenutzer des Knotens 15 eine geänderte Verteilung aufweisen. Soll auch für andere Benutzer eine verschobene Verteilung gelten, so wird der beschriebene Effekt noch deutlicher. Es ist auffällig, daß die Vorteile der dynamischen Szenarien in der Variante 4 stärker ausgeprägt sind als in der Variante 3. Dieser Sachverhalt untermauert nochmals die These, daß eine verschobene Verteilung, die natürlich bei der Variante 4 stärker zur Geltung kommt, die dynamischen Szenarien vorteilhafter erscheinen läßt.

5.3.2 Einfluß der Anrufparameter

5.3.2.1 Untersuchte Alternativen

Die Anrufparameter bestimmen die Verteilung der Anrufe im gesamten UMTS-Gebiet. Es wurden insgesamt 6 Parameter im Anrufmodell eingeführt:

p_{cHD}, p_{cVD}	bestimmen das Verhalten zwischen UMTS-Teilgebieten
loc_b, p_{h_r}, p_{h_v}	beschreiben das Verhalten innerhalb eines Heimteilgebietes
$p_{v_r/v}$	beschreibt das Verhalten innerhalb des Besuchsteilgebietes

Es werden folgende Varianten untersucht:

- unterschiedliches Verhalten zwischen den Teilgebieten:
 1. starke Verteilung der Anrufe auf andere Teilgebiete (Diagramme B.13 und B.14):

$$p_{cHD} = p_{cVD} = 0,3$$

$$loc_b = 0,5$$

$$p_{h_r}(i, i) = p_{h_v}(i, i) = p_{v_r/v}(i, i) = 0,7$$
 2. starke Konzentration der Anrufe auf das eigene Teilgebiet (Diagramme B.15 und B.16):

$$p_{cHD} = p_{cVD} = 0,9$$

$$loc_b = 0,5$$

$$p_{h_r}(i, i) = p_{h_v}(i, i) = p_{v_r/v}(i, i) = 0,7$$
- unterschiedliches Verhalten innerhalb eines Teilgebietes, speziell innerhalb des Heimteilgebietes:
 3. starke Konzentration der Anrufe auf das eigene DN-Heimgebiet (Diagramme B.17 und B.18):

$$p_{cHD} = p_{cVD} = 0,7$$

$$loc_b = 0,1$$

$$p_{h_r}(i, i) = p_{h_v}(i, i) = p_{v_r/v}(i, i) = 0,7$$
 4. starke Konzentration der Anrufe auf das aktuell besuchte DN-Gebiet (Diagramme B.19 und B.20):

$$p_{cHD} = p_{cVD} = 0,7$$

$$loc_b = 0,9$$

$$p_{h_r}(i, i) = 0,9, p_{h_v}(i, i) = 0,9, p_{v_r/v}(i, i) = 0,7$$
 5. starke Verteilung der Anrufe auf das gesamte Teilgebiet (Diagramme B.21 und B.22):

$$p_{cHD} = p_{cVD} = 0,7$$

$$loc_b = 0,9$$

$$p_{h_r}(i, i) = 0,1, p_{h_v}(i, i) = 0,1, p_{v_r/v}(i, i) = 0,7$$

5.3.2.2 Ergebnisse der Auswertung

Nachstehende Folgerungen lassen sich aus den erhaltenen Ergebnissen herleiten:

- Die 1. und 2. Variante unterscheiden sich von der Grundkonfiguration im wesentlichen durch die Netzbelastungsanteile, die durch die Kommunikation innerhalb des Heimteilgebietes und zwischen den Teilgebieten entstanden sind. Die Verhältnisse sowohl innerhalb des Heimteilgebietes als auch zwischen den Teilgebieten entsprechen den bereits bekannten aus der Grundkonfiguration, d.h., das ST-Szenario und die *passing*-Technik sind die bevorzugten Lösungen. Weiter ist zu beobachten, daß die Zunahme der Knotenbelastung, die durch die gestiegene Zwischengebietkommunikation entstanden ist, nicht durch die Abnahme der internen Kommunikation aufgewogen wird. Die Gesamtbelastung der Knoten nimmt unabhängig von dem angewendeten Szenario und der Abfragetechnik stark zu. Das gleiche gilt offensichtlich auch für die Gesamtbelastung der Verbindungen.
- Die 3. und 4. Variante untersuchen die Fälle, in denen die Anrufverteilung der Benutzer während des Besuchsknotenaufenthaltes stärker bzw. schwächer der Anrufverteilung ähnelt, die die Benutzer aufweisen, wenn sie sich im Heimknotengebiet aufhalten. Hier ist zu beobachten, daß eine Zunahme des heimähnlichen Verhaltens, wie in der 3. Variante, zu einem allgemeinen Belastungsanstieg führt. Dieser Anstieg ist bei den statischen Szenarien stärker ausgeprägt. Aus diesem Grund werden die Vorteile der ST- und R-Szenarien geringer, wobei die bereits aus der Grundkonfiguration bekannten Verhältnisse ihre Gültigkeit weiterhin beibehalten. Daraus folgt, daß in einer Situation wie in der 3. Variante die dynamischen Szenarien, besonders das RV-Szenario, durchaus als sinnvolle Möglichkeit zu bewerten sind.
- Die 5. Variante, in der die Anrufe stark auf das gesamte Teilgebiet verteilt sind, zeigt in der untersuchten Form einige Unterschiede zu den aus der Grundkonfiguration bekannten Folgerungen. Ein Anstieg der gesamten Netzbelastung ist zu beobachten, und die Proportionen der einzelnen Szenarien zueinander ändern sich wesentlich. Die Belastungen der Datenknoten und/oder des Wurzelknotens in den dynamischen Szenarien ist jetzt denen in den statischen Szenarien ähnlich. Daraus folgt wiederum, daß in einer Situation wie in der Variante 5 die dynamischen Szenarien eine sinnvolle Möglichkeit bieten, besonders wenn die direkten Zeiger in den V- und RV-Szenarien angewendet werden.

5.3.3 Einfluß der DIR-Architektur

5.3.3.1 Untersuchte Alternativen

Die bisherigen Auswertungen wurden für eine symmetrische vollständige Grundarchitektur (36-4-1) durchgeführt. Hier werden folgende Varianten der Architektur untersucht:

- symmetrische vollständige Architektur mit unterschiedlicher Baumhöhe:
 1. Variante: hohe Verzeichnishierarchie, d.h. 36-9-3-1 (Diagramme B.23 und B.24)
 2. Variante: flache Verzeichnishierarchie, d.h. 36-1 (Diagramme B.25 und B.26)
- symmetrische unvollständige Architektur mit unterschiedlicher Struktur des Baumes und zwei Optionen der Zeigerverwaltung auf der Wurzelebene:
 - a) keine Suche auf der Wurzelebene - die Zeiger werden bereits bei der Aktualisierungsanfrage mit neuen Inhalten versehen,
 - b) Suche auf der Wurzelebene unter Verwendung der *referral*-Abfragetechnik;
- 3. Variante: mittlere Verzeichnishierarchie, d.h. 36-4 :
 - a) Diagramme B.27 und B.28
 - b) Diagramme B.29 und B.30
- 4. Variante: hohe Verzeichnishierarchie, d.h. 36-9-3 :
 - a) Diagramme B.31 und B.32
 - b) Diagramme B.33 und B.34
- 5. hohe Verzeichnishierarchie, d.h. 36-4-2 :
 - a) Diagramme B.35 und B.36
 - b) Diagramme B.37 und B.38
- unsymmetrische Architektur, die dem Verteilungsmuster der Benutzer angepaßt ist, d.h., der Heimknoten und seine Nachbarn haben einen gemeinsamen Verzeichnisknoten:
 6. Variante: Benutzer stark konzentriert um den Heimknoten (Diagramme B.39 und B.40)
 7. Variante: Benutzer schwach konzentriert um den Heimknoten (Diagramme B.41 und B.42)
 8. Variante: Anrufe verteilt auf die lokale Umgebung (Diagramme B.43 und B.44)
 9. Variante: Anrufe stark verteilt im Teilgebiet (Diagramme B.45 und B.46)

5.3.3.2 Ergebnisse der Auswertung

Nachstehende Folgerungen lassen sich aus den erhaltenen Ergebnissen herleiten:

- Die 1. Variante untersucht die Anwendung eines höheren DIR-Baumes, dessen symmetrische Struktur ohne Rücksicht auf das Benutzerverhalten festgelegt worden ist.
 Aus den Zahlen ist offensichtlich, daß die Suchanfragen jetzt eine höhere Gesamtknoten- und Verbindungsbelastung als in der Grundkonfiguration verursachen. Dies gilt sowohl innerhalb des Heimteilgebietes als auch zwischen den Teilgebieten.
 Die erwartete Reduktion der Wurzelknotenbelastung bleibt innerhalb des Heimteilgebietes fast vollständig aus, was sich auf die gewählte Struktur des

Baumes im Verhältnis zur Verteilung der Benutzer und deren Anrufverhalten zurückführen läßt. Es wird hier keine Rücksicht darauf genommen, daß der Hauptknoten 15, in dem die betrachteten Benutzer zu Hause sind, an der Grenze der DIR-Gebiete aller Ebenen liegt. Diese Lage führt jedoch zur Notwendigkeit, die Suche sehr oft bis in den Wurzelknoten fortzuführen. Darüber hinaus ist die Anrufverteilung derartig strukturiert, daß die Anzahl der Suchprozesse, die über den Wurzelknoten gehen, hier und im Grundscenario ähnlich sind.

Die Gesamtbelastung wird durch die Baumerhöhung immer ansteigen.

Die durch die Suche zwischen den Teilgebieten verursachte Belastung ist eindeutig höher als in der Grundkonfiguration. Hier ist keine Verbesserung zu erwarten, auch nicht in den einzelnen Knoten.

- Die 2. Variante stellt die Ergebnisse der Untersuchung einer möglichst flachen DIR-Struktur dar, nämlich mit nur einem Verzeichnisknoten. Dieser einzige Verzeichnisknoten ist gleichzeitig der Wurzelknoten des Baumes. Aus den erhaltenen Zahlen folgt, daß die Gesamtbelastung sowohl innerhalb des Heimteilgebietes als auch zwischen den Teilgebieten abnimmt. Ein Nachteil dieser Lösung ist die starke Zunahme der Belastung des Wurzelknotens aufgrund der Suche innerhalb des Heimgebietes. Das wiederum stellt hohe Anforderungen an die Leistungsfähigkeit und Zuverlässigkeit dieses einzelnen Knotens. Für den Fall der Suchanfrage zwischen den Teilgebieten ist keine erhöhte Belastung der einzelnen Knoten zu erwarten. Hier ist eine flache DIR-Struktur eindeutig von Vorteil.
- Die Varianten 3 - 5 betrachten die sog. unvollständigen DIR-Bäume, d.h. Bäume ohne einen gemeinsamen Wurzelknoten. Dabei wurden von den im Kapitel 3.4.2.3 beschriebenen Alternativen nur zwei numerisch untersucht:
 - a) Der Baum wird nach der Benutzerbewegung komplett aktualisiert: Während der Suchanfrage kann der zuständige Wurzelknoten direkt angesprochen werden.
 - b) Der Baum ist nach der Benutzerbewegung nur bis zum eigenen Wurzelknoten aktualisiert: Während der Suchanfrage muß der verantwortliche Wurzelknoten gesucht werden (hier mit der *referral*-Technik).

Die Varianten 3a - 5a sind bei der Suchanfrage immer dem vollständigen Baum mindestens äquivalent:

- innerhalb des Heimgebietes sind sie vorteilhafter, und
- zwischen den Teilgebieten sind sie bei V- und RV-Szenarien vorteilhafter und bei ST- und R- Szenarien äquivalent.

Der eventuelle Mehraufwand der Variante a wird erst beim Aktualisieren des Baumes nach einer Benutzerbewegung sichtbar (vgl. Kapitel 5.4.2).

Die Varianten 3b - 5b sind differenzierter zu betrachten:

- Für die Variante 3b, die mit der Grundkonfiguration verglichen werden muß, sind folgende Beobachtungen zu machen:

- Die Gesamtbelastungen innerhalb des Heimteilgebietes sind jeweils für das ST- und R-Szenario mit dem aus der Grundkonfiguration identisch; dies folgt aus der Anwendung der direkten Zeiger und aus der Existenz von vier Wurzelknoten in dieser Variante. Der Aufwand für die V- und RV-Szenarien ist geringer, da hier indirekte Zeiger verwendet werden.
- Zwischen den Teilgebieten gilt, daß die Gesamtbelastung nur im Vergleich mit dem ursprünglichen V-Szenario reduziert worden ist.
- Für die Variante 4b, die mit der 1. Variante verglichen werden muß, gelten gleiche Folgerungen wie für die Variante 3b.
- Für die Variante 5b, die ebenfalls mit der 1. Variante zu vergleichen ist, gilt, daß sie bezüglich der Knoten- und Verbindungsbelastung gegenüber der 1. Variante immer vorteilhafter ist.

Die hohe Belastung der Wurzelknoten während der Suchanfrage zwischen Teilgebieten ergibt sich aus der Tatsache, daß für die von außerhalb ankommenden Abfragen der zuständige Wurzelknoten erst gefunden werden muß.

- In den Varianten 3a - 5a ist dann maximal eine zusätzliche Nachfrage notwendig; der Mehraufwand gegenüber der Ein-Baum-Struktur beträgt:

$$f = \frac{n_{\text{Wurzel}} - 1}{n_{\text{Wurzel}}} \cdot A_T, \quad (5.1)$$

gleichmäßige Benutzerverteilung im Teilgebiet vorausgesetzt.

- In den Varianten 3b - 5b sind ggf. mehrere zusätzliche Nachfragen notwendig; der Mehraufwand gegenüber der Ein-Baum-Struktur beträgt:

$$f = \sum_{i=1}^{n_{\text{Wurzel}}-1} i/n_{\text{Wurzel}} \cdot A_T, \quad (5.2)$$

ebenfalls gleichmäßige Benutzerverteilung im Teilgebiet vorausgesetzt.

In beiden Fällen steht n_{Wurzel} für die Anzahl der Wurzelknoten im betrachteten Teilgebiet und A_T für die Gesamtanzahl der Suchanfragen zwischen den Teilgebieten.

Die Ergebnisse in den Diagrammen ziehen obige Überlegungen in Betracht.

- Die Varianten 6 - 9 betrachten eine DIR-Struktur, die der Verteilung der Benutzer angepaßt ist.

Als Folgerung läßt sich hier festhalten, daß die Vorteile einer benutzerangepaßten DIR-Struktur bescheiden ausfallen. Der Grund dafür ist offensichtlich: Die DIR-Struktur ist den Benutzerbewegungen angepaßt worden und zieht nur unzureichend das Anrufverhalten in Betracht. Die Vorteile einer angepaßten DIR-Struktur werden erst bei der Untersuchung der Aktualisierungsanfragen sichtbar (vgl. Kapitel 5.4.2). Als wichtige Konsequenz läßt sich jedoch festhalten, daß keine Verschlechterung der Leistungsfähigkeit zu verzeichnen ist.

5.3.4 Einfluß der Gebietstopologie

5.3.4.1 Untersuchte Alternativen

In den bisherigen Untersuchungen wurde stets die Teilgebietstopologie wie in der Grundkonfiguration vorausgesetzt. Nun werden folgende Varianten untersucht:

- Die Größe des Teilgebietes wird geändert:
 1. Variante: ein kleines Gebiet mit 16 DN-Gebieten und einer mittleren Verzeichnisstruktur, d.h. 16-4-1 (Diagramme B.47 und B.48)
 2. Variante: ein großes Gebiet mit 64 DN-Gebieten und einer hohen Verzeichnisstruktur, d.h. 64-16-8-4-2-1 (Diagramme B.49 und B.50)
- Die Form des Teilgebietes wird geändert:
 3. Variante: ein längliches Gebiet mit 10 mal 3 DN-Gebieten und einer mittleren Verzeichnisstruktur, d.h. 30-5-1 (Diagramme B.51 und B.52)

5.3.4.2 Ergebnisse der Auswertung

Nachstehende Folgerungen lassen sich aus den erhaltenen Ergebnissen herleiten:

- Die 1. Variante, die ein kleines Teilgebiet unter Beibehaltung aller sonstigen Charakteristika der Grundkonfiguration untersucht, erlaubt, gleiche Schlüsse bezüglich der Szenarien und Techniken wie in der Grundkonfiguration zu ziehen. Dies kann man verallgemeinern: Die Auswertung verschiedener Datenverteilungs- und Strukturalternativen in einem kleineren Gebiet führt zu den gleichen Konsequenzen wie in den bereits untersuchten Fällen in einem größeren Gebiet.
- Die 2. Variante untersucht ein großes Teilgebiet, wobei im Unterschied zur Grundkonfiguration die direkten Zeiger in allen Szenarien und eine sehr hohe Verzeichnisstruktur (5 Ebenen) angewendet worden sind.
Es wird hier deutlich, daß bei der Verwendung eines hohen Baumes und der direkten Zeiger das RV-Szenario eine durchaus interessante Möglichkeit neben dem R-Szenario bietet (als zweitbestes nach dem ST-Szenario).
- Die 3. Variante untersucht ein Teilgebiet, das eine längliche Form statt einer quadratischen besitzt. Die aus der Grundkonfiguration bekannten Folgerungen bezüglich der Szenarien und Techniken lassen sich hier ebenfalls herleiten. Die Anwendung der direkten Zeiger im RV-Szenario erweist sich genauso vorteilhaft und das sogar in stärkerem Maße als in der Grundkonfiguration, was sich auf eine stärkere Streckung des Gebietes zurückführen läßt.

Aus dem Vergleich der Ergebnisse für Teilgebiete mit drei unterschiedlichen Größen und variabler Form folgt, daß Größe und Form des Teilgebietes keinen wesentlichen Einfluß auf die Vorteilhaftigkeit der Techniken, Szenarien und der Struktur des Baumes haben. Wesentlich sind lediglich die Benutzer- und Anrufverteilungsmuster, wie dies in den untersuchten Varianten mehrmals gezeigt worden ist.

5.3.5 Verzögerungszeit bei Suchanfragen

Die Verzögerungszeit, die beim Suchen in unterschiedlichen Szenarien und in unterschiedlichen Abfragetechniken zustande kommt, wird lediglich analytisch erfaßt. Danach werden an einigen Beispielen die numerischen Ergebnisse präsentiert. Wie es sich zeigen wird, weichen die absoluten Werte der einzelnen Varianten nicht sehr stark voneinander ab. Das ist der Grund, warum von einer generellen Wiedergabe der Verzögerungszeiten abgesehen worden ist.

Folgende Abkürzungen werden hier verwendet:

$DN(t)$	Verzögerung einzelner Operationen im DN-Knoten; t steht für den Operationstyp: l bedeutet das Lesen und b das Bearbeiten,
$DIR(t)$	Verzögerung einzelner Operationen im DIR-Knoten; t steht für den Operationstyp: l bedeutet das Lesen und b das Bearbeiten,
$L(s)$	Verzögerung zwischen zwei Knoten aufgrund begrenzter Bandbreite; s bedeutet die Größe der Nachricht (<i>size</i>).

Für die Verzögerung D zwischen zwei Knoten innerhalb eines Teilgebietes gilt unter Verwendung der *passing*-Technik:

$$D = L(s_{\text{req}}) + \text{DIR}(l_{\text{DN}}) + L(s_{\text{req}}) + \text{DIR}(l_{\text{DIR}}) + \dots + L(s_{\text{req}}) + \text{DIR}(l_{\text{DIR}}) + L(s_{\text{req}}) + \text{DIR}(l_{\text{DN}}) + DN(l) + D_R + L(s_{\text{res}}) + DN(b) + L(s_{\text{res}}) \quad (5.3)$$

Der Wert von D_R ist vom verwendeten Szenario abhängig:

- Für R- oder RV-Szenarien, d.h. für die Szenarien, in denen die Suche nach dem *resident*-DN durchgeführt wird, gilt:

$$D_R = L(s_{\text{req}}) + DN(l) + L(s_{\text{res}}) + DN(b), \quad \text{und}$$

- für V- oder ST-Szenarien gilt:

$$D_R = 0.$$

Der notwendige Suchweg ist durch die jeweiligen Standorte des Anrufenden und des Angerufenen festgelegt. Die Häufigkeit, mit der ein bestimmter Suchweg gewählt wird, gehorcht den Wahrscheinlichkeitsformeln, die bereits bei der formalen Beschreibung der Suchanfrage präsentiert wurden.

Aus der obigen Formel für die Verzögerung D ist es leicht, die minimale und maximale Verzögerungszeit zu berechnen:

$$D_{\min} = L(s_{\text{req}}) + \text{DIR}(l_{\text{DN}}) + DN(l) + L(s_{\text{res}}) \quad (5.4)$$

$$D_{\max} = D_{\min} + k \cdot L(s_{\text{req}}) + \text{DIR}(l_{\text{DN}}) + (k - 1) \cdot \text{DIR}(l_{\text{DIR}}) + D_R + L(s_{\text{res}}) + DN(b) \quad (5.5)$$

Dabei bedeutet k die maximale Suchpfadlänge im DIR-Baum.

Daraus folgt, daß der Minimalwert für alle Prozeduren gleich ist - er entspricht dem Fall, daß die Daten sich lokal befinden. Der Maximalwert ist durch die Höhe des

Baumes, damit durch die maximale Anzahl der Suchschritte im Verzeichnisbaum, und gegebenenfalls durch D_R bestimmt.

Der Mittelwert D_{mean} ergibt sich aus der jeweiligen Benutzerverteilung und aus dem Anrufverhalten und muß numerisch bestimmt werden. In allen untersuchten Varianten mit der gleichen Datenbankstruktur 36-4-1 (vgl. Kapitel 5.3.1 und 5.3.2) variiert der Wert D_{mean} um den Faktor 2,7 :

$$F_{max}(D_{mean}) = \text{Max}(D_{mean})/\text{Min}(D_{mean}) = 2,7.$$

Wenn jede einzelne Variante separat untersucht wird, gilt

$$F_{max}(D_{mean}) = 1,73 .$$

Der Minimalwert von D_{mean} ist für das ST-Szenario und der Maximalwert von D_{mean} für das V-Szenario jedesmal zu beobachten.

Wird in der Grundkonfiguration der Verzeichnisbaum erhöht (von 36-9-1 auf 36-9-3-1), steigt $F_{max}(D_{mean})$ von 1,37 auf 1,52. Das entspricht ca. einem 10%igen Anstieg des Mittelwertes bei Erhöhung des Verzeichnisbaumes um eine Ebene.

Wenn die Werte für Operationen in Knoten in der Größenordnung von einigen Millisekunden liegen /Kluß93/, so sind auch die absoluten Differenzen für die gesamte Anfrage maximal eine Zehnerpotenz größer und liegen im Bereich von 10 - 40 Millisekunden. Dieser Wert ist jenseits der Benutzerwahrnehmung und braucht somit nicht als ein Kriterium herangezogen werden, wenn die Entscheidung über die Zweckmäßigkeit einer Variante getroffen werden soll (vgl. /Rons94/).

5.3.6 Folgerung bezüglich der Datenverteilungsszenarien

Aus den oben ausgewerteten Varianten der einzelnen Alternativen lassen sich nachstehende Schlußfolgerungen bezüglich der Vorteile der einzelnen Datenverteilungsszenarien während der Suchanfrage herleiten:

- Die statischen Szenarien, d.h. ST und R, erweisen sich als vorteilhaft, wenn die Benutzer sich hauptsächlich im und um das Heimgebiet konzentrieren und die Anrufe meist lokal oder in der lokalen Umgebung durchgeführt werden.
- Die dynamischen Szenarien, d.h. V und RV, sind unterschiedlich zu bewerten: Wenn das V-Szenario in fast allen Situationen schlechter als alle anderen Szenarien ist, so ist das RV-Szenario in solchen Situationen von Vorteil, wenn die Benutzer sich um einen Besuchsknoten konzentrieren oder viele Anrufe einen regionalen Charakter, d.h. nicht lokalen, haben. Dieser Vorteil wird durch die Anwendung von direkten Zeigern verstärkt.
- Der Grund für die hohe Belastung bei Verwendung des V-Szenarios liegt darin, daß in den meisten Szenarien (Ausnahme ist die 5. Variante im Kapitel 5.3.2) die Anrufe zu einem großen Teil im eigenen DN-Heimgebiet durchgeführt werden. Gleichzeitig befindet sich aber ein relativ großer Anteil der Benutzer außerhalb des eigenen DN-Heimgebietes. Im V-Szenario ist es notwendig, die Su-

che nach dem Besuchsknoten auch dann durchzuführen, wenn die Anrufe für die Benutzer des eigenen Heimgebietes bestimmt sind. In allen anderen Szenarien ist bei solcher Konstellation keine Suche notwendig. Aus diesem Grund erweisen sich die verschobenen Benutzerverteilungen als günstig für die Anwendung von dynamischen Szenarien, inklusive des V-Szenarios.

5.3.7 Folgerung bezüglich der Abfragetechnik

Aus den durchgeführten Auswertungen der Suchalternativen können nachstehende Schlußfolgerungen bezüglich der Abfragetechnik gezogen werden:

- Die *passing*-Technik ist in jedem Fall zu bevorzugen, wenn es gilt, nur die Leistungsfähigkeit der Abfragetechnik zu beurteilen.
- Unter den sonstigen Techniken erscheint die Kombination aus *chaining* und *referral* als eine mögliche optionale Lösung, die die Belastung der DDB-Knoten gleichmäßiger als die *chaining*- oder *referral*-Technik alleine auf die Datenbankknoten aufteilt. Jede dieser Techniken verlangt einen erhöhten Leistungsbedarf sowohl in den Knoten als auch auf den Verbindungsleitungen, wobei zu bemerken ist, daß der erhöhte Knotenbedarf ausschließlich auf die sonstigen Operationen zurückzuführen ist, d.h., die kritischen Speicherzugriffe sind in allen Techniken identisch.
- Sollen sonstige Aspekte außer der Leistungsfähigkeit, z.B. die Fehlertoleranz, beachtet werden, so ist die *passing*-Technik der *chaining*/*referral*-Technik deutlich unterlegen. Dies ergibt sich aus der mangelnden Überwachung des Informationsflusses.

Da die Problematik der Fehler hier nicht quantifiziert wird, so bleibt als einziges Entscheidungskriterium nur die Leistungsfähigkeit. Die Fehleranfälligkeit soll lediglich als eine zusätzliche Information dienen, wenn die Entscheidung für eine bestimmte Technik vom Standpunkt der Leistungsfähigkeit und des Protokollaufbaus getroffen werden soll.

5.4 Auswertung der Aktualisierungsanfragen

In diesem Unterkapitel werden die Ergebnisse der Auswertung der Aktualisierungsanfrage dargestellt. Ähnlich wie bei der Auswertung der Suchanfrage werden die Änderungen der einzelnen Alternativen bezüglich der Grundkonfiguration erläutert. Alle referenzierten Diagramme sind im Anhang B zu finden.

5.4.1 Einfluß der Mobilitätsparameter

5.4.1.1 Untersuchte Alternativen

Die Beeinflussung der Aktualisierungsanfrage durch die Mobilitätsparameter ist vielseitig. Zum einen ist die Verteilungsdichte ein wichtiger Punkt, zum anderen ist die Beweglichkeit der Benutzer, die durch die Übergangswahrscheinlichkeiten aus-

gedrückt werden, ebenfalls von großer Bedeutung. Die Grundkonfiguration betrachtet eine **mittlere** Konzentration der Benutzer um den Wohnort zusammen mit einer **niedrigen** Beweglichkeit. Hier werden folgende Varianten untersucht:

- Benutzer verteilt um ihren Heimknoten:
 1. Variante: **mittlere** Benutzerkonzentration verbunden mit einer **hohen** Beweglichkeit (Diagramm B.53)
 2. Variante: **starke** Benutzerkonzentration verbunden mit einer **niedrigen** Beweglichkeit (Diagramm B.54)
 3. Variante: **starke** Benutzerkonzentration verbunden mit einer **hohen** Beweglichkeit (Diagramm B.55)
 4. Variante: **schwache** Benutzerkonzentration verbunden mit einer **niedrigen** Beweglichkeit (Diagramm B.56)
 5. Variante: **schwache** Benutzerkonzentration verbunden mit einer **hohen** Beweglichkeit (Diagramm B.57)
- Benutzer verteilt um einen Besuchsknoten (hier den DN-Knoten 23):
 6. Variante: **mittlere** Benutzerkonzentration verbunden mit einer **niedrigen** Beweglichkeit (Diagramm B.58)
 7. Variante: **mittlere** Benutzerkonzentration verbunden mit einer **hohen** Beweglichkeit (Diagramm B.59)
 8. Variante: **starke** Benutzerkonzentration verbunden mit einer **niedrigen** Beweglichkeit (Diagramm B.60)
 9. Variante: **starke** Benutzerkonzentration verbunden mit einer **hohen** Beweglichkeit (Diagramm B.61)
 10. Variante: **schwache** Benutzerkonzentration verbunden mit einer **niedrigen** Beweglichkeit (Diagramm B.62)
 11. Variante: **schwache** Benutzerkonzentration verbunden mit einer **hohen** Beweglichkeit (Diagramm B.63)
- nichtstationärer Fall, d.h., die Benutzer bewegen sich geordnet aus dem Zustand, wo sie um ihren Heimknoten konzentriert sind, in einen neuen Zustand, wo sie um einen Besuchsknoten konzentriert sind:
 12. Variante: **starke** Benutzerkonzentration um den jeweiligen Knoten (Diagramm B.64)
 13. Variante: **mittlere** Benutzerkonzentration um den jeweiligen Knoten (Diagramm B.65)

Die Begriffe schwache, mittlere und starke Benutzerkonzentration entsprechen zahlenmäßig denen, die bei der Suchanfrage verwendet wurden. Die hohe und niedrige Beweglichkeit P_{Bew} haben von Variante zu Variante unterschiedliche Werte, weil in Abhängigkeit von der jeweiligen Benutzerkonzentration die Beweglichkeit nur in bestimmten Grenzen variierbar sind. Dies ist durch das verwendete Modell und die Zustandsgleichungen bedingt. Das Beispiel in Abbildung 5.4 erklärt den Zusammenhang zwischen der Benutzerverteilung (Benutzerkonzentration) und dem möglichen

Bereich der Beweglichkeitswerte (vgl. Kapitel 4.2.1.6 zur Erklärung der Beweglichkeit).

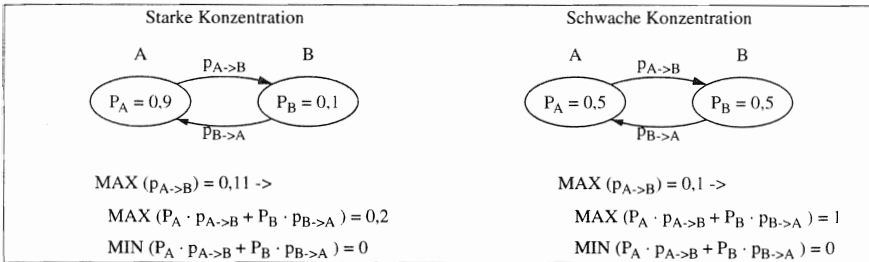


Abb. 5.4 Zusammenhang zwischen der Benutzerverteilung und den Übergangswahrscheinlichkeiten

5.4.1.2 Ergebnisse der Auswertung

Nachstehende Folgerungen lassen sich aus den erhaltenen Ergebnissen herleiten:

- Die 1. Variante ist ähnlich der Grundkonfiguration mit dem einzigen Unterschied, daß die Benutzer jetzt beweglicher geworden sind. Die Ergebnisse der Auswertung zeigen ein gänzlich unterschiedliches Bild der einzelnen Szenarien. Das ST-Szenario erweist sich als das beste Szenario, gefolgt durch das R-Szenario und später durch die viel schlechteren V- und RV-Szenarien. Ein interessantes Bild stellt auch die Situation im Besuchsgebiet dar: Aufgrund der höheren Beweglichkeit ist das ST-Szenario jetzt nur geringfügig schlechter als das R-Szenario.
Die Anwendung der direkten Zeiger in den dynamischen Szenarien ist in dieser Variante nicht vorteilhaft. Dabei gilt ähnliches, wie in der Grundkonfiguration zu beobachten war: Bei der Anwendung der direkten Zeiger ist die Schreibzunahme gegenüber der Leseabnahme geringfügig stärker.
- In der 2. Variante ist der Benutzer ähnlich schwach mobil wie in der Grundkonfiguration, ist aber stark um den Heimknoten konzentriert. Die Ergebnisse sind prinzipiell wie in der Grundkonfiguration, wobei die Nachteile des ST-Szenarios nicht so stark ausgeprägt sind, wie dies in der Grundkonfiguration zu beobachten war.
- In der 3. Variante sind die Absolutwerte niedriger als in der 1. Variante. Die Aussagen, die für die 1. Variante gemacht worden sind, behalten jedoch ihre Gültigkeit.
- In der 4. Variante sind die Ergebnisse ähnlich wie in der Grundkonfiguration, wobei die Nachteile des ST-Szenarios viel stärker ausgeprägt sind als in der Grundkonfiguration.
- In der 5. Variante sind die Absolutwerte höher als in der 1. Variante. Die grundsätzlichen Aussagen, die für die 1. Variante gemacht worden sind, behalten jedoch ihre Gültigkeit. Es stellt sich allerdings heraus, daß das ursprünglich sehr

schlechte V-Szenario jetzt nur geringfügig schlechter als das R-Szenario erscheint.

- Die absolute Belastung des Netzes, d.h. der Knoten und der Verbindungen, in den Varianten 2 und 4 bei Verwendung der R-, V- und RV-Szenarien ist sehr ähnlich der Situation in der Grundkonfiguration. Das ST-Szenario erfährt die größten Änderungen.
- Die Varianten 6 bis 11 untersuchen die Fälle, in denen die Benutzer sich um ein Besuchsgebiet konzentrieren. Folgendes ist dabei zu beobachten:
 - In allen Varianten mit einer niedrigen Beweglichkeit, also in den Varianten 6, 8 und 10, ist das ST-Szenario im Vergleich zu allen anderen Szenarien beim Lesen und bei den sonstigen Operationen in den Datenbankknoten sehr schlecht. Beim Schreiben ist das ST-Szenario gegenüber allen anderen leicht im Vorteil (ca. 10%). Diese Unterschiede sind um so stärker, je konzentrierter sich die Benutzer um einen Knoten versammeln.
Die Verbindungsbelastung des ST-Szenarios ist hier nur geringfügig niedriger als in allen anderen Szenarien.
Die absolute Belastung des Netzes, d.h. der Knoten und der Verbindungen, in den R-, V- und RV-Szenarien ist sehr ähnlich der Situation in der Grundkonfiguration. Ähnliche Eigenschaften wurden in Variante 2 und 4 beobachtet.
Die Anwendung der direkten Zeiger verschlechtert nur geringfügig die Eigenschaften der dynamischen Szenarien.
 - In allen Varianten mit einer hohen Beweglichkeit, also in Variante 7, 9 und 11, sind die Ergebnisse für das ST-Szenario bezüglich der Knotenbelastung praktisch äquivalent zu den entsprechenden mit einer niedrigen Beweglichkeit.
Die Ergebnisse für die sonstigen Szenarien sind jedoch gänzlich unterschiedlich: Die Belastung der Knoten und Verbindungen steigt in den hoch mobilen Varianten im Vergleich zu den entsprechenden mit einer niedrigen Beweglichkeit sehr stark an.
Die Szenarien ST und R sind diejenigen, die bezüglich der Knotenbelastung insgesamt ziemlich äquivalent sind: Das ST-Szenario zeigt mäßige Vorteile in den Datenknoten, das R-Szenario ähnliche Vorteile in den Verzeichnisknoten.
Die Verbindungsbelastung ist beim ST-Szenario deutlich niedriger.
Die V- und RV-Szenarien sind wegen der Notwendigkeit, die Verzeichnisknoten zu aktualisieren, sehr schlecht im Vergleich zu den beiden statischen Szenarien, die keine Schreiboperationen in den Verzeichnisknoten verlangen.
 - Die Anwendung der direkten Zeiger in den Varianten 7, 9 und 11 ist nicht sinnvoll, da der Anstieg der Schreibaktionen gegenüber den Verbesserungen beim Lesen und bei sonstigen Knotenoperationen stark überwiegt.
- Die Varianten 12 und 13 untersuchen die nichtstationären Fälle. Die Benutzer bewegen sich aus der Verteilung um den Heimknoten 15 in eine neue Verteilung um den Besuchsknoten 23 und rufen dabei 3148 bzw. 3303 Aktualisierungsaktionen hervor (in den Diagrammen wurde nur die nichtstationäre Komponente dargestellt). Folgendes ist dabei zu beobachten:

- Das ST-Szenario ist in beiden Varianten eindeutig am besten. Das Ergebnis der Untersuchung ist dadurch so extrem günstig für das ST-Szenario, weil in diesem Szenario kein Datentransport bei Bewegungen innerhalb eines Teilgebietes stattfindet.
- Das R-Szenario erweist sich als zweitbestes Szenario, wobei in der untersuchten Konfiguration der Unterschied zum ST-Szenario sehr stark ist.

5.4.2 Einfluß der DIR-Architektur

5.4.2.1 Untersuchte Alternativen

Die bisherigen Auswertungen wurden für die symmetrische Struktur wie in der Grundkonfiguration (36-4-1) durchgeführt. Im folgenden werden verschiedene, u.a. auch unsymmetrische Verzeichnisstrukturen bewertet. Als Grundlage wird ein hochmobiles Benutzerverhalten angenommen, wie in der vorherigen 1. Variante, da bei diesem Verhalten die Belastung des Verzeichnisses höher ist und deswegen auch die etwaigen Unterschiede stärker ausgeprägt sind. Als Strukturalternativen werden die gleichen Varianten untersucht wie zuvor bei der Suchanfrage:

- symmetrische vollständige Architektur:
 1. Variante: hohe Verzeichnishierarchie:
 - a) 36-9-3-1 : vgl. Diagramme B.66 und B.67
 - b) 36-4-2-1 : vgl. Diagramm B.68
 2. Variante: flache Verzeichnishierarchie, d.h. 36-1 (Diagramme B.69 und B.70)
- symmetrische unvollständige Architektur - hier werden zwei Optionen der Zeigerverwaltung untersucht:
 - a) keine Suche auf der Wurzelebene - die Zeiger werden bei der Aktualisierungsanfrage mit neuen Inhalten versehen
 - b) Suche auf der Wurzelebene unter Verwendung der *referral*-Abfragetechnik - d.h. keine Aktualisierung
- 3. Variante: flache Verzeichnishierarchie, d.h. 36-4 :
 - a) vgl. Diagramm B.71
 - b) vgl. Diagramm B.72
- 4. Variante: hohe Verzeichnishierarchie, d.h. 36-9-3 :
 - a) vgl. Diagramm B.73
 - b) vgl. Diagramm B.74
- 5. Variante: hohe Verzeichnishierarchie, d.h. 36-4-2 :
 - a) vgl. Diagramm B.75
 - b) vgl. Diagramm B.76
- unsymmetrische Architektur - hier ist die Verzeichnisstruktur dem Benutzerverteilungsmuster angepaßt, d.h., der Heimknoten und seine Nachbarn haben einen gemeinsamen Verzeichnisknoten. Es werden vier Varianten der Benutzerverteilung untersucht:

6. Variante: Benutzer **stark** konzentriert um den Heimknoten, **niedrige** Mobilität (Diagramm B.77)
7. Variante: Benutzer **stark** konzentriert um den Heimknoten, **hohe** Mobilität (Diagramm B.78)
8. Variante: Benutzer **schwach** konzentriert um den Heimknoten, **niedrige** Mobilität (Diagramm B.79)
9. Variante: Benutzer **schwach** konzentriert um den Heimknoten, **hohe** Mobilität (Diagramm B.80)

5.4.2.2 Ergebnisse der Auswertung

Nachstehende Folgerungen lassen sich aus den erhaltenen Ergebnissen herleiten:

- Die Varianten 1a und 1b verwenden eine höhere DIR-Struktur als die Grundkonfiguration. Die Ergebnisse zeigen deutlich, daß die Anwendung eines höheren DIR-Baumes immer durch einen Anstieg der Gesamtnetzbelastung begleitet wird. Die Belastungsänderungen der einzelnen Knoten bei Bewegungen innerhalb eines Teilgebietes ist gegenüber der Grundkonfiguration davon abhängig, inwieweit sich die verglichenen DIR-Strukturen ähneln:
 - Stellt der höhere Baum eine direkte Erweiterung eines niedrigeren Baumes dar, so wird der Wurzelknoten entlastet (Variante 1b).
 - Ist jedoch die Ähnlichkeit der Bäume nicht so groß, so ist es sogar möglich, daß der Wurzelknoten stärker belastet wird (Variante 1a).

Eine ähnliche Überlegung gilt für Knoten jeder Baumebene, die in zwei unterschiedlichen Strukturen verglichen werden.

Zwischen den Teilgebieten wird durch die Baumerhöhung keine Reduktion der Belastung in einzelnen Knoten zu erwarten sein.

Der Einsatz von direkten Zeigern in den dynamischen Szenarien ist nicht empfehlenswert, da die Gesamtbelastung des Netzes aufgrund der gestiegenen Anzahl an Schreiboperationen stark erhöht wird.

- Die 2. Variante benutzt eine maximal flache DIR-Struktur, die nur aus einem Knoten besteht. Aus den Ergebnissen ist sichtbar, daß einerseits die Gesamtbelastung der Knoten und Verbindungen gegenüber der Grundkonfiguration gesunken ist. Andererseits ist die maximale Belastung des einzigen DIR-Knotens (hier des Wurzelknotens) fast um 100% angestiegen.
In einem realen Netz ist also zu prüfen, inwieweit der Anstieg der notwendigen Leistungsfähigkeit in den verbleibenden Verzeichnisknoten noch zulässig ist, oder anders ausgedrückt, ob die Belastung der Verzeichnisknoten in einer flachen Struktur noch zu vertreten ist.
- Die Varianten 3, 4 und 5 untersuchen DIR-Strukturen, in denen der gemeinsame Wurzelknoten fehlt. Sie werden entsprechend mit der Grundkonfiguration und den Varianten 1a und 1b verglichen. Folgende Schlußfolgerungen für den Fall der Bewegungen innerhalb eines Teilgebietes lassen sich hier ermitteln:
 - Die Varianten 3a - 5a, die das Suchen auf der Wurzelebene überflüssig machen, haben den Nachteil, daß sie sehr viele zusätzliche Schreibzugriffe ver-

langen, um den gesamten Baum zu aktualisieren. Dies ist um so aufwendiger, je mehr Wurzelknoten vorhanden sind. Die Varianten 3a, 4a und 5a mit entsprechend 4, 3 und 2 Wurzelknoten zeigen diese Abhängigkeit deutlich:

- In der Variante 3a ist die Belastung im Vergleich mit der Ein-Baum-Struktur sehr hoch.
- Dagegen ist in der Variante 5a die Gesamtschreibbelastung geringer und die Belastung der einzelnen DIR-Knoten gegenüber der Ein-Baum-Struktur kaum angestiegen.

Der Vorteil der obigen Varianten ist, daß die entstandene Lesebelastung bei Bewegungen innerhalb eines Teilgebietes bei den dynamischen Szenarien geringer und bei den statischen Szenarien nur geringfügig höher ist als in der vergleichbaren Ein-Baum-Struktur.

Die Verbindungsbelastung der Varianten 3a - 5a ist im Vergleich zu den entsprechenden Varianten mit vollständigen Bäumen beinahe äquivalent.

- Die Varianten 3b - 5b, die keine Aktualisierung unter den Wurzelknoten verlangen, sind offensichtlich besser bezüglich der Schreiboperationen, haben jedoch im Regelfall (in Varianten 3b und 4b) eine höhere Anzahl Lese- und sonstiger Operationen. Der letzte Sachverhalt ist bei Anwendung von nur zwei Wurzelknoten umgekehrt: in der Variante 5b ist die Anzahl sowohl der Lese- als auch der Schreiboperationen geringer als in der vergleichbaren Variante 1b.

Diese Abhängigkeit von der Anzahl der Wurzelknoten ist ähnlich zu der, die bei der Auswertung der Varianten 3a, 4a und 5a beobachtet wurde.

Die Verbindungsbelastung ist in jeder Variante beinahe äquivalent zu der des vollständigen Baumes.

- Der obige Vergleich gilt für die Bewegungen innerhalb eines Teilgebietes. Bei Bewegungen zwischen den Teilgebieten (Teilgebietsaktualisierung) kommt aufgrund des Vorhandenseins von mehreren Wurzelknoten eine Komplikation hinzu, die in den entsprechenden Diagrammen nicht berücksichtigt wurde, die aber einfach separat berechnet werden kann:
- In den Varianten 3a - 5a, die eine Aktualisierung aller Wurzelknoten voraussetzen, müssen immer alle Wurzelknoten in beiden, d.h. im neuen und im alten Teilgebiet, aktualisiert werden. Das resultiert in einer Zusatzbelastung der Wurzelknoten um

$$B = 2 \cdot (n - 1) \cdot x_{\text{move}} \quad (4.6)$$

wobei n die Wurzelknotenanzahl und x_{move} die Anzahl der Bewegungen zwischen den Teilgebieten ist.

In den dargestellten Varianten 3a, 4a und 5a betragen die Werte der Zusatzbelastung entsprechend 180, 120 und 60, da $x_{\text{move}} = 30$ gilt.

- In den Varianten 3b - 5b, die keine Aktualisierung vorsehen, ist die Situation einfacher. Da man annehmen darf, daß beim Teilgebietswechsel die zuständigen Wurzelknoten direkt angesprochen werden (dies ist ein wesentlicher Unterschied zu der Situation, die sich auf der Wurzelebene wäh-

rend der Suchanfrage ergibt), braucht keine zusätzliche Suche unter den Wurzelknoten durchgeführt werden, und somit ist keine Zusatzbelastung zu erwarten.

Die obige Überlegung ist dann entscheidend, wenn die Anzahl der Bewegungen zwischen den Teilgebieten deutlich höher ist, als in dem untersuchten Fall, z.B. wenn die Teilgebiete ineinander 'verflochten' sind, d.h., wenn zwei Netzbetreiber auf dem gleichen geographischen Gebiet koexistieren und ein Wechsel zwischen den beiden Netzen regelmäßig vorkommt.

- Die Varianten 6 bis 9 betrachten die Belastung des DIR-Baumes, wenn dessen Struktur an die Benutzerbewegungen angepaßt worden ist. Sie sind entsprechend mit den Varianten 2, 3, 4 und 5 aus dem Kapitel 5.4.1 zu vergleichen, die ohne eine Strukturanpassung ein identisches Benutzerverhalten aufweisen. Folgende Schlußfolgerungen lassen sich jetzt herleiten:
 - Wenn die Benutzer stark um das Gebiet konzentriert sind, so ergibt sich eine deutliche Reduktion der Knotenbelastung, die um so stärker ist, je beweglicher die Benutzer sind (Variante 6 und 7 gegenüber der Variante 2 und 3). Die Reduktion der Belastung ist sowohl insgesamt als auch in jedem einzelnen Knoten erreicht worden.
Die Verbindungsbelastung ist in allen Situationen beinahe äquivalent, mit Ausnahme der *chaining*-Technik, wo eine Reduktion der Verbindungsbelastung auf 50 % des ursprünglichen Wertes erreicht wurde.
Die Verwendung von direkten Zeigern in den dynamischen Szenarien ist nicht zu empfehlen.
 - Wenn die Benutzer stark auf das Gebiet verteilt sind, so liegt offensichtlich keine Anpassung des DIR-Baumes an die Bewegungen vor. Dies ist in den Varianten 8 und 9 durch eine leicht angestiegene Knoten- und Verbindungsbelastung zu beobachten.
 - Der leichte Anstieg der Belastung im zweiten Fall gegenüber der starken Reduktion im ersten Fall lassen den Schluß zu, daß bei starker Benutzerverteilung jede DIR-Struktur beinahe gleich gut ist, aber bei starker Konzentration nur eine angepaßte Struktur sinnvoll erscheint.

5.4.3 Einfluß der Gebietstopologie

5.4.3.1 Untersuchte Alternativen

In den bisherigen Untersuchungen wurde stets vorausgesetzt, daß das UMTS-Teilgebiet immer die gleiche Topologie wie in der Grundkonfiguration aufweist. Im folgenden werden andere Topologien ausgewertet. Die jetzt gewählten Topologien entsprechen denen, die bereits bei der Suchanfrage untersucht wurden. Bezüglich des Benutzerverhaltens wird angenommen, daß die Benutzer mittlere Konzentration aufweisen und eine hohe Beweglichkeit zeigen - die 1. Variante im Kapitel 5.4.1. Folgende Varianten werden jetzt untersucht:

- Die Größe des Teilgebietes wird geändert:

1. Variante: ein kleines Gebiet mit 16 DN-Gebieten (Diagramm B.81)
2. Variante: ein großes Gebiet mit 64 DN-Gebieten (Diagramm B.82)
- Die Form des Teilgebietes wird geändert:
 3. Variante: ein längliches Gebiet mit 10 mal 3 DN-Gebieten (Diagramm B.83)

5.4.3.2 Ergebnisse der Auswertung

Nachstehende Folgerungen lassen sich aus den erhaltenen Ergebnissen herleiten:

- In allen Varianten gelten prinzipiell die bereits aus der 1. Variante im Kapitel 5.4.1 bekannten Schlußfolgerungen.
Daraufhin kann man die These aufstellen, daß die Ergebnisse der Aktualisierungsanfrage von der Größe und Form des UMTS-Teilgebietes unabhängig sind, sofern sich die Benutzercharakteristika infolge der Topologieänderung nicht verändern. Das bedeutet auch, das die obige Folgerung nur dann gilt, wenn keine zu starken Abweichungen von der untersuchten Grundform gegeben sind.
- Folgende geringfügige Unterschiede sind im einzelnen zu beobachten:
 - Im kleineren Teilgebiet (1. Variante) sind die Vorteile des ST-Szenarios stärker und im größeren Teilgebiet (2. Variante) schwächer ausgeprägt.
 - Im länglichen Teilgebiet (3. Variante) sind die Vorteile des ST-Szenarios stärker ausgeprägt.

5.4.4 Folgerungen bezüglich der Datenverteilungsszenarien

Aus den oben ausgewerteten Varianten der einzelnen Alternativen lassen sich zusammenfassend Schlußfolgerungen bezüglich der Vorteile der einzelnen Datenverteilungsalternativen in der Aktualisierungsanfrage herleiten:

- Es gibt zwei Aspekte, die darüber entscheiden, welches Szenario bezüglich der Knotenbelastung vorteilhaft ist:
 - die Benutzerbeweglichkeit und
 - das Anziehungsgebiet, d.h. dieses DN-Gebiet, um das sich die Benutzer konzentrieren.
- Wenn die Benutzer sich um das eigene Heimgebiet konzentrieren, so bietet sich das folgende Bild:
 - Im Falle der hohen Beweglichkeit ist das ST-Szenario eindeutig die beste Option, und das R-Szenario folgt an zweiter Stelle.
 - Im Falle einer niedrigen Beweglichkeit ist das R-Szenario die beste Option, gefolgt durch das V-Szenario an zweiter Stelle - das V-Szenario ist geringfügig besser beim Lesen, aber gleichzeitig viel schlechter beim Schreiben.
Die Unterschiede zwischen dem besten R-Szenario und dem ST-Szenario sind vom Konzentrationsgrad abhängig und nehmen mit steigender Konzentration ab.

- Wenn die Benutzer sich um ein Besuchsgebiet konzentrieren, so stellt sich folgende Situation dar:
 - Betrachtet man Fälle mit einer hohen Beweglichkeit, so sind die ST- und R-Szenarien ziemlich äquivalent.
 - Betrachtet man dagegen die Fälle mit einer niedrigen Beweglichkeit, so ist das ST-Szenario mit Abstand das schlechteste, da die Nachteile bei Lese- und sonstigen Operationen die geringen Vorteile beim Schreiben weit übertreffen. Die R- und V-Szenarien sind jetzt diejenigen, die die niedrigsten Belastungen hervorrufen.
- Bezüglich der Verbindungsbelastung ist immer das ST-Szenario vorteilhaft.
- Die Hauptursache für die Nachteile der dynamischen Szenarien (V und RV) ist die Tatsache, daß dort der DIR-Baum aktualisiert werden muß. Hier ist das Schreiben in die Verzeichnisknoten beim Aktualisieren notwendig. Dieser Nachteil ist nur schwer durch die Vorteile beim Lesen aufzuholen. Bei der Anwendung von flachen Bäumen sind allerdings die Unterschiede nicht so gravierend.

5.4.5 Folgerungen bezüglich der Abfragetechnik

Aus den durchgeführten Auswertungen können gleiche Schlußfolgerungen bezüglich der Abfragetechnik hergeleitet werden wie bereits im Kapitel 5.3.7 für die Suchanfrage. Der einzige Unterschied im Vergleich zur Suchanfrage ist dieser, daß hier die Vorteile der *passing*-Technik geringer sind: Der Anteil der Lese- und Schreiboperationen an der Gesamtanzahl der Operationen in Knoten ist höher als bei Suchanfragen, wo keine Schreiboperationen auftreten. Damit sind die Vorteile der *passing*-Technik, die sich lediglich aus der unterschiedlichen Anzahl sonstiger Operationen ergeben, notwendigerweise geringer (in Relation zur gesamten Anzahl an Operationen).

5.4.6 Einfluß des Vorwissens auf die Auswertung

In den obigen Untersuchungen wurde stets vorausgesetzt, daß das Endgerät dem Netz keine Information darüber liefert, in welchem DN-Gebiet sich der Benutzer zuvor befunden hat. Dies erzwingt jedesmal während der Aktualisierungsanfrage einen Suchvorgang nach dem alten Standort. Erst danach findet die tatsächliche Aktualisierung statt. Im folgenden soll die Situation betrachtet werden, in der das Endgerät die Kennung des alten DN-Gebietes an das Festnetz liefert, d.h., die Datenbank verfügt zu Beginn jeder Aktualisierungsanfrage über die Information, die den erwähnten Suchvorgang überflüssig macht. Diese Information wird hier als Vorwissen bezeichnet.

Die notwendige Voraussetzung für die Durchführbarkeit dieser Alternative ist die Fähigkeit der Datenbankknoten, die sonstigen DN-Knoten des eigenen Teilgebietes aufgrund ihrer Nummern zu adressieren und die DN-Knoten der anderen Teilgebiete als solche zu erkennen.

Zwei Varianten mit unterschiedlichen Benutzerverhalten werden im folgenden untersucht:

1. **mittlere** Benutzerkonzentration, verbunden mit einer **niedrigen** Beweglichkeit, entspricht der Grundkonfiguration (Diagramme B.84 und B.85)
2. **mittlere** Benutzerkonzentration, verbunden mit einer **hohen** Beweglichkeit, entspricht der 1. Variante aus dem Kapitel 5.4.1 (Diagramm B.86)

Nachstehende Folgerungen, die aus den erhaltenen Ergebnissen herleitet wurden, konzentrieren sich auf die Unterschiede zu den entsprechenden Vergleichsszenarien:

- Der Hauptunterschied zu den davor untersuchten Varianten liegt hier in der Entlastung der Verzeichnisknoten aufgrund fehlender Suche nach dem alten Aufenthaltsgebiet, d.h., es gibt keine Leseoperationen in den Verzeichnisknoten.
- Da nicht nach dem alten Aufenthaltsgebiet gesucht wird, ist die Anwendung von direkten Zeigern in den dynamischen Szenarien eindeutig von Nachteil.
- Die Unterschiede zwischen den einzelnen Abfragetechniken sind jetzt kaum noch relevant, da alle Schreiboperationen identisch durchgeführt werden.

5.5 Speicherbedarf bei verschiedenen Alternativen

Der gesamte Speicherbedarf in allen Alternativen ist durch den Speicherbedarf in den DN-Knoten und in den DIR-Knoten bestimmt.

5.5.1 Speicherbedarf in den DN-Knoten

Der Speicherbedarf in den DN-Knoten ist ausschließlich von der verwendeten Datenverteilungsstrategie abhängig, wenn man von der generellen Problematik der Sicherheitskopien absieht. Dabei sind die einzelnen Szenarien folgendermaßen zu bewerten:

- Das ST-Szenario hat den geringsten Speicherbedarf, da die Daten nur einmal im Heimknoten gehalten werden.
- Das R-Szenario verlangt gegenüber dem ST-Szenario zusätzlichen Speicherplatz, um die Zeiger zu halten, die auf den Besuchsknoten gerichtet sind.
- Das V-Szenario verlangt mehr Speicherplatz als das R-Szenario, da hier Zweitkopien von *subscription*- und *routing*-Daten im Besuchsknoten gehalten werden.
- Das RV-Szenario verlangt gegenüber dem V-Szenario zusätzlichen Speicherplatz, um die Zeiger zu halten, die auf den Besuchsknoten gerichtet sind.

Bei Bewegungen zwischen Teilgebieten oder innerhalb des Besuchsteilgebietes gibt es keinen Unterschied zwischen dem V- und RV-Szenario, da hier im Besuchsteilgebiet kein Heimknoten vorhanden ist.

Die tatsächliche Menge an zusätzlichem Speicherplatz aufgrund der Zweitkopien in den Besuchsknoten ist vom Benutzeranteil außerhalb des Heimknotens linear abhängig.

5.5.2 Speicherbedarf in den DIR-Knoten

Der Speicherbedarf in den DIR-Knoten ist sowohl von der Baumstruktur als auch vom Datenverteilungsszenario abhängig:

- Ein hoher Baum verstärkt die Unterschiede zwischen den einzelnen Szenarien, und ein flacher Baum gleicht sie aus.
- Für die Szenarien ergibt sich folgender Zusammenhang:
 - Die ST- und R-Szenarien sind identisch und verlangen ausschließlich Zeiger auf den Heimknoten; die Zeiger werden immer als direkte Zeiger organisiert.
 - Die V- und RV-Szenarien sind ebenfalls identisch und verlangen sowohl Zeiger auf den Heimknoten als auch Zeiger auf den Besuchsknoten; damit benötigen sie mehr Speicherplatz als die beiden statischen Szenarien.

5.5.3 Numerisches Beispiel

Die geschilderten Differenzen sind grundsätzlicher Art. Bei der Beurteilung der Leistungsfähigkeit einer Alternative werden sie jedoch geringere Bedeutung haben, weil sie kein hohes kommerzielles Gewicht besitzen. Folgendes Beispiel soll die zahlenmäßige Größenordnung der Unterschiede zwischen den Szenarien demonstrieren und die obige These untermauern:

- Es werden 1 Mio. Benutzer in einem UMTS-Teilgebiet betrachtet.
- Die Datenknoteneinträge bestehen aus Benutzer- und Endgerätedaten. Die Größe der *subscription*- und *routing*-Daten sei zusammen 5 kByte und die restlichen Benutzer- und Endgerätedaten seien 10 kByte (hauptsächlich aufgrund der *registration*-Daten). Dies ist eine sehr datenintensive Annahme.
- Die Verzeichnisknoteneinträge bestehen aus Zeigern auf die Benutzer- und Endgerätedaten. Jeder Zeigereintrag setzt sich maximal aus vier Elementen zusammen: einfacher oder zweifacher Schlüssel, eine Typbezeichnung und eine Knotenliste (/Schi95/). Die Größe der einzelnen Elemente sei 8 Byte. Damit ist jeder Eintrag im Verzeichnisknoten maximal 32 Byte lang.
- 75 % der Benutzer verlassen ihr Heimatgebiet, d.h., die Benutzer sind sehr mobil.

Daraus ergibt sich in den DN-Knoten unter Verwendung von V- oder RV- Szenarien ein Extrabedarf an Speicher von insgesamt 3,75 GByte ($= 750\,000 \cdot 5\text{kByte}$), d.h. 25 % der ursprünglich notwendigen 15 GByte ($= 1\,000\,000 \cdot 15\text{kByte}$). Unter Beachtung des heutigen Preises von ca. 50 DM pro 1 MByte ergibt sich damit ein Zusatzaufwand von 187,5 TDM. Dies ist angesichts der Kosten des Gesamtsystems (einiger Hundert Mio. DM) ein bescheidener Betrag. Dies gilt trotz der angenommenen starken Benutzerverteilung und der beträchtlichen Größe des Benutzerprofils.

Der zusätzliche Speicherbedarf in den DIR-Knoten (unter der Annahme von 3 Verzeichnisebenen) wird 72 MByte ($= 750\,000 \cdot 32 \text{ Byte} \cdot 3$) nicht überschreiten, also eine Größe, die keine nennenswerten Kosten verursacht.

5.6 Relation zwischen UMTS-Konfigurationen und GSM

Es ist schwierig, das GSM-Konzept direkt mit den UMTS-Alternativen zu vergleichen, weil mindestens zwei grundsätzliche Unterschiede zwischen den beiden Systemen bezüglich der Datenkonzepte vorliegen:

- Die GSM-Nummer ist gleichzeitig die Leitwegnummer des für den Benutzer verantwortlichen HLR-Knotens.
- Die DN-Knoten in der UMTS-Datenbank unterscheiden nicht zwischen physikalisch getrennten Besuchs- und Heimknoten.

Trotz der Unterschiede stellt sich die Frage, inwieweit eine GSM-ähnliche Konfiguration der UMTS-Datenbank die UMTS-spezifischen Aufgaben erfüllen kann, und wenn ja, ob die GSM-Lösung wegen ihrer bereits erworbenen Reife den UMTS-Konzepten nicht vorzuziehen ist. Um auf diese Frage eine Antwort zu geben, wird eine speziell gewählte UMTS-Datenbankalternative diskutiert (von hier an als GSM-Alternative bezeichnet) und mit den bereits vorliegenden Ergebnissen verglichen.

Die GSM-Alternative wird wie folgt spezifiziert:

1. Da im UMTS-Konzept jeder DN-Knoten für manche Benutzer den Heimknoten und für andere den Besuchsknoten darstellt, gibt es keine physikalische Trennung zwischen den HLR- und VLR-ähnlichen Funktionen.
2. Der DIR-Baum besteht aus einer Minimalkonfiguration, nämlich aus einem Wurzelknoten. Damit ist die Suche nach dem Heimknoten des Benutzers in maximal zwei Schritten erfolgreich abgeschlossen. Es ist wegen der personenbezogenen Nummer in UMTS nicht möglich, auf das Verzeichnis gänzlich zu verzichten.
3. Für die Datenverteilung wird das R-Szenario angewandt: Der Heimknoten speichert allgemeine Benutzerdaten und hält den Zeiger auf den Besuchsknoten; dies entspricht der Relation zwischen den Inhalten von HLR- und VLR-Knoten.
4. Als interne Abfragetechnik wird die *chaining*-Technik verwendet, da diese dem GSM-Informationsfluß sehr ähnlich ist.

Aus der obigen Spezifikation folgt, daß die sog. GSM-Alternative mit der bereits untersuchten 2. Variante im Kapitel 5.3.3 und 5.4.2 übereinstimmt. Angesichts der bereits vorliegenden Ergebnisse kann für die GSM-Alternative folgendes geschlossen werden:

1. Die Unterbringung der beiden GSM-Knotentypen, HLR und VLR, in einem DN-Knoten bringt keine prinzipiellen Komplikationen mit sich. Die Benutzeranzahl und die Benutzerverteilung, die der heutigen Dimensionierung des GSM-Netzes zugrunde liegen, machen eine physikalische Kombination von HLR- und

VLR-Knoten ineffizient. Im künftigen Benutzerszenario ist dagegen eine Kombination von Heim- und Besuchsknotenfunktion durchaus sinnvoll. Somit liegt hier kein grundsätzlicher Widerspruch zwischen dem GSM-Konzept und der GSM-Alternative vor.

2. Die Anwendung des R-Szenarios ist neben dem ST-Szenario in den meisten der untersuchten Fälle eine akzeptable Option, so daß auch hier kein Widerspruch zu GSM vorliegt.
3. Die Verzeichnisstruktur, auch wenn sie aus einem Wurzelknoten besteht, stellt eine prinzipielle Erweiterung des heutigen GSM-Konzeptes dar. Diese minimale Verzeichnisstruktur bietet eine optimale UMTS-Lösung, vorausgesetzt, daß die Leistungsfähigkeit des einzelnen Wurzelknotens für das gesamte Teilgebiet ausreichen wird. Das bedeutet, daß eine auf GSM aufgebaute Lösung direkt die optimale UMTS-Alternative hervorbringt. Die Funktion des Wurzelknotens könnte in GSM eine speziell für diesen Zweck eingerichtete Übersetzungstabelle übernehmen.
4. Die Anwendung der *chaining*-Abfragetechnik ist keine empfehlenswerte Lösung, besonders, wenn nur eine Verzeichnisebene vorhanden ist und der Wurzelknoten (oder die Übersetzungstabelle) ohnehin leicht überlastet werden kann. Hier bietet sich die *referral*-Abfragetechnik an.

Aus dem obigen folgt, daß das GSM-Datenkonzept durchaus erweiterbar ist, wobei eine Integration und Erweiterbarkeit der Verzeichnisstruktur vorgesehen werden müssen, wenn die Skalierbarkeit der künftigen Lösung gewährleistet werden soll.

5.7 Folgerungen und Empfehlungen

Die unterschiedlichen Varianten, die ausgewertet wurden, geben einen guten Überblick über den Einfluß von verschiedenen Verhaltensparametern, Verzeichnisstrukturen und Datenverwaltungsmethoden auf die Leistungsfähigkeit der UMTS-Datenbank. In den bisherigen Untersuchungen wurden die verschiedenen Einflußgrößen separat ausgewertet. Jetzt werden eine Zusammenfassung der Folgerungen gegeben und einige Empfehlungen ausgesprochen.

Es hat sich herausgestellt, daß sich eine flache Verzeichnishierarchie in allen untersuchten Konfigurationen als vorteilhaft erweist, d.h., die insgesamt benötigte Knotenleistung ist deutlich kleiner. Die dabei entstehende gesamte Belastung der Verbindungsleitungen ist lediglich geringfügig niedriger. Diese Verbesserung der Gesamtleistungsfähigkeit hat jedoch einen gravierenden Nachteil: die maximale Belastung der verbliebenen Verzeichnisknoten steigt an (in einigen Situationen sogar erheblich). In der Konsequenz bedeutet das, daß bei der Dimensionierung eines Netzes das erwartete Datenverkehrsaufkommen als Grundlage der Entscheidung über die Struktur des Verzeichnisses genommen werden muß: Wenn es möglich ist, den Anforderungen mit einem einzigen Verzeichnisknoten gerecht zu werden, soll diese Lösung bevorzugt werden. Wenn die Anforderungen zu hoch sind, ist zuerst eine Multi-Baum-Struktur bis ca. 2-3 Wurzelknoten zu bevorzugen. Erst danach soll der zweiten Verzeichnisebene Vorzug gewährt werden. Die Verwendung von Multi-

Baum-Strukturen mit mehr als 3 Wurzelknoten ist nicht zu empfehlen: Der Aufwand, der für die Suche oder für die Aktualisierung mehrerer Wurzelknoten notwendig ist, übersteigt deutlich den Extraaufwand, der aufgrund des Einsatzes eines zusätzlichen gemeinsamen Wurzelknotens entsteht. Dies gilt sowohl für die Bewegungen innerhalb eines Teilgebietes als auch zwischen den Teilgebieten.

Wenn es zur Anwendung von mehreren Verzeichnisknoten kommen soll, ist es von Vorteil, die Struktur des Baumes dem Bewegungsverhalten der Benutzer anzupassen, d.h., die Baumstruktur wird zu einer Benutzereigenschaft. Da unterschiedliche Verhaltensmuster zu erwarten sind, ist es empfehlenswert, mehrere Strukturklassen zu spezifizieren und jeden Benutzer jeweils einer Klasse zuzuordnen. Eine eventuell notwendige Änderung der Zuordnung kann am einfachsten dann durchgeführt werden, wenn sich der jeweilige Benutzer in seinem Heimknoten befindet. Es wird dann lediglich eine Änderung im Benutzerprofil durchgeführt, indem ein Kürzel für die ab jetzt geltende Struktur eingetragen wird. Zusätzlich wird der alte Pfad bis zum Wurzelknoten gelöscht und der neue angelegt. Durch die Verwendung von benutzerangepaßten Bäumen erreicht man eine deutliche Reduktion des Aktualisierungsaufwandes. Die dabei erreichte Verbesserung des Suchaufwandes ist relativ geringer als beim Aktualisieren und hängt stark von der Korrelation der Benutzerbewegungen und der Anrufverteilung ab. Die im folgenden Kapitel dargestellte Implementierung eines Benutzerprotokolls erlaubt eine benutzerspezifische Baumdefinition. Es soll stets beachtet werden, daß die benutzerangepaßte Verzeichnisstruktur nur dann physikalisch realisierbar ist, wenn das Transportnetz grundsätzlich von der logischen Struktur der Datenbankknoten unabhängig ist. Dies entspricht der Annahme, die im Kapitel 3.2.1 aufgestellt wurde.

Bezüglich der Abfragetechnik gilt, daß die *passing*-Technik die geringste Knoten- und Verbindungsbelastung verursacht. Diese Aussage bedarf jedoch einer Relativierung, da die Belastungsunterschiede lediglich unter den sog. sonstigen Knotenoperationen zu finden sind. Bei der Aktualisierung ist beispielsweise der sonstige Beitrag im Vergleich zu den anderen Beiträgen, die durch Lese- und Schreiboperationen verursacht werden, geringer als bei der Suchanfrage. Darüber hinaus hat die *passing*-Technik den gravierenden Nachteil, daß sie eine schlechte Fehlertoleranz aufgrund mangelnder Flußüberwachung aufweist. Die eventuellen Fehler werden erst sehr spät entdeckt und können im allgemeinen nicht lokalisiert werden. Eine wegen guter Belastungsverteilung günstige Kompromißlösung stellt die Kombination von *chaining*- und *referral*-Technik dar. Für eine praktische Implementierung ist es empfehlenswert, grundsätzlich die *passing*-Technik zu verwenden und im Falle eines Fehlers diese Prozedur z.B. durch *chaining-referral*-Technik zu ersetzen. Das bedeutet natürlich, daß beide Protokolloptionen implementiert werden müssen, was einen höheren Implementierungsaufwand erfordert.

Die Wahl des optimalen Datenverteilungsszenarios stellt die schwierigste Aufgabe dar. Aus den bisher präsentierten Varianten geht nicht hervor, daß ein bestimmtes Szenario eindeutig die beste Lösung bietet:

- Bei der Suchanfrage ist der Konzentrationspunkt, um den sich die Benutzer bewegen, das entscheidende Kriterium (siehe die Zusammenfassung im Kapitel 5.3.6).
- Bei der Aktualisierungsanfrage sind sowohl der Konzentrationspunkt als auch die Beweglichkeit der Benutzer für die Wahl des optimalen Szenarios entscheidend.

Sollte ein bestimmtes Szenario gewählt werden, so bietet das R-Szenario eine Kompromißlösung, die in einigen Situationen die beste und niemals die schlechteste ist. Wenn eine weitere Verbesserung der Leistungsfähigkeit gefordert wird, kann diese nur erreicht werden, indem das gewählte Szenario dem Mobilitätsverhalten des Benutzers angepaßt wird, d.h., das benutzte Datenverteilungsszenario wird ähnlich wie die Verzeichnisstruktur zu einer Benutzereigenschaft. Unter der Voraussetzung, daß sich das Verhalten im Vergleich zu den Benutzeraktionen (Bewegungen und Anrufen) langsam ändert, kann aus der Vergangenheit auf das zukünftige Verhalten der Benutzer geschlossen werden. Der notwendige Wechsel der verwendeten Datenverteilungsstrategie kann nur dann problemlos durchgeführt werden, wenn sich der Benutzer in seinem Heimgebiet befindet, da in diesem Fall alle Szenarien äquivalent sind. Die einzige notwendige Änderung muß in einem solchen Fall im Benutzerprofil durchgeführt werden, indem das Kürzel für die ab jetzt geltende Verteilung neu eingetragen wird. Es sind keine sonstigen Änderungen der Daten oder Zeiger notwendig. Die neue Verteilungsstrategie wird erst wirksam, wenn der Benutzer seinen Heimknoten verläßt. Selbstverständlich ist eine Änderung des Szenarios ebenfalls möglich, wenn der Benutzer sich in einem Besuchsknoten befindet, verlangt aber eine aufwendige Aktualisierung des Baumes und der Inhalte der Datenknoten. Die im nächsten Kapitel dargestellte Implementierung eines Benutzerprotokolls erlaubt eine solche benutzerspezifische Festlegung des Szenarios.

Bei der diskutierten Struktur- und Strategieanpassung soll folgendes beachtet werden:

1. Die Erfassung der charakteristischen Benutzerdaten ist nicht ein Teil des dargestellten Benutzerprotokolls und beeinflußt somit nicht die Leistungsfähigkeit dieses Protokolls. Die Benutzerdaten werden ohnehin zu statistischen Zwecken erfaßt, verarbeitet und gespeichert.
2. Da die Anpassung nur relativ selten durchgeführt wird, sind auch die damit verbundenen Aktionen für die gesamte Leistungsfähigkeit der Datenbank nicht relevant.

Aus dem obigen folgt, daß es nicht notwendig ist, eine Aufwand-Nutzen-Analyse durchzuführen, da der zusätzliche Aufwand nur aus dem obigen Punkt 2 resultiert und dadurch sehr gering sein wird.

6 Applikationsprotokoll für die UMTS-Datenbank

Die dargestellten Konzepte der UMTS-Datenbank wurden exemplarisch in einem Transportprotokoll implementiert. Dadurch wurde es möglich, den Aufwand abzuschätzen, der mit der Realisierung der Szenarien und Techniken verbunden ist.

Im folgenden werden die Anforderungen an das Protokoll, dessen Spezifikation und die praktische Realisierung vorgestellt.

6.1 Anforderungen an das Protokoll

Das Protokoll formalisiert die Beschreibung der UMTS-Benutzerprozeduren, die im Kapitel 3.3 definiert wurden. Dabei wurden sowohl die Flexibilität als auch die Zuverlässigkeit besonders beachtet.

Das Verwenden eines allgemeinen Protokolls (z.B. X.500) ist durchaus möglich, aber wegen der speziellen UMTS-Situation (vgl. Kapitel 3 und 5) nicht effizient.

6.1.1 Formale Informationsflüsse

Im Kapitel 3.3 wurden bereits alle Prozeduren und deren Zugriffe auf die Datenbank erläutert. Jeder Zugriff entspricht einer Datenbankanfrage. In Erweiterung zur numerischen Auswertung wurden im Protokoll nicht nur die zwei für die Leistungsfähigkeit wichtigsten UMTS-Datenbankanfragen, sondern alle für die Benutzerprozeduren relevanten Anfragen implementiert. Die Anfragen sind aus unabhängigen Grundelementen (*primitives*) konstruiert. Diese wiederum sind aus Nachrichten aufgebaut. Alle Nachrichten einer Anfrage ergeben den Informationsfluß dieser Anfrage. Zusammenfassend ergibt sich folgende Spezifikationshierarchie:

1. Alle betrachteten Benutzeraktionen werden als Prozeduren bezeichnet.
2. Jede Prozedur besteht aus einer oder mehreren Anfragen.
3. Jede Anfrage wird aus Grundelementen aufgebaut (vgl. Kapitel 6.2).
4. Jedes Grundelement wird als Aneinanderreihung von Nachrichten dargestellt.

Aus Platzgründen wurde hier auf die Darstellung aller implementierten Prozeduren und Anfragen verzichtet (vgl. /Schi95/). Die hier explizit wiedergegebenen Informationsflüsse beschränken sich lediglich auf die zwei numerisch ausgewerteten Anfragen, die Such- und die Aktualisierungsanfrage.

6.1.1.1 Suchanfrage

Die Suchanfrage benutzt lediglich das Auslesen-Grundelement (*retrieve*), wobei der übergebene Schlüssel der Benutzernummer des Angerufenen entspricht. Die Suche wird zuerst innerhalb des eigenen Teilgebietes durchgeführt. Im Falle des Mißerfol-

ges wird die Suche im Heimatgebiet des Angerufenen und danach im aktuellen Besuchsteilgebiet fortgesetzt (vgl. Kapitel 6.2).

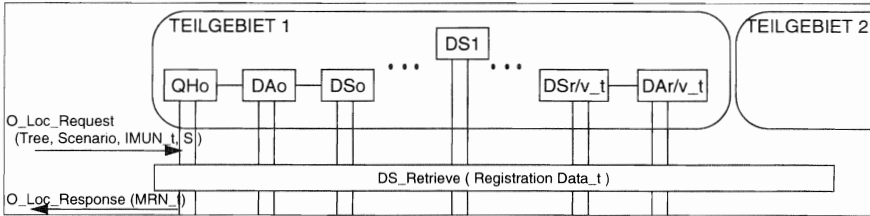


Abb. 6.1 Suchanfrage - die vertikalen Linien repräsentieren die Funktionseinheiten: QH, DA und DS; die Lage der DSx-Funktionseinheiten entspricht der Ebene im Verzeichnisbaum

6.1.1.2 Aktualisierungsanfrage

Die Aktualisierungsanfrage benutzt das Bewegen-Grundelement (*move*), das entsprechend der Benutzerbewegung entweder innerhalb eines Teilgebietes oder zwischen den Teilgebieten unterschiedlich durchgeführt wird (vgl. Kapitel 6.2).

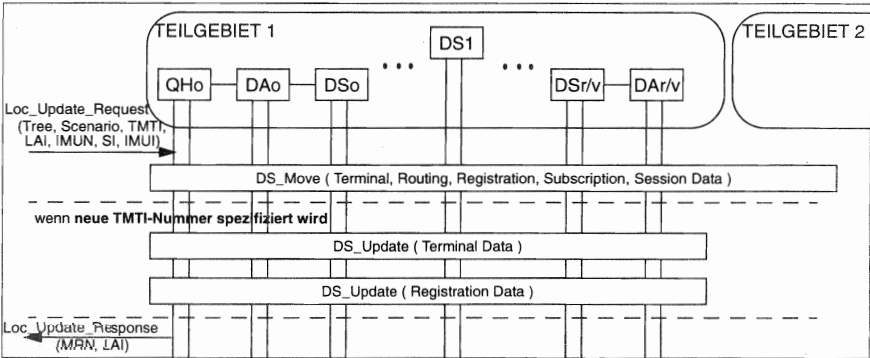


Abb. 6.2 Aktualisierungsanfrage

Wenn die Aktualisierung lokal durchgeführt wird, d.h. innerhalb eines DN-Gebietes, ist es lediglich notwendig, die Endgerätedaten zu aktualisieren, um die Zuordnung zwischen dem Endgerät und dem aktuellen Aufenthaltsgebiet wiederherzustellen.

Wenn die Benutzerdaten nicht bewegt werden, z.B. wie im ST-Szenario, so wird lediglich das *registration*-Profil mit der neuen TMTI-Nummer aktualisiert.

6.1.2 Flexibilität

Das UMTS-Konzept verlangt vom Protokoll ein hohes Maß an Flexibilität, da sehr unterschiedliche Alternativen durch das Protokoll unterstützt werden sollen:

- unterschiedliche Abfragetechniken
- unterschiedliche Datenverteilungsszenarien
- verschiedene Zeigerverwaltungsmechnismen
- verschiedene Verzeichnisstrukturen:
 - variable Baumhöhe
 - variable Anzahl der Wurzelknoten

Aufgrund der durchgeführten Auswertungen stellte sich heraus, daß die Leistungsfähigkeit der einzelnen Alternativen vom Benutzerverhalten sehr stark abhängig ist. Daraus resultiert das Bestreben, die durch das Protokoll realisierte Alternative an das Verhalten des Benutzers anzupassen und auf diese Art die Leistungsfähigkeit des Protokolls zu verbessern. Damit ist das ausgeführte Protokoll keine globale Datenbankeigenschaft mehr: Es wird zu einer Benutzereigenschaft.

Im entwickelten Protokoll wurde die Benutzermobilität als Grundlage für die Festlegung einer Protokolloption gewählt, wobei folgende zwei Merkmale als optional gewählt wurden:

- die Verzeichnisstruktur und
- das Datenverteilungsszenario.

Die Struktur des Verzeichnisses kann in einem System beispielsweise aufgrund der ohnehin statistisch erfaßten Benutzerverteilungsdaten festgelegt werden. Es ist somit keine zusätzliche Meßdatenerfassung erforderlich. Da zu erwarten ist, daß viele Benutzer aufgrund ähnlicher Bewegungsmuster gleicher Verzeichnisstrukturen bedürfen, sind sog. Verzeichnisklassen zu definieren. Jeder Benutzer besitzt dann in seiner Profilbeschreibung lediglich die Kennung der für ihn am günstigsten Verzeichnisklasse. Die Möglichkeit, verschiedene Verzeichnisklassen zu verwenden, ist nur dann gegeben, wenn der Ansatz eines universellen Transportnetzes realisiert wird (vgl. die Erklärung im Kapitel 3.2.1).

Das Datenverteilungsszenario kann ebenfalls aufgrund der erfaßten Benutzerdaten festgelegt werden. Es sei zu beachten, daß die Information darüber, welche Verteilungsstrategie anzuwenden ist, lediglich während der Aktualisierungsanfrage benötigt wird. Alle anderen Anfragen folgen den bereits angelegten Zeigern, ohne wissen zu müssen, welches Szenario aktuell verwendet wird.

Bei Verwendung von benutzerabhängigen Strukturen (oder Szenarien) ist ein Wechsel der verwendeten Struktur (oder des Szenarios) nur dann einfach möglich, wenn sich der Benutzer in seinem Heimgebiet aufhält. Der dafür notwendige Aufwand ist im Vergleich mit einem, der durch andere Prozeduren verursacht wird, sehr gering.

6.1.3 Datenintegrität und Datenzuverlässigkeit

Für die Anwendungen in verteilten Datenbanken sind Mechanismen entwickelt worden, die die Datenintegrität bei verteilten Transaktionen garantieren (vgl. Kapi-

tel 2). Zusätzlich können Sicherheitskopien in separaten Knoten angelegt werden, um die Zuverlässigkeit der Datenspeicherung zu erhöhen. Beide Methoden, Transaktionsmechanismen und Sicherheitskopien, sind von der Applikation unabhängig und werden hier deswegen nicht betrachtet.

Abgesehen von den obigen Methoden kann das Protokoll selbst die Datenintegrität unterstützen. Dies soll genauer untersucht werden. Während des Protokollablaufs können zwei prinzipielle Fehlerarten auftreten:

- die Knotenfehler und
- die Übertragungsfehler.

Die Knotenfehler in einem DN-Knoten mit Hilfe des Applikationsprotokolls zu beheben, ist eine aufwendige Aufgabe, die nur mit Hilfe von expliziten Sicherheitskopien zu bewerkstelligen ist. Diese Aufgabe wird auch durch die allgemeinen Sicherheitsprozeduren erfüllt - eine UMTS-Sonderlösung ist überflüssig.

Die Knotenfehler in einem DIR-Knoten können dagegen einfacher behoben werden, indem die Verzeichnisstruktur nachträglich so modifiziert wird, daß der fehlerhafte Knoten ausgeschlossen wird, d.h., die geltende Verzeichnisstruktur wird für alle betroffenen Benutzer geändert werden. Um dies zu realisieren, ist es notwendig, eine Suchaktion durchzuführen: Die Daten der direkt Betroffenen, d.h. der Benutzer, deren Zeiger im ausgefallenen Knoten bereits angelegt wurden, und der potentiell Betroffenen, d.h. der Benutzer, deren Zeiger bei einer Aktualisierung im ausgefallenen Knoten angelegt worden wären, müssen der neuen Verzeichnisstruktur angepaßt werden. Der Aufwand ist erheblich und verlangt deswegen eine adaptive Strategie. Es wäre beispielsweise vereinfachend möglich, die Anpassung für alle Benutzer durchzuführen, nicht nur für die direkt Betroffenen, vorausgesetzt, daß der ausgefallene Knoten rechtzeitig wieder in Betrieb genommen wird. Diese Problematik ist sehr komplex und überschreitet den Rahmen dieser Arbeit.

Im Gegensatz zu den Knotenfehlern können die Übertragungsfehler durch eine zweckmäßig gewählte Reihenfolge der Protokollnachrichten leicht behoben werden. Das implementierte Protokoll zieht dabei folgendes in Betracht:

- Die Nachrichtenfolge wird über eine Zeitüberwachung (*timer*) kontrolliert.
- Die Fehler, die während der Anfragebearbeitung in einem Knoten entdeckt werden (z.B. Inkonsistenz der gespeicherten Information), führen zu einer Fehlermeldung, die an den initiiierenden Knoten zurückgeschickt wird.
- Die Modifizierungsprozesse im Verzeichnis, d.h. Löschen und Erstellen von Dateneinträgen, werden durch den betroffenen Datenknoten kontrolliert (siehe Abbildung 6.4 und 6.6). Dabei wird die Reihenfolge der Operationen in den Knoten unterschiedlich sein:
 - Das Löschen wird im höchsten Knoten angefangen und nach unten fortgesetzt.
 - Das Erstellen wird im niedrigsten Knoten angefangen und nach oben fortgesetzt.

Dadurch wird es immer möglich sein, eine durch einen Fehler unterbrochene Nachrichtenfolge am Ende der zuvor erstellten Zeigerkette fortzusetzen.

- Das Bewegen der Daten wird in folgender Reihenfolge durchgeführt:
 1. Daten verriegeln
 2. Daten kopieren
 3. Daten löschen

Unter Beachtung der obigen Regeln ist der totale Zusammenbruch einer verteilten Transaktion selten. Eine aufwendige Wiederherstellungsverfahren braucht selten durchgeführt zu werden. Die Leistungsfähigkeit der Datenbank wird dadurch erhöht.

6.2 Protokollspezifikation

Die Anfragen werden mittels unabhängiger Grundelemente realisiert, denen prozedur- und anfragespezifische Parameter übergeben werden. Die Implementierung des gesamten Protokolls ist dadurch stark vereinfacht: Nachdem einzelne Grundelemente spezifiziert worden sind, beschränkt sich die Implementierung der einzelnen Anfragen und Prozeduren auf eine Aneinanderreihung dieser Grundelemente. Der Unterschied zwischen Aktionen innerhalb eines einzelnen Teilgebietes und zwischen den Teilgebieten wird lediglich durch einen unterschiedlichen Aufbau der einzelnen Grundelemente zum Vorschein kommen.

Das Kriterium für die Festlegung der Grundelemente ist der Zugriff auf die Datenknoten. Als unabhängig wurden folgende Grundelemente spezifiziert /Mone61/:

- Auslesen
- Erstellen
- Aktualisieren
- Löschen
- Bewegen
- Verriegeln

Die vier ersten sind bereits aus den Ein-Knoten-Datenbanken bekannt (/Voss91/) und werden hier für die verteilte Datenbank angewendet. Die zwei letzten wurden wegen der Datenverteilung eingeführt:

- Das Bewegen der Daten ist kein bloßes Aneinanderreihen von Lesen, Schreiben usw. Es verlangt eine spezielle Reihenfolge der Aktionen.
- Das Verriegeln ist notwendig, um in einer verteilten Umgebung die Datenkonsistenz während der Durchführzeit der sonstigen Grundelemente zu bewahren.

6.2.1 Grundelemente innerhalb eines Teilgebietes

Im folgenden werden die genannten Grundelemente für die Kommunikation innerhalb eines Teilgebietes detailliert beschrieben.

Die in den Nachrichtenflußdiagrammen verwendeten Indizes haben folgende Bedeutung:

- XXo - steht für den initiiierenden Knoten (*origin*), z.B. DAo für *DAta node origin*, DSo für *Directory Service origin*, QHo für *Query Handler origin*,
- XXr - steht für den Heimknoten (*resident*) und
- XXv - steht für den Besuchsknoten (*visited*).

Folgendes soll beachtet werden:

1. Die graphisch dargestellten Nachrichtenflüsse kommen unter Anwendung der *passing*-Abfragetechnik zustande. Die sonstigen Techniken sind ähnlich aufzubauen, wobei die charakteristischen Merkmale jeder Technik zu beachten sind.
2. Die Nachrichtenflüsse wurden in einer Ein-Baum-Struktur präsentiert. Der Multi-Baum-Fall unterscheidet sich von dem dargestellten durch eine zusätzliche Suche (oder Aktualisierung) unter den Knoten der Wurzelebene. Die Aufwärts- und Abwärtspfade im Verzeichnis sind mit den dargestellten identisch.
3. Der Zugriff auf den entfernten Heimknoten (oder auf den alten Besuchsknoten) wird unterschiedlich durchgeführt, je nachdem, ob die Adresse des Knotens bereits bekannt ist (sog. Alternative mit Vorwissen) oder noch nicht.
4. Für jedes einzelne Grundelement werden mehrere Fälle betrachtet. Das tatsächliche Auftreten einzelner Fälle ist von der aktuellen Konfiguration zwischen dem aktuellem Aufenthaltsort und der Speicherstelle der gesuchten Daten abhängig und wird hier nicht bis ins Detail erörtert (vgl. /Schi95/).

6.2.1.1 Auslesen

Das Auslesen der Daten erfolgt unterschiedlich in Abhängigkeit davon, ob der Heimknoten (in den ST- und R-Szenarien) oder der Besuchsknoten (im V-Szenario) oder der nähere von beiden (im RV-Szenario) gesucht wird.

In Abbildung 6.3 sind die gemeinsame Aufwärtsfolge und die drei möglichen Abwärtsfolgen nacheinander dargestellt.

6.2.1.2 Erstellen

Beim Erzeugen der Daten werden vier Fälle unterschieden, und zwar werden erstellt:

1. nur Daten im lokalen DN-Knoten,
2. Zeiger in den DIR-Knoten im Pfad zum lokalen DN-Knoten,

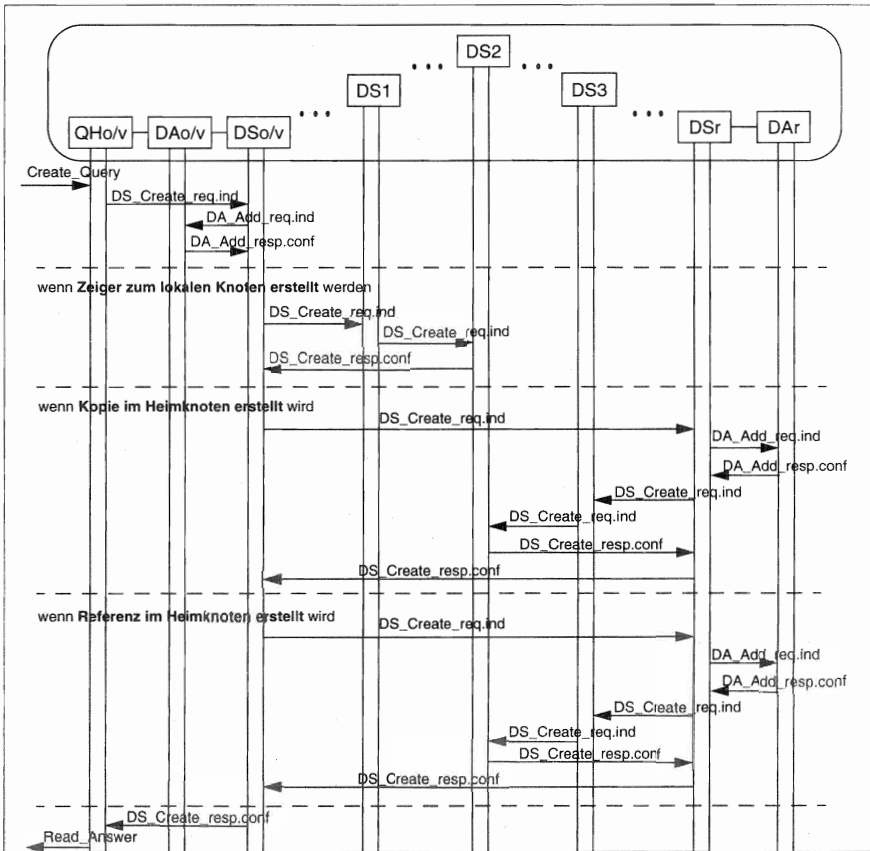


Abb. 6.4 Das Erstellen-Grundelement

6.2.1.4 Löschen

Beim Löschen der Daten werden vier Fälle unterschieden:

1. Das Löschen erfolgt nur lokal, und es existierten keine Zeiger zu den Daten.
2. Zeiger zu den Daten existierten nur lokal im DN-Knoten.
3. Zeiger in den DIR-Knoten zum lokalen DN-Knoten werden gelöscht.
4. Daten oder eine Referenz im Heimknoten werden gelöscht.

Die Abbildung 6.6 stellt die einzelnen Fälle nacheinander dar.

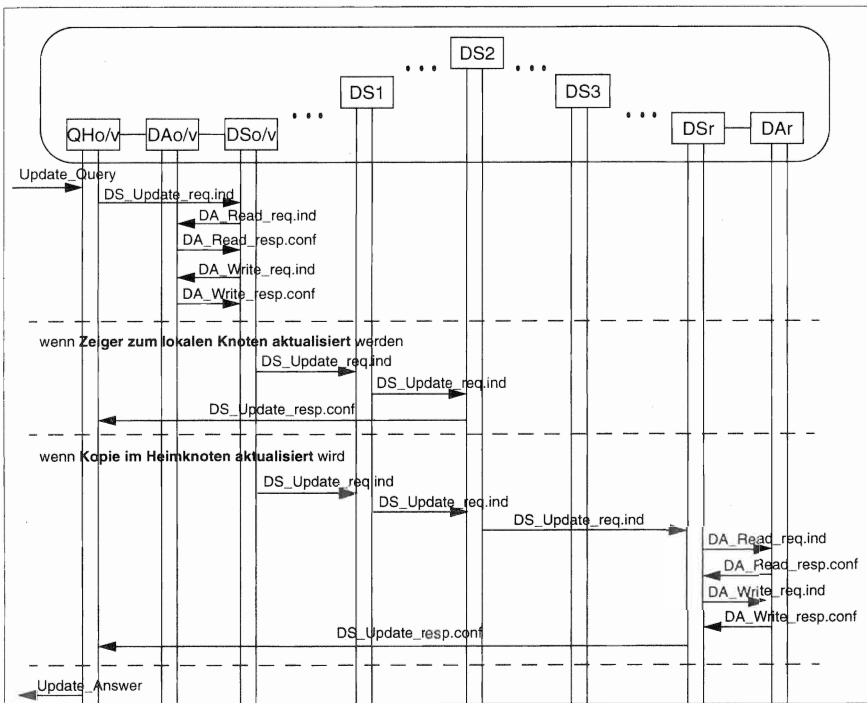


Abb. 6.5 Das Aktualisieren-Grundelement

6.2.1.5 Bewegen

Das Bewegen der Daten wird in zwei Phasen durchgeführt:

1. Die Daten werden aus dem alten DN-Knoten in den neuen kopiert.
2. Die Daten und Zeiger werden auf den aktuellen Stand gebracht.

In der 1. Phase werden zwei Fälle unterschieden:

1. Die kopierten Daten werden direkt im alten Besuchsknoten gefunden.
2. Die kopierten Daten werden indirekt im alten Besuchsknoten gefunden, d.h. mittels eines Zeigers im Heimknoten.

Die Abbildung 6.7 stellt die einzelnen Fälle nacheinander dar.

In der 2. Phase werden mehrere Fälle betrachtet, je nachdem, wie der alte und der neue Knoten zueinander in Relation stehen (Abbildung 6.8). Gegebenenfalls wird die

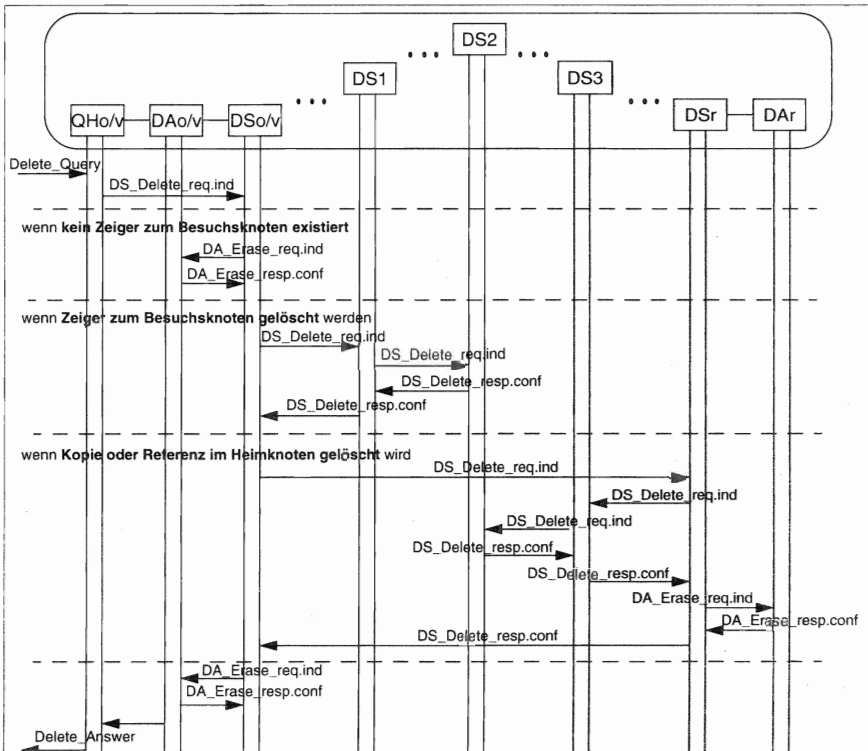


Abb. 6.6 Das Löschen-Grundelement

Referenz im Heimknoten aktualisiert und die Zeiger der neuen Datenverteilung angepasst.

6.2.1.6 Verriegeln

Das Verriegeln der Daten wird gleichzeitig mit dem Bewegen-Grundelement durchgeführt und bedarf keiner separaten Informationsflüsse.

6.2.2 Grundelemente zwischen den Teilgebieten

Die Grundelemente für die Durchführung der Prozeduren zwischen den Teilgebieten unterscheiden sich von denen innerhalb eines Teilgebietes durch das Vorhandensein zusätzlicher Nachrichten, die zwischen den einzelnen Teilgebieten ausgetauscht werden. Diese Nachrichten werden im folgenden dargestellt. Die restlichen Auf-

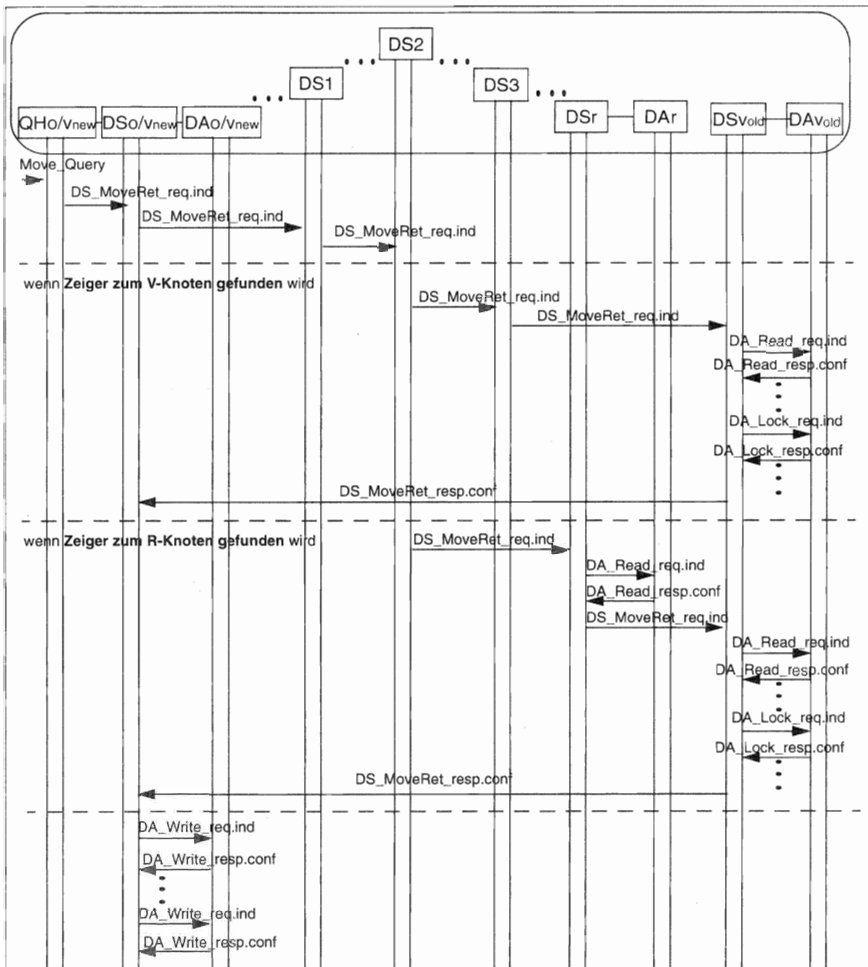


Abb. 6.7 Das Bewegen-Grundelement (1. Phase)

wärts- und Abwärtsfolgen innerhalb jedes einzelnen Teilgebietes sind mit den zuvor dargestellten identisch.

Es ist zu beachten, daß die Kommunikation zwischen den Teilgebieten immer über einen Netzübergangsknoten mit einer speziellen Funktionalität erfolgt. Hier sind es die Wurzelknoten und deren QH-Funktion. Eine direkte Kommunikation zwischen zwei DN-Knoten aus unterschiedlichen Teilgebieten ist nicht erlaubt. Dies ist ein

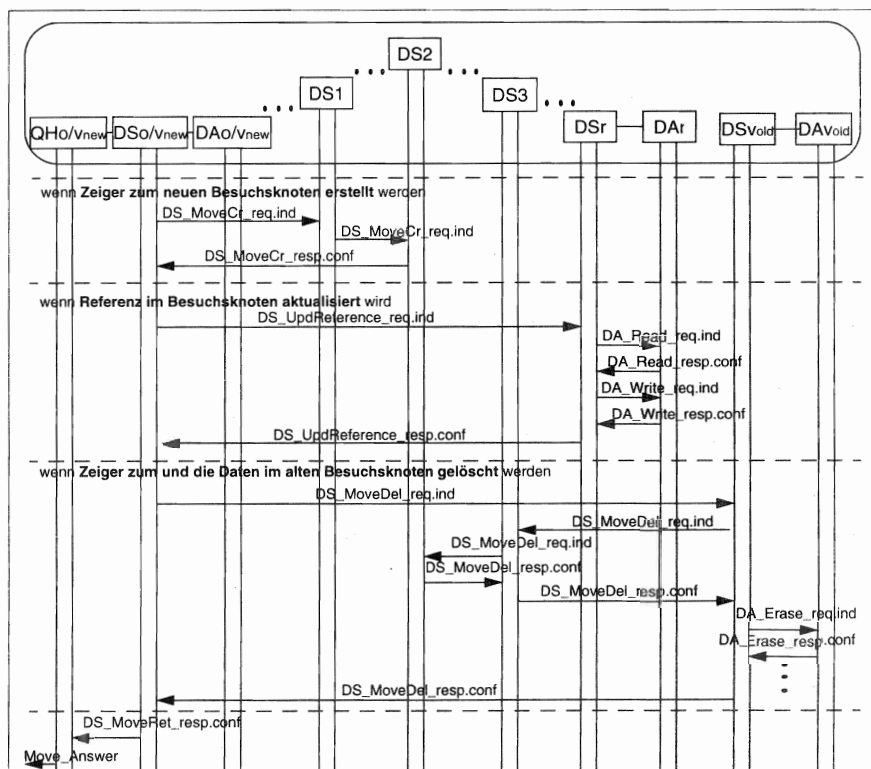


Abb. 6.8 Das Bewegen-Grundelement (2. Phase)

wichtiger Unterschied zur Kommunikation zwischen zwei DN-Knoten innerhalb einer Multi-Baum-Struktur.

Es wird im weiteren *angenommen*, daß die Kommunikation zwischen den Teilgebieten mit Hilfe der *referral*-Abfragetechnik durchgeführt wird. Im kompliziertesten Fall, d.h., wenn das aktuelle Besuchsteilgebiet unbekannt ist, wird zuerst das Heimteilgebiet angesprochen, da dieses aufgrund der Benutzernummer eindeutig identifiziert werden kann (vgl. /Mone14/). Dort wird mit Hilfe der Einträge im *routing*-Profil das gesuchte Besuchsteilgebiet ermittelt. Mit diesem Wissen wird dann im zweiten Schritt die Anfrage an das jeweilige Teilgebiet gerichtet.

6.2.2.1 Auslesen

Das in Abbildung 6.9 dargestellte Szenario stellt den kompliziertesten Fall dar: Es werden Daten eines anderen Benutzers ausgelesen, der sich außerhalb seines Heim-

teilgebietes und außerhalb des initiiierenden Gebietes aufhält. Sollten lediglich eigene Daten ausgelesen werden, so sind sie meistens im Heimatgebiet gespeichert (z.B. die *subscription*- und die *routing*-Daten) und dort direkt abrufbar.

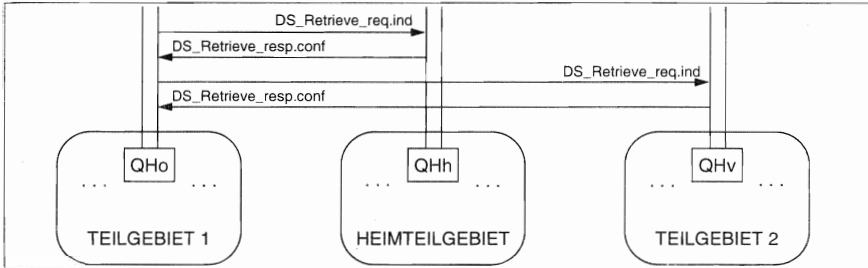


Abb. 6.9 Das Auslesen-Grundelement

6.2.2.2 Erstellen

Das Erstellen der Dateneinträge wird grundsätzlich im eigenen Teilgebiet durchgeführt. Im Heimatgebiet wird gelegentlich eine Erstellung durchgeführt, z.B. wird beim Registrieren neuer Dienste ein Eintrag im *routing*-Profil gemacht.

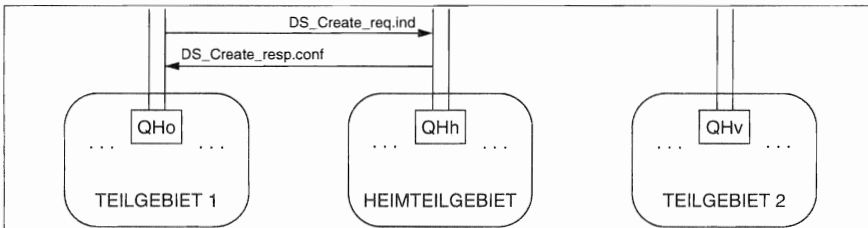


Abb. 6.10 Das Erstellen-Grundelement

6.2.2.3 Aktualisieren

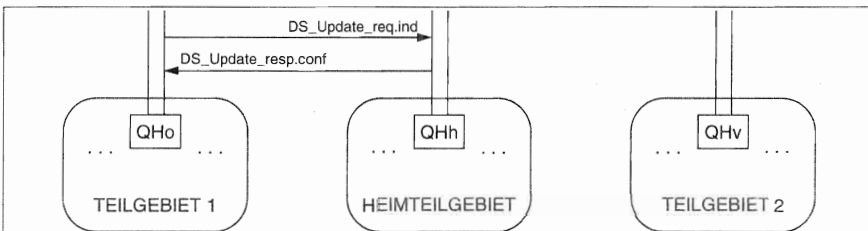


Abb. 6.11 Das Aktualisieren-Grundelement

Das Aktualisieren wird ähnlich wie das Erstellen grundsätzlich im eigenen Teilgebiet durchgeführt. Lediglich bei einigen Aktionen wird es notwendig sein, das Heimprofil zu aktualisieren (z.B. beim Ändern der Registrierung, vgl. 6.2.2.2).

6.2.2.4 Löschen

Das Löschen wird grundsätzlich im eigenen Teilgebiet durchgeführt. Lediglich bei einigen Aktionen wird es notwendig sein, das Heimprofil zu löschen (z.B. beim Austragen nicht mehr benutzter Dienste, vgl. 6.2.2.2).

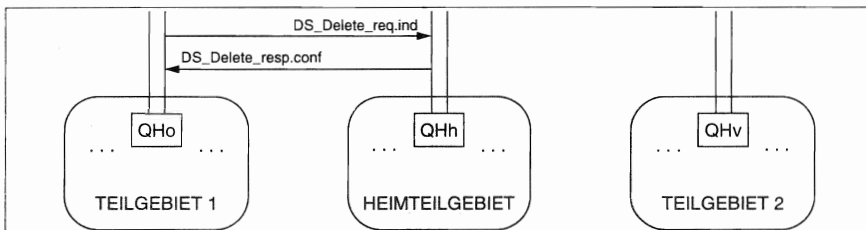


Abb. 6.12 Das Löschen-Grundelement

6.2.2.5 Bewegen

Das Bewegen der Daten wird grundsätzlich zwischen dem alten und dem neuen Teilgebiet durchgeführt. Im allgemeinen wird es nicht möglich sein, das alte Teilgebiet direkt zu ermitteln. In solchem Fall muß zuerst das Heimteilgebiet gefragt werden.

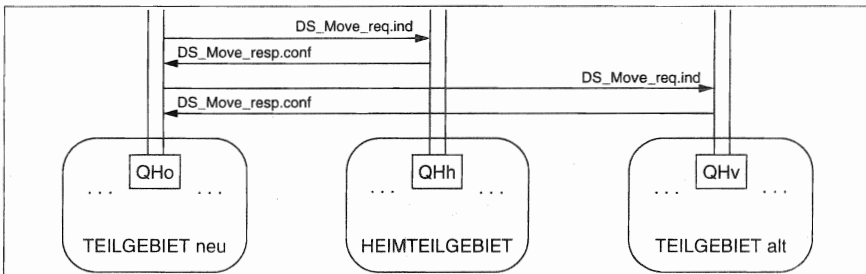


Abb. 6.13 Das Bewegen-Grundelement

6.3 Protokollimplementierung

Im folgenden wird ein kurzer Überblick über die realisierte Implementierung des UMTS-Datenprotokolls gegeben. Zuerst wird die gewählte formale Beschreibungssprache dargestellt und danach die Struktur der erstellten Software erläutert.

6.3.1 SDL-Sprache

Die Abkürzung SDL steht für *Specification and Description Language*, eine standardisierte Sprache, die durch das frühere CCITT (heute ITU) im Jahre 1988 endgültig spezifiziert wurde (vgl. /Z.100/). Die Sprache ermöglicht eine exakte Beschreibung der verteilten Telekommunikationssysteme und deren Schnittstellen, die als eine Ansammlung von SDL-Prozessen modelliert werden, die mittels ausgetauschter Signale miteinander kommunizieren /Hogr89/.

Für eine systematische Beschreibung eines Systems und eine bessere Übersicht werden die einzelnen SDL-Prozesse zu Blöcken zusammengefaßt, die wiederum rekursiv in Blöcke gruppiert werden können. Auf der obersten Hierarchieebene steht dann das gesamte System, das sich aus mehreren Blöcken zusammensetzt, siehe Abbildung 6.14. In /Sara89/ und /Beli91/ findet man eine gute Übersicht der möglichen Anwendungen und eine benutzerfreundliche Beschreibung.

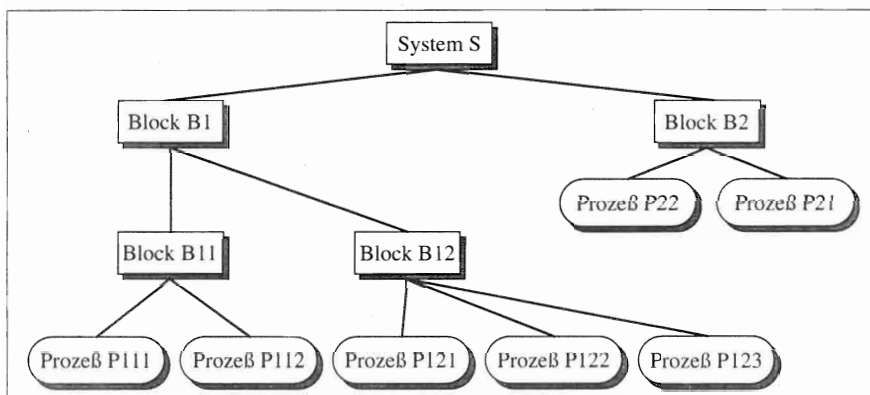


Abb. 6.14 Hierarchische Struktur einer SDL-Spezifikation

Um mit der SDL-Sprache arbeiten zu können, wurden verschiedene Tools entwickelt. Für die vorliegende Implementierung wurde das SDT™ (SDL-Tool) verwendet - ein graphisches Werkzeug, das eine Erstellung von SDL-Spezifikationen und deren syntaktische und semantische Analyse erlaubt. Darüber hinaus erlaubt das SDT mit Hilfe des C-Code-Generators, ein lauffähiges Programm zu erstellen und dadurch die entwickelte Systemspezifikation zu verifizieren.

6.3.2 Funktionsblöcke des UMTS-Datenprotokolls

Die generelle Struktur des spezifizierten Protokolls wird anhand der Abbildung 6.15 erläutert. Auf der Systemebene wurden 5 Blöcke definiert:

1. QH - Anfrageverwalter oder Anfragenmanager (*query handler*)
2. DS - entspricht der DIR-Funktion

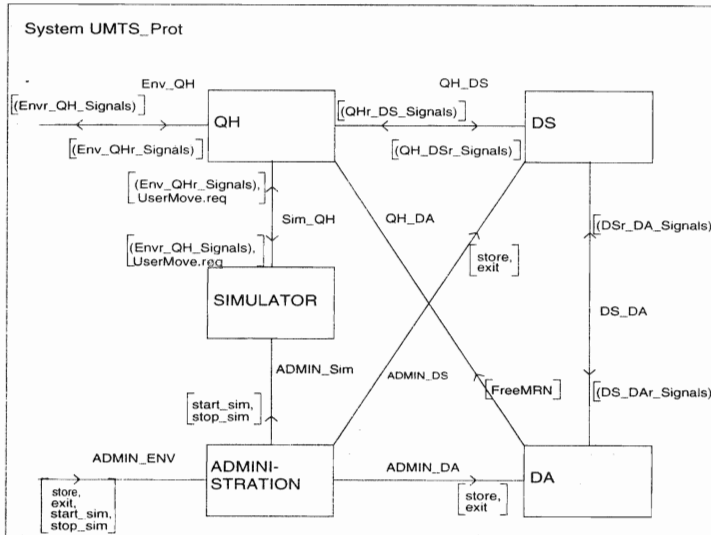


Abb. 6.15 UMTS-Datenbankprotokoll - Systemebene

3. DA - entspricht der Datenspeicher-Funktion
4. Administration
5. Simulator

Die drei ersten Blöcke repräsentieren die bereits bekannten Funktionseinheiten der UMTS-Datenbank (vgl. Kapitel 3.2.2). Die zwei letzten haben die Aufgabe, die Validierung des Protokolls zu ermöglichen: ein Vermittlungsteilnetz wird durch den Simulator modelliert und die Überwachung durch die Administration bewerkstelligt.

Die Realisierung der einzelnen UMTS-Funktionsblöcke wird im folgenden genauer präsentiert.

6.3.2.1 Anfragemanager (QH)

Der Anfragemanager wurde mit Hilfe eines SDL-Prozesses realisiert. Seine Aufgabe ist es, die Anfragen, die vom Simulator kommen, in einzelne Grundelemente zu zerlegen und gemäß dem Protokoll an die DS-Funktion des eigenen DN-Knotens zu schicken.

Es soll hier beachtet werden, daß die physikalischen Aspekte, d.h. die Knoten und Verbindungen der Datenbank, nicht direkt sichtbar sind:

- Die Knoten setzen sich aus einzelnen Funktionen zusammen.

- Die Verbindungselemente kommen gar nicht erst vor, da sie im UMTS-Ansatz durch eine unabhängige Infrastruktur zur Verfügung gestellt werden und somit hier nicht separat modelliert werden müssen.

6.3.2.2 Datenverzeichnis (DS)

Das Datenverzeichnis wurde mittels mehrerer SDL-Prozesse realisiert, die jeweils für unterschiedliche Grundelemente verantwortlich sind. Diese Aufteilung wurde aus Komplexitätsgründen notwendig, um die Generierungszeiten des SDT-C-Code-Generators zu reduzieren.

Die DS-Funktionalität ist verantwortlich für eine korrekte Durchführung der Protokollaufgaben gemäß der gewählten Technik, dem Szenario und der Verzeichnisstruktur. In Abbildung 6.16 ist der DS-Block dargestellt.

6.3.2.3 Datenspeicher (DA)

Der Datenspeicher wurde ebenfalls wie der QH-Anfragemanager mit Hilfe eines SDL-Prozesses realisiert. Seine Aufgabe ist es, die Benutzer- und Endgerätedaten zu speichern. In der Implementierung wurde eine Schnittstelle zwischen den DA-Funktionen und dem Speicher festgelegt, um auf diese Weise das Protokoll vom jeweiligen Speichermedium unabhängig zu machen. Beispielsweise ist in der implementierten Protokollversion der Datenspeicher als eine lokale SDL-Struktur realisiert. Die Anwendung von einem Unix-Dateiensystem würde lediglich eine Anpassung der Datenzugriffsfunktionen verlangen, ohne die DA-Funktionen zu beeinflussen.

6.3.2.4 Administration und Simulator

Die Administration hat die Aufgabe, den Ablauf des Validierungs- oder Simulationsprozesses zu überwachen. Dabei werden solche Aspekte, wie der Anfang und das Ende der Simulation sowie die Datenspeicherung in den einzelnen Unterfunktionen des Protokolls kontrolliert.

Der Simulator dagegen übernimmt die Aufgabe, das Verhalten des Vermittlungsteilnetzes zu simulieren. Im einzelnen werden hier die Benutzerverhaltensmodelle repräsentiert und ausgewertet sowie die Aktionen der Benutzer initiiert, die mittels eines endlichen Automaten modelliert werden (/Schi95/).

6.4 Zusammenfassung

Die durchgeführte Implementierung erlaubt es, Einsichten in die Funktionsweise eines realen UMTS-Datenbankprotokolls zu gewinnen und dessen Komplexität zu bewerten. Dadurch ist es möglich geworden, den Aufwand abzuschätzen, der

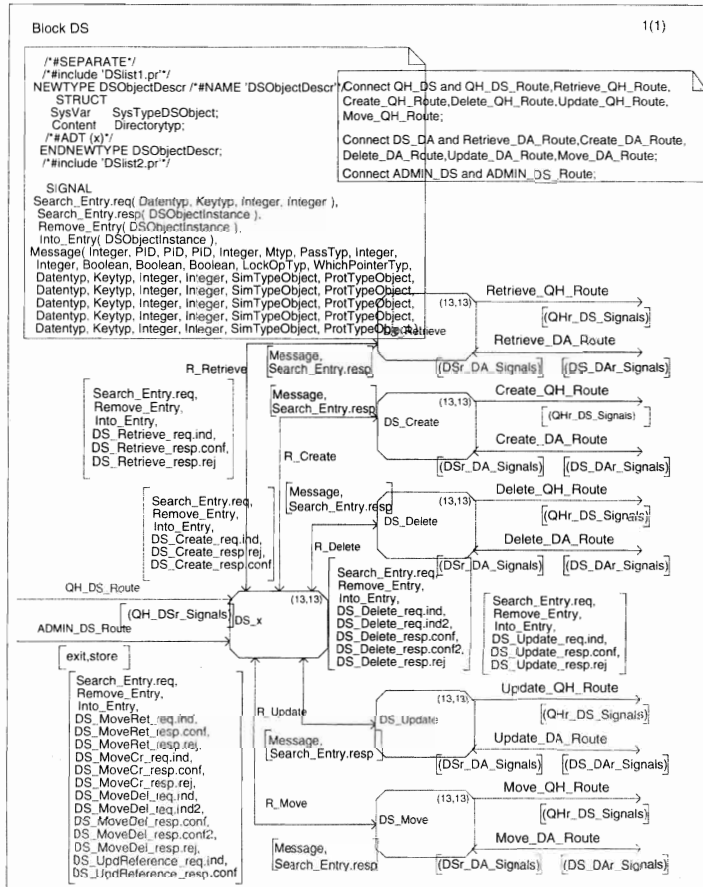


Abb. 6.16 DS-Block

programmtechnisch nötig sein wird, um die einzelnen Techniken, Szenarien und Strukturen der UMTS-Datenbank zu realisieren.

Folgende Schlußfolgerungen lassen sich aus der gewonnenen Erfahrung herleiten:

1. Der Programmieraufwand zur Unterstützung von unterschiedlichen Szenarien differiert, d.h., die Komplexität der einzelnen Szenarien ist vom programmtechnischen Standpunkt her unterschiedlich ($ST < R < V < RV$). Dies kann durch folgende Beobachtungen begründet werden:

- Die Überwachung der Zeigerkonsistenz in den Verzeichnisknoten bei den dynamischen Szenarien (V und RV) verlangt einen Zusatzaufwand.

- Die Überwachung der Datenkopien und Referenzen in den Datenknoten verursacht ebenfalls einen Zusatzaufwand (R und RV).
2. Die Abfragetechniken verlangen lediglich einen geringfügig unterschiedlichen Aufwand.
- Folgende Beobachtungen führen zu diesem Ergebnis:
- Die Entscheidungslogik bezüglich der Daten und Zeiger ist in jeder Technik identisch.
 - Die einzigen Unterschiede ergeben sich aufgrund
 - der unterschiedlichen Handhabung von *Follow-Up*-Nachrichten und
 - der Notwendigkeit, beim Verketteten der Nachrichtenfolge (*chaining*) auf die zugehörige Antwort zu warten.
3. Der Aufwand, ein Protokoll zu implementieren, das eine beliebige Verzeichnisstruktur unterstützt, ist naturgemäß höher im Vergleich zu einem, das lediglich eine Struktur erlaubt. Die Flexibilität des Verzeichnisbaumes ist aber eine notwendige Voraussetzung für die Verwendbarkeit des UMTS-Protokolls und muß deswegen unterstützt werden.
- Auf dieser Grundlage stellte es sich als sehr einfach heraus, den nächsten Schritt zu vollziehen und eine benutzerspezifische Struktur zu definieren. In der Implementierung ist man davon ausgegangen, daß es nur einige Benutzerklassen geben wird, die unterschiedliche Verzeichnisstrukturen verlangen werden. Deswegen ist der einfachere Weg eingeschlagen worden: Es werden einige Strukturen vordefiniert (die maximale Anzahl ist theoretisch unbegrenzt), und die einzelnen Benutzer werden lediglich einer dieser Strukturen zugeordnet. Bei diesem Ansatz muß jeder Verzeichnisknoten die Information darüber enthalten, wie seine Umgebung in jeder Klasse aussieht. Jeder Datenknoten muß die komplette Information darüber besitzen, wie jede Klasse aufgebaut ist, um die Benutzer entsprechend einordnen zu können.
- Der umgekehrte Ansatz würde verlangen, daß jeder Benutzer die Information darüber besitzt, wie seine Verzeichnisstruktur auszusehen hat. Der dafür notwendige Speicheraufwand ist wegen der Existenz von lediglich einigen Varianten der Struktur nicht gerechtfertigt.
4. Der Programmieraufwand, der sich aufgrund der Szenarioflexibilität ergibt, ist sehr gering. Jeder Benutzer wird einem der vier Szenarien zugeordnet und enthält in seinem Profil lediglich eine Kennung des aktuell für ihn gültigen Szenarios. Bei der Aktualisierung werden die Zeiger und Daten entsprechend den Erfordernissen des jeweiligen Szenarios angelegt. Bei allen anderen Anfragen wird lediglich den Zeigern gefolgt, ohne die Besonderheiten der jeweiligen Datenverteilung zur Kenntnis zu nehmen.
- Es ist natürlich notwendig, daß alle Szenarien implementiert werden: dies verursacht offensichtlich einen Zusatzaufwand, der jedoch aufgrund der erwarteten Vorteile durchaus zu vertreten ist.

7 Schlußbemerkung und Ausblick

Die in der Arbeit untersuchten Datenbankkonzepte zur Verwaltung der Benutzerdaten in UMTS geben Aufschluß darüber, in welchen Situationen bestimmte Alternativen besser oder schlechter sind. Die Untersuchung wurde basierend auf Datenbankankfragen durchgeführt, die erstens in ihrer Leistungsfähigkeit entscheidend durch die Verzeichnisstruktur und durch die Datenverwaltungsalternativen beeinflußt werden und zweitens einen entscheidenden Beitrag zur Gesamtnetzbelastung liefern (Such- und Aktualisierungsanfrage).

Die im Kapitel 3 dargestellten Alternativen der Verzeichnisstrukturen, Zeigeroptionen, Abfragetechniken und Datenverteilung bieten eine systematische Beschreibung von Einflußfaktoren, die die Leistungsfähigkeit der künftigen UMTS-Datenbank entscheidend bestimmen werden. Eine derartig umfassende Ausarbeitung wurde erstmalig in dieser Arbeit durchgeführt. Darüber hinaus stellen die beschriebenen Alternativen den Ausgangspunkt für die numerische Analyse dar.

Um die Relationen zwischen dem Benutzerverhalten und der Netzbelastung zu quantifizieren, wurde ein formales Modell entwickelt. Es vermag mit Hilfe einiger Parameter das Mobilitäts- und Anrufverhalten der Benutzer zu beschreiben. Die dabei getroffenen Annahmen stellen ein vereinfachtes Bild der Wirklichkeit dar, bieten jedoch genügend Flexibilität, um ein breites Spektrum des Benutzerverhaltens zu modellieren. Im einzelnen ist es möglich, sowohl stationäre als auch nichtstationäre Bewegungen zu beschreiben und die Anrufverteilung detailliert zu spezifizieren.

Die gewählten Datenbankankfragen wurden formal beschrieben, indem die möglichen Konstellationen innerhalb eines Teilgebietes und zwischen den Teilgebieten erfaßt wurden. Dabei ist die Wahrscheinlichkeit jeder Datenbankaktion zwischen zwei beliebigen DN-Knoten formelmäßig erfaßt. Das Wissen, zwischen welchen zwei DN-Knoten und mit welcher Häufigkeit eine Aktion durchgeführt wird, ist entscheidend, um den Aufwand innerhalb der Datenbank zu bewerten.

Beide, das Benutzermodell und die formale Beschreibung der Anfragen, stellen die theoretische Grundlage der quantitativen Auswertung dar.

Unabhängig von den Wahrscheinlichkeitsbetrachtungen wurde das Datenbankkommunikationsprotokoll für die untersuchten Anfragen spezifiziert. Auf der gegebenen mathematischen Grundlage und unter Verwendung des Protokolls wurde letztlich mittels Matlab™ ein Programm implementiert, das die numerische Auswertung erlaubt. Dabei wurden solche Kenngrößen wie Anfragen in den einzelnen Knoten, die Menge der transportierten Informationen und die Verzögerungen ermittelt, die wiederum zur Bewertung der einzelnen Alternativen verwendet wurden.

Die präsentierte Leistungsanalyse der Knotenbelastung beschränkt sich darauf, die Anzahl der Anfragen im Knoten zu vergleichen. Dabei wird lediglich unterschieden, ob es sich um Lese-, Schreib- oder sonstige Operationen handelt. Der Grund für die-

se Abstraktion ist die Tatsache, daß die Anzahl der Anfragen ein technologieunabhängiges Kriterium der Knotenbelastung ist. Ähnlich wurde bei der Bewertung der Verbindungsbelastung von der Betrachtung der einzelnen Verbindungen abstrahiert und nur die Summe aller transportierten Nachrichten betrachtet, eventuell mit der Entfernung gewichtet. Die Verzögerungen, die die Anfragen in der Datenbank erfahren, wurden kaum beachtet, da wegen der zu erwartenden Größenordnung die Differenzen zwischen den einzelnen Alternativen eine untergeordnete Rolle spielen.

Die Ergebnisse der Auswertung werden hier nicht wiederholt, da sie zusammenfassend am Ende des Kapitels 5 präsentiert wurden. Es soll jedoch nicht unerwähnt bleiben, daß hier zwei Neuerungen gegenüber den existierenden Protokollansätzen vorgeschlagen wurden, um die Leistungsfähigkeit der Datenbank zu erhöhen:

- eine benutzerangepaßte Verzeichnisstruktur und
- ein benutzerangepaßtes Datenverteilungsszenario.

Zusätzlich zu der Analyse der Leistungsfähigkeit wurde im Rahmen der Arbeit eine praktische Implementierung des spezifizierten Protokolls mittels SDT™ durchgeführt und verifiziert. Das Protokoll wurde systematisch aufgebaut, d.h., die Benutzerprozeduren wurden in Anfragen zerlegt und die Anfragen wiederum in Grundelemente. Damit ist die Implementierung der Prozeduren auf eine Aneinanderreihung von Grundelementen reduziert worden.

Das Ziel der Implementierung war, den Aufwand zu bewerten, der durch die Anwendung von verschiedenen Datenverteilungsszenarien und Abfragetechniken in den Datenbankknoten entsteht. Die dabei gewonnene Erfahrung erlaubt die Folgerung, daß die programmtechnischen Unterschiede zwischen den einzelnen Szenarien und Abfragetechniken von untergeordneter Bedeutung sind und nicht in Betracht gezogen werden brauchen. Außerdem stellte sich heraus, daß die Verwendung von benutzerangepaßten Bäumen und Verteilungsszenarien programmtechnisch sehr einfach zu realisieren ist und keine grundsätzlichen Schwierigkeiten verursacht.

Die an einigen Stellen der Arbeit erwähnte Erweiterung des UMTS-Datenbankkonzeptes, die eine ausgereifte Personalisierung der dem Benutzer zur Verfügung stehenden Dienste erlaubt, stellt ein interessantes Gebiet für weitergehende Studien dar. Hierbei gilt es, die Abhängigkeit zwischen der Mobilität und der Personalisierung zu erfassen. Die Leistungsbewertung eines solchen Systems würde sicherlich interessante Ergebnisse im Vergleich zu den in der Arbeit vorliegenden liefern.

Eine andere Weiterentwicklung bietet die vollständige Implementierung des Kommunikationsprotokolls für die UMTS-Datenbank. Diese Entwicklung kann über das Stadium der Simulation hinausgehen und sich sowohl mit den realen Daten der Benutzer und der Endgeräte befassen als auch mit einer realen Datenverteilung im Netz.

Letztendlich könnte eine Kombination beider obigen Aspekte angestrebt werden, um eine Lösung der Datenhaltungsprobleme für künftige mobile und personalisierte Dienste zu erhalten.

8 Literaturverzeichnis

- /Alna94/ L. Alnas, P. T. Fjuk, S. Wallinder, *The Current Status of Universal Personal Telecommunication (UPT)*. Proc. of 3rd International Conference on Intelligence in Networks, pp. 301-304, Bordeaux, Oktober 1994.
- /Arms94/ G. Armstrong et. al., *An Analysis of Protocols for the UMTS Distributed Database*. Proc. of RACE Mobile Telecommunications Workshop, pp. 455-459, Amsterdam, Mai 1994.
- /Ash93/ C. Ash, *The probability tutoring book*. IEEE Press, 1993.
- /Awer91/ B. Awerbuch and D. Peleg, *Concurrent Online Tracking of Mobile Users*. Proc. ACM SIGCOMM '91, Conference on Communications, Architectures and Protocols, pp. 221-233, Zürich, September 1991.
- /Åker92/ D. Åkerberg, *Novel Radio Access Principles Useful for Third Generation Mobile Radio Systems*. Proc. of IEEE PIMRC, Boston, pp. 1.2.1.1-9, Oktober 1992.
- /Beli91/ F. Belina, D. Hogrefe and A. Sarma, *SDL with Applications from Protocol Specification*. Prentice Hall International Ltd, 1991.
- /Bell92/ D. Bell, *Distributed Database Systems*, University of Ulster, Addison-Wesley Publishing Company, 1992.
- /Bell93/ Bellcore, *Service Management System (SMS)/800 Mechanized Generic Interface Specification*, in TA-NWT-001247, Juni 1993.
- /Bern87/ P. A. Bernstein, *Concurrency Control and Recovery in Database Systems*. Addison-Wesley Publishing Company, 1987.
- /Boer93/ H. de Boer, *RACE Mobile Communications*. Electronics & Communication Engineering Journal, pp. 157-158, Juni 1993.
- /Boyl90/ S.M. Boyles, et. al., *Common Channel Signalling: The Nexus of an Advanced Communications Network*. IEEE Communications Magazine, Vol. 28, No. 7, pp. 57-63, Juli 1990.
- /Bolc91/ G. Bolch, *Leistungsbewertung von Rechensystemen*, B. G. Teubner Stuttgart, 1991.
- /Bowe91/ T. F. Bowen, et. al., *A Scalable Database Architecture for Network Services*. IEEE Communications Magazine, Vol. 29, No. 1, pp. 52-59, Januar 1991.

- /Bowe92/ T.F. Bowen, *The Datacycle Architecture*. Communications of the ACM, Vol. 35, No. 12, pp. 71-81, Dezember 1992.
- /Broe93/ W. van den Broek, et al., *Functional models of UMTS and integration into future networks*. Electronics & Communication Engineering Journal, pp. 165-172, Juni 1993.
- /Broe94/ W. van den Broek, K. Georgokitsos, *Impact of UMTS developments on IN standardisation*. Proc. of the RACE Mobile Telecommunications Workshop, pp. 295-299, Amsterdam, Mai 1994.
- /Call94/ M. H. Callendar, *Future Public Land Mobile Telecommunication Systems (FPLMTS)*. IEEE Personal Communications, Vol. 1, No. 4, pp. 18-22, Fourth Quarter 1994.
- /Carr94/ L. Carrasco, C. Eynard, et. al., *BSS Architecture: from Applications to Physical Entities*. Proc. of the RACE Mobile Telecommunications Workshop, pp. 91-97, Amsterdam, Mai 1994.
- /Casa94/ V. Casares, et al., *Evaluation studies of DDB Architecture with hierarchical Structure*. Proc. of the 3rd ICUPC, pp. 345-349, San Diego, September 1994.
- /Chan93/ D. Chan, et.al., *Network Architecture for Personal Mobility*. Proc. of the 4th International Symp. Indoor & Mobile Radio Communications, Yokohama, September 1993.
- /Chen93/ S. Chen, M. Fujioka, et. al., *Intelligent Networking for the Global Marketplace*, IEEE Communications Magazine, Vol. 31, No. 3, pp. 86-92, März 1993.
- /Chia91/ S.T.S. Chia, *Handover Techniques for a Third Generation Mobile System*. Proc. of Mobile Radio Conference, pp.243-249, Nice, November 1991.
- /Chia92/ S. Chia, *The Universal Mobile Telecommunication System*. IEEE Communications Magazine, Vol. 30, No. 12, pp. 54-62, Dezember 1992.
- /Chon94/ Hing Fai (Lousi) Chong, *UPT Implementation and Capabilities in the GTE IN Testbed*. Proc. of 3rd International Conference on Intelligence in Networks, pp. 296-300, Bordeaux, Oktober 1994.
- /Come88/ D. Comer, *Interworking with TCP/IP: Principles, Protocols and Architecture*. Prentice-Hall, 1988.
- /CS-2/ *Intelligent Network Capability Set 2. Interworking between IN-structured Networks for CS-2*, ETSI Report NA 603-01, April 1994.

- /CS-3/ *CS2/CS3 Baseline Documents*. ITU, Study Group 11, Working Party 4, Washington, 21-30 Juni 1994.
- /Date85/ C.J. Date, *An Introduction to Database* - Vol. II. Addison- Wesley Publishing Company, 1985.
- /Date91/ C.J. Date, *An Introduction to Database* - Vol. I. Addison- Wesley Publishing Company, 1991.
- /Dean94/ R. A. Dean, *Satellite/Terrestrial Interoperation in Personal Communication Networks*. Proc. of the 3rd ICUPC, pp. 450-454, San Diego, September 1994.
- /DECT92/ *Digital European Cordless Telephone (DECT) - Common Interface*, ETSI ETS 300 175-1 to 9, 1992.
- /Dini94/ M. Dinis, et al., *A Satellite Network Architecture for a MEO Constellation in a Global Personal Communication System Scenario*. 3rd ICUPC pp. 402-406, San Diego, September 1994.
- /EIA91/ *Cellular Radiotelecommunications Intersystem Operations: Functional Overview IS-41.1* (Revision B). EIA/TIA Project Number 2078, Juli 1991.
- /Fara94/ F. F. Faraci, et al., *IN-Based Approach for the Provision of PCS Services Using PSTN Infrastructures*. Proc. of 3rd International Conference on Intelligence in Networks, pp. 326-331, Bordeaux, Oktober 1994.
- /Fort85/ P. J. Fortier, *Design and Analysis of Distributed Real-Time Systems*. [ntertext Publications Inc., McGraw-Hill, 1985.
- /Fuji90/ M. Fujioka, et al., *Hierarchical and Distributed Information Handling for UPT*. IEEE Network Magazine, Vol. 4, No. 6, pp. 61-69, November 1990.
- /Fuhr90/ W. Fuhrmann, V. Brass, *Performance Evaluation of a GSM Public Land Mobile Network*. IEEE Magazines, Special Issue on PCS, pp. 295-302, 1990.
- /Garr93/ J. J. Garrahan, P. A. Russo, et. al., *Intelligent Network Overview*. IEEE Communications Magazine, Vol. 31, No. 3, pp. 30-36, März 1993.
- /Gius91/ P. P. Giusto, *Towards Personal Communications*. CSELT Technical reports, Vol. 19, No. 1, pp. 1-6, Februar 1991.
- /Gran93/ M. Grant, M. Garner, *Personal Numbering Services: The Business Opportunity for UPT*. Ovum report, Ovum Ltd., 1993.

- /Gril92/ D. Grillo, S.T.S. Chia, *UMTS - mobile communications beyond the Year 2000*. Electronics & Communication Engineering Journal, pp. 331-340, Oktober 1992.
- /Groe94/ W. Groenen, *GSM und danach - Digitale zellulare Mobilfunksysteme auf dem Weg zu globalen Netzen*. Mobile Kommunikationssysteme - 8. Aachener Kolloquium Signaltheorie, pp. 309-316, Aachen, März 1994.
- /Gros85/ D. Gross, C. M. Harris, *Fundamentals of Queueing Theory*. John Wiley & Sons, 1985.
- /GSM-Spec/ *GSM Technical Specifications*, Phase 1 und Phase 2, ETSI, 1988-92.
- /Gunt93/ M. Guntermann, et al., *Integration of Advanced Communication Services in the PSCS - a Realisation Study*. Int. Conference on Intelligence in Broadband Services and Networks (IS&N '93), pp. II/1, Paris, November 93.
- /Hans92/ S. Hansen, *The Standardization of UMTS in ETSI SMG5*. 5th Nordic Seminar on Digital Mobile Radio Communication, pp. 185-194, Helsinki, Dezember 1992.
- /Hans94/ M. S. Hansen, *Intelligent Network Support for Personal Communication Services (PCS) Access Services*. Proc. of 3rd International Conference on Intelligence in Networks, pp. 304-307, Bordeaux, Oktober 1994.
- /Hato90/ T. Hattori et al., *Personal Communications - Concept and Architecture*. Proc. of the IEEE International Communication Conf., pp. 1351-1357, Atlanta, April 1990.
- /Hogr89/ D. Hogrefe, *Estelle, Lotus and SDL*, Springer Verlag Berlin, 1989.
- /Horn94/ G. Horn, *Security Aspects of New UMTS Features - Requirements and Solutions*. Proc. of the RACE Mobile Telecommunications Workshop, pp. 350-355, Amsterdam, Mai 1994.
- /Ioan92/ J. Ioannidis, et. al., *Protocols for Mobile Internetworking*. IETF Mobile Working Group, Juni 1992.
- /IN95/ *IN/ UMTS Framework Document - Version 6*. ETSI NA-61301, Januar 1995.
- /Isbe94/ J. Isberg, M.C. Ciancetta, *Combined personal and terminal mobility in IN environment*. Proc. of 3rd International Conference on Intelligence in Networks, pp. 287-292, Bordeaux, Oktober 1994.

- /IS-54/ *Cellular System Dual-Mode Mobile Station-Base Station Compatibility Standard*. TIA/EIA IS-54, Telecommunications Industry Association, April 1992.
- /Jain91/ R. Jain, *The Art of Computer System Performance Analysis*. John Wiley & Sons, 1991.
- /Jans90/ H. Jansson, J.S. Swerup, S. Wallinder, *The Future of Cellular Telephony*. Ericsson Review, pp. 42-52, No. 1, 1990.
- /John94/ J.S. Johnson, C. Nixon, *Personal Number Services Using the AIN - Canada*. 3rd International Conference on Intelligence in Networks, pp. 293-296, Bordeaux, Oktober 1994.
- /Keat94/ J-P. Keaton, A. Saidi, I. Baccaro, *A UMTS Network Architecture*. Proc. of the RACE Mobile Telecommunications Workshop, pp. 285 - 290, Amsterdam, Mai 1994.
- /Klein76/ L. Kleinrock, *Queueing Systems* - Vol. 1 und 2. Wiley & Sons, 1976.
- /Kluß93/ N. Klußmann, *Implementierung eines flexiblen Datenbankkonzepts für personalisierte Telekommunikationsdienste*. Diplomarbeit, Lehrstuhl für Kommunikationsnetze RWTH Aachen, September 1993.
- /Kort91/ H. F. Korth, A. Silberschatz, *Database System Concepts* - 2nd edition. University of Texas at Austin, McGraw - Hill Inc., 1991.
- /Kung91/ R. Kung, *Rationale for Intelligent Nertwork*. Proc. of the IFIP TC6/WG6.4 International Workshop on Open Distributed Processing, pp. 69-80, Berlin, Oktober 1991.
- /Laih92/ K. Laihonen, *A Conceptual Model of UMTS Service Provision and Use*. 5th Nordic Seminar on Digital Mobile Radio Communication, pp. 195-202, Helsinki, Dezember 1992
- /Laue94/ G. S. Lauer, *IN Architectures for Implementing Universal Personal Telecommunications*. IEEE Network, Vol. 8 No. 2, pp. 6-16, März/April 1994.
- /Lawn94a/ J. Ericsson, D. Lawniczak, J. Dahlström, *Evaluation of Directory Node Architectures for the DDB Network in UMTS*. Proc. of RACE Mobile Telecommunications Workshop, pp. 459-462, Amsterdam, Mai 1994.
- /Lawn94b/ E. Geulen, D. Lawniczak, *PSCS Service Control*. Proc. of RACE Mobile Telecommunications Workshop, pp. 438-443, Amsterdam, Mai 1994.

- /Lawn95/ D. Lawniczak, J. Ericsson, J. Dahlström, *Design Options for the DDB Network for Mobile Applications in UMTS*. Proc. of ISS, pp. 431-435, Berlin, April 1995.
- /LiNi93/ M.-S. Li, E. Nichols, *Intelligent Networks: Strategies for Customised Global Services*. Ovum Report, Ovum Ltd, 1993.
- /Lock93/ P. C. Lockemann, *Telekommunikation und Datenhaltung*, Universität Karlsruhe, Carl Hanser Verlag, München-Wien, 1993.
- /Loui94/ M. Louie, *Worldwide Integration of Satellite PCS and Terrestrial systems Globstar: LEO Satellites with CDMA*. 3rd ICUPC, pp. 465-469, San Diego, September 1994.
- /Lyck92/ E. Lycksell, *Network Architectures for FPLMT*. 5th Nordic Seminar on Digital Mobile Radio Communication, pp. 203-212, Helsinki, Dezember 1992.
- /Mall88/ B. Mallinder, *An overview of the GSM system*. Proc. of the 3rd Nordic Seminar on Digital Land Mobile Radio Communication, pp. 1-13, Copenhagen, September 1988.
- /Mart94/ M. Martin, *IN Architecture: An Opportunity for Providing New Services in GSM Network*. Proc. of 3rd International Conference on Intelligence in Networks, pp. 311-315, Bordeaux, Oktober 1994.
- /Maru94/ Y. Maruyama, K. Yamagata, M. Yabusaki, *Intelligent Network with Mobility*. Proc. of 3rd International Conference on Intelligence in Networks, Bordeaux, Oktober 1994.
- /MatR92// *Reference Guide - MATLAB® High-Performance Numeric Computation and Visualisation Software*, The MathWorks, Inc., August 1992.
- /MatU92/ *User's Guide - MATLAB® High-Performance Numeric Computation and Visualisation Software*, The MathWorks, Inc., August 1992.
- /Mobi4/ *PSCS Concept*, RACE Deliverable R2003/EED/CT1/DS/P/004/b1, April 1993.
- /Mobi9/ *PSCS Manager Demonstration - Draft Results*, RACE Deliverable R2003/EED/CT4/DS/P/009/b1, September 1993.
- /Mobi12/ *PSCS Specification and CFS: Management Components - Final Version*. RACE Deliverable R2003/ETM/CT2/DS/P/012/b1, Dezember 1993.
- /Mobi16/ *PSCS Specification and CFS: Architectural Framework - Draft Version*. RACE Deliverable R2003/TVT/CT2/DS/P/016/b1, April 1994.
- /Mobi94/ *Technical Annex for 1995*, RACE Project R2003, September 1994.

- /Mone14/ *Stage 2 specification of communications between databases*. RACE Deliverable R2066/PTTNL/MF1/DS/P/014/a1, Oktober 1992.
- /Mone54/ *Signalling traffic requirements for UMTS*. RACE Deliverable R2066/SESA/GA2/DS/P/054/a1, Dezember 1993.
- /Mone61/ *Implementation Aspects of the UMTS Database*. RACE Deliverable R2066/PTTNL/MF1/DS/P/061/b1, Mai 1994.
- /Moul92/ M. Mouly, M.-B. Pautet, *The GSM System for Mobile Communications*. Michel Mouly and Marie-Bernadette Pautet, 1992.
- /Mura89/ K. Murakami and M. Katoh, *Control Architecture for Next-Generation Communication Networks Based on Distributed Databases*. IEEE Journal of Selected Areas in Communication, Vol. 7, No. 3, pp. 418-423, April 1989.
- /Nieb91/ N. Niebert, *Struktur und Algorithmen eines Formelauswertesystems für die Datenverkehrstheorie*. Dissertation, RWTH Aachen, 1991.
- /Nieb94/ N. Niebert, E. Geulen, *Personal Communication - What is Beyond Radio?*. Proc. of 2nd Int. Conference on Intelligence in Broadband Service and Networks (IS&N '94), pp. 247-257, Aachen, September 1994.
- /Nied93/ J. Nieder, *Digital kommunizieren mit DECT*, Funkschau, pp. 74-77, No. 10, 1993.
- /Norp94a/ T. Norp, A. J. M. Roovers, *UMTS Integrated with B-ISDN*. IEEE Communications Magazine, Vol. 32, No. 11, pp. 60-65, November 1994.
- /Norp94b/ T. Norp, *Protocol stacks for an integrated UMTS user access network*, Proc. of the RACE Mobile Telecommunications Workshop, pp. 310-314, Amsterdam, Mai 1994.
- /Ogaw94/ K. Ogawa, K. Kohiyama, T. Kobayashi, *Toward the Personal Communication Era - A Proposal of the 'Radio Access' Concept from Japan*. International Journal Wireless Networks, pp. 17-27, Vol. 1, No. 1, Januar 1994.
- /Padg95/ J. E. Padgett, C. G. Günther, T. Hattori, *Overview of Wireless Personal Communications*. IEEE Communications Magazine, Vol. 33, No. 1, pp. 28-41, Januar 1995.
- /Pers93/ Bellcore, *Personal Communication Network Access Demand Model & Capacity Analysis*. Special report SR-INS-002603, April 1993.
- /Q.1200/ Q-Series Intelligent Network Recommendations, *CCITT Recommendations Q.1200*, 1992.

- /R1D67/ *Final Report on Fixed Network Aspects in UMTS*. RACE Deliverable R1043/RNL/FN12/DS/A/067/b1, Dezember 1991.
- /Rams94/ P.A. Ramsdale, *Personal Communication in the UK - Implementation of PCN Using DCS 1800*. Internat. Journal Wireless Info. Network, pp. 29-36, Vol. 1, No. 1, 1994.
- /Rayn88/ M. Raynal, *Distributed Algorithms and Protocols*. University of Rennes, John Wiley & Sons, 1988.
- /Res91/ Research & Development Center for Radio Systems (RCR), *Digital Cellular Telecommunication Systems*. RCR STD-27, Dezember 1991.
- /Res93/ Research & Development Center for Radio Systems (RCR), *Personal Handy Phone Systems*. RCR STD-28, Dezember 1993.
- /Rons94/ M. Ronström, *Computer Architectures for Evaluation of the UMTS Network*. Proc. of RACE Mobile Telecommunications Workshop, pp. 462-467, Amsterdam, Mai 1994.
- /Sara89/ R. Saracco, J. R. W. Smith, R. Reed, *Telecommunications Systems Engineering using SDL*. Elsevier Science Publishers B.V., 1989.
- /Schi95/ A. Schieder, *Development and implementation of a flexible and fault tolerant data exchange protocol for UMTS*. Diplomarbeit, Lehrstuhl für Kommunikationsnetze RWTH Aachen, Februar 1995.
- /SMG5-4/ *The Relationship between UMTS and IN*. ETSI/SMG5 #4 report from the meeting on 11-15 Januar 1993 in Munich, Appendix 9.
- /SSD94/ *UMTS System Structure Document*. RACE Deliverable R2066/BT/PM2/DS/P/070/b2, Dezember 1994.
- /Stee92/ R. Steele, *Mobile Radio Communication*, Pentech Press Publishers, London, 1992.
- /Syka94/ E. D. Sykas, E. P. Adamidis, *On the Communication among UMTS DDB Domains*. Proc. of RACE Mobile Telecommunications Workshop, Amsterdam, pp. 450-458, Mai 1994.
- /T1P1/ *System Engineering Working Document for Personal Communication*. T1P1.2/92-002R3, Juli 1992.
- /Tera91/ F. Teraoka, et. al., *A Network Architecture Providing Host Migration Transparency*. Proc. of ACM SIGCOMM Conference on Communications, Architectures and Protocols, pp. 209-220, Zürich, September 1991.

- /Tera93/ F. Teraoko and M. Tokoro, *Host Migration Transparency in IP Networks: The VIP Approach*, ACM SIGCOMM, Vol. 23, No. 1, pp. 45-65, Januar 1993.
- /UPT1/ *Universal Personal Telecommunication Phase 1, Service Aspects: Guidelines - Version 1*. ETSI Draft TC-TR/NA-70101, September 1991.
- /UPT2/ *A Standardisation Programme for Universal Personal Telecommunication (UPT)*. ETSI/STC NA7 (91)5, Report from STC NA7 to the ETSI Technical Assembly, Sophia Antipolis, Februar 1991.
- /UPT3/ *Network Aspects (NA): Universal Personal Telecommunication (UPT): General Service Description - Version 1.2.0*. ETSI Draft TC-TR/NA-70201, Juli 1992.
- /UPT4/ *Network Aspects (NA): Universal Personal Telecommunication (UPT): Supplementary Services - Version 0.3.1*. ETSI DTR NA-70210, November 1992.
- /UPT5/ *Network Aspects (NA): Universal Personal Telecommunication (UPT): UPT subscriptions and services profiles - Version 0.3.0*. ETSI DTR NA-70205, August 1992.
- /Voss91/ G. Vossen, *Data Models, Database Languages and Database Management Systems*. Addison-Wesley Publishing Company, 1991.
- /Walk78/ B. Walke, *Realzeitrechner-Modelle: Theorie und Anwendung*. Oldenburg Verlag, 1978.
- /Walk91/ B. Walke, *Kommunikationsnetze und Verkehrstheorie I, Dienste und Protokolle von Rechnernetzen*. Vorlesungsunterlagen, Lehrstuhl für Kommunikationsnetze RWTH Aachen, 1991.
- /Walk92a/ B. Walke, *Kommunikationsnetze und Verkehrstheorie II, Verkehrstheoretische Modelle von Echtzeitsystemen und Rechnernetzen*. Vorlesungsunterlagen, Lehrstuhl für Kommunikationsnetze RWTH Aachen, 1992.
- /Walk92b/ B. Walke, *Angewandte Informatik 2, Rechnerorganisation und Systemleistung*. Vorlesungsunterlagen, Lehrstuhl für Kommunikationsnetze RWTH Aachen, 1992.
- /WanD93/ D.C.C. Wang, *A Network Architecture for Number and Service Mobility*. Proc. of 2nd ICUPC, pp. 118-122, Ottawa, Oktober 1993.
- /WanJ93/ J.Z. Wang, *A Fully Distributed Location Registration Strategy for Universal Personal Communication Systems*. IEEE Journal of Selected Areas in Communication, Vol. 11, No. 6, pp. 850-860, August 1993.

- /Wild94/ C.G. Wildey, *Satellite and Cellular Intergration: A Terminal Manufacturer's Perspective*. Proc. of 3rd ICUPC, pp. 392-396, San Diego, September 1994.
- /Wizg94/ M. Wizgall, *PCS Implementation - A Merger of American and European Mobile Systems*. Proc. of ICUPC, pp. 612-616, San Diego, September 1994.
- /Woin94/ M. Woinsky, *PCS Standards Development - Comittee T1*. Proc. of 3rd ICUPC, pp. 573-577, San Diego, September 1994.
- /X.500/ *Data Communication Networks Directory*, Blue Book, CCITT Recommendations X.500-X.521, 1988.
- /YuMe94/ C. Yu, W. Meng, *Progress in Database Search Strategies*, IEEE Software, Vol. 11, No 3, p. 11-19, Mai 1994.
- /Z.100/ *Functional Specification and Description Language (SDL)*. Blue Book, CCITT Recommendations Z.100 and Annexes A, B, C, and E, 1988.

Anhang A - Glossar

A.1. Begriffe

Abfrage	jede Nachricht, die an einen Datenbankknoten geschickt wird (= Datenabfrage), verursacht im Knoten eine Operation
Anrufverhalten	Charakteristik des Benutzerverhaltens in Hinblick auf die Verteilung der Anrufe
Anfrage	eine von außerhalb der Datenbank ankommende Nachricht, die nach einer Information in der Datenbank sucht
Aufenthaltsgebiet	physikalisches Gebiet, in dem der Benutzer beim Anrufaufbau gleichzeitig gesucht wird (<i>location area - LA</i>)
Aufenthaltsort	Ort, in dem sich der Benutzer momentan befindet (<i>location</i>)
Benutzer	natürliche Person, die mit Erlaubnis des Teilnehmers einen Dienst in Anspruch nimmt (<i>user</i>)
Benutzerklasse	Menge aller Benutzer, die dem gleichen DN-Heimgebiet angehören
Benutzermigration	eine geordnete Bewegung der Benutzer, die eine Veränderung der Benutzerverteilung zur Folge hat
Besuchsteilgebiet	Teilgebiet, das der Benutzer nur temporär besucht (<i>visited domain</i>)
Beweglichkeit	Charakteristik des Benutzerverhaltens, die die Häufigkeit der Bewegungen im stationären Zustand beschreibt
DA-Funktion	DAtenspeicherfunktion in den DN-Knoten
DBMS	Datenbank-Management-System
Dienstanbieter	juristische Person, die UMTS-Dienste anbietet (<i>service provider</i>)
DIR-Gebiet	ein Teil des UMTS-Teilgebietes, das von einem DN-Knoten kontrolliert und aus mehreren DN-Gebieten aufgebaut ist (<i>directory node</i>)
DIR-Knoten	UMTS-Datenbankknoten, der die Zeiger speichert; auch Verzeichnisknoten genannt (<i>directory node</i>)
DIR-Struktur	Verzeichnisstruktur; logisch gesehen, eine baumartig aufgebaute Hilfsknotenstruktur, die das Ziel hat, die Datenverwaltung und die Datenkonsistenz in DDB zu unterstützen
DIR-Wurzelknoten	der höchste DIR-Knoten im UMTS-Verzeichnis (<i>root node</i>)
DN-Besuchsgebiet	zum DN-Besuchsknoten gehörendes DN-Gebiet (<i>visited area</i>)
DN-Besuchsknoten	DN-Knoten, in dessen DN-Gebiet der aktuelle Aufenthaltsort des Benutzers liegt; üblicherweise ist dort die Zweitkopie der Benutzerdaten abgespeichert; auch Besuchsdatenknoten genannt (<i>visited node</i>)

DN-Gebiet ¹⁾	ein Unterteil des Teilgebietes, das von einem DN-Knoten kontrolliert wird; auch Datenknotengebiet genannt (<i>data node area</i>)
DN-Heimknoten	DN-Knoten, in dessen DN-Gebiet der heimische Aufenthaltsort des Benutzer liegt; üblicherweise ist dort die Hauptkopie der Daten abgespeichert; auch Heimdatenknoten genannt (<i>resident node</i>)
DN-Heimgebiet	zum DN-Heimknoten gehörendes DN-Gebiet (<i>resident area</i>)
DN-Knoten	ein physikalischer Knoten der Datenbank, in dem die Profile der Benutzer und Endgeräte gespeichert sind; auch Datenknoten genannt (<i>data node</i>)
DS-Funktion	Verzeichnisfunktion in den DN- und DIR-Knoten
dynamische Sz.	Szenarien mit dynamischen Zeigern, d.h. V- und RV-Szenario
Grundelement	ein prozedurunabhängiger Teil der Anfrage, z.B. Lesen
Heimteilgebiet	Teilgebiet, in dem der Benutzer sich zu Hause befindet (<i>home domain</i>)
Konzentration	Verteilung der Benutzer einer Klasse um ein bestimmtes DN-Gebiet
Mobilität	Charakteristik des Benutzerverhaltens, die sich aus seiner Beweglichkeit und seiner Verteilungswahrscheinlichkeit zusammensetzt
Netzwerkbetreiber	juristische Person, die das Eigentum über die physikalischen Transportnetze besitzt (<i>network operator</i>)
Nachricht	eine einzelne Abfrage, die von einem Knoten zum anderen gesendet wird
Operation	eine Aktion im Knoten, die sich aufgrund der Abfrage in diesem Knoten ergibt - Lesen, Schreiben (Modifizieren), sonstiges
Prozedur	eine Anreihung von Anfragen im Netz, z.B. Anrufaufbau
QH-Funktion	Anfrageverwaltungsfunktion, die die Anfrage in eine Nachrichtenfolge zerlegt
statische Sz.	Szenarien mit statischen Zeigern, d.h. ST- und R Szenario
Teilgebiet	ein organisatorischer Verwaltungsteil des UMTS-Gebietes, der sich durch die Zugehörigkeit zu einem Dienstanbieter und durch einheitliche Struktur des UMTS-Netzes auszeichnet (<i>domain</i>)
Teilnehmer	juristische Person, die einen Dienst im voraus oder direkt kauft und sich zur Zahlung verpflichtet (<i>subscriber</i>)
UMTS-Dienste	Dienste, die dem Benutzer des UMTS-Systems zur Verfügung gestellt werden

1. DN-Knoten wird manchmal gleichbedeutend mit DN-Gebiet verwendet

UMTS-DTN	Datenteilnetz des UMTS-Netzes: der Datenbankteil der UMTS-Infrastruktur
UMTS-Funktionen	Funktionseinheiten, die als wesentlich für die Funktionsweise des UMTS-Systems angesehen werden; sie sind dem funktionalen Modell zu entnehmen
UMTS-Gebiet	ein physikalisches Gebiet, auf dem die UMTS-Funktionalität den Benutzern angeboten wird
UMTS-Netz	physikalische Infrastruktur, die die spezifischen UMTS-Funktionen unterstützt
UMTS-System	= UMTS
UMTS-VTN	Vermittlungsteilnetz des UMTS-Netzes: die gesamte UMTS-Infrastruktur mit Ausnahme der Datenbankeigenschaft

A.2. Abkürzungen

ACTS	<i>Advanced Communications Technologies and Services</i>
ADC	<i>American Digital Cellular (= IS-54)</i>
ANSI	<i>American National Standards Institute</i>
B-ISDN	<i>Broadband ISDN</i>
BSC	<i>Base Station Controller</i>
C-450	der analoge zellulare Standard in Deutschland und Portugal
CCITT	<i>Consultative Committee for Int. Telephones & Telegraphs</i>
CT1+	<i>Cordless Telephone Plus</i>
CT2	<i>Cordless Telephone 2</i>
CS-1, CS-2	<i>Capability Set 1, 2, ... u.s.w</i>
DA	<i>Data Storage</i>
DECT	<i>Digital European Cordless Telephone</i>
DTN	Datenteilnetz
ETACS	<i>European Telephony Analog Cellular System</i>
ETSI	<i>European Telecommunications Standards Institute</i>
FPLMTS	<i>Future Public Land Mobile Telecommunications System</i>
GSM	<i>Global System for Mobile Communication</i>
HLR	<i>Home Location Register</i>
IMUN	<i>International Mobile User Number: die Nummer, unter der ein Benutzer von einem anderen angerufen wird</i>
IN	<i>Intelligent Network</i>
IS-41	der analoge zellulare Standard in USA
IS-54	= ADC
ISDN	<i>Integrated Services Digital Network</i>

ITU	<i>International Telecommunication Union</i>
LA	<i>Location Area</i> = Aufenthaltsgebiet
LE	lokale Vermittlungsstelle (<i>Local Exchange</i>)
MSC	<i>Mobile Switching System</i>
NMT	<i>Nordic Mobile Telecommunication</i>
PCN	<i>Personal Communication Network</i>
PCS	<i>Personal Communication System</i>
PDC	<i>Personal Digital Cellular</i>
PHS	<i>Personal Handyphone System</i>
PIN	<i>Personal Identification Number</i>
PSTN	<i>Public Switched Telephone Network</i>
PSCS	<i>Personal Service Communication Space</i>
TACS	<i>Total Access Communication System</i>
TCP/IP	<i>Transport Communication Protocol / Interchange Protocol</i>
TDMA	<i>Time Division Multiple Access</i>
TMTI	<i>Temporary Mobile Terminal Identifier</i> : die Kennung, die einem Endgerät lokal vergeben wird, um seine Verwaltung innerhalb eines DN-Gebietes oder eines Aufenthaltsgebietes zu vereinfachen
TP	<i>Terminal Profile</i> - Endgerätprofil
TX	Durchgangsvermittlungsstelle (<i>Transit Exchange</i>)
RACE	<i>Research and technology development for Advanced Communications technologies in Europe</i>
RTMS	der analoge zellulare Standard in Italien
SIM	<i>Subscriber Identity Module</i>
SPP	<i>Service Provider Profile</i>
SS7	<i>Signalling System No 7</i>
SUP	<i>Subscriber Profile</i> - Teilnehmerprofil
TCP/IP	<i>Transmission Control Protocol / Internet Protocol</i>
UMTS	das Universale Mobile Telekommunikationssystem (<i>Universal Mobile Telecommunication System</i>)
UP	<i>User Profile</i> - Benutzerdaten, Benutzerprofil
UPT	<i>Universal Personal Telecommunications</i>
VLR	<i>Visited Location Register</i>
VPN	<i>Virtual Private Network</i>
VTN	Vermittlungsteilnetz

A.3. Parameter und Variablen des Benutzermodells

A.3.1. Allgemeines

N	Anzahl der DN-Gebiete innerhalb eines UMTS-Teilgebietes
$P_r(i)$	Aufenthaltswahrscheinlichkeit der Benutzer im DN-Gebiet i im Heimteilgebiet
$P_{vis_r}(i)$	Aufenthaltswahrscheinlichkeit der Benutzer im DN-Gebiet i im Besuchsteilgebiet
U_r	Anzahl der Benutzer innerhalb der Klasse r ²⁾

A.3.2. Mobilitätsmodell

$A_r(i)$	Ankunftsrate in das DN-Gebiet i
B_{TN}	Anzahl aller Teilgebiete in einem UMTS-Gebiet
$cross$	Relation zwischen den Bewegungen über die Heimteilgebietsgrenze und allen Bewegungen im Heimteilgebiet (oder in Besuchsteilgebieten)
$cross_{vis}$	Relation zwischen den Bewegungen über die Besuchsteilgebietsgrenze und allen Bewegungen in Besuchsteilgebieten
$D_r(i)$	Abgangsrate aus dem Gebiet i
$f_r(j)$	Proportionalitätsfaktor zwischen den Bewegungen und Aufenthaltswahrscheinlichkeiten
k_M	Proportionalitätsfaktor zwischen der Anzahl der Benutzer und der Anzahl der Bewegungen
k_{CS}	Proportionalitätsfaktor zwischen der Anzahl der Benutzer und der Anzahl der initiiierenden Anrufe
$len(i)$	Anteil der Teilgebietsgrenze im DN-Gebiet i
$p_r(i, j)$	Übergangswahrscheinlichkeit aus dem Gebiet i in das Gebiet j
p_{out_r}	Wahrscheinlichkeit, daß Benutzer sich außerhalb des Heimteilgebietes befinden
x_r	Kompensationsparameter, der den Einfluß der Verteilungswahrscheinlichkeit auf die Bewegungshäufigkeit steuert
$S(i)$	Nachbargebiete des DN-Gebietes i
P_{Bew}	Parameter der Benutzerbeweglichkeit
$P_{ext, r, 0}$	Anfangsverteilung der Benutzer, die das Teilgebiet verlassen wollen
$P_{inter, r, k}$	Verteilung der Benutzer nach dem Migrationsschritt k

2. Der Index r bezieht sich immer auf die Benutzer der Klasse r .

$P_{mig, r}$	Anfangsverteilung der migrationswilligen Benutzer
$P_{new, r}$	Zielverteilung der Benutzer nach der Migration
$P_{old, r}$	Anfangsverteilung der Benutzer vor der Migration
P_h	Wahrscheinlichkeit, daß die Benutzer sich im Heimteilgebiet befinden
P_{in}	Wahrscheinlichkeit, daß die Benutzer sich innerhalb des eigenen Heimteilgebietes befinden
P_v	Wahrscheinlichkeit, daß die Benutzer sich in einem Besuchsteilgebiet befinden
p_{h_v}	Übergangswahrscheinlichkeit aus dem Heim- ins Besuchsteilgebiet
p_{v_h}	Übergangswahrscheinlichkeit aus dem Besuchs- ins Heimteilgebiet
λ_{h_h}	Übergangsrate aus dem Heim- ins Heimteilgebiet
λ_{h_v}	Übergangsrate aus dem Heim- in ein Besuchsteilgebiet
λ_{v_h}	Übergangsrate aus einem Besuchs- ins Heimteilgebiet
λ_{v_i}	Übergangsrate innerhalb eines Besuchsteilgebietes
λ_{v_n}	Übergangsrate aus einem Besuchsteilgebiet in ein anderes
λ_{v_x}	Abgangsrate aus einem Besuchsteilgebiet

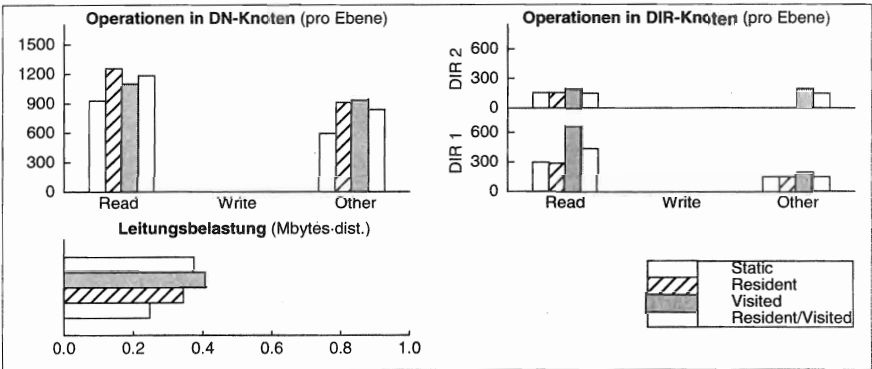
A.3.3. Anrufmodell

loc_b	Parameter, der das sog. lokale Verhalten beschreibt
$P_{orig, r}(i)$	Verteilung der anrufinitiierenden Benutzer
$P_{term, r}(i)$	Verteilung der anrufempfangenden Benutzer
p_{cHD}	Wahrscheinlichkeit der Anrufe außerhalb eines Heimteilgebietes
p_{cVD}	Wahrscheinlichkeit der Anrufe außerhalb eines Besuchsteilgebietes
$p_{h_r}(r, i)$	Wahrscheinlichkeit der Anrufe aus dem DN-Heimgebiet r ins DN-Gebiet i (innerhalb des Heimteilgebietes)
$p_{h_v}(j, i)$	Wahrscheinlichkeit der Anrufe aus dem DN-Besuchsgebiet j ins DN-Gebiet i (innerhalb des Heimteilgebietes)
$p_{h_v_{tot}}(r, j, i)$	überlagerte Wahrscheinlichkeit der Anrufe aus dem DN-Gebiet j ins DN-Gebiet i (innerhalb des Besuchsteilgebietes)
$p_{v_r/v}(j, i)$	Wahrscheinlichkeit der Anrufe aus dem DN-Gebiet j ins DN-Gebiet i (innerhalb des Besuchsteilgebietes)

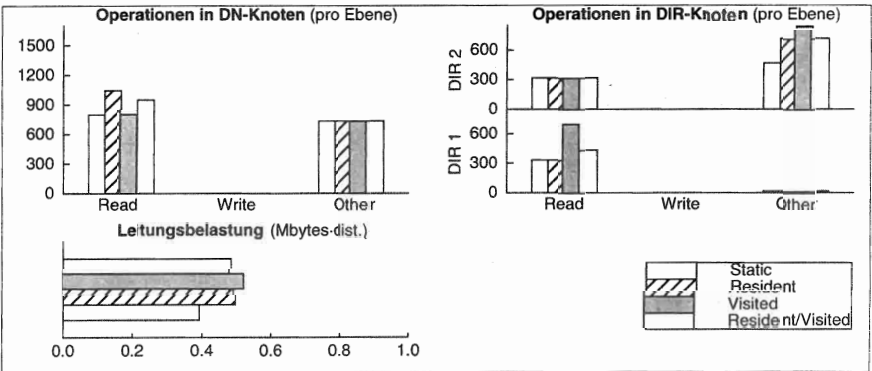
Anhang B - Numerische Ergebnisse

B.1. Grundkonfiguration_2

B.1.1. Suchanfrage

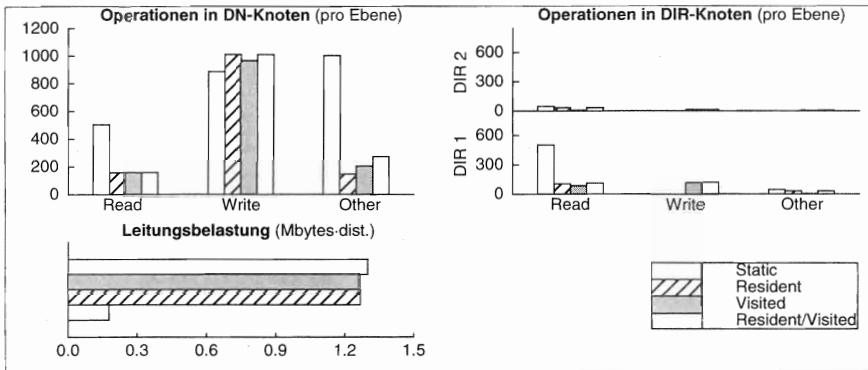


Diag. B.1 Grundkonfiguration_2 - Suchanfrage innerhalb des Heimteilgebietes (chaining/referral)

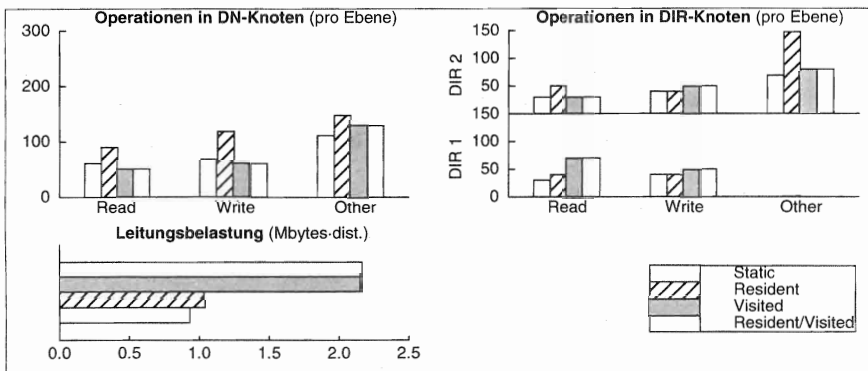


Diag. B.2 Grundkonfiguration_2 - Suchanfrage zwischen Teilgebieten (chaining/referral)

B.1.2. Aktualisierungsanfrage



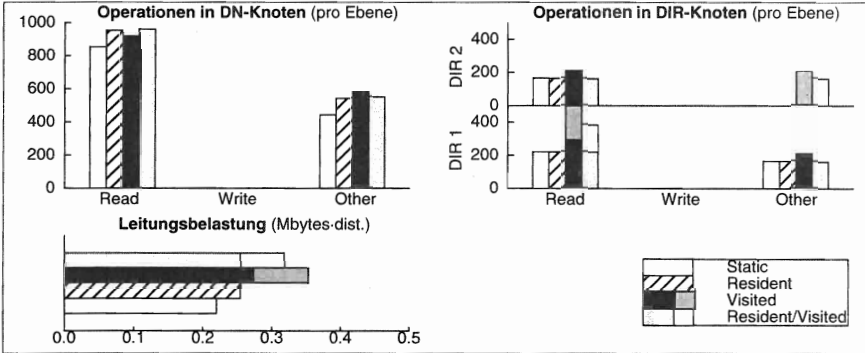
Diag. B.3 Grundkonfiguration_2 - Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)



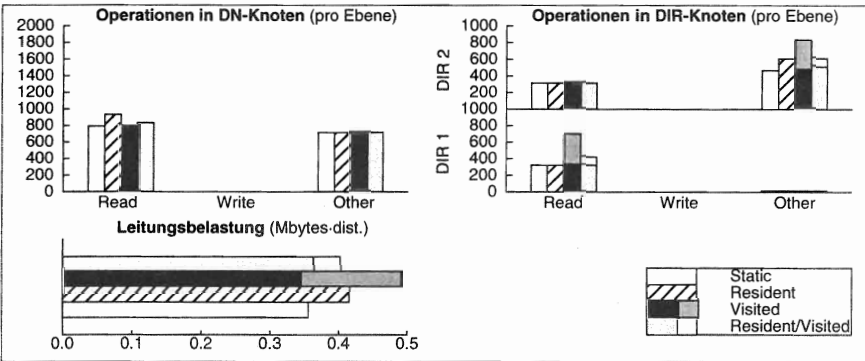
Diag. B.4 Grundkonfiguration_2 - Aktualisierung zwischen Teilgebieten (*chaining/referral*)

B.2. Alternativen der Suchanfrage

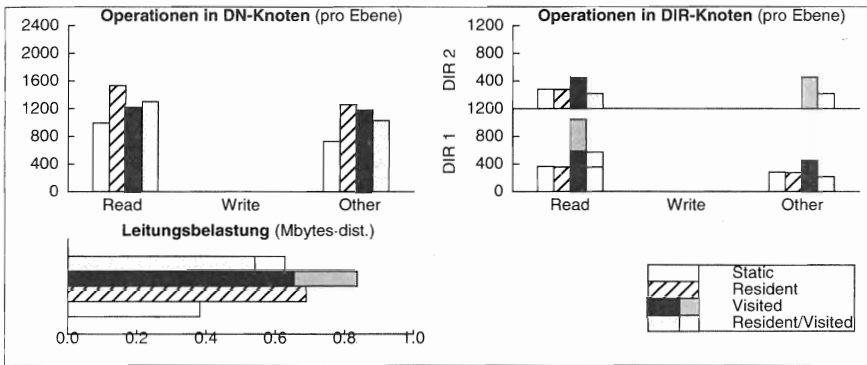
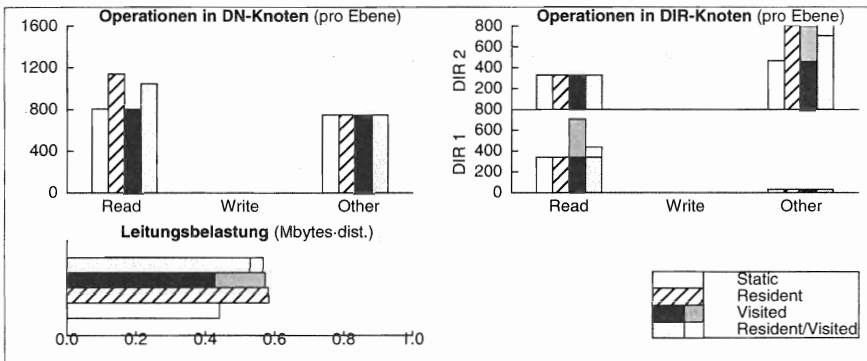
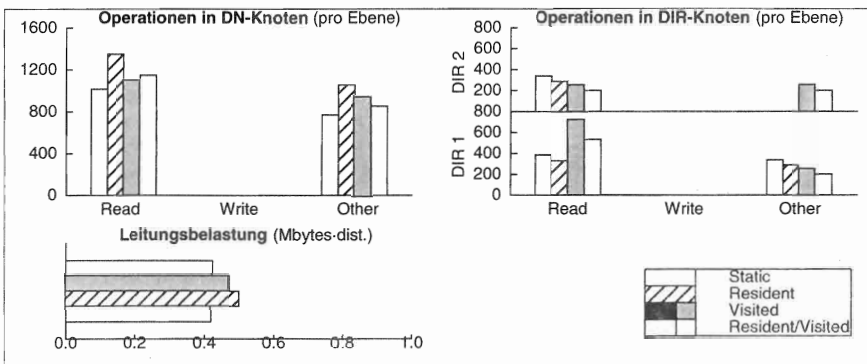
B.2.1. Einfluß der Mobilitätsparameter

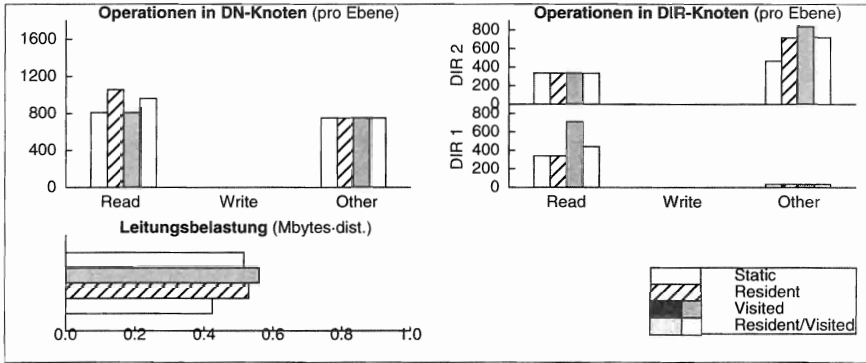


Diag. B.5 Variante 1: Suchanfrage innerhalb des Heimteilgebietes (chaining/referral)

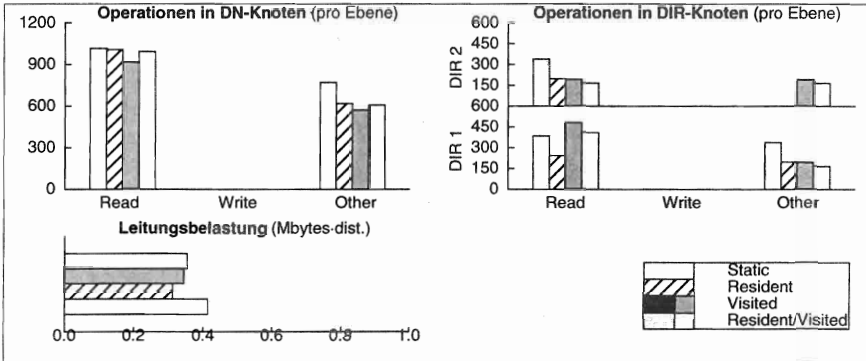


Diag. B.6 Variante 1: Suchanfrage zwischen Teilgebieten (chaining/referral)

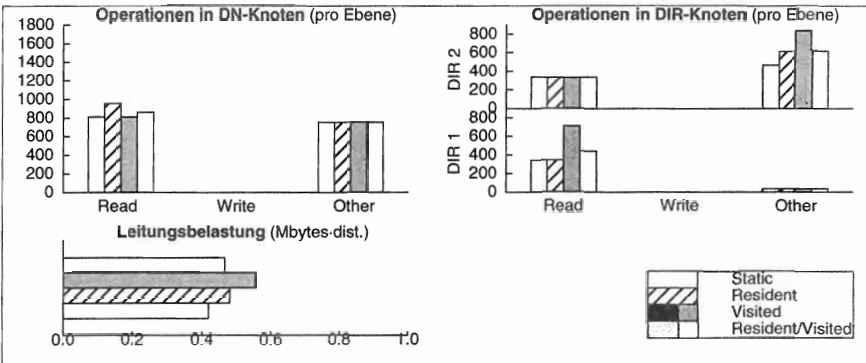
Diag. B.7 Variante 2: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)Diag. B.8 Variante 2: Suchanfrage zwischen Teilgebieten (*chaining/referral*)Diag. B.9 Variante 3: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)



Diag. B.10 Variante 3: Suchanfrage zwischen Teilgebieten (*chaining/referral*)

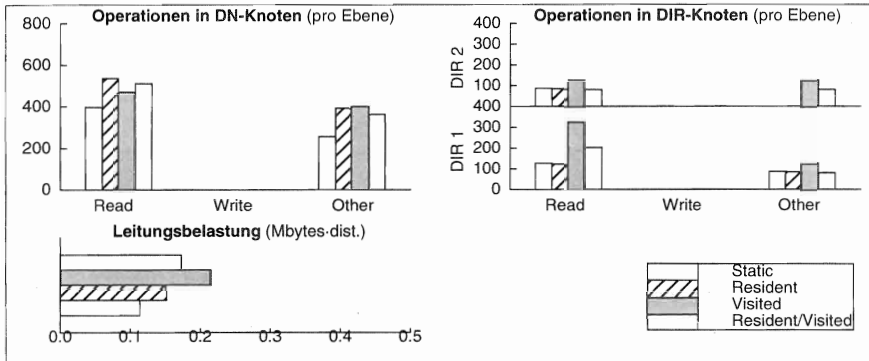


Diag. B.11 Variante 4: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)

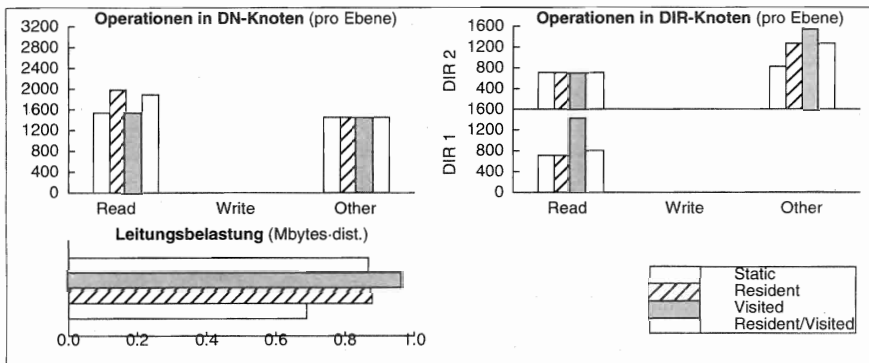


Diag. B.12 Variante 4: Suchanfrage zwischen Teilgebieten (*chaining/referral*)

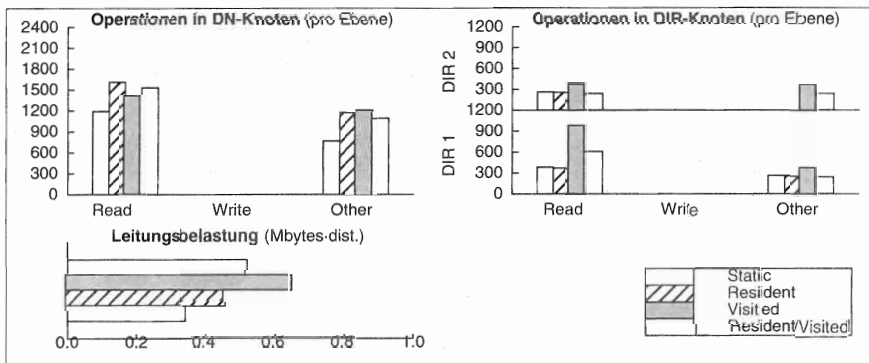
B.2.2. Einfluß der Anrufparameter



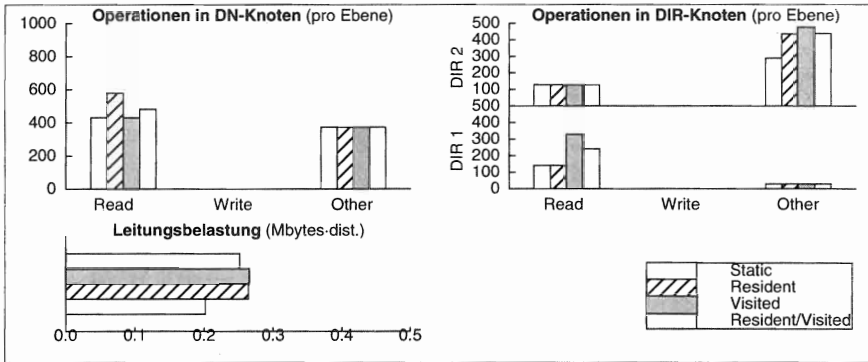
Diag. B.13 Variante 1: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)



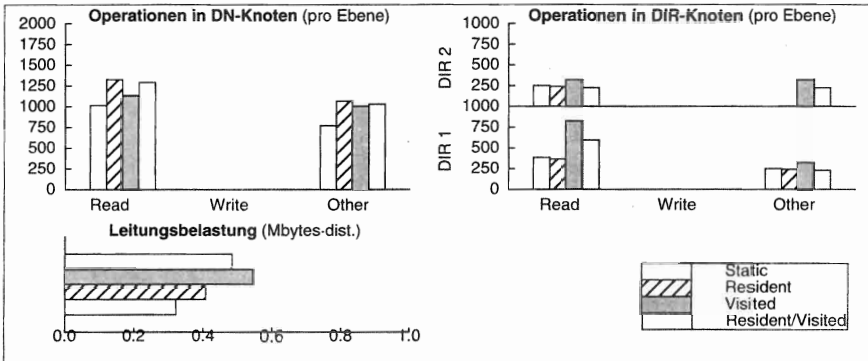
Diag. B.14 Variante 1: Suchanfrage zwischen Teilgebieten (*chaining/referral*)



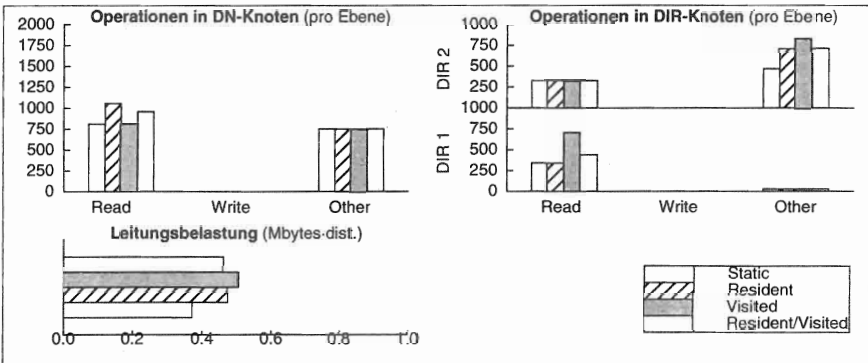
Diag. B.15 Variante 2: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)



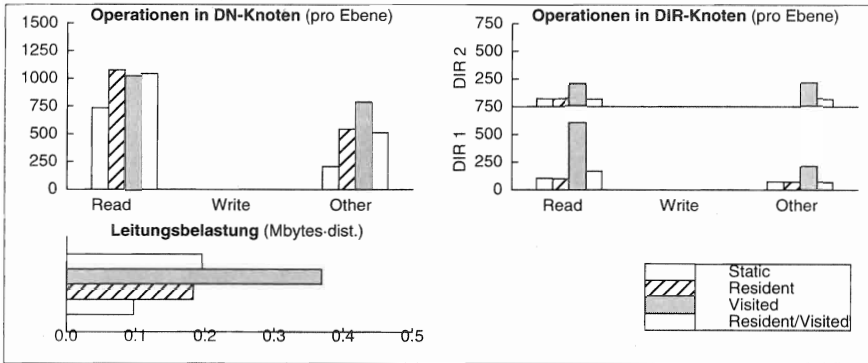
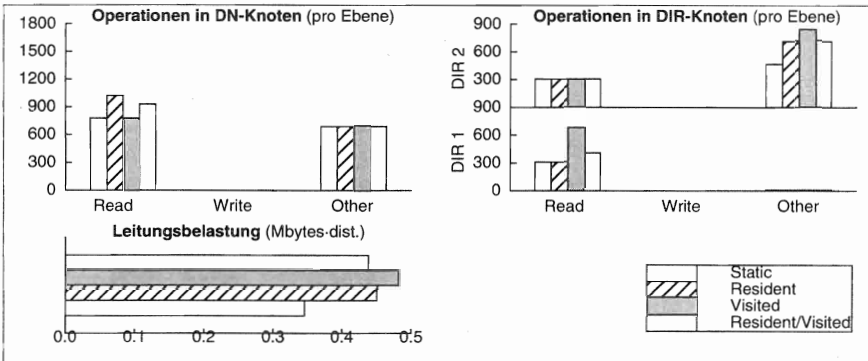
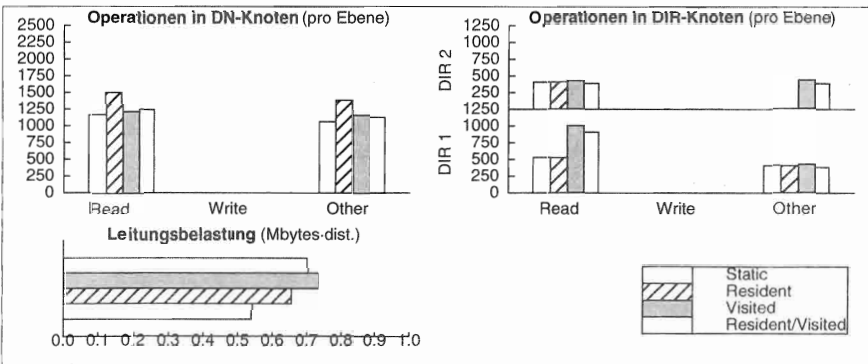
Diag. B.16 Variante 2: Suchanfrage zwischen Teilgebieten (chaining/referral)

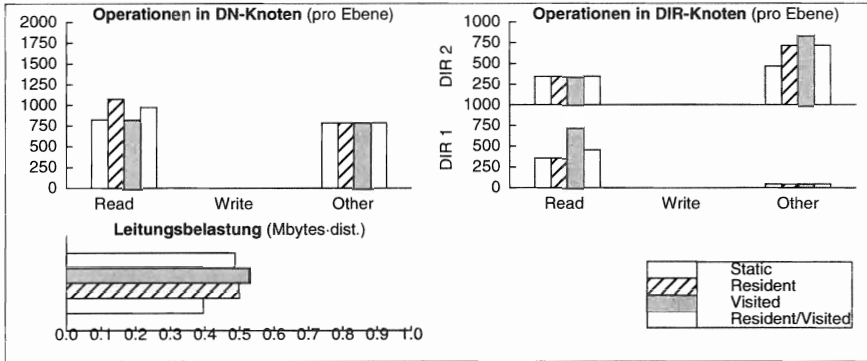


Diag. B.17 Variante 3: Suchanfrage innerhalb des Heimteilgebietes (chaining/referral)



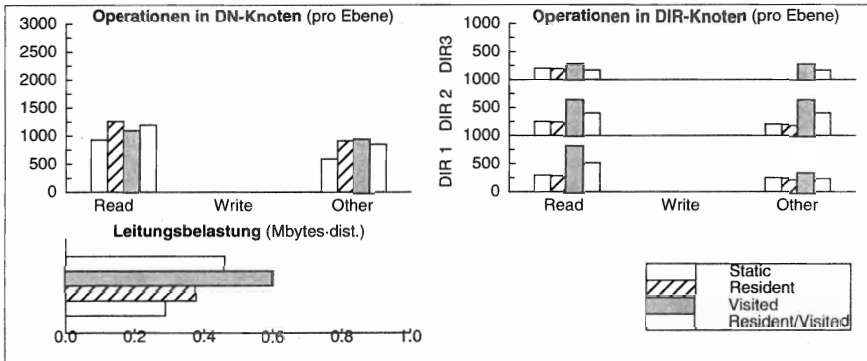
Diag. B.18 Variante 3: Suchanfrage zwischen Teilgebieten (chaining/referral)

Diag. B.19 Variante 4: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)Diag. B.20 Variante 4: Suchanfrage zwischen Teilgebieten (*chaining/referral*)Diag. B.21 Variante 5: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)



Diag. B.22 Variante 5: Suchanfrage zwischen Teilgebieten (*chaining/referral*)

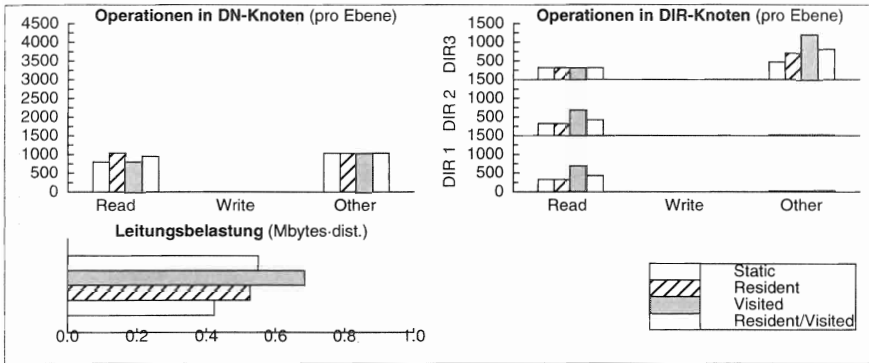
B.2.3. Einfluß der DIR-Architektur



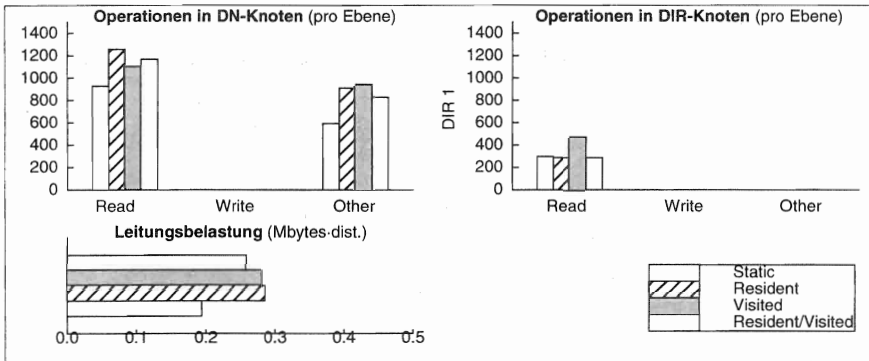
Diag. B.23 Variante 1: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)

Bemerkung:

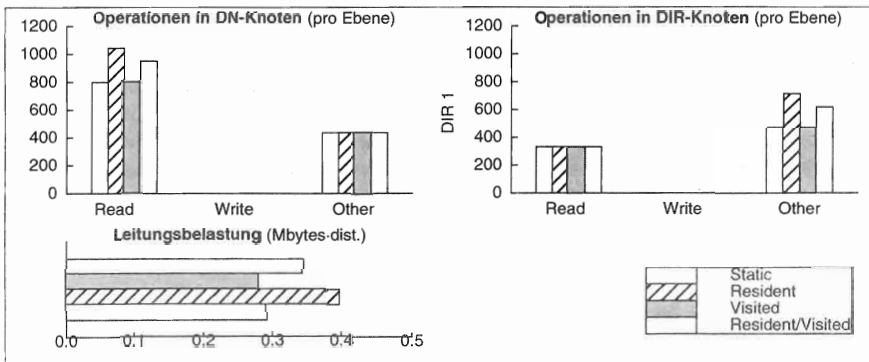
Es ist zu beachten, daß die Wurzelknotenbelastung in den Ergebnissen für die Suchanfrage zwischen Teilgebieten sich zu einem großen Teil aus der Kommunikation zwischen den Teilgebieten zusammensetzt. Es werden insgesamt 398 Lese- und 368 sonstige Nachrichten zwischen den Teilgebieten ausgetauscht - vorausgesetzt, die Anrufverteilung ist spezifiziert wie in der Grundkonfiguration.



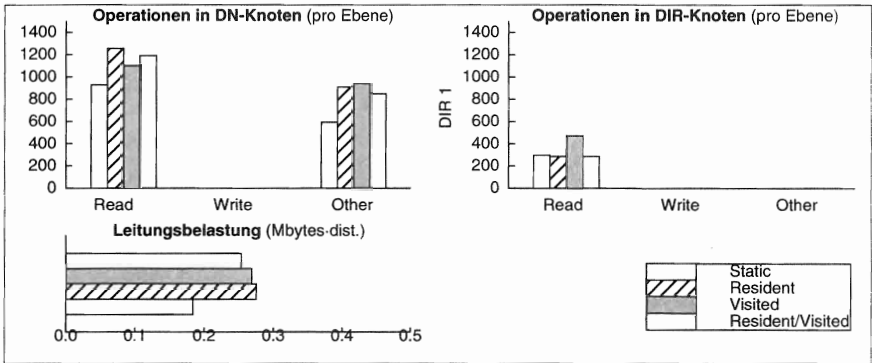
Diag. B.24 Variante 1: Suchanfrage zwischen Teilgebieten (*chaining/referral*)



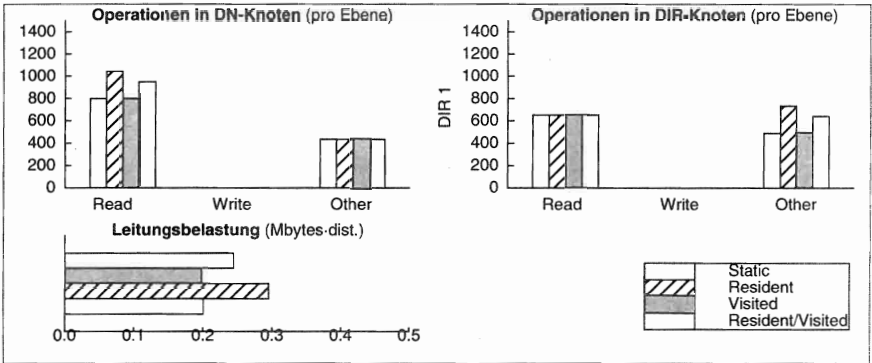
Diag. B.25 Variante 2: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)



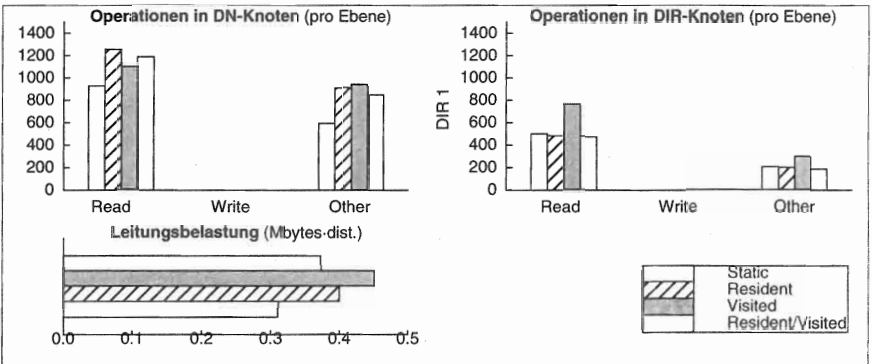
Diag. B.26 Variante 2: Suchanfrage zwischen Teilgebieten (*chaining/referral*)



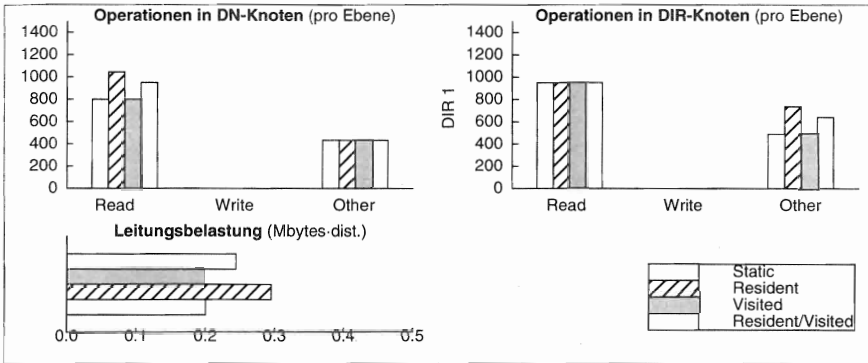
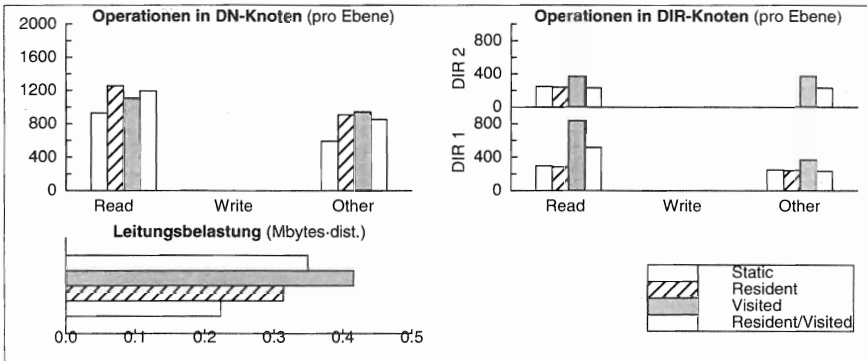
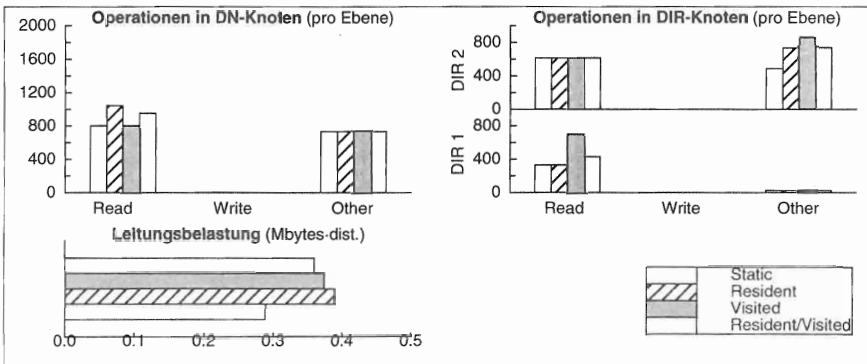
Diag. B.27 Variante 3a: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)

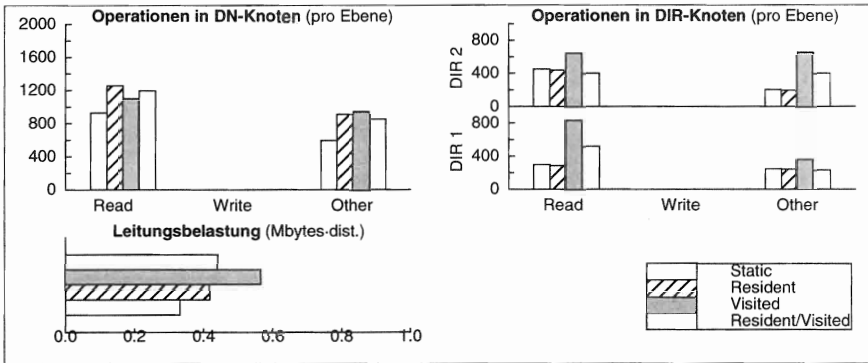


Diag. B.28 Variante 3a: Suchanfrage zwischen Teilgebieten (*chaining/referral*)

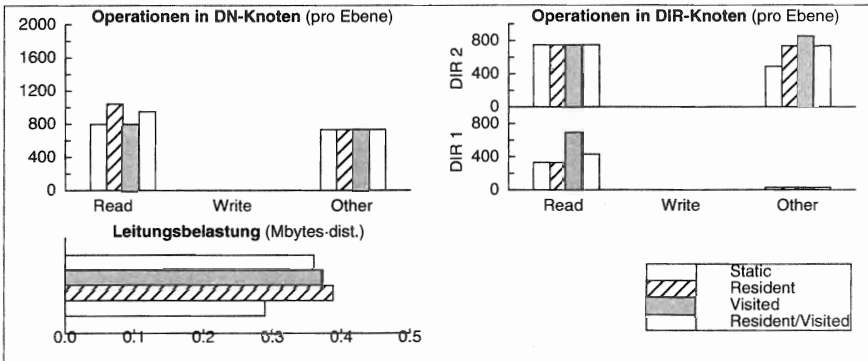


Diag. B.29 Variante 3b: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)

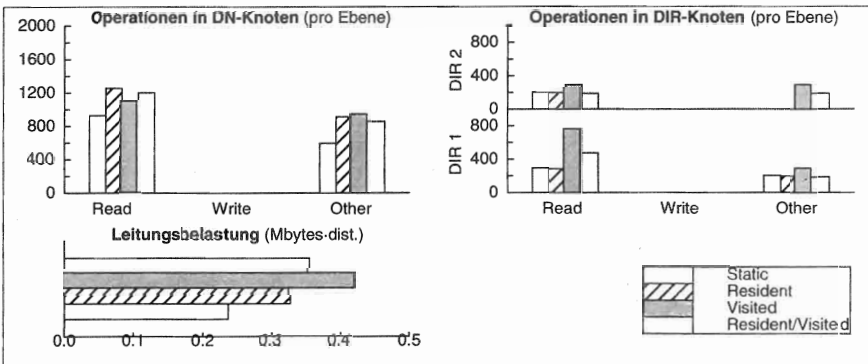
Diag. B.30 Variante 3b: Suchanfrage zwischen Teilgebieten (*chaining/referral*)Diag. B.31 Variante 4a: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)Diag. B.32 Variante 4a: Suchanfrage zwischen Teilgebieten (*chaining/referral*)



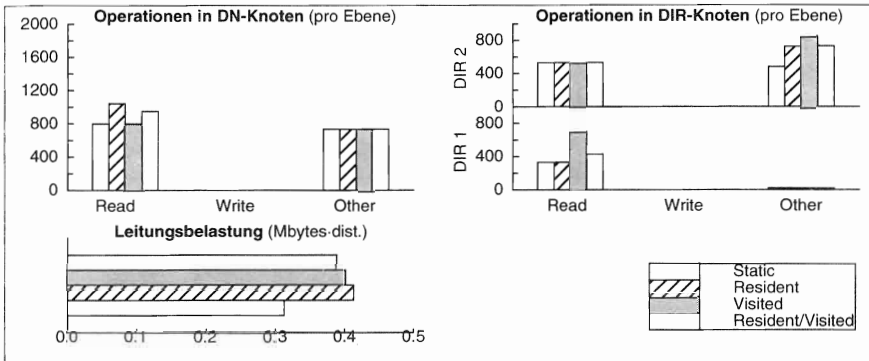
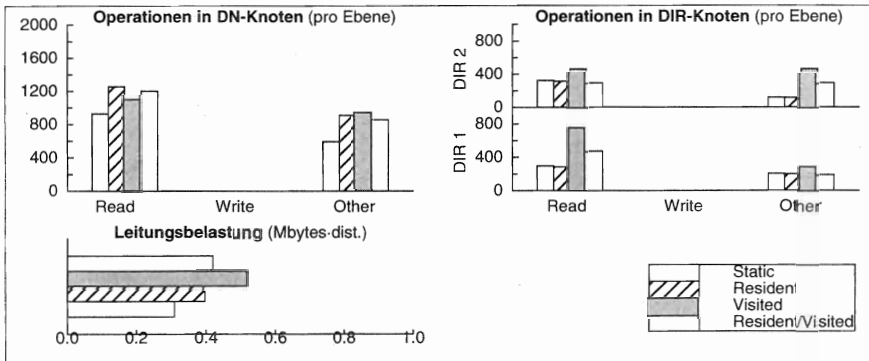
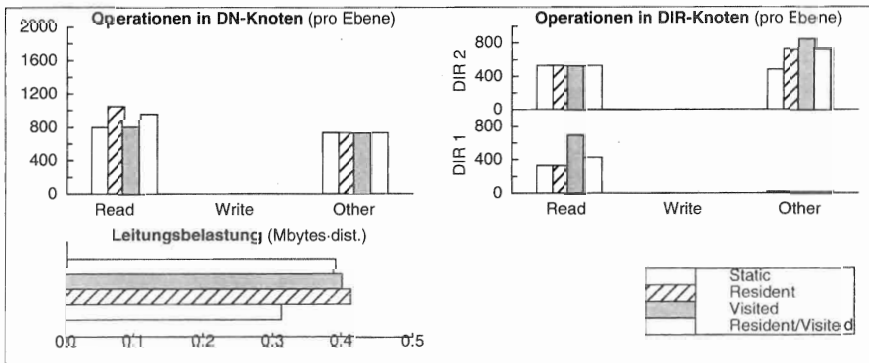
Diag. B.33 Variante 4b: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)

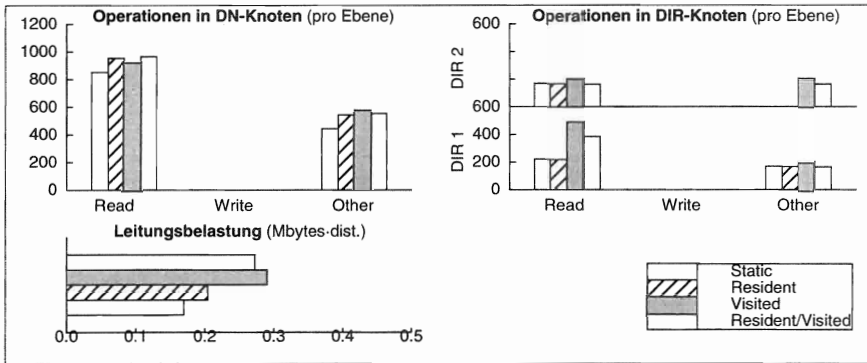


Diag. B.34 Variante 4b: Suchanfrage zwischen Teilgebieten (*chaining/referral*)

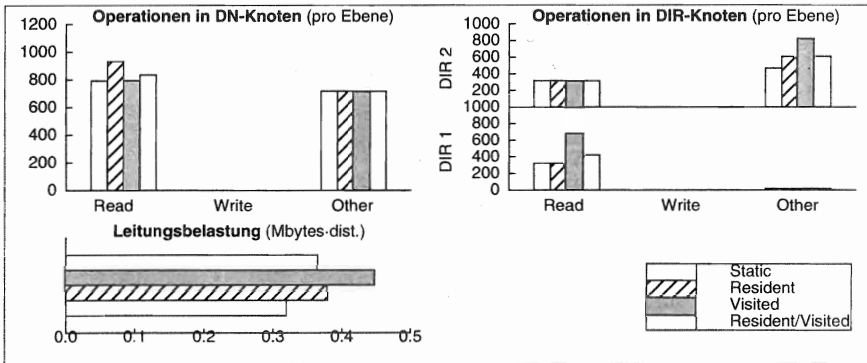


Diag. B.35 Variante 5a: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)

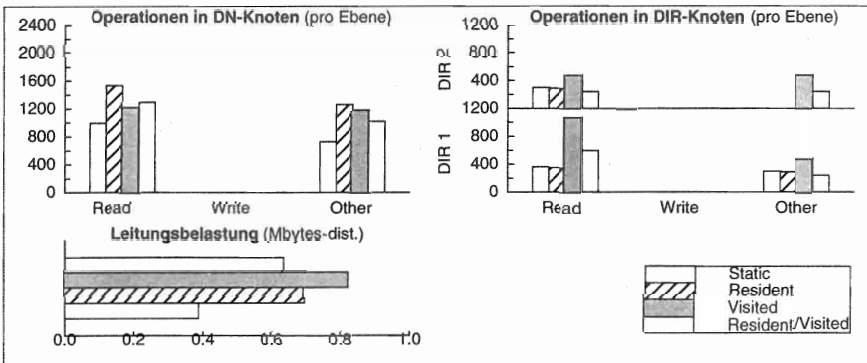
Diag. B.36 Variante 5a: Suchanfrage zwischen Teilgebieten (*chaining/referral*)Diag. B.37 Variante 5b: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)Diag. B.38 Variante 5b: Suchanfrage zwischen Teilgebieten (*chaining/referral*)



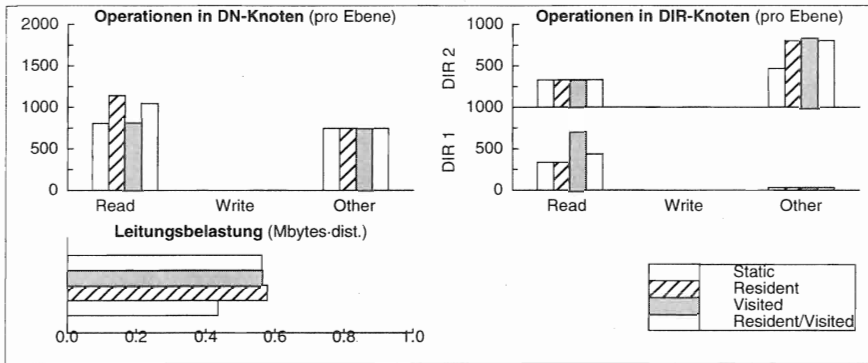
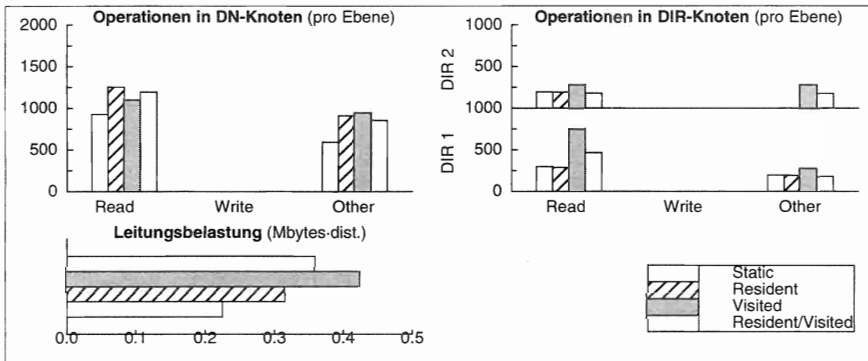
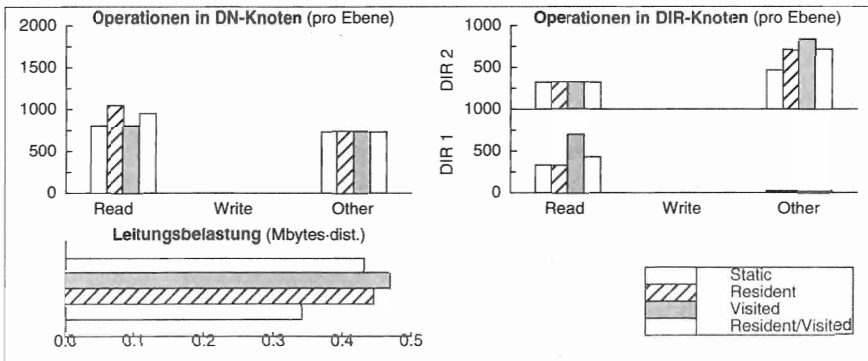
Diag. B.39 Variante 6: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)

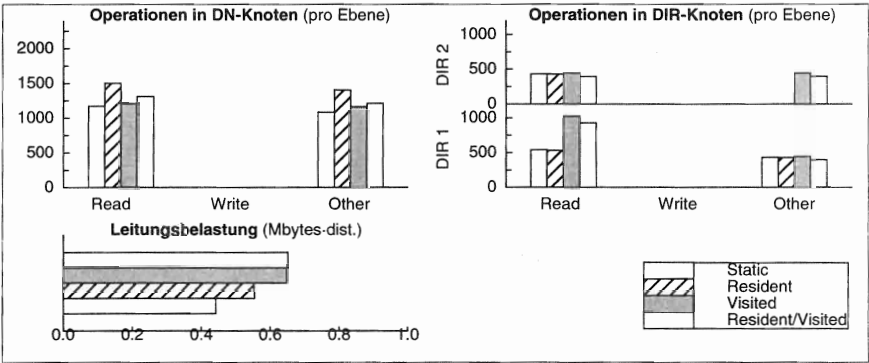


Diag. B.40 Variante 6: Suchanfrage zwischen Teilgebieten (*chaining/referral*)

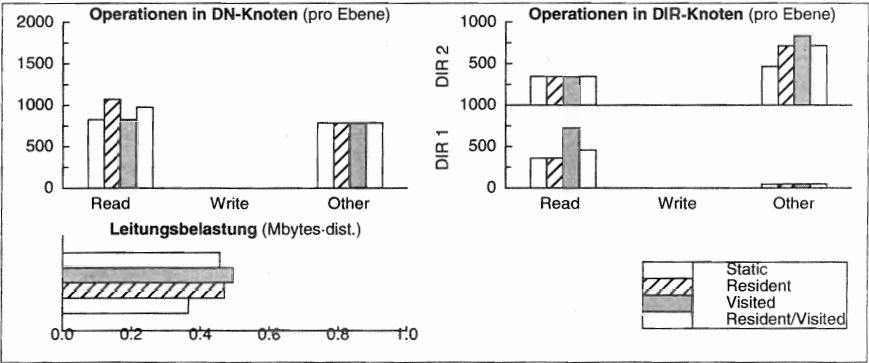


Diag. B.41 Variante 7: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)

Diag. B.42 Variante 7: Suchanfrage zwischen Teilgebieten (*chaining/referral*)Diag. B.43 Variante 8: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)Diag. B.44 Variante 8: Suchanfrage zwischen Teilgebieten (*chaining/referral*)

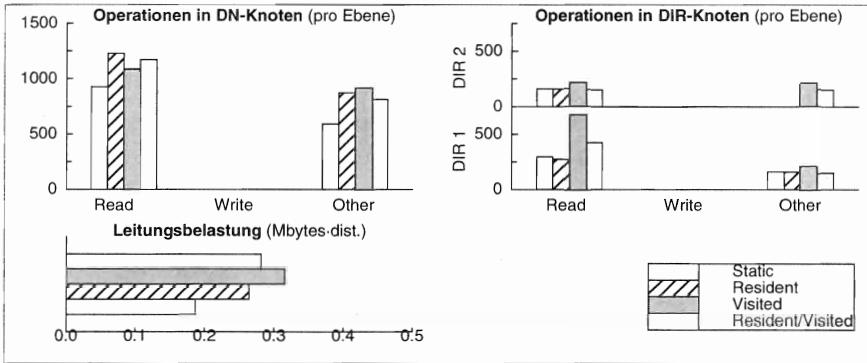


Diag. B.45 Variante 9: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)

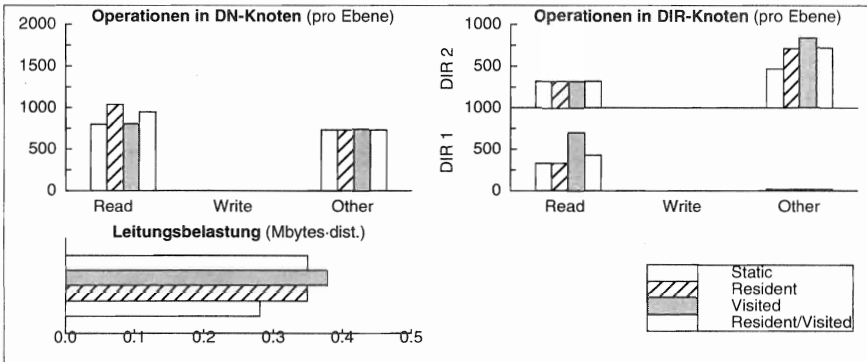


Diag. B.46 Variante 9: Suchanfrage zwischen Teilgebieten (*chaining/referral*)

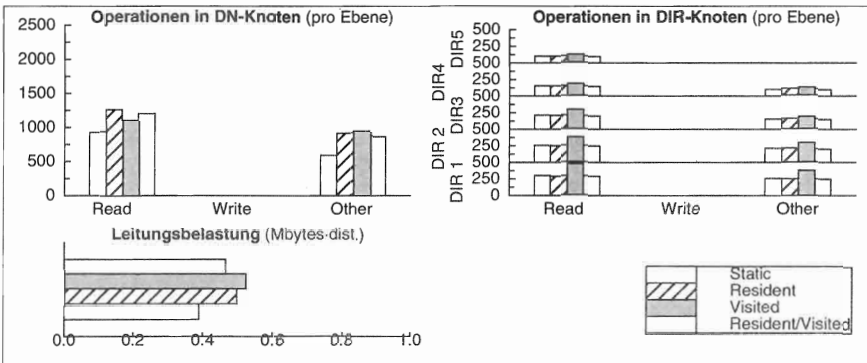
B.2.4. Einfluß der Gebietstopologie



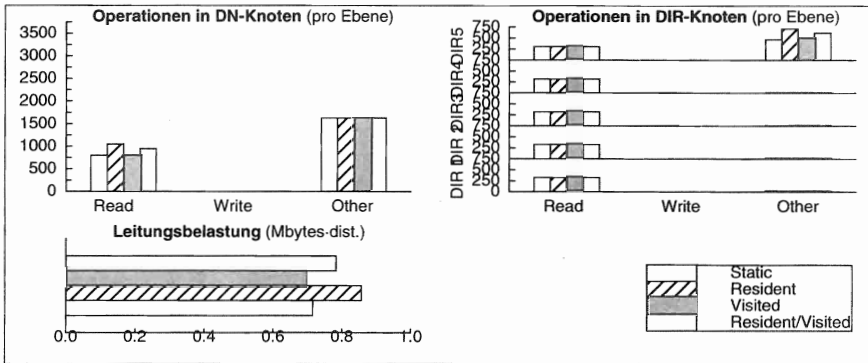
Diag. B.47 Variante 1: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)



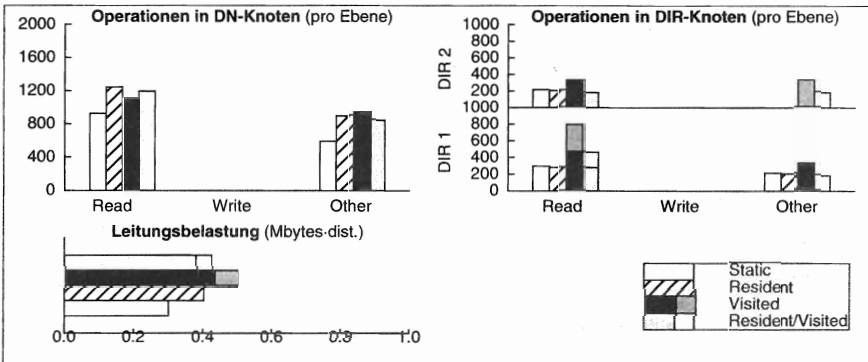
Diag. B.48 Variante 1: Suchanfrage zwischen Teilgebieten (*chaining/referral*)



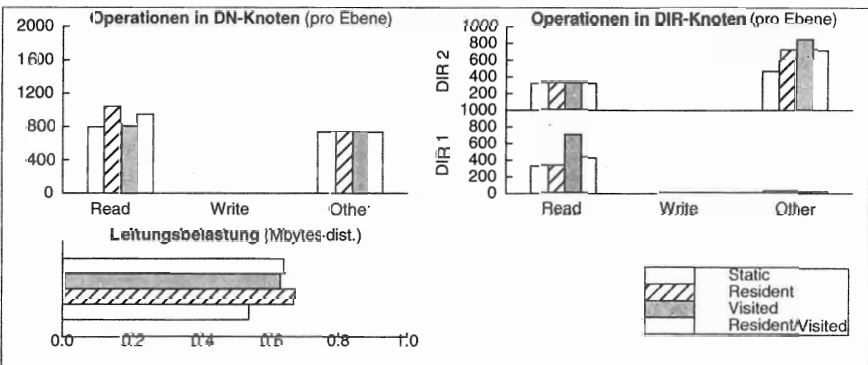
Diag. B.49 Variante 2: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)



Diag. B.50 Variante 2: Suchanfrage zwischen Teilgebieten (*chaining/referral*)



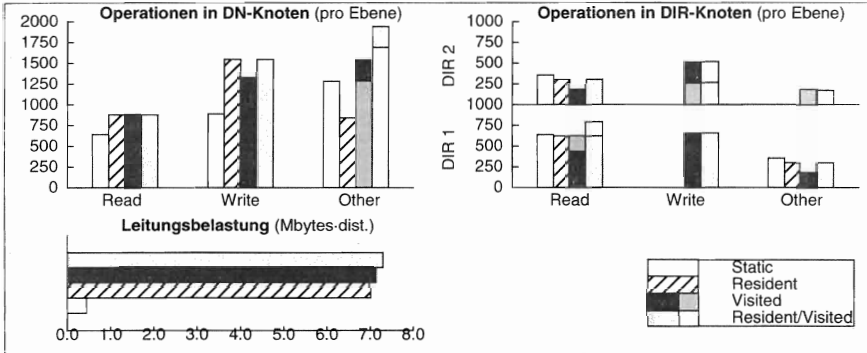
Diag. B.51 Variante 3: Suchanfrage innerhalb des Heimteilgebietes (*chaining/referral*)



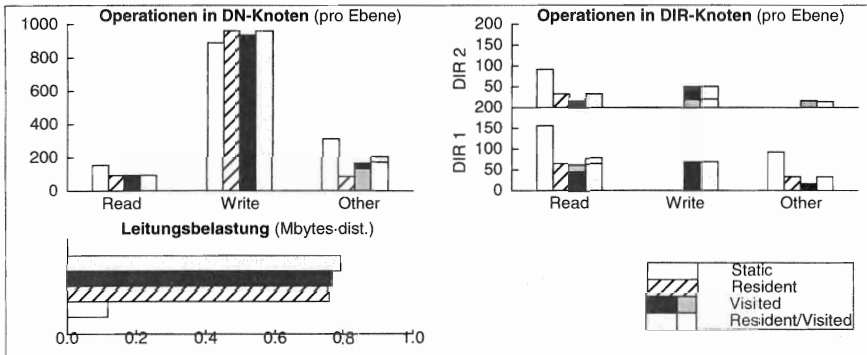
Diag. B.52 Variante 3: Suchanfrage zwischen Teilgebieten (*chaining/referral*)

B.3. Aktualisierungsanfrage

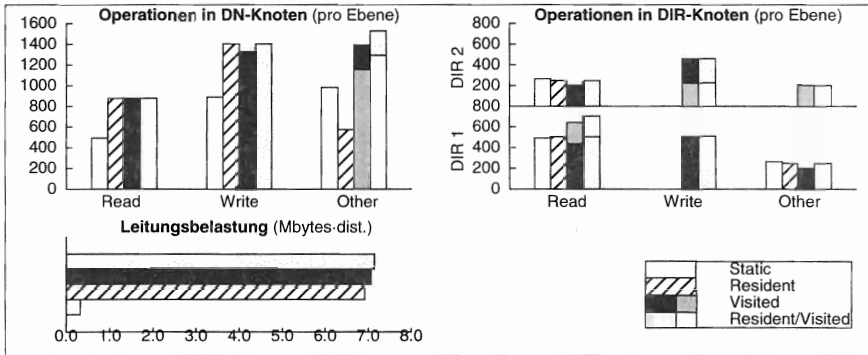
B.3.1. Einfluß der Mobilitätsparameter



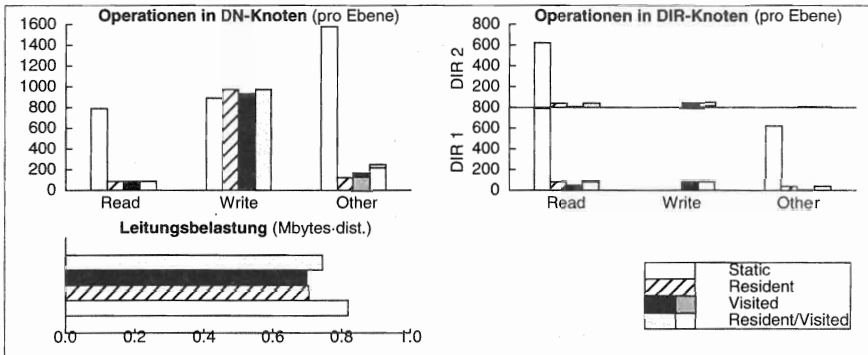
Diag. B.53 Variante 1: Aktualisierung innerhalb des Heimteilgebietes (chaining/referral)



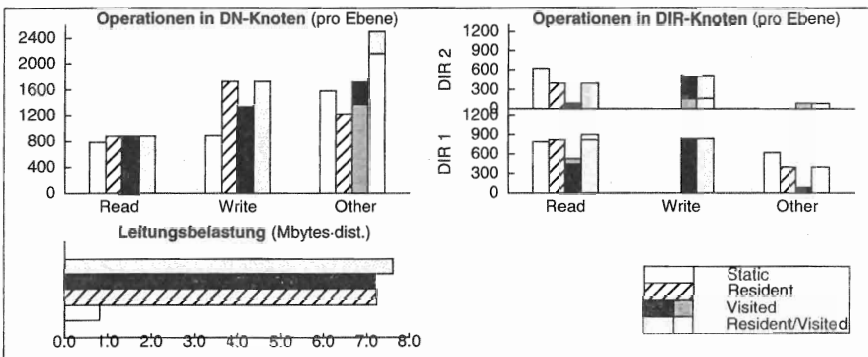
Diag. B.54 Variante 2: Aktualisierung innerhalb des Heimteilgebietes (chaining/referral)



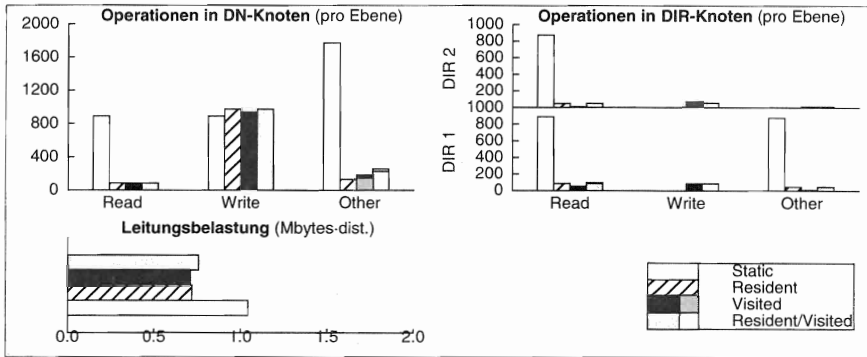
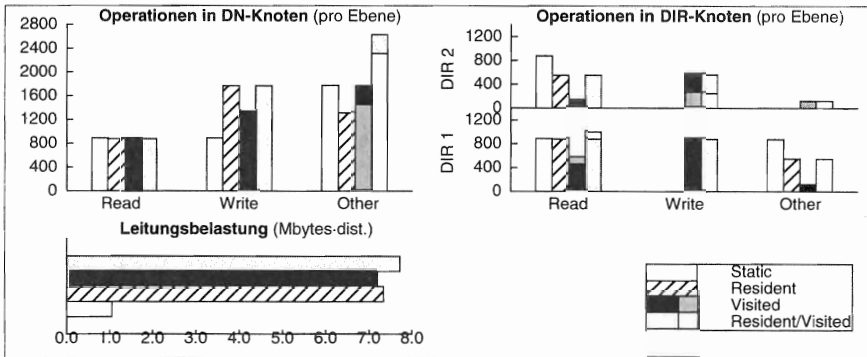
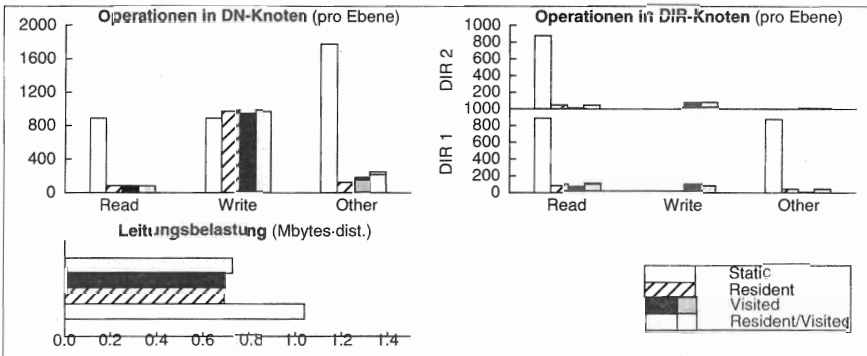
Diag. B.55 Variante 3: Aktualisierung innerhalb des Heimteilgebietes (chaining/referral)

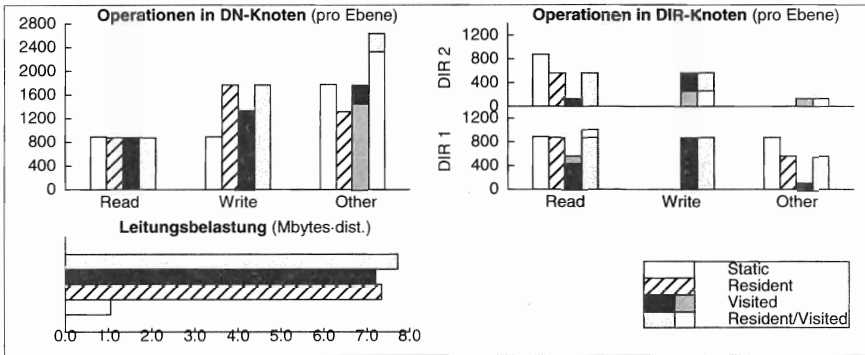


Diag. B.56 Variante 4: Aktualisierung innerhalb des Heimteilgebietes (chaining/referral)

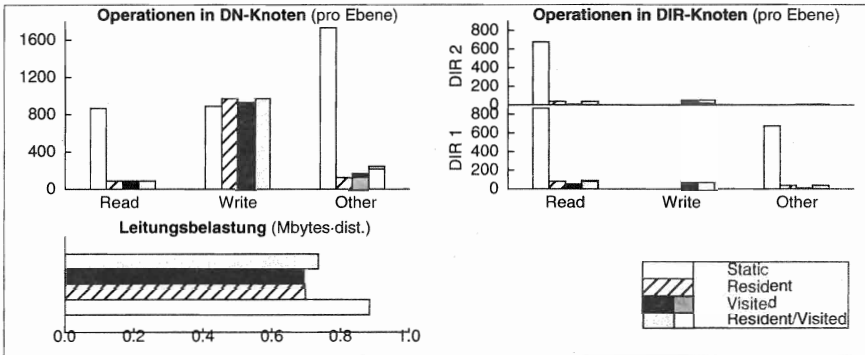


Diag. B.57 Variante 5: Aktualisierung innerhalb des Heimteilgebietes (chaining/referral)

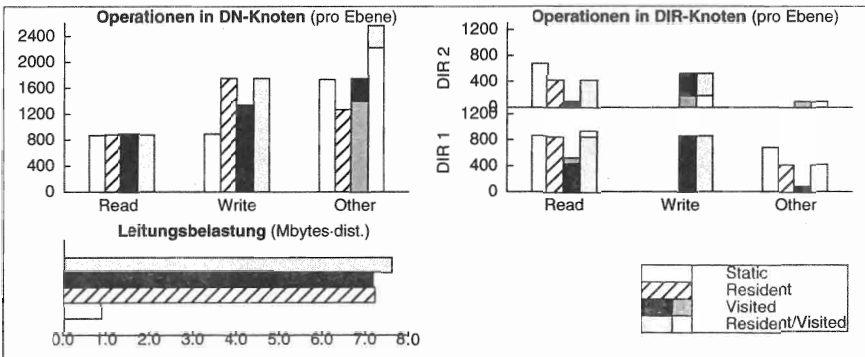
Diag. B.58 Variante 6: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)Diag. B.59 Variante 7: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)Diag. B.60 Variante 8: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)



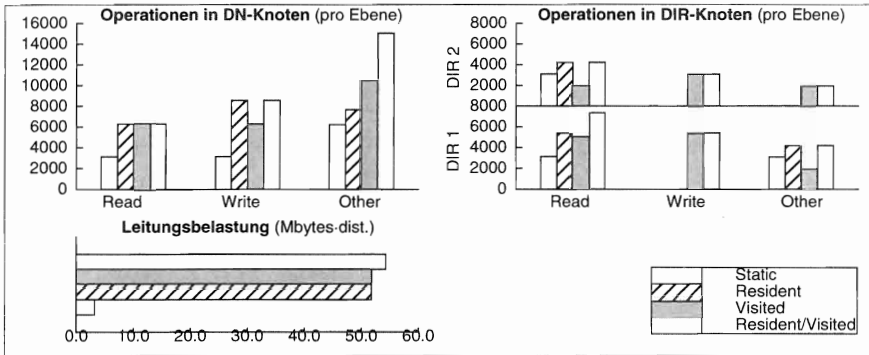
Diag. B.61 Variante 9: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)



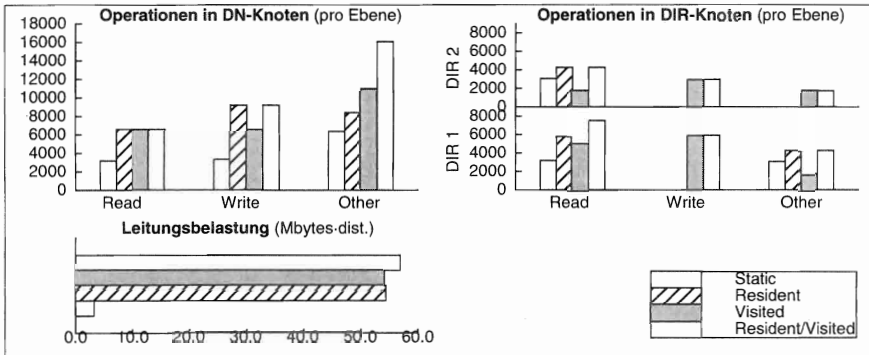
Diag. B.62 Variante 10: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)



Diag. B.63 Variante 11: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)

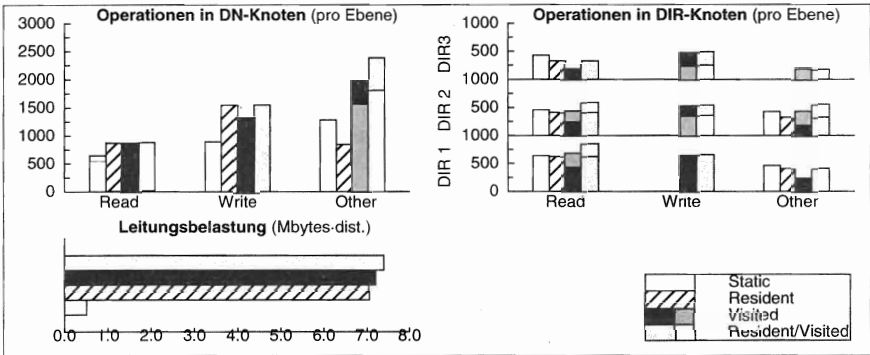


Diag. B.64 Variante 12: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)

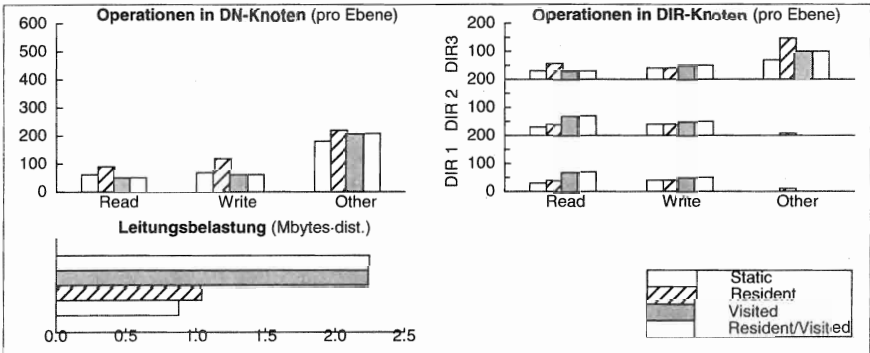


Diag. B.65 Variante 13: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)

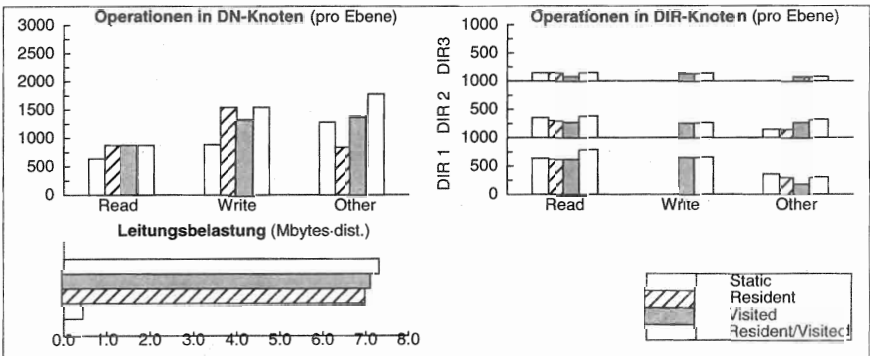
B.3.2. Einfluß der DIR-Architektur



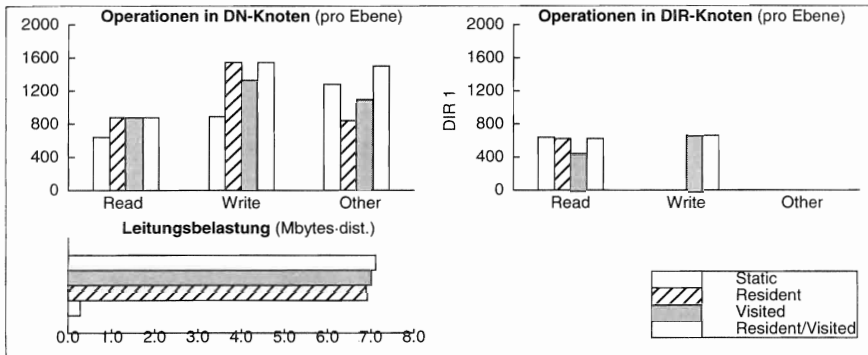
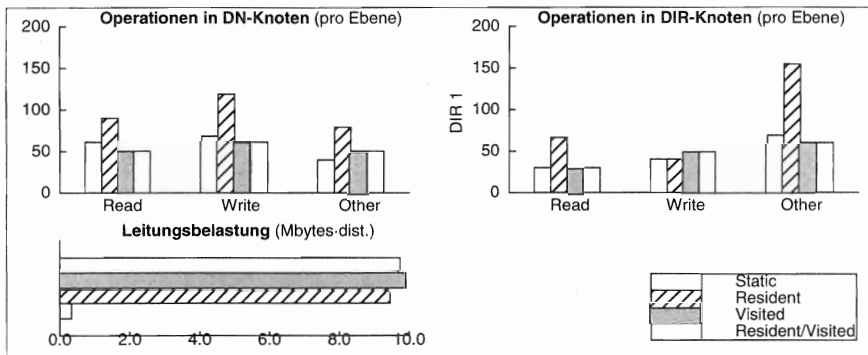
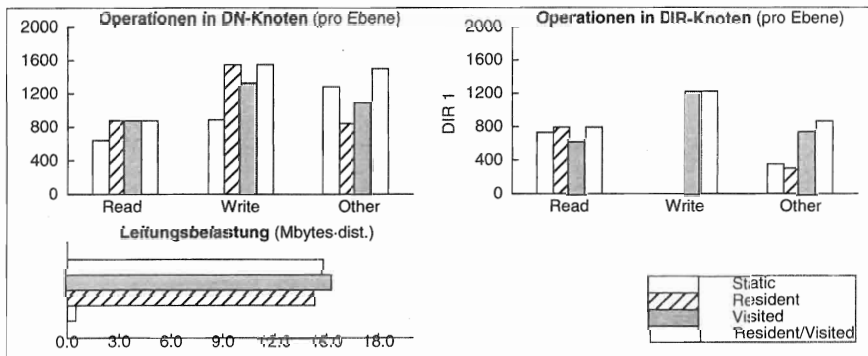
Diag. B.66 Variante 1a: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)

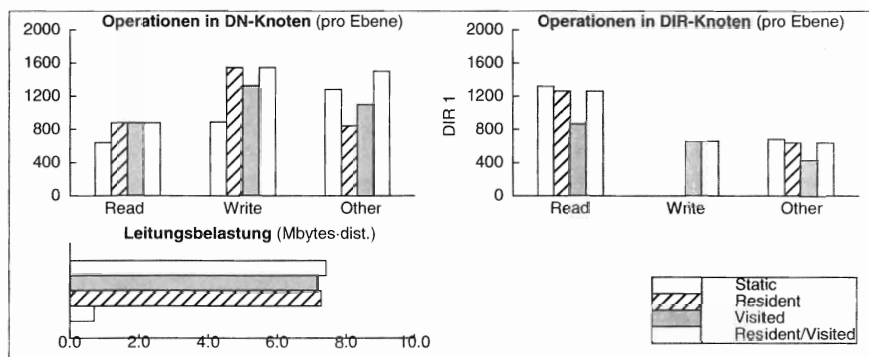


Diag. B.67 Variante 1a: Aktualisierung zwischen Teilgebieten (*chaining/referral*)

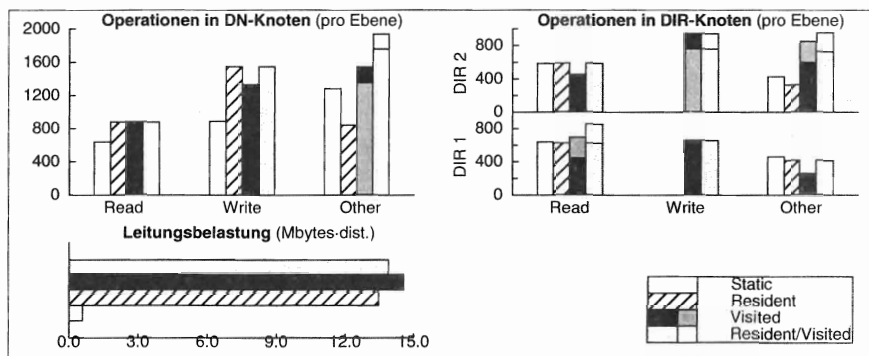


Diag. B.68 Variante 1b: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)

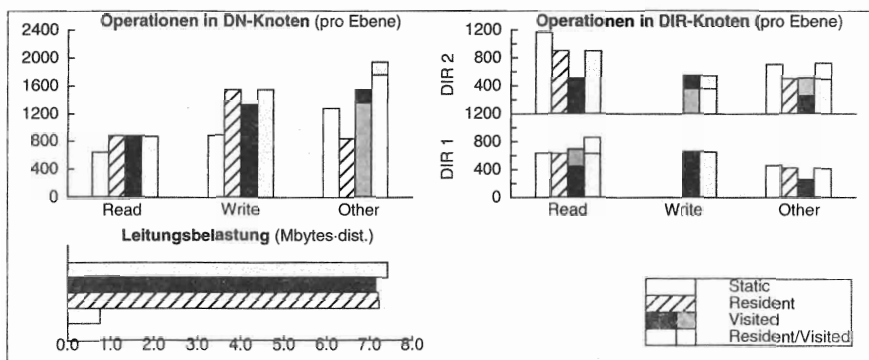
Diag. B.69 Variante 2: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)Diag. B.70 Variante 2: Aktualisierung zwischen Teilgebieten (*chaining/referral*)Diag. B.71 Variante 3a: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)



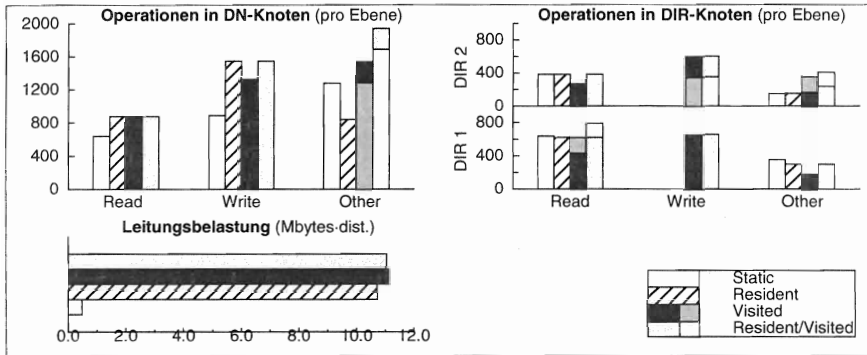
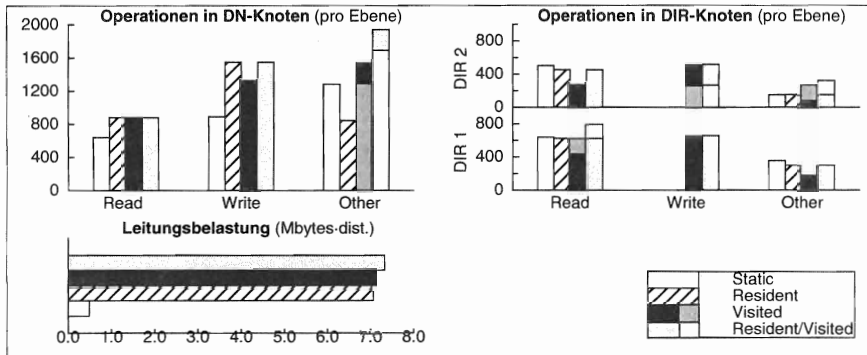
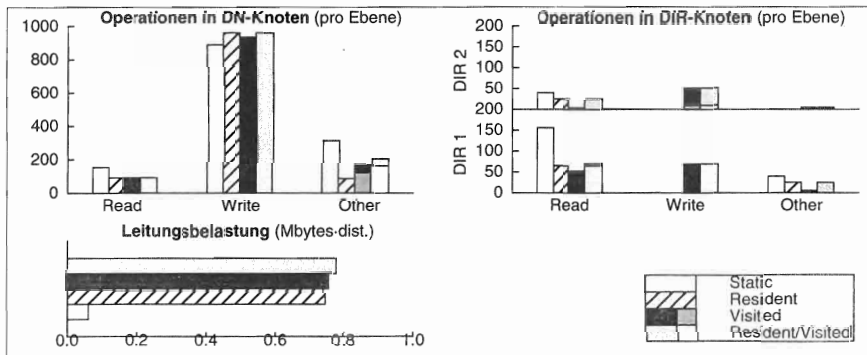
Diag. B.72 Variante 3b: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)

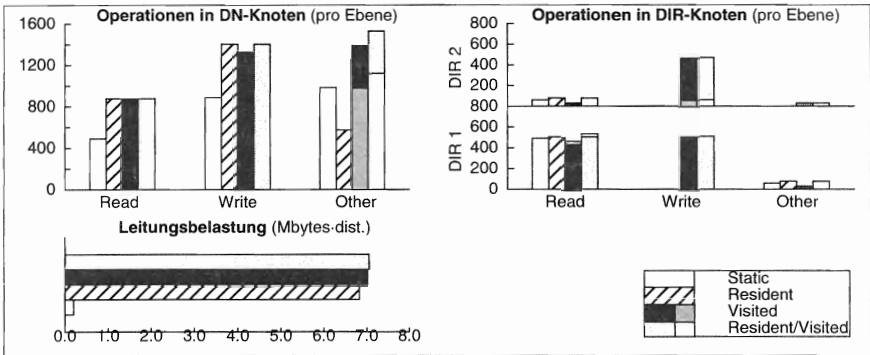


Diag. B.73 Variante 4a: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)

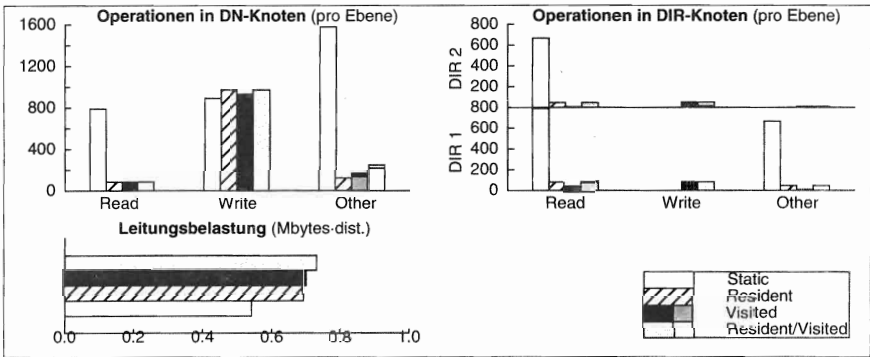


Diag. B.74 Variante 4b: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)

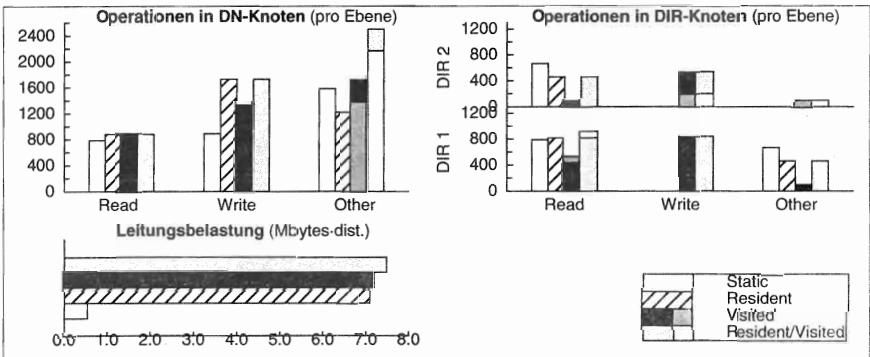
Diag. B.75 Variante 5a: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)Diag. B.76 Variante 5b: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)Diag. B.77 Variante 6: Aktualisierung innerhalb des Heimteilgebietes (*chaining/referral*)



Diag. B.78 Variante 7: Aktualisierung innerhalb des Heimteilgebietes (chaining/referral)

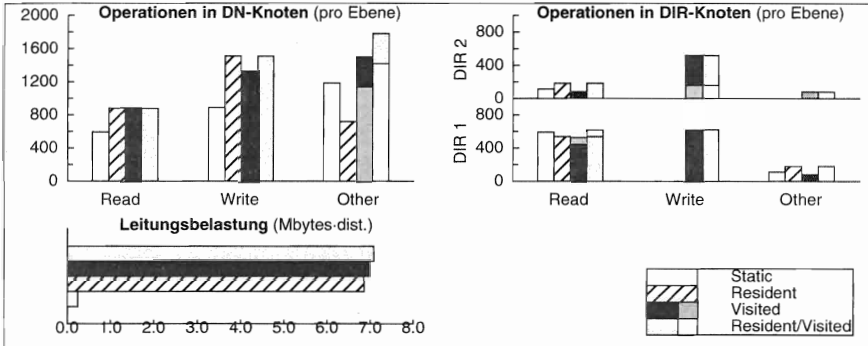


Diag. B.79 Variante 8: Aktualisierung innerhalb des Heimteilgebietes (chaining/referral)

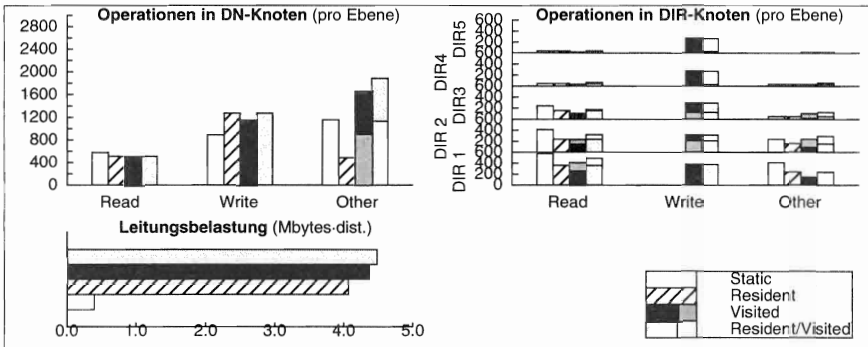


Diag. B.80 Variante 9: Aktualisierung innerhalb des Heimteilgebietes (chaining/referral)

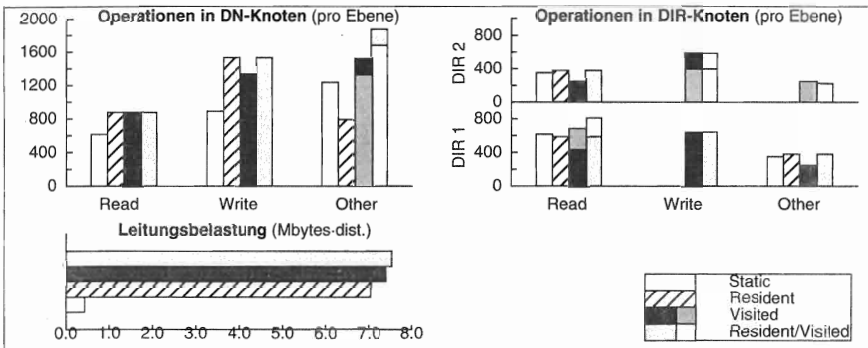
B.3.3. Einfluß der Gebietstopologie



Diag. B.81 Variante 1: Aktualisierung innerhalb des Heimteilgebietes (chaining/referral)

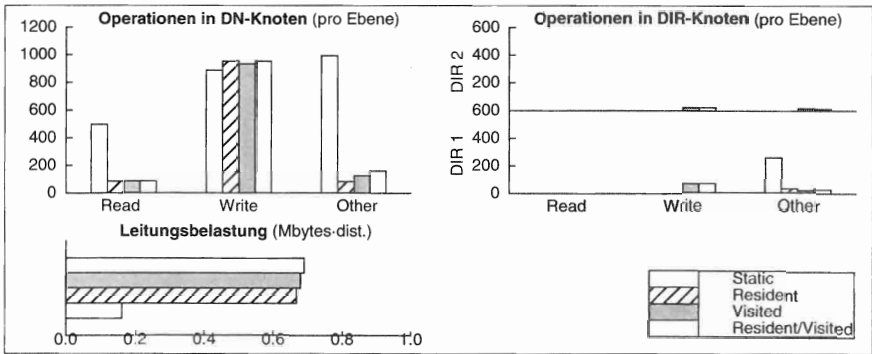


Diag. B.82 Variante 2: Aktualisierung innerhalb des Heimteilgebietes (chaining/referral)

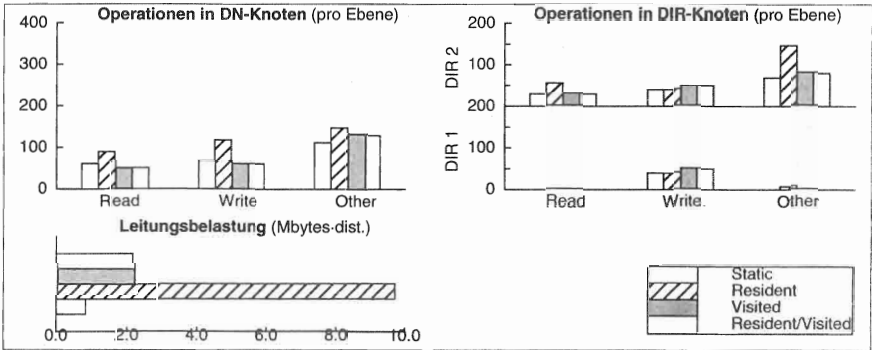


Diag. B.83 Variante 3: Aktualisierung innerhalb des Heimteilgebietes (chaining/referral)

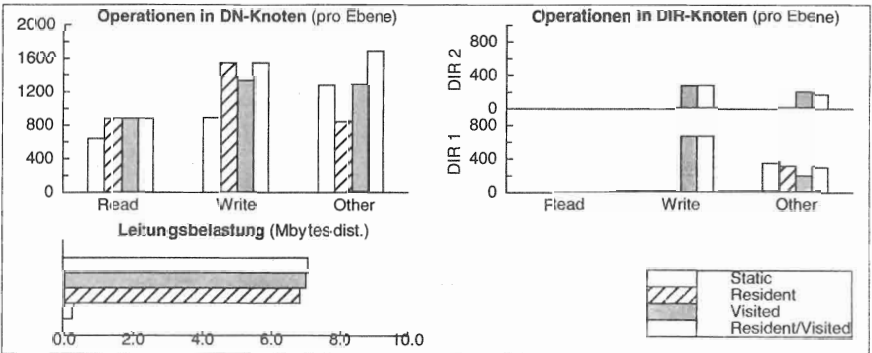
B.3.4. Einfluß des Vorwissens



Diag. B.84 Variante 1: Aktualisierung innerhalb des Heimteilgebietes (chaining/referral)



Diag. B.85 Variante 1: Aktualisierung zwischen Teilgebieten (chaining/referral)



Diag. B.86 Variante 2: Aktualisierung innerhalb des Heimteilgebietes (chaining/referral)

Lebenslauf

Damian Lawniczak

Persönliche Angaben:

Geburtsdatum 23. Januar 1964

Geburtsort Driesen

Ausbildung:

Herbst 1970 - Sommer 1978 Besuch der Grundschule in Driesen

Herbst 1978 - Sommer 1982 Besuch des Gymnasium in Driesen, abgeschlossen mit der Allgemeinen Hochschulreife

Herbst 1982 - Ende 1983 Studium der angewandten Mathematik an der Technischen Hochschule in Posen, unterbrochen durch die Aussiedlung in die Bundesrepublik Deutschland am 22. Dezember 1983

Februar 1984 - August 1984 Sprachkurs in Anrath/Mönchengladbach

Herbst 1984 - Sommer 1985 Sonderlehrgang für Spätausgesiedelte Abiturienten in Geilenkirchen zur Erlangung der Allgemeinen Hochschulreife

Oktober 1985 - Juli 1990 Studium der Elektrotechnik an der RWTH Aachen, Fachrichtung Nachrichtentechnik; im Frühling 1988 Aufnahme in die Studienstiftung des Deutschen Volkes

Oktober 1993 - Studium der Wirtschaftswissenschaften an der FernUniversität - Gesamthochschule in Hagen

Beruflicher Werdegang:

August, September 1985 Praktikum bei ASEAG in Kohlscheid

April, Mai 1986 Praktikum bei AUCOS GmbH in Aachen

15. August 1990 - 30. September '91 Beschäftigt bei FEV Motorentechnik GmbH in Aachen als Entwicklungsingenieur

Seit 1. Oktober 1991 Beschäftigt bei Ericsson Eurolab Deutschland GmbH in Herzogenrath als Forschungsingenieur

