

**Modellierung und Bewertung von  
Verfahren zur Last- und Leistungsregelung  
in Steuereinheiten von  
B-ISDN/ATM-Vermittlungssystemen**

Von der Fakultät Elektrotechnik und Informationstechnik  
der Universität Stuttgart  
zur Erlangung der Würde eines Doktor-Ingenieurs (Dr.-Ing.)  
genehmigte Abhandlung

vorgelegt von  
**Albrecht B. Schwarz**  
geboren in Stuttgart

Hauptberichter: Prof. Dr.-Ing. Dr. h.c. mult. P. J. Kühn  
Mitberichter: Prof. Dr.-Ing. P. Tran-Gia  
Tag der mündlichen Prüfung: 8. Mai 2002

Institut für Nachrichtenvermittlung und Datenverarbeitung  
der Universität Stuttgart  
2002

# **Modelling and Evaluation of Load and Performance Control Mechanisms of B-ISDN/ATM Switching Systems**

## **Summary**

Objective of this thesis are load control and performance control concepts for broadband switching systems. Focus is the services integrating network technology B-ISDN using ATM as transfer mode. The studied mechanisms and concepts are principally of generic nature. Specifically they are designed within the envisaged context of B-ISDN, due to its extensive vision with respect to service integration, Quality of Service (QoS) support and ATM bearer capabilities.

Area of application is implicitly the network control plane, but interactions between user and control plane have to be considered, too. The prime scope are switching nodes between access and core network domain, i.e., B-ISDN switches which have to provide user-to-network and network-to-network signalling protocol functions. Thus, beside service distinction call type differentiation is also covered due to the considered network positioning.

Chapter 1 provides a brief introduction and a general motivation for the need of overload control schemes in broadband switching nodes.

Chapter 2 summarises the protocol specific and technological background, which has to be taken into account. Services and signalling protocol capabilities are crucial in the context of B-ISDN load regulation functions because they are directly associated with the characterisation of the control plane traffic. Signalling traffic is the decisive input parameter of overload control loops. A crucial node-internal architecture property is the specific design of a control path in B-ISDN nodes. Appropriate processor cluster structures are categorised and discussed for a generic ATM switch.

Chapter 3 discusses comprehensively all necessary details, which are required for the design of closed control loops. Many overload control mechanisms and features are presented due to the long tradition of this important function in telephone and computer communication networks. New aspects and challenges in the context of broadband networks are evaluated and elaborated. New requirements for the design of load control methods are identified.

The presentation of a novel framework for load and performance control in B-ISDN nodes is subject of chapter 4. The basic idea of that new concept are multiple cooperating control loops with different spatial and temporal control horizons. Focus are three specific algorithms with different, but complementary scope. The assumed control platform shall be a generic multi-processor architecture organised in a distributed hierarchical manner.

Chapter 5 deals with performance evaluation and validation aspects. In order to investigate overload phenomenons and to evaluate the performance of load control strategies, analytical and simulation methods are employed. The event-driven simulation technique for nonstationary studies is briefly outlined. Transient and time-dependent behaviour are the major motivation for using the instationary simulation as basic evaluation tool. Further reasons are service modelling and complexity, non-linearity, and detailed modelling level of the processor cluster and control loop system.

Chapter 6 presents results of a load control scheme targeting new incoming service request messages. This signalling plane related admission control mechanism is based on service specific metrics. A set of similar services is represented by a certain traffic class. Each class is using a rough estimation of CPU processing resources and expected call respectively session holding time. The underlying principle is comparable to the user plane connection admission control (CAC). The key difference is that there is much more uncertainty because there doesn't

exist any "traffic contract" in the control plane. Controllability and robustness are achieved by bounding incoming requests in a preventive manner already under normal load conditions. The specific mechanism uses the rejection possibility in layer 3 signalling protocols. It has to be noted that the controller adapts the used reference parameters repetitively according the difference between estimated and measured load.

Chapter 7 illustrates a second controller aiming another time and load range. Objective is to optimise the performance by means of a specific compromise metric, called *Power*. This trade-off strategy is taking into account contradictory customer and operator specific expectations using the quotient of throughput and response time. In order to be able to support different service requirements this metric is extended by consideration of multiple traffic classes. The feedback controller is therefore labelled as Power-Product Control.

The spatial control domain of both above mentioned mechanisms is limited on single processor units. High capacity broadband switching nodes are providing typically hierarchical organised multiprocessor systems in order to be scalable in functional as well as on load level. Such complex distributed CPU structures should be complemented by a global scheme besides local limited control instruments. Chapter 8 analyses a totally distributed load balancing mechanism with focus on the overall switch control platform. Further specialities of this globally acting load distribution algorithm are multiple traffic class support and avoidance of blind load by separating the control modules from the controlled system.

Finally, the conclusions appear in chapter 9 followed by the list of used references. The concluding remarks emphasise again the significant properties of the novel load and performance control schemes. Major results of this report are summarised and a brief outlook is given.

# Inhaltsverzeichnis

<b>Abkürzungen</b> .....	<b>XIII</b>
<b>Formelzeichen</b> .....	<b>XV</b>
<b>1 Einleitung</b> .....	<b>1</b>
1.1 Dienstintegrierende Breitbandkommunikation .....	1
1.2 Überlastproblematik und Laststeuerung .....	1
1.3 Motivation und Zielsetzung dieser Arbeit .....	2
1.4 Gliederung der Arbeit .....	3
<b>2 Rechnergesteuerte Vermittlungssysteme und Breitbandkommunikation</b> .....	<b>5</b>
2.1 Wichtige Merkmale von Vermittlungssystemen .....	5
2.1.1 Typische Funktionen .....	6
2.1.2 Einsatzmöglichkeiten von Vermittlungssystemen .....	7
2.1.3 Aufbau und Systemsteuerung .....	8
2.1.4 Anforderungen und Leistungskennwerte .....	9
2.2 Klassifizierung von Vermittlungssystemen .....	9
2.2.1 Leitungsvermittelnde Netzknoten .....	9
2.2.2 Speichervermittelnde Netzknoten .....	9
2.2.3 Leistungsgrößen von Vermittlungssystemen .....	10
2.3 Vermittlungssteuerungen .....	11
2.3.1 Anforderungen und grundsätzliche Steuerungsfunktionen .....	11
2.3.2 Software – Anforderungen und Aufbau .....	12
2.3.3 Der Vermittlungsprozess .....	13
2.3.4 Rechenzeitintensive Steuerungsfunktionen .....	15
2.3.4.1 Signalisierung .....	15
2.3.4.2 Verkehrslenkung .....	15
2.3.4.3 Rufbearbeitung .....	16
2.3.5 Last- und Funktionsteilung .....	16
2.3.6 Architekturen und Strukturen von Steuerungseinheiten .....	17
2.4 Dienste und Protokolle des Breitband-ISDN .....	17
2.4.1 Der Dienstbegriff .....	17
2.4.2 Klassifizierungsmerkmale von Diensten .....	18
2.4.3 Dienstintegration .....	18

2.4.4	Tele- und Übermittlungsdienste	19
2.4.5	Neue und zukünftige Dienste	20
2.4.5.1	„Gesprächs-“ bzw. Kommunikationsbeziehungen	21
2.4.5.2	Internet – Datendienste	22
2.4.5.3	ATM-Übermittlungsdienst	23
2.4.6	Breitband-ISDN im Kontext des Intelligenten Netzes	23
2.4.7	Charakterisierung von Diensten hinsichtlich ihres Steuerungsaufwandes	23
2.5	Zeichengabe im Breitband-ISDN	24
2.5.1	Grundlegende Merkmale	24
2.5.2	Anforderungen im Breitband-ISDN	25
2.5.3	Schnittstellen und Protokollübersicht	26
2.5.4	ATM-Anpassungsschicht für die Signalisierung (SAAL)	28
2.5.4.1	Wichtige Funktionen des <i>Service Specific Connection Oriented Protocol</i>	28
2.5.4.2	Überwachung der Signalisierstrecken	28
2.5.4.3	Realisierung der Schicht-2-Signalisierung in Netzelementen	29
2.5.5	Zeichengabe am Netzzugang: Teilnehmersignalisierung	29
2.5.5.1	Meldungstypen und -format	29
2.5.5.2	Anforderung von neuen Rufen und Verbindungen	30
2.5.5.3	Modifikation von Verbindungsparametern	32
2.5.5.4	Modifikation der Verbindungszahl	32
2.5.5.5	Zustandsautomat und Protokolltimer	32
2.5.6	Zeichengabe im Netz: Zwischenamts-Signalisierung	32
2.5.6.1	Öffentliche Schnittstellen	32
2.5.6.2	Private Schnittstellen	33
2.5.7	Signalisierszenarien	33
2.6	B-ISDN/ATM-Vermittlungssysteme	33
2.6.1	ATM-Vermittlungstechnik	34
2.6.2	Generische Architektur einer B-ISDN/ATM-Vermittlungsstelle	35
2.6.2.1	Spezifische Funktionen	35
2.6.2.1.1	Nutzer-Ebene	35
2.6.2.1.2	Steuerungs-Ebene	36
2.6.2.1.3	Management-Ebene	37
2.6.2.1.4	Schnittstellenunterstützung	38
2.6.2.2	Bearbeitungsschritte beim Ruf- und Verbindungsaufbau	38
2.6.2.3	Funktionsblöcke und generische Architektur	39
2.6.3	Kennwerte von ATM-Vermittlungssystemen	40
2.6.4	Aufbau von Steuerungseinheiten	40
2.6.4.1	Implementierungskonzepte	41
2.6.4.2	Hardwareaspekte der Steuerung	41
2.6.4.3	Softwareaufbau	42
2.6.4.4	Verteilung von Aufgaben: Last- und Funktionsteilung	42
2.6.4.5	Interne Steuerungskonzepte	42

2.6.5	Klassifizierung von Steuerungseinheiten .....	43
2.7	ATM-Knoten für heterogene Netzumgebungen .....	45
<b>3</b>	<b>Problematik der Lastregelung in B-ISDN-Vermittlungsrechnern .....</b>	<b>47</b>
3.1	Einführung .....	47
3.1.1	Wichtige Begriffe .....	48
3.1.2	Das Überlastphänomen .....	49
3.1.2.1	Leistungseinbruch bei hohen Lasten .....	49
3.1.2.2	Wechselwirkung zwischen Teilnehmer und Vermittlungssystem .....	50
3.1.2.3	Weitere Effekte .....	50
3.1.3	Leistungsmaße .....	51
3.1.3.1	Elementare Kenngrößen .....	51
3.1.3.2	Zusammengesetzte Kenngrößen .....	53
3.2	Ressourcenbedarf .....	55
3.2.1	Mittel zur Charakterisierung von Rufbeziehungen .....	55
3.2.2	Zustandsraum .....	56
3.3	Überlastursachen und -situationen .....	56
3.3.1	Einführung .....	57
3.3.2	Überlast aus Sicht des Architekturmodells .....	57
3.3.3	Überlast in verschiedenen Netztypen .....	57
3.3.3.1	Paket- bzw. Rechnernetze, Internet .....	57
3.3.3.2	Fernsprech- bzw. Telekommunikationsnetze .....	58
3.3.3.3	Signalisiernetze .....	58
3.3.3.4	Intelligentes Netz .....	58
3.3.4	Überlast im Normalbetrieb .....	58
3.3.5	Überlast durch spezielles Teilnehmerverhalten .....	59
3.3.6	Überlast aufgrund Systemausfällen und Naturkatastrophen .....	59
3.4	Überlasterkennung und Regeleinrichtungen .....	60
3.4.1	Regelkreiskomponenten .....	61
3.4.2	Allgemeine Regelziele .....	61
3.4.3	(Über-)Lastindikatoren .....	61
3.4.4	Beobachtung und Messung von Kenngrößen .....	62
3.4.4.1	Problematik der Messdatenerfassung und -aufbereitung .....	62
3.4.4.2	Eignung und Einsatzbereiche bestimmter Kenngrößen .....	64
3.4.4.3	Allgemeine Definition der Systembelastung .....	64
3.4.4.4	Messproblematik von Systemzustandsgrößen .....	64
3.4.5	Zustandsgrößenschätzung .....	65
3.4.6	Störsignalfilterung .....	66
3.4.7	Festlegung der Zeitkonstanten – Mess-, Schätz- und Regelintervalle .....	67
3.4.8	Anforderungen und Kriterien an Überlastabwehreinrichtungen .....	68

3.4.9	Lastregel- und Überlastabwehrverfahren	70
3.4.9.1	Klassifizierungsschema für Wirkungsbereich	70
3.4.9.2	Konventionelle Methoden	70
3.5	Beispiele aus etablierten Netztechnologien	71
3.6	Neue Aspekte und Herausforderungen durch die Dienstintegration in Breitbandnetzen	73
3.6.1	Dienstintegration und Verkehrscharakteristik	73
3.6.1.1	Allgemeine Aspekte	73
3.6.1.2	Konsequenzen für eine Vermittlungssteuerung	74
3.6.1.3	Weitere Dienstaspekte	74
3.6.1.4	Auswirkung Internet	75
3.6.2	Abschätzung des Ressourcenbedarfs	75
3.6.2.1	Identifizierung eines bestimmten Dienstes	75
3.6.2.2	Nennkapazität als BHCA-Wert	76
3.6.2.3	Ansätze für äquivalente „BHCA“-Definitionen in Breitbandnetzen	77
3.6.3	Vergabe und Verwaltung von Ressourcen	77
3.6.4	Wechselwirkungen zwischen Nutzer- und Steuerungs-Ebene	79
3.6.4.1	Ressourcenverwaltungen	79
3.6.4.2	Durch aktive Nutzverbindungen induzierte Steuerungslast	80
3.6.5	Steuerungsarchitektur	81
3.7	Vorschläge und bekannte Lösungsansätze	81
<b>4</b>	<b>Neues Konzept einer mehrstufigen hierarchischen Regelkreisstruktur zur Lastkontrolle und Leistungsoptimierung</b>	<b>83</b>
4.1	Randbedingungen und Merkmale	83
4.1.1	Verkehr	83
4.1.2	Vermittlungsrechner	84
4.1.3	Regelkreisstruktur	84
4.1.4	Bewertungskriterien	84
4.2	Generische Steuerungsarchitektur	85
4.2.1	Übersicht	85
4.2.2	Gewählte Funktionsverteilung	87
4.2.3	Merkmale der Lastverteilung	87
4.3	Regelungstechnischer Ansatz – Mehrstufige Regelung	88
4.3.1	Räumliche Ebenen	88
4.3.2	Berücksichtigte Zeitebenen	89
4.3.3	Lastebenen	90
4.3.4	Verfahren zur Last- und Leistungsregelung	90
4.4	Annahmekontrolle für neue Dienstanforderungen	93

4.4.1	Problematik und Anforderungen	93
4.4.2	Zentrale Ressource	95
4.4.2.1	Lastcharakterisierung	95
4.4.2.2	Effektive Prozessorbelastung	97
4.4.3	Algorithmus	99
4.4.3.1	Übersicht	99
4.4.3.2	Annahmeregulung	99
4.4.3.3	Nachführung von Parametern	101
4.4.3.4	Weitere Diskussion	104
4.4.4	Abgrenzung zu bekannten Verfahren	104
4.5	Behandlung neuer Dienstanforderungen	106
4.5.1	Motivation – Prinzip der Leistungsregelung	106
4.5.2	Neuer Algorithmus	106
4.5.2.1	Übersicht	106
4.5.2.2	Berechnung der Zurückweisungswahrscheinlichkeiten	108
4.5.2.2.1	Motivation	109
4.5.2.2.2	Variante I – Schwellwertverfahren	109
4.5.2.2.3	Variante II – Differenzenverfahren	110
4.5.2.2.4	Variante III – Tripelverfahren	110
4.5.3	Diskussion der Power-Produkt-Funktion	113
4.6	Algorithmen zur Lastzuteilung von A- an B-Prozessorknoten	114
4.6.1	Einführung	114
4.6.2	Bekanntes Verfahren für Vermittlungsrechner	114
4.6.3	Besonderheiten der neu entwickelten Algorithmen	115
4.6.4	Distributed Load Balancing – Algorithmen	117
4.6.4.1	Erfassung der Lastzustandsinformationen	117
4.6.4.2	Verfahren auf Basis der Antwortzeit	119
4.6.4.3	Verfahren auf Basis des Durchsatzes	120
4.6.4.4	Verfahren auf Basis von Power-Kenngrößen	121
4.6.5	Weitere Diskussion	122
<b>5</b>	<b>Modellierung und angewandte Analyseverfahren</b>	<b>124</b>
5.1	Prinzipien der verkehrstheoretischen Modellierung und Analyse	124
5.2	Modellierung von Anwendungen und Diensten	125
5.2.1	Prinzipien der Quellmodellierung	125
5.2.1.1	Modellierungshierarchie in ATM-Systemen	125
5.2.1.2	Quellmodelle für die Steuerungs-Ebene	125
5.2.2	Modellierung klassischer Fernsprechrufe	126
5.2.3	Bekanntes Ansätze zur Modellierung neuer Dienste	126
5.2.3.1	Verkehrsklassen	126
5.2.3.2	Szenariodetaillierung	127
5.2.3.3	Verteilung der Belegungszeiten	127



5.2.4	Modellierung des Signalisierverkehrs unter Berücksichtigung neuer Dienste	128
5.2.5	Anforderungen an die Dienstmodellierung in Breitbandnetzen	128
5.2.6	Modellierung des Steuerungsverkehrs in dieser Arbeit	129
5.2.6.1	Problematik	129
5.2.6.2	Signalisierszenarien ausgewählter Dienste	129
5.2.6.3	Repräsentative Auswahl bestimmter Dienste	130
5.2.6.4	Lastmodellierung	130
5.2.6.5	Gewählte Dienstmodellierung und Verkehrsklassen	131
5.3	Teilnehmerverhalten, Endgeräteaspekte und Protokollmechanismen	132
5.3.1	Problematik	132
5.3.2	Modellierung von Rückkopplungseffekten	133
5.4	Modellierung von Vermittlungsrechnern	133
5.4.1	Einführung	133
5.4.2	Einfache Modelle für analytische Untersuchungen	134
5.4.3	Komplexere Modelle	134
5.4.4	Entworfene und verwendete Simulationsmodelle	135
5.4.4.1	Anforderungen	135
5.4.4.2	Der <i>N</i> -Server als Elementarmodell eines Prozessorknotens	135
5.4.5	Bedienzeitverteilung	137
5.5	Methoden zur Leistungsuntersuchung	138
5.5.1	Analytische Verfahren	138
5.5.2	Simulationstechniken	138
5.5.2.1	Ereignisgesteuerte Simulation	138
5.5.2.2	Simulation stationärer Vorgänge	139
5.5.2.3	Simulation transients Vorgänge	139
5.6	Technik der instationären Simulation	139
5.6.1	Übersicht	139
5.6.2	Erfahrungen auf dem Gebiet der Überlastabwehr	140
5.6.3	Flexibles Werkzeug für die instationäre Untersuchung von Vermittlungsrechnern	141
<b>6</b>	<b>Untersuchung der Annahmekontrolle für neue Kommunikationsbeziehungen</b>	<b>142</b>
6.1	Fallstudien im Überblick	142
6.1.1	Zielsetzung	142
6.1.2	Fixe Modell- und Verkehrsparameter	143
6.2	Automatische Parameternachführung	143
6.2.1	Lernen der sogenannten Originalparameter	143
6.2.1.1	Stationäre Normallast	144
6.2.1.2	Verhalten unter starker Überlast	145

6.2.2	Einfluss der Adaptionintervalldauer und anderer Parameter	146
6.2.3	Mess- und Schätzverfahren	147
6.2.4	Lastprognosealgorithmus	148
6.2.4.1	Effektive Prozessorbelastung	148
6.2.4.2	Ausblick – Weitere Planungsparameter	148
6.3	Annahmealgorithmus	149
6.3.1	Zurückweisung neuer Dienstanforderungen	149
6.3.2	Zusätzliche Verwerfung	150
6.3.3	Verhalten unter „extremer Überlast“	151
6.4	Bewertung und Ausblick	152
<b>7</b>	<b>Untersuchung der Algorithmen für die lokale Leistungsregelung eines Dienstbearbeitungsprozessors</b>	<b>153</b>
7.1	Übersicht	153
7.2	Führungsgröße ‚Power-Produkt‘	153
7.2.1	Arbeitsbereich	153
7.2.2	Schnittbilder des Leistungsprofils und stationäre Werte	155
7.3	Optimumsuche	158
7.4	Kurze Hochlastimpulse	159
7.4.1	Annahme aller Anforderungen	159
7.4.2	Zurückweisung neuer Anforderungen	160
7.5	PPC als reine Überlastabwehr	162
7.5.1	Vereinfachte Berechnung der Zurückweisungswahrscheinlichkeiten	162
7.5.2	Messungen und Bewertung	163
7.6	Bewertung und Ausblick	164
<b>8</b>	<b>Untersuchung der Verfahren zur globalen Lastverteilung</b>	<b>165</b>
8.1	Übersicht	165
8.1.1	Feste Parametereinstellungen	165
8.1.2	Bewertungskriterien für die Gesamtleistung	166
8.2	Verfahren auf Basis der Antwortzeit	168
8.3	Verfahren auf Basis der Durchsatzkenngröße	170
8.4	Verfahren auf Basis von Power-Maßzahlen sowie Vergleich der drei Algorithmen	170
8.4.1	Inhomogener Verkehr und lokale Überlastabwehr	170
8.4.2	Inhomogener Verkehr und deaktivierte Überlastabwehr	172
8.4.3	Interne Lastschwankungen	172
8.5	Bewertung und Ausblick	173

<b>9</b>	<b>Zusammenfassung und Ausblick</b> .....	<b>175</b>
	<b>Literaturverzeichnis</b> .....	<b>180</b>
	Lehrunterlagen und Fachbücher .....	180
	Veröffentlichungen und wissenschaftliche Arbeiten .....	183
	Patente .....	214
	Empfehlungen, Normen, Standards .....	216
	Datenblätter, Produktschriften .....	221
<b>A</b>	<b>Steuerungsstrukturen ausgewählter ATM-Netzknoten</b> .....	<b>223</b>
A.1	Entwicklungstendenzen in der Steuerungsarchitektur .....	223
A.2	Auswahl repräsentativer Systeme .....	225
A.2.1	Zentraler Vermittlungsrechner .....	225
A.2.1.1	A1000 Multimedia Multiservice Switch (Alcatel) .....	225
A.2.1.2	AXD 301 (Ericsson) .....	225
A.2.1.3	Magellan Concorde (Nortel) .....	226
A.2.1.4	FETEX-150 ATM ESP B-160 (Fujitsu) .....	226
A.2.1.5	MainstreetXpress (Siemens/Newbridge) .....	227
A.2.2	Verteilte Vermittlungssteuerungen .....	227
A.2.2.1	TDX-ATM (ETRI) .....	227
A.2.2.2	GlobeView-2000 (Lucent) .....	228
A.2.2.3	ATOMNET/M20 (NEC) .....	228
A.2.2.4	Hyper-Distributed System (Hitachi) .....	229
A.2.3	Weitere Informationen zu ATM-Systemen .....	229
A.2.3.1	Bekannte Steuerungsstruktur .....	229
A.2.3.2	Unbekannte Steuerungsstruktur .....	230
A.3	Beispiele konkreter Lastregelverfahren .....	231
A.3.1	Load Balancing – Siemens/Newbridge .....	231
A.3.2	Mehrstufige Regelkreisstruktur – Lucent/AT&T .....	233
<b>B</b>	<b>Leistungsverhalten einfacher Warteschlangensysteme</b> .....	<b>234</b>
B.1	M/M/N-System – Multiprozessorstruktur .....	234
B.2	M/M/1/∞-System – Begrenzte Ressource ‚CPU‘ .....	234
B.2.1	Grundgleichungen .....	234
B.2.2	System mit Entmutigung .....	235
B.3	M/M/1/K-System – Begrenzte Ressource ‚Speicher‘ .....	236
B.3.1	Ohne Überlastabwehr .....	237
B.3.2	Überlastabwehr auf Basis der Warteschlangenbelegung .....	240
B.3.2.1	Stationäres Verhalten .....	241
B.3.2.2	Dynamisches Verhalten .....	244

B.4	Stationäre Leistungsmaße des $M_1M_2/M_1M_2/1/K$ -Systems	245
<b>C</b>	<b>Dienstszenarien und Simulationsmodelle</b>	<b>255</b>
C.1	Verkehrsmodelle	255
C.1.1	Modell für einen Dualprozessor	255
C.1.2	Multiprozessor mit Last- und Funktionsteilung	256
C.2	Simulationsmodelle	258
C.3	Signalisierszenarien	258
C.3.1	Anforderungen für die Modellierung	258
C.3.2	Signalisierabläufe für reale Dienste	258
C.3.2.1	Basisruf	258
C.3.2.2	IN-Dienst	260
C.3.3	Exemplarische Szenarien für die Leistungsuntersuchung	260
C.3.3.1	Dualprozessorstruktur	260
C.3.3.2	(2, 3, 1)-Steuerungsstruktur	262
C.4	Gewählte Bediendauerverteilung	262
<b>D</b>	<b>Exemplarische Messdaten und Ergebnisse der DLB-Algorithmen</b>	<b>263</b>
D.1	Sensorik beim Power-Produkt-Verfahren	263
D.2	Aktorik und Vertrauensintervalle	265
D.3	Leistungswerte beim Power-Produkt-Verfahren	266
D.4	Bewertung der drei verschiedenen Algorithmen	268

## Abkürzungen

AAL	ATM Adaptation Layer	DBR	Deterministic Bit Rate
ABR	Available Bit Rate	DIN	Deutsche Industrie Norm
ABT	ATM Block Transfer	DLB	Distributed Load Balancing
ACM	Association for Computing Machinery	DSS	Digital Subscriber Signalling System
AI	Adaptation Interval (Adaptionsintervall)	DSSx	DSS No. x (digitales Signalisierungsverfahren Nr. x auf der Kunden-Anschlussleitung)
ANSI	American National Standards Institute	ETRI	Electronics and Telecommunications Research Institute (Südkorea)
APS	Anlagenprogrammssystem	ETSI	European Telecommunications Standards Institute
ATM	Asynchroner Transfermodus	FDDI	Fiber Distributed Data Interface
ATM-F	ATM-Forum	FIFO	First-In, First-Out
AT&T	American Telephone & Telegraph Company	GLOBECOM	IEEE Global Communications Conference
BHCA	Busy Hour Call Attempts (Belegungsversuche in der Hauptverkehrsstunde)	GLR	Globale Last-/Leistungsregelung
B-ICI	Broadband Inter-Carrier Interface	GoS	Grade of Service (Verkehrsgüte)
B-ISDN	Broadband Integrated Services Digital Network	ICC	International Conference on Communications
B-ISSI	Broadband Inter-Switching System Interface	ICCC	International Conference on Computer Communications (→ INFOCOM) bzw. International Conference on Computer Communication <sup>1</sup>
B-ISUP	Broadband Integrated Services User Part	ICI	Inter-Carrier Interface
BL	Basic Load (Grundlast) oder Background Load (Hintergrundlast)	IEEE	Institute of Electrical and Electronic Engineers
B-SSP	Broadband Service Switching Point	IEICE	The Institute of Electronics, Information and Communication Engineers
CAC	Connection Admission Control	IETF	Internet Engineering Task Force
CAM	Content Addressable Memory	IFIP	International Federation for Information Processing
CAPS	Call Attempts per Second	IISP	Interim Inter-Switch Signalling Protocol (P-NNI Phase 0)
CBR	Constant Bit Rate	ILMI	Interim Local Management Interface
CCITT	Comité Consultatif International Téléphonique et Télégraphique bzw. Consultative Committee on International Telephone and Telegraph (jetzt ITU)	IN	Intelligentes Netz <sup>2</sup>
CCS7	Common Channel Signalling No. 7 (Zentralkanal-Zeichengabeverfahren Nr. 7), siehe auch SS7	INAP	Intelligent Network Application Part
CES	Circuit Emulation Service	INFOCOM	IEEE International Conference on Computer Communications
CHILL	CCITT High Level Language	IP	Internet Protocol
CI	Control Interval (Regelintervall)	ISDN	Integrated Services Digital Network (Integration von PSTN, CSPDN, PSPDN u. a. Dienste), diensteintegrierendes digitales Kommunikationsnetz
CLP	Cell Loss Priority	ISO	International Organization for Standardization
CLR	Cell Loss Ratio	ISP	Internet Service Provider
CO	Central Office	ISS	International Switching Symposium
CP	Control Processor (Steuerprozessor)		
CPCS	Common Part Convergence Sublayer		
cps	Calls per Second		
CPU	Central Processing Unit		
CR	Call Reference		
CS	Capability Set (B-ISDN Signalling)		
CSPDN	Circuit Switched Public Data Network (leitungsvermittelltes, öffentl. Datennetz)		
CTD	Cell Transfer Delay		
DATEX	Data-Telegraph-Exchange		

<sup>1</sup> Ausrichter: *International Council for Computer Communication*.

<sup>2</sup> Trennung von Verbindungs- und Dienststeuerung in einem Netz. Die wichtigsten Netzelemente sind die *dienstneutralen Vermittlungsknoten (SSP)* und der zentrale *Dienststeuerungsknoten (SCP)*.

ISSI	Inter-Switching System Interface	PPC	Power-Product Control (Leistungsregelungsverfahren basierend auf der Kenngröße <i>Power-Produkt</i> )
ITC	International Teletraffic Congress	PSPDN	Packet Switched Public Data Network (paketvermitteltes, öffentl. Datennetz)
ITG	Informationstechnische Gesellschaft	PSTN	Public Switched Telephone Network (öffentliches Telefonvermittlungsnetz)
ITT	International Telephone & Telegraph Corporation	PVC	Permanent Virtual Circuit
ITU	International Telecommunication Union (früher CCITT)	PVCC	Permanent Virtual Channel Connection
ITU-TSS	ITU Telecommunication Standardization Sector	PVPC	Permanent Virtual Path Connection
LAN	Local Area Network	QoS	Quality of Service (Dienstgüte)
LAPD	Link Access Procedure D	RA	Regelalgorithmus
LIFO	Last-In, First-Out	RE	Regeleinrichtung
LLR	Lokale Last-/Leistungsregelung	RFC	Request for Comments
LX	Local Exchange (→ TInVSt)	RM	Ressource Management
M	Memory (Speicher)	RPC	Repeated Call Attempt
MAN	Metropolitan Area Network	RS	Regelstrecke
ME	Messeinrichtung	SAAL	Signalling ATM Adaptation Layer
MFV	Mehrfrequenz-Verfahren	SAP	Service Access Point
MI	Measurement Interval (Messintervall)	SAR	Segmentation and Reassembly
MPLS	Multiprotocol Label Switching	SBR	Statistical Bit Rate
Msg	Message	SCCP	Signalling Connection Control Part
MTP	Message Transfer Part	SCP	Service Control Point (Dienststeuerungsknoten)
μP	Mikroprozessor	SDH	Synchronous Digital Hierarchy
NEC	Nippon Electric Company	SDL	Specification and Description Language
NNI	Network-to-Network Interface oder Network Node Interface (ATM Forum)	SDP	Service Data Point
N-ISDN	Narrowband ISDN (Schmalband-ISDN)	SE	Stelleinrichtung
NP	Network Performance (Netzgüte)	SMDS	Switched Multi-Megabit Data Service
NPC	Network Parameter Control oder Network Policing Control	SP	Service Processor
NRAC	New Request Admission Control (Annahmeverfahren für neue Dienstansforderungen)	SPC	Stored-Program Control
NTG	Nachrichtentechnische Gesellschaft	SS7	Signalling System Number 7
NTT	Nippon Telegraph and Telephone Corporation	SSCF	Service Specific Coordination Function
OAM	Operations, Administration, Maintenance (Betrieb, Verwaltung, Wartung)	SSCS	Service Specific Convergence Sublayer
PABX	Private Automatic Branch Exchange (→ TKAnl)	SSCOP	Service Specific Connection Oriented Protocol
PC	Personal Computer	SSP	Service Switching Point (Dienstvermittlungsknoten)
PCS	Personal Communication System	STM	Synchronous Transfermodus
PDH	Plesiochronous Digital Hierarchy	STM- <i>n</i>	Synchronous Transport Module- <i>n</i>
POCS	Plain Old Computer System [2]	STP	Signaling Transfer Point
POTS	Plain Old/Ordinary Telephone Service	STS- <i>n</i>	Synchronous Transport Signal- <i>n</i>
POTV	Plain Old TV Service	SVC	Switched Virtual Circuit oder Signalling Virtual Channel
PMP	Point to Multipoint	SVCC	Switched Virtual Channel Connection
PN	Processor Node (Prozessorknoten)	SVPC	Switched Virtual Path Connection
PNNI	Private Network-to-Network Interface oder Private Network Node Interface	TCAP	Transaction Capabilities Application Part
PoP	Point of Presence	TCP	Transmission Control Protocol
		TDM	Time Division Multiplexing
		TEDE	Teledialogeinrichtung <sup>3</sup> (auch kurz TED)

Th	Threshold (Schwellwert)
TKAnl	Telekommunikationsanlage (= Nebenstellenanlage, NStAnl)
TInVSt	Teilnehmervermittlungsstelle
TMN	Telecommunication Management Network
TX	Transit Exchange (→ Transitvermittlungsstelle)
UBR	Unspecified Bit Rate
UNI	User-Network Interface
UPC	Usage Parameter Control oder User Policing Control
URL	Uniform Resource Locator
UUS	User-to-User Signalling
VBR	Variable Bit Rate
VC	Virtual Circuit

<sup>3</sup> Abstimmdienst für Rundfunkanstalten zur aktiven Beteiligung des Publikums an Hörfunk- und Fernsehsendungen.

VCC	Virtual Channel Connection
VCFC	Virtual Channel Flow Control
VCI	Virtual Channel Identifier
VDE	Technisch-Wissenschaftlicher Verband der Elektrotechnik, Elektronik und Informationstechnik e. V.
VoD	Video on Demand
VoIP	Voice over IP
VP	Virtual Path
VPC	Virtual Path Connection
VPCI	Virtual Path Connection Identifier
VPI	Virtual Path Identifier
VPN	Virtual Private Network
VSt	Vermittlungsstelle
WAN	Wide Area Network
WWW	World-Wide Web
ZE	Zeiteinheit
ZV	Zufallsvariable
ZZK	Zentraler Zeichengabekanal (CCS7)

## Formelzeichen

$A$	Verkehrsangebot [Erlang] ( $= \lambda \cdot t_H$ )
$B$	Blockier-/Verlustwahrscheinlichkeit oder Büschelfaktor
$C$	Verkehrsmenge
$E[X]$	Erwartungswert der Zufallsvariable $X$
$F(x)$	Verteilungsfunktion (VF)
$f(x)$	Verteilungsdichtefunktion (VDF)
$g$	Gewichtungsfaktor
$i, j$	ganzahlige Zählvariable
$K$	Anzahl der Zustände eines Warteschlangensystems, Anzahl der Verkehrsklassen
$p$	Wahrscheinlichkeit
$P(X=x)$	Wahrscheinlichkeit, dass die Zufallsvariable $X$ den Wert $x$ annimmt
$p(x)$	Wahrscheinlichkeit für den Zustand $x$
$r$	Maß für Ressourcenbedarf z.B. $\hat{r}_{p,v}^{org}$ für Ressourcenbedarf an CPU-Zeit für Verkehrsklasse $v$ als sogenannter Originalschätzparameter (siehe Abschnitt 4.4.3.3, S. 101)

$R$	Gesamtbedarf an Ressourcen
$\mathfrak{R}$	Menge bestimmter Ressourcen (Zustandsraum)
$s$	Anzahl von Pufferplätzen einer Warteschlange (Warteplätze) <sup>4</sup>
$T_A$	Zufallsvariable für den Ankunftsabstand
$T_H$	Zufallsvariable für die Bediendauer
$T_S$	Zufallsvariable für die Dauer einer Kommunikationsbeziehung
$t$	Zeit
$t_H, h$	mittlere Bediendauer ( $= E[T_H]$ )
$X$	Zufallsvariable
$x$	tatsächlicher Wert einer Kenngröße, allgemeine Variable
$\lceil x \rceil$	kleinste ganze Zahl größer oder gleich $x$
$\lfloor x \rfloor$	größte ganze Zahl kleiner oder gleich $x$
$\bar{x}$	Mittelwert
$\tilde{x}$	Messwert
$\hat{x}$	Schätzwert

<sup>4</sup> z.B.  $s = K - 1$  beim  $M/D/1/K$ -System

$X(t_i)$	Zufallsvariable für den Systemzustand zum Zeitpunkt $t_i$
$x_i$	Systemzustand zum Zeitpunkt $t_i$
$Y$	Verkehrswert
$\alpha$	Dämpfungskonstante
$\delta$	Grenzwert
$\Delta$	Intervall, Differenz
$\phi$	Rate (Fluss)
$\gamma$	mittlerer Durchsatz
$\lambda$	mittlere Ankunftsrate oder allgemein als mittlere Meldungsrate
$\Lambda$	Leistungsmaß (z.B. Nutzdurchsatz $\Lambda_\phi$ , Antwortzeit $\Lambda_\tau$ , Power $\Lambda_P$ , Power-Produkt $\Lambda_{P^II}$ , Power-Summe $\Lambda_{P^Z}$ , Power-Produkt-Produkt $\Lambda_{P^II, II}$ u. a.)
$\mu$	mittlere Bedienrate
$\pi$	Zuweisungswahrscheinlichkeit
$\Theta$	Belastungsgrad (normiertes Verhältnis vergebener Betriebsmittel)
$\rho$	mittlere Auslastung (einer Prozessoreinheit)
$\sigma$	Standardabweichung
$\tau$	mittlere Durchlaufzeit <sup>5</sup> für ein System (auch Antwort- oder Aufenthaltszeit)
$\Omega$	mittlere Warteschlangenlänge
$\xi$	Glättungs- oder Alterungsfaktor
$\zeta$	Überlastfaktor oder Anzahl aktiver Kommunikationsbeziehungen

### Zusätzliche Indizes

a) Verkehrsströme (Signalisiernachrichten)	
$AC$	Index zur Kennzeichnung von angenommenen Rufen ( <i>Accepted Calls</i> )
$BC$	Index zur Kennzeichnung von blockierten Rufen ( <i>Blocked Calls</i> ) <sup>6</sup>
$CC$	Index zur Kennzeichnung von komplettierten Rufen ( <i>Completed Calls</i> )
$DC$	Index zur Kennzeichnung von gelöschten Rufen ( <i>Discarded Calls</i> )

<sup>5</sup> gleich Summe aller mittleren Bedien- und Wartezeiten

<sup>6</sup> Der Ruf geht aufgrund einer vollen Warteschlange verloren, ohne Belegung des Betriebsmittels ‚Prozessor‘ und damit ohne Inanspruchnahme von CPU-Zeit.

$FC$	Index zur Kennzeichnung von neuen Rufen ( <i>Fresh Calls</i> )
$OL$	Index zur Kennzeichnung von angebotenem Verkehr ( <i>Offered Load</i> ) <sup>7</sup>
$RC, Rej$	Indizes zur Kennzeichnung abgelehnter Rufe ( <i>Rejected Calls</i> ) <sup>8</sup>
$RM$	Index zur Kennzeichnung neuer Dienstanforderungen ( <i>REQUEST Message</i> )
$v$	Index zur Kennzeichnung einer Verkehrsklasse

### b) Lastregelung

$M$	Index zur Kennzeichnung von Messvorgängen (z.B. $t_{M,i}$ für $i$ -ter Messzeitpunkt)
$R$	Index zur Kennzeichnung von Regelvorgängen (z.B. $\Delta_R$ für Regelintervall)
$S$	Index zur Kennzeichnung von Schätzvorgängen

### c) Sonstige

$C$	Index zur Kennzeichnung der Steuerungsebene ( <i>Control Plane</i> )
$HL$	Index zur Kennzeichnung von Hochlast
$NL$	Index zur Kennzeichnung von Normal- oder Niedriglast
$U$	Index zur Kennzeichnung der Nutzer-Ebene ( <i>User Plane</i> )
$ÜL$	Index zur Kennzeichnung von Überlast
$VS$	Index zur Kennzeichnung eines Verlustsystems
$WS$	Index zur Kennzeichnung eines Wartesystems

<sup>7</sup> Z.B. Überlagerung der Verkehrsströme neuer Rufe, Signalisiermeldungen innerhalb eines Rufes, Rufwiederholungen, knoteninterne Steuermeldungen für die Prozessoreinheit eines Vermittlungsrechners.

<sup>8</sup> Der Unterschied zwischen blockierten, gelöschten und abgelehnten Rufanforderungen repräsentiert sich in den entsprechend notwendigen Prozessorbearbeitungszeiten:  $t_{H,BC} (= 0) < t_{H,DC} < t_{H,RC}$ .



# Kapitel 1

## Einleitung

### 1.1 Dienstintegrierende Breitbandkommunikation

1965 erfolgte die Inbetriebnahme des *Electronic Switching System No. 1* (1ESS), dem ersten Vermittlungssystem mit elektromechanischem Koppelfeld *und rechnerunterstützter Verbindungssteuerung* [31, 118, 224, 300]. Damit begann das Zeitalter der rechnergesteuerten Vermittlungstechnik. In der Folgezeit wurde die durchgehende Digitalisierung sowohl der Übertragungstechnik auf den Verbindungsleitungen als auch der Vermittlungstechnik in den Netzknoten vollzogen. Begleitet wurde diese Entwicklung durch den zunehmenden Einsatz von Mikroprozessoren in Netzelementen und damit dem rapiden Wachstum an verfügbarer Rechenleistung.

Neben der Telekommunikation mit Fernsprechen als zentralem Dienst begann Ende der sechziger Jahre der Aufbau von Rechnernetzen und damit das Zeitalter der Rechner- und Datenkommunikation. Betrachtet man die vermittlungstechnischen Aspekte, dann basieren diese Rechnernetze vornehmlich auf dezentralen Vermittlungssteuerungen (LAN-Protokolle) und Routern als Paketvermittlungsknoten auf einer höheren Netzebene. Die aktuelle Popularität des Internets und seiner Dienste hat mittlerweile dazu geführt, dass das dadurch erzeugte Verkehrsvolumen das der traditionellen Sprachtelefonie übersteigt.

Nachfolgende Schlüsseltechnologien waren und sind auch zukünftig noch die wichtigsten antreibenden Kräfte für die technische Entwicklung der Kommunikationsnetze: zunehmende Integrationsdichte auf dem Gebiet der Mikroelektronik und der Halbleitertechnik, Verfügbarkeit leistungsfähiger optischer und opto-elektronischer Komponenten sowie Fortschritte in der Softwaretechnik und der Bereitstellung leistungsfähiger Entwicklungswerkzeuge.

Die zunehmende Verfügbarkeit an Prozessorleistung in Vermittlungssystemen ermöglichte auch die Dezentralisierung lokaler Systemfunktionen, die Implementierung umfangreicher Signalisierungsprotokolle, die Bereitstellung einer Vielzahl von Dienst- und Leistungsmerkmalen und die Integration verschiedener Dienste in eine Netztechnologie. Die logische Konsequenz dieser technischen Möglichkeiten war die Definition dienstintegrierender digitaler Netze wie das ISDN sowie das Breitband-ISDN als dessen Fortführung. Als Zielsetzung für das B-ISDN wurde die Bereitstellung einer universellen Kommunikations-Infrastruktur formuliert, auf der alle heutigen und zukünftigen Kommunikationsdienste und -anwendungen aufgesetzt werden können.

### 1.2 Überlastproblematik und Laststeuerung

Für die Nutzung eines bestimmten Dienstes sind von einem Kommunikationsnetz gewisse Ressourcen bereitzustellen. Da sowohl vermittlungs- als auch steuerungstechnische Betriebs-

mittel nur endlich verfügbar sind, müssen zur Einhaltung garantierter Dienstgütern vielfältige Steuerungs- und Überwachungsfunktionen vorgesehen werden. Während sich diese Aufgaben primär direkt auf die individuellen Kommunikationsbeziehungen (z.B. Ruf oder Verbindung) selbst beziehen, sind in einem Vermittlungssystem oder auch für einen größeren Netzbereich zusätzlich übergeordnete Steuerungsfunktionen notwendig. So sind auch für den zuverlässigen Betrieb rechnergesteuerter Vermittlungssysteme leistungsfähige Lastregelmechanismen unverzichtbar. Besonders die Erkennung von und Reaktion auf Überlasten stellt hohe Anforderungen an diese Systemfunktion. Dabei geht es in hier weniger um den Verkehr der Nutzer-Ebene (*User Plane*), sondern vielmehr um die Last in der Steuerungs-Ebene (*Control Plane*).

Moderne Vermittlungssysteme zeichnen sich durch immer umfangreichere Funktionalität und zunehmende Komplexität aus. Durch den vermehrten Einsatz von Mikrorechner-Komponenten entstehen aus struktureller und aus steuerungstechnischer Sicht immer leistungsfähigere und komplexere Systeme, sowohl in der Hard- als auch in der Software. Allgemein hat der Einzug der Rechnertechnik zu einem Anstieg an maschineller Intelligenz, aber auch zu einem dramatischen Anwachsen des Softwareanteils in einem Vermittlungssystem, gerade in neuen Netztechnologien geführt. Zusätzlich lässt sich die allgemeine Tendenz einer quantitativen Zunahme an Diensten mit zunehmend verbesserter qualitativer Ausgestaltung beobachten. Die dadurch entstehende Vielfalt an Diensten und Anwendungen führt aber auch zu stark unterschiedlichen Anforderungen in Bezug auf Bandbreite oder einzuhaltende Dienstgüte, was wiederum die Komplexität der zuständigen Steuerungsfunktionen erhöht.

### 1.3 Motivation und Zielsetzung dieser Arbeit

Die Arbeit beschäftigt sich mit Verfahren zur Überlastabwehr<sup>1</sup> und Leistungsregelung<sup>1</sup> in Vermittlungsrechnern<sup>2</sup> von Netzknoten. Aus der Vielfalt derzeit existierender Netztechnologien wird hier der Schwerpunkt auf Vermittlungssysteme des ATM-basierten B-ISDN gelegt. Neben den einfühend erwähnten, mehr allgemeinen Entwicklungstendenzen sind deshalb auch die spezifischen Anforderungen des B-ISDN zu betrachten.

Diese neuartige Problematik entsteht hauptsächlich durch die Vielfalt bereitzustellender Dienste [319], durch Änderungen in der Verkehrszusammensetzung und der zeitlichen Lastprofile, durch die Vielfalt von Zeichengabeschnittstellen und deren Protokollkomplexität, durch die aufwendigen Vermittlungs- und Dienststeuerfunktionen sowie durch zusätzliche Anforderungen an die nichtvermittlungstechnischen Aufgaben. All dies bedingt entsprechend ausgelegte Vermittlungssteuerungen, wobei neben der Vermittlungsleistung hauptsächlich die Eigenschaften Echtzeitfähigkeit, Zuverlässigkeit und Robustheit zu nennen sind. Da in der Konzeption einer Lastregelung immer auch die zugrundeliegende Steuerungsstruktur eine wichtige Randbedingung darstellt, sind auch Architekturmerkmale für diese neue Kategorie von Vermittlungssystemen zu berücksichtigen.

Betrachtet man die Verkehrsquellen, dann treten neben dem menschlichen Benutzer zunehmend rechnerbasierte Endgeräte als Lasterzeuger auf. Diese Endsystemtypen können im Vergleich zum Menschen ein grundsätzlich anderes ‚Teilnehmerverhalten‘ aufweisen, so dass beispielsweise bekannte Quellmodelle nicht mehr verwendet werden können.

---

<sup>1</sup> Die Begriffe *Überlastabwehr*, *Last-* und *Leistungsregelung* unterscheiden sich hauptsächlich hinsichtlich ihrer wirklichen Lastebene und des zu regelnden Kenngrößentyps (siehe Kapitel 3 für eine genauere Begriffsfassung). Im allgemeinen Kontext dagegen werden diese hier synonym verwendet.

<sup>2</sup> Unter einem *Vermittlungsrechner* wird die zentrale Steuerungseinheit eines Vermittlungssystems verstanden.

Im einzelnen werden in dieser Arbeit folgende *Ziele* verfolgt:

- Da der Entwurf von Überlastabwehrinrichtungen ein Forschungsgebiet großer Tradition ist, soll zu Beginn der Stand der Wissenschaft auf dem Gebiet der Überlastproblematik dargestellt werden.
- Der Mittelpunkt soll die Darstellung der neuen An- und Herausforderungen für die Last- und Leistungsregelung im Kontext von Breitband-ISDN und ATM bilden. Begleitet werden soll diese Diskussion durch eine Systematisierung gängiger ATM-Vermittlungssteuerungen vornehmlich anhand ihrer topologischen Struktur.
- Aus den entstandenen Anforderungen sollen neue Lösungsansätze entsprechender Regelkreisstrukturen aufgezeigt werden. Mehrere kooperierende Regelkreise ergänzen sich dabei zu einer mehrstufigen, hierarchischen Regelkreisstruktur.
- Aufgrund der Vielfalt möglicher Algorithmen sollen wenige ausgewählte Konzepte mit besonders innovativem Charakter näher ausgeführt und auch eingehend analysiert werden.
- Als Methode zur Leistungsuntersuchung ist für dieses Anwendungsfeld die instationäre Simulationstechnik prädestiniert. Aus Ermangelung eines entsprechenden Werkzeuges müssen für diese Arbeit Erweiterungen einer vorhandenen Simulationsplattform vorgenommen werden. Dies betrifft vornehmlich Komponenten zur Organisation einer instationären Simulation, die Lasterzeugung, die Messtechnik sowie Elemente für Regelkreise. Dabei soll das Werkzeug so flexibel gestaltet werden, so dass es für die Untersuchung generischer Vermittlungsrechner und Regelkreisstrukturen weiterverwendet werden kann.
- Weiterhin sollen geeignete Modelle von Vermittlungsrechnern und exemplarischer Dienstabläufe entworfen werden. In diesem Zusammenhang soll auch die Modellierung der Wechselwirkung zwischen Teilnehmer und Netzelementen diskutiert werden.
- Den Abschluss soll die simulative Untersuchung, Optimierung und Bewertung repräsentativer Algorithmen bilden. Infolge der Parameter Vielfalt können in diesen Fallstudien nur ausgewählte Effekte eingehender betrachtet werden.

Hinsichtlich der Vielschichtigkeit des betrachteten Problems ist keine erschöpfende Behandlung dieser Thematik im Rahmen einer abgeschlossenen Arbeit möglich, sondern es können nur ausgewählte Aspekte näher beleuchtet werden.

## 1.4 Gliederung der Arbeit

Die Strukturierung der Arbeit orientiert sich an den oben formulierten Zielen.

**Kapitel 2** gibt einen kurzen Abriss über die Entwicklung rechnergesteuerter Vermittlungssysteme sowie die relevanten Aspekte hinsichtlich der Breitbandkommunikation. Im Mittelpunkt stehen dabei Dienste und Zeichengabeprotokolle, da sie auf der betrachteten Zeitebene hauptsächlich den erzeugten Steuerungsverkehr charakterisieren. Ergänzt wird dieses Kapitel durch die Beschreibung eines generischen ATM-Knotens sowie einer einfachen Klassifizierung gängiger Steuerungsstrukturen.

Überlastabwehrverfahren sind schon seit Jahrzehnten, nämlich seit der Entwicklung der ersten Computer und rechnergesteuerter Kommunikationsnetze, Gegenstand von Leistungsuntersuchungen. **Kapitel 3** gibt eine Übersicht von Problemstellungen zu Überlastabwehrverfahren

sowie einen Einblick in dabei verwendete Lösungsansätze und erzielte Ergebnisse. Insbesondere sollen auch die neuen Aspekte und Fragestellungen, die durch Breitbandnetze aufgeworfen werden, diskutiert werden.

Aufbauend auf den herausgearbeiteten Randbedingungen und Anforderungen wird ein Rahmenwerk zur Last- und Leistungsregelung für B-ISDN-Vermittlungsrechner in **Kapitel 4** entworfen. Dieses Konzept basiert auf einer mehrstufigen Regelkreisstruktur für eine generische Vermittlungssteuerung. Weiterhin werden für drei ausgewählte Verfahren mit unterschiedlichen Zielsetzungen konkrete Algorithmen und Realisierungsansätze vorgestellt.

Funktion und Leistungsfähigkeit von Konzepten und Algorithmen zur Überlastabwehr müssen nachgewiesen werden. In der Verkehrstheorie haben sich hierfür Warteschlangenmodelle mit ihren Analyseverfahren bewährt. In **Kapitel 5** werden Modellierung von Verkehr und Vermittlungsrechnern erläutert sowie Methoden zur Leistungsuntersuchung vorgestellt. Die zugrundeliegenden Regelungseinrichtungen gehören zur Klasse der stochastischen, nichtlinearen Mehrgrößensysteme. Für die Untersuchung aufwendigerer Modelle mit ihren Regelalgorithmen muss deshalb die Simulationstechnik eingesetzt werden. Ein wichtiger Aspekt bei Regelkreisen stellt ihr dynamisches Verhalten, z.B. transiente Zustandsübergänge, dar. Aspekte der instationären Simulation sind deshalb ebenfalls Gegenstand dieses Kapitels. Ein wesentlicher Punkt dieser Arbeit war die Bereitstellung eines Werkzeuges für die Untersuchung von Lastregelalgorithmen für generische Steuerungsstrukturen. Hierzu wurde eine bestehende objektorientierte Simulationsbibliothek zum einen um Komponenten für die instationäre Simulation, zum anderen um Klassen für Regeleinrichtungen, erweitert.

In **Kapitel 6** erfolgt die Analyse eines *Annahmeverfahrens für neue Dienstanforderungen* (*New Request Admission Control*, NRAC), dessen Algorithmus auf dienstspezifischen Maßzahlen steuerungstechnischer Betriebsmittel basiert. Dieser präventiv wirkende Lastregelkreis hat eine ähnliche Zielsetzung wie die Verbindungsannahmesteuerung (*Connection Admission Control*, CAC) für die Nutzer-Ebene.

Für eine andere Zeit- und Lastebene ist die Leistungsregelung, die in **Kapitel 7** untersucht wird, konzipiert. Dieses Verfahren verwendet die abgeleitete Kenngröße *Power-Produkt*, die zum einen die Leistungsdaten verschiedener Verkehrsklassen berücksichtigt, und zum anderen einen Kompromiss hinsichtlich der Optimierung des Antwortzeitverhaltens und des Nutzdurchsatzes bildet.

Während sich der räumliche Wirkungskreis der ersten beiden Verfahren auf einen lokalen Prozessorknoten begrenzt, erfolgt in **Kapitel 8** die Leistungsuntersuchung eines globalen Lastverteilverfahrens für eine Mehrprozessorarchitektur. Wesentliche Merkmale dieses ebenfalls neu entworfenen Verfahrens sind der völlig verteilt arbeitende Algorithmus, die Berücksichtigung verschiedener Verkehrsklassen und die Vermeidung von Blindlast.

## Kapitel 2

# Rechnergesteuerte Vermittlungssysteme und Breitbandkommunikation

Ein Nachrichtennetz ist die Gesamtheit von End-, Übertragungs- und Vermittlungssystemen, die für den Aufbau und die Bereitstellung von individuellen Verbindungen eingerichtet sind und gemeinsam dem Nachrichtenaustausch dienen. Nachrichtennetze lassen sich nach den unterschiedlichsten Gesichtspunkten einteilen. Dabei spielen die Strukturierungsaspekte *Vermittlungsverfahren* und *Dienste* eine besondere Rolle im Rahmen dieser Arbeit.

Die Vergangenheit war dadurch gekennzeichnet, dass für verschiedene Dienste spezialisierte Netze, beispielsweise Fernsprechnetze für Sprachkommunikation, Datennetze für Rechnerkommunikation<sup>1</sup> oder Verteilnetze für Fernseh- und Rundfunksignale, bereitgestellt wurden. Durch die Dienste- und Netzintegration wird dieser Heterogenität entgegenengewirkt und eine Vereinheitlichung angestrebt. Zentrale Netzelemente sind hierbei die Vermittlungssysteme<sup>2</sup>. Diese haben die Aufgabe Nachrichten anhand von Zielinformationen durch ein Netz zu leiten.

Folgende Punkte repräsentieren das Umfeld und die speziellen Randbedingungen für die in dieser Arbeit erörterten Überlastabwehrproblematik und die entwickelten Verfahren zur Last- und Leistungsregelung: Rechnergesteuerte Vermittlungssysteme, Vermittlungssteuerungen (Aufgaben und Strukturen), Dienstintegration und Breitbandkommunikation (B-ISDN, Zeichengabe) und B-ISDN/ATM-Vermittlungssysteme. Diese Punkte bilden auch die wesentliche Gliederung dieses Kapitels.

## 2.1 Wichtige Merkmale von Vermittlungssystemen

Die Kommunikation zwischen zwei oder mehreren Instanzen (z.B. Mensch, Endgeräten wie PC oder Fernsehgerät) erfolgt durch den Austausch von Nachrichten über Verbindungen, die auch als Kommunikationskanal bezeichnet werden. Eine Verbindung kann dauerhaft oder temporär, als physikalischer oder logischer Übertragungspfad, eingerichtet werden. Für die Steuerung einer Verbindung sowie zur Koordinierung der kommunizierenden Instanzen, müssen neben den Nutz- auch Steuerungsinformationen ausgetauscht werden.

Bezüglich der Steuerung und Ressourcenreservierung werden zwei prinzipiell gegensätzliche Kommunikationsformen unterschieden [41]: verbindungsorientierte (physikalische oder virtuelle Verbindung) und verbindungslose Kommunikation.

---

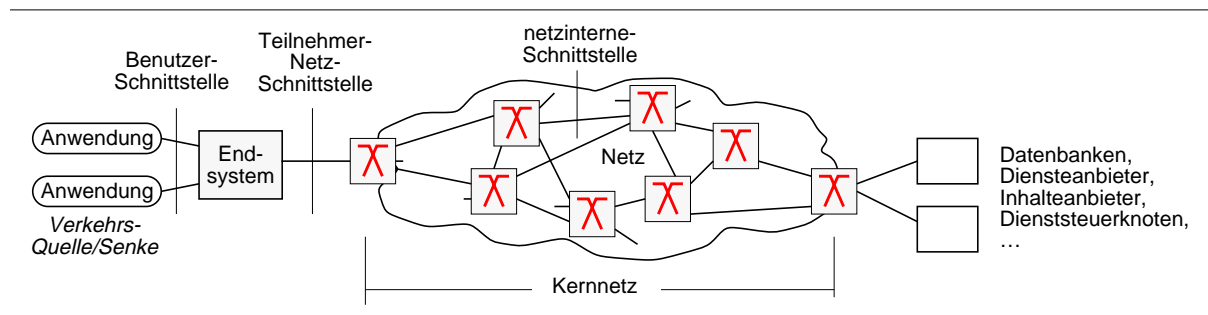
<sup>1</sup> Beispiele: DATEX-L (öffentliches, leitungsvermittelteres Datennetz), DATEX-P (öffentliches, paketvermittelteres Datennetz), private Rechnernetze.

<sup>2</sup> Nach dem allgemeinen Sprachgebrauch ist unter einem *Vermittlungssystem* die Gesamtheit der einem bestimmten technischen Konzept folgenden Einrichtungen zur Bildung von *Vermittlungsstellen* zu verstehen [6].

Bei der verbindungsorientierten Kommunikation lassen sich die Phasen Verbindungsaufbau, Nachrichtenaustausch und Verbindungsabbau während der „Lebensdauer“ einer Verbindung unterscheiden. Für viele Anwendungen der Datenverarbeitung sind aber die relativ langen Verbindungsaufbauzeiten störend. Bei verbindungsloser Kommunikation wird deshalb aus Zeit- und Aufwandsgründen keine Verbindung aufgebaut.<sup>3</sup> Unter Vermittlung wird das wahlweise Herstellen von Verbindungen (bei verbindungsorientierter Kommunikation) oder das Übermitteln von Nachrichten zwischen wechselnden Anschlusspunkten oder Benutzern verstanden. Der Ort der Vermittlung kann entweder zentralisiert in Vermittlungstellen lokalisiert sein, oder wie in bestimmten Nahbereichsnetztechnologien (*LANs* auf Basis eines *shared medium*) dezentral in den Endgeräten.

In den letzten Jahren haben sich folgende Tendenzen hinsichtlich der Integration in Kommunikationsnetzen abgezeichnet, die wiederum einen starken Einfluss auf Funktionalität und Architektur von Vermittlungssystemen haben:

- *Netzintegration*: Vereinheitlichung von Schnittstellen, Codewandlungen, Protokollumsetzungen usw. zwischen übertragungs- und vermittlungstechnischen Einrichtungen im Netz.<sup>4</sup>
- *Diensteintegration*: Ein Netz wird für verschiedene Dienste genutzt, es wird nicht für einen bestimmten Dienst optimiert. Dies bedingt Vorteile hinsichtlich Aufwand, Wartung, Betrieb usw. sowie eine bessere Ausnutzung des relativ teureren Teilnehmeranschlussnetzes. Aber erst ATM bietet eine wirtschaftliche Technologie, die unterschiedlichsten Dienste mit ihren spezifischen Charakteristika effizient zusammenzufassen.



**Bild 2-1:** Kommunikationsmodell mit Schnittstellen und Netzkomponenten (ohne Anschluss- bzw. Zugangsnetz)

Bild 2-1 zeigt ein allgemeines Kommunikationsmodell. Im Rahmen dieser Arbeit wird ein Netzknoten am Kernnetzrand, d.h. mit Teilnehmer-Netz- und netzinternen Schnittstellen, betrachtet.

## 2.1.1 Typische Funktionen

Die Hauptaufgabe einer Vermittlungsstelle bildet die Zusammenschaltung von Nutzkanälen bzw. die zielgerichtete Weiterleitung von Nachrichtenblöcken. Die Vielfalt technischer Lösungen für Vermittlungsstellen erschwert eine allgemein gültige Beschreibung. Es hat sich daher

<sup>3</sup> Für manche Dienste erhebt sich allgemein die prinzipielle Frage, ob sie verbindungslos oder verbindungsorientiert realisiert werden sollten.

<sup>4</sup> Je besser Übertragungs- und Vermittlungstechnik aufeinander abgestimmt sind, desto wirtschaftlicher wird das Netz.

durchgesetzt, Vermittlungsstellen nicht nach ihrer technischen Ausführung, sondern nach der Aufgabe in Funktionsblöcke einzuteilen [6, 646]. Mit dieser Abstrahierung lassen sich folgende Funktionsblöcke identifizieren:

- *Vermittlungstechnische Grundfunktionen* (rufbezogene Steuerungsaufgaben)  
Signalisierung<sup>5</sup>, Rufbearbeitung und Verbindungssteuerung, Dienststeuerung, Adressenauswertung bzw. -umsetzung, Wegesuche (intern: Koppelnetz; extern: bestimmter Netzbereich), Verwaltung und Vergabe von Betriebsmitteln u. a.
- *Zusätzliche Vermittlungsfunktionen*  
Bereitstellung ergänzender Dienst- bzw. Leistungsmerkmale sowie Sonderfunktionen.
- *Übergeordnete Steuerungsfunktionen* wie z.B. Überlastabwehr.
- *Funktionen für den Teilnehmerzugang sowie Hilfsfunktionen*  
Synchronisierung, Takterzeugung, Erzeugung von Hör- oder Ansagen.
- *Schnittstellenfunktionen in den Anschlussgruppen* wie z.B. übertragungstechnische Einrichtungen.
- *Sicherungstechnik und Wartungsfunktionen*<sup>6</sup>  
Beseitigung von Betriebsstörungen (Fehlererkennung, -lokalisierung, -beseitigung sowie Diagnose), vorbeugendes Prüfen und Warten (Routinetests, Eigenüberwachung, Sicherung von Zuständen).
- *Administrative Funktionen, Überwachungsfunktionen und Betriebsführungstechnik*  
Führung der Vermittlungsstelle (Möglichkeit zur Mensch-Maschine-Kommunikation oder Zugriffsmöglichkeit über Netz-Management), Teilnehmerverwaltung (logische Einrichtungen), Verwaltung von Systemkomponenten, Einbringen von Softwareänderungen, Austausch von Hardwarekomponenten, Entgelterfassung und -abrechnung, Verkehrsmessungen oder Netzbeobachtung sind insbesondere für den Netzbetreiber wichtige Funktionen.

Diese Aufstellung erhebt keinen Anspruch auf Vollständigkeit. Eine weitergehende Beschreibung der einzelnen Funktionen ist beispielsweise in [5, 6, 8, 23, 38, 49, 52, 57, 281, 549, 556] zu finden. In Abschnitt 2.3 werden Funktionen der Vermittlungssteuerung detaillierter betrachtet. In Abschnitt 2.6 werden speziell die Funktionen und Aufbau eines B-ISDN-Netzknoten näher erläutert.

## 2.1.2 Einsatzmöglichkeiten von Vermittlungssystemen

Vermittlungssysteme sind in der Regel so konzipiert, dass durch Austausch und Ausbau von Hardwarekomponenten oder anderen Softwareversionen ein breites Spektrum an Vermittlungstypen, geforderter Verkehrsleistung und unterschiedlichste Anwendungen abgedeckt werden können. Folgende Klassifizierung kann diesbezüglich vorgenommen werden:

- *Typen* (Orts-, Ortsdurchgangs-<sup>7</sup>, Fern- oder Auslandsvermittlungsstellen),

---

<sup>5</sup> In der Literatur (und in dieser Arbeit) wird in Anlehnung an den engl. Begriff *„signalling“* immer häufiger der Begriff *Signalisierung* synonym mit dem Begriff der *Zeichengabe* verwendet. Der vermittlungstechnische Begriff *Zeichengabe* steht für den Austausch vermittlungstechnischer Information.

<sup>6</sup> Ziel: Aufrechterhaltung der Betriebsbereitschaft.

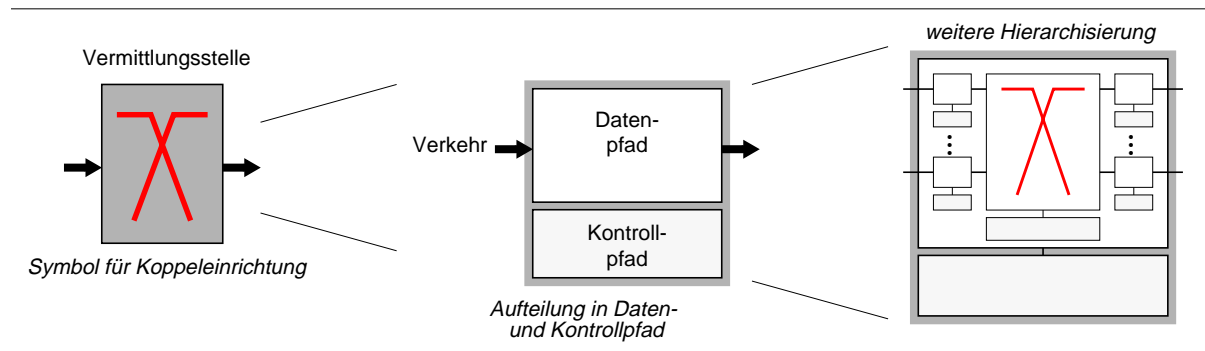
<sup>7</sup> sogenannte Tandem-Vermittlungsstellen

- *Ausbaustufe* bzw. *Amtsgröße*<sup>8</sup> (Konzentrator, Multiplexer, Kleinvermittlungsstellen, mittlere und große Vermittlungsstellen),
- *Netzbereich* (öffentliche, private oder Sondernetze),
- *Anwendungen* (z.B. Protokolle für Mobilfunknetze, Netzknoten in einem reinen Signali- sierungs- oder Verwaltungsnetz (z.B. *TMN*), Intelligentes Netz<sup>9</sup>, Paketnetz usw.).

Der Einsatzbereich und -zweck hat mit einem Einfluss auf Anforderungen und Struktur für die Überlastabwehr eines Vermittlungssystems.

### 2.1.3 Aufbau und Systemsteuerung

Vermittlungssysteme gehören zur Kategorie der informationsverarbeitenden Systeme. Die Informationstypen, welche verarbeitet werden sind Nutz- und Steuerungsnachrichten<sup>10</sup>. Beim Aufbau dieser Systeme kann eine allgemeine Unterscheidung in zwei Teilsysteme vorgenommen werden: *Daten-* und *Kontrollpfad* (Bild 2-2). Der Kontrollpfad ist für die Steuerung, d.h. Einstellung (Konfiguration), Überwachung, Test usw., des Datenpfades verantwortlich. Steuerungsnachrichten die von einer Vermittlungsstelle mit ihrer Umgebung, d.h. mit benachbarten Netzelementen (Endgeräte, Nachbarvermittlungsstellen) ausgetauscht werden, werden als Signalisierungsdaten bezeichnet. Neben dieser externen Steuerungsinformation werden noch interne Steuerungssignale unterschieden. Diese werden innerhalb der Vermittlungsstelle zwischen Daten- und Kontrollpfad ausgetauscht. Bei einem realen System existiert eine weitere Verfeinerung des Systemaufbaus. Der Datenpfad kann in unterschiedlich funktionelle Teilsysteme (z.B. Koppelfeld, Anschlussgruppen) aufgeteilt werden, wobei jedes Teilsystem seinen eigenen Kontrollpfad besitzt. In sogenannten teilzentralen Systemen gibt es eine übergeordnete Steuerung für die Koordinierung aller Steuerungsaufgaben.



**Bild 2-2:** Vermittlungsstelle als informationsverarbeitendes System

Mit Beginn der rechnergesteuerten Vermittlungssysteme haben sich folgende Begriffe im Rahmen der Systemsteuerung eingebürgert. Der Steuerungsrechner für die Bearbeitung zentraler Vermittlungsfunktionen wird als *Vermittlungsrechner* bezeichnet, der (unter Umständen abgesetzte) Rechner für die Einrichtung von Festverbindungen und Verwaltungsaufgaben stellt den *Bedienrechner* dar.

<sup>8</sup> Anschlusszahl, Vermittlungsleistung, Funktionsumfang

<sup>9</sup> Einsatz als Dienststeuerungs- (SCP) oder Dienstvermittlungsknoten (SSP)

<sup>10</sup> Nachrichten zum Betrieb, Verwaltung und Wartung von Vermittlungssystemen (sogenannte OAM-Meldungen) werden hier als spezieller Teil der Steuerungsinformation betrachtet.



## 2.1.4 Anforderungen und Leistungskennwerte

Allgemein ist die *öffentliche Vermittlungstechnik* für ihre extremen Anforderungen an die Zuverlässigkeit und die hohe Verfügbarkeit bekannt (z.B. zwei Stunden Totalausfall in 40 Jahren was 99,9994% Verfügbarkeit entspricht [502]). Allein diese Anforderung bedingt die Notwendigkeit einer Überlastabwehr. Weitere allgemeine Aspekte betreffen die Einhaltung garantierter Dienstgütern oder Kompatibilität bzw. Erweiterbarkeit.

Betrachtet man die *Nutzinformationsebene*, so stellen hier beispielsweise Verkehrsleistung, Signalverzögerung, Verzögerungsschwankung (Jitter) oder Verlustwahrscheinlichkeiten wichtige Leistungskenngrößen dar. In der *Steuerungsebene* spielen dagegen die Belegungsbelastung (BHCA), Antwortzeiten auf Signalisier Nachrichten, Wahrscheinlichkeit für innere Blockierung durch vollständige Belegung und Vergabe von Betriebsmitteln oder Echtzeit-Eigenschaften eine besondere Rolle.

In der *privaten Vermittlungstechnik* liegt der Schwerpunkt mehr auf Funktionsumfang oder Betriebsmöglichkeiten als aufwendigen Redundanzkonzepten für hohe Zuverlässigkeit.

## 2.2 Klassifizierung von Vermittlungssystemen

Das einem Netz zugrundeliegende *Vermittlungsverfahren* beeinflusst in entscheidendem Maße die Architektur eines Vermittlungssystems. Anhand dieses Prinzips kann nachfolgende Klassifizierung vorgenommen werden. Der Hauptunterschied zwischen Leitungs- und Speichervermittlung besteht darin, dass bei leitungsvermittelnden Verbindungen eine feste Zuordnung von übertragungstechnischen Betriebsmitteln zu einer Kommunikationsbeziehung existiert, während bei Paketvermittlung diese bedarfsorientiert angefordert werden.

### 2.2.1 Leitungsvermittelnde Netzknoten

Bei der Durchschalte- oder Leitungsvermittlung wird ein durchgehender physikalischer Übermittlungskanal zur exklusiven Nutzung des Netzes zur Verfügung gestellt. Bevor Nutzinformation übertragen werden kann, muss dieser Übertragungspfad zwischen den jeweiligen Kommunikationsinstanzen bestehen. Typisch durchschaltetermittelnde Netze sind das Telefonnetz (PSTN) sowie das diensteintegrierende Digitalnetz (ISDN).

Bereits im analogen Fernsprechnet wurde eine Vielfalt von Diensten (z.B. Fax, Modemzugang für Datenübermittlung, Funkruf, Bildschirmtext) für den Teilnehmer angeboten [84]. Diese Dienste wurden auf den Übermittlungsdienst für Sprachtelefonie abgebildet. Es findet somit keine dedizierte Unterstützung anderer Endanwendungen durch das PSTN statt.

### 2.2.2 Speichervermittelnde Netzknoten

Bei der Speichervermittlung wird die zu übertragende Information in Blöcke mit einer festgeschriebenen begrenzten Länge aufgeteilt, denen eine zusätzliche Steuerungsinformation vorangestellt wird. Nutz- und Steuerungsinformation bilden zusammen ein sogenanntes Paket. Die

Vermittlung erfolgt abschnittsweise mit Zwischenspeicherung in den Netzknoten. Die Übermittlung kann dabei entweder verbindungsorientiert oder verbindungslos erfolgen. Ein wesentliches Merkmal dieses Prinzips stellt die dynamische Belegung der übermittlungstechnischen Betriebsmittel dar. Neben der *Sendungsvermittlung* haben nachfolgende Realisierungsformen der Speichervermittlung eine große Bedeutung erlangt.

- *Paketvermittelnde Netzknoten*

Bei der Paketvermittlung wird eine Nachricht in Pakete mit variabler Länge aufgeteilt, die beim Empfänger wieder zusammengesetzt werden. Es wird für die Paketübertragung keine Verbindung durchgeschaltet. Man spricht in diesem Zusammenhang bei verbindungsorientiertem Betrieb von *virtuellen Verbindungen* bzw. bei verbindungslosem Betrieb von *Datagrammen*.

- *Zellvermittelnde Netzknoten (ATM-Systeme)*

Diese Vermittlungstechnologie soll die Vorteile von Leitungs- und Paketvermittlung kombinieren. Alle Pakete besitzen eine fixe Länge und werden als Zellen bezeichnet. Der Asynchrone Transfer-Modus bildet die Basis dieser Arbeit. Relevante Eigenschaften werden in den Abschnitten 2.4 bis 2.6 näher ausgeführt.

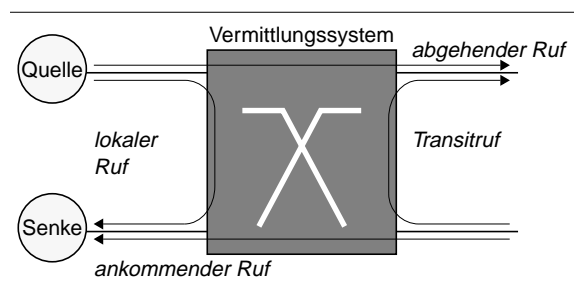
### 2.2.3 Leistungsgrößen von Vermittlungssystemen

Die physikalische Definition von Leistung ist Nutzarbeit pro Zeit. Die Leistung, die ein durchschaltvermittelndes System erbringt, schlägt sich in der Anzahl von Rufen, die pro Zeiteinheit erfolgreich auf- und wieder abgebaut werden können, nieder. Diese Größe wird auch als Nutzdurchsatz, Ruffleistung oder Komplettierungsrate bezeichnet. In Paketnetzen wird als Leistungsangabe für einen Netzknoten die Anzahl von Paketen, die pro Zeiteinheit erfolgreich weitervermittelt werden können, spezifiziert. Weitere Parameter für die Charakterisierung der Leistung von Vermittlungssystemen können z.B. [170] oder Datenblättern entnommen werden.

Dabei lassen sich Nutzer- und Steuerungs-Ebene ihre spezifischen Leistungskenngrößen zurechnen. Bei leitungsvermittelnden Netzknoten wird als Maß für die *Leistungsfähigkeit der Steuerung* üblicherweise der *BHCA-Kennwert* angegeben [541]. In der Paketvermittlung wird die Steuerungsleistung wiederum an der Zahl von vermittelnden Paketen pro Zeiteinheit gemessen.

#### Der BHCA-Kennwert

Die Ruffleistung eines Netzknotens wird in *Busy Hour Call Attempts* (BHCA) angegeben. Darunter versteht man die Anzahl von Belegungsversuchen während der Hauptverkehrsstunde [616]. Die Hauptverkehrsstunde ist der Tagesabschnitt von 60 aufeinanderfolgenden Minuten, in welchem die betrachtete Verkehrsgröße maximal ist. Der BHCA-Kennwert wird auch als Nennkapazität<sup>11</sup> eines Vermittlungssystems bezeichnet.



**Bild 2-3:** Grundsätzliche Rufftypen

<sup>11</sup> Mittleres maximales Rufangebot, welches vom Vermittlungssystem unter Einhaltung der geforderten Güte, bearbeitet werden kann.

Allerdings gibt es nun nicht den Standardruf schlechthin, sondern aus Netzsicht eine Vielfalt unterschiedlichster Rufarten. Wenn man die Netzschnittstellen betrachtet, über die Signalisierungsinformation für die Steuerung eines Basisrufes ausgetauscht wird, dann können bereits vier verschiedene Ruftypen<sup>12</sup> in einem Netzknoten unterschieden werden (vgl. Bild 2-3).

Jeder Ruftyp hat unterschiedliche Anforderungen an Betriebsmitteln (z.B. CPU-Zeit oder Speicher) im Vermittlungsrechner. Die Rufe eines Typs lassen sich selbst wiederum hinsichtlich weiterer Kriterien, z.B. Nutzung von Leistungsmerkmalen (*supplementary services*), Verkehrsklassen<sup>13</sup> oder Fehlerszenarien während der Rufbearbeitung, differenzieren. Mit diesem Grund wurden in der ITU-Empfehlung Q.543 [648] Referenzlasten für ISDN-Systeme (u. a.) definiert, um die Leistungsfähigkeit unterschiedlicher Architekturen verschiedener Hersteller objektiv vergleichen zu können. In dieser Empfehlung werden u. a. auch zwei Verfahren zur Ermittlung der Nennkapazität eines Vermittlungsrechners vorgeschlagen.

Die Vermittlungsleistung als BHCA-Angabe basiert deshalb üblicherweise auf einem Mix verschiedener Rufarten *eines Dienstes*, dem *Fernsprechen*. Ob diese Methode der Nennkapazitätsbestimmung für Breitbandssysteme sinnvoll ist, erscheint äußerst fraglich, insbesondere wenn der BHCA-Wert als Parameter in die Lastregelung mit eingeht (siehe auch Abschnitt 3.6.2).

## 2.3 Vermittlungssteuerungen

### 2.3.1 Anforderungen und grundsätzliche Steuerungsfunktionen

In den Anfängen des Telefonverkehrs musste jede Gesprächsverbindung von einer Vermittlungskraft an entsprechend ausgestatteten Vermittlungsschränken manuell mit Kabelschnüren zusammenschaltet werden. Die Signalisierung erfolgte über Steuerspannungen, -ströme und der Durchsage des gewünschten Rufteilnehmers. Durch die Entwicklung steuerbarer Koppelanordnungen (z.B. Wähler) konnten die Vermittlungen automatisiert werden.

Die allerersten Vermittlungsrechner wurden noch mit *festverdrahteter Logik* aufgebaut. Dies resultierte bei Einbringung neuer Leistungsmerkmale oder bei sonstigen Modifikationen einen immensen Änderungsaufwand [5]. Durch die Einführung *speicherprogrammierter Steuerungen*, auch SPC-Systeme (*Stored-Program Control*) genannt, konnten Änderungen des Vermittlungssystems durch Programmänderungen durchgeführt werden. Die ersten Zentraleinheiten in SPC-Vermittlungssystemen wurden aus diskreten Bauelementen (Gatterebene) aufgebaut. Nach Einführung des Mikroprozessors als integriertem Halbleiterschaltkreis 1971 (Intel 4004, [48]) wurden handelsübliche Prozessoren bzw. dedizierte Eigenentwicklungen eingesetzt. Die speziell für diesen Anwendungszweck entwickelten Prozessoren hatten Vorteile hinsichtlich Leistungsfähigkeit und Betriebssystemunterstützung. Erst aber der konsequente Einsatz von Prozessorfamilien die auch im Massenmarkt zum Einsatz kamen und kommen, führte zu wirtschaftlichen Lösungen und der Einsatz der Rechnersteuerung wurde auch für kleine Vermittlungsstellen rentabel.

Gleichzeitig ging man dazu über, für die Steuerprogrammentwicklung höhere Programmiersprachen (z.B. C, CHILL, ADA, PEARL) einzusetzen. Damit war man nicht mehr abhängig

---

<sup>12</sup> Begriffsdefinitionen entsprechend Verkehrsarten wie in [616].

<sup>13</sup> Ein privater Ruf hat eine andere Verkehrscharakteristik als ein geschäftliches Gespräch. Genauso gibt es Unterschiede bei Rufen im Nebenstellen- (privat) oder Weitverkehrsbereich (öffentlich).

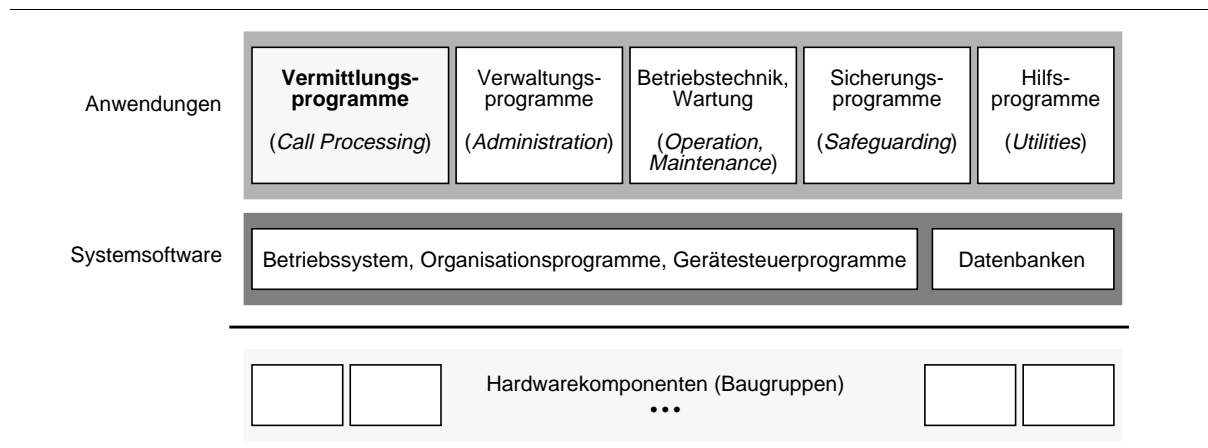
vom Zielrechner, sondern konnte Programme mit entsprechender Unterstützung (Entwicklungsumgebungen) auf kommerziellen Rechnern entwickeln.

Folgende allgemeinen Aufgabengruppen werden von JOEL [274] für die Steuerung des Vermittlungsvorganges unterschieden: Übertragung von Steuerungsinformation („*Call Signal Transmission*“), Empfangen und Senden von Steuerungsinformation („*Call Signal Processing*“, entspricht Signalisierung) und Verarbeitung von Steuerungsinformation („*Call Information Processing*“, entspricht Rufbearbeitung). Neben den vermittlungstechnischen Aufgaben existieren noch eine Vielzahl nichtvermittlungstechnischer Steuerungsfunktionen wie z.B. Verwaltungs- und Wartungsaufgaben (*Administration and Maintenance*)<sup>14</sup> oder die Sicherungstechnik [390].

Ein programmgesteuertes Vermittlungssystem besitzt allgemein einen Steuerungsrechner, der sich wiederum in einen Vermittlungs- und einen Bedienrechner trennen lässt, für die Bearbeitung dieser Aufgaben.

### 2.3.2 Software – Anforderungen und Aufbau

Speicherprogrammgesteuerte Vermittlungssysteme arbeiten als Echtzeitverarbeitungssysteme (*Realtime System*), da z.B. ein Verbindungswunsch „möglichst bald“ nach Auftreten des Ereignisses bearbeitet werden muss. Grundsätzlich werden folgende Anforderungen an die Steuerungssoftware gestellt [39,54]: Echtzeitfähigkeit, modulare Erweiterbarkeit, Trennung von anwendungs- und hardwareorientierter Software, Unabhängigkeit von Software und Steuerungshardware, Wiederverwendbarkeit, hohe Verfügbarkeit und eine effiziente Unterstützung der Nebenläufigkeit<sup>15</sup>. Bei neuen Vermittlungssystemen umfasst der Software-Entwicklungsaufwand mehr als 80% des Gesamtentwicklungsaufwandes [390].



**Bild 2-4:** Typischer Softwareaufbau einer Vermittlungssteuerung

Bild 2-4 zeigt eine typische Strukturierung der Software einer Vermittlungssteuerung. Spezielle Anforderungen werden an das Betriebssystem hinsichtlich Prozessverwaltung (bzw. Steue-

<sup>14</sup> Einrichten bzw. Abschalten von Teilnehmeranschlüssen, Rufnummernänderungen, Berechtigungsänderungen, Gesprächsdatenerfassung, Entgeltberechnung, allgemeine Prüfungsaufgaben, Fehlerdiagnose usw.

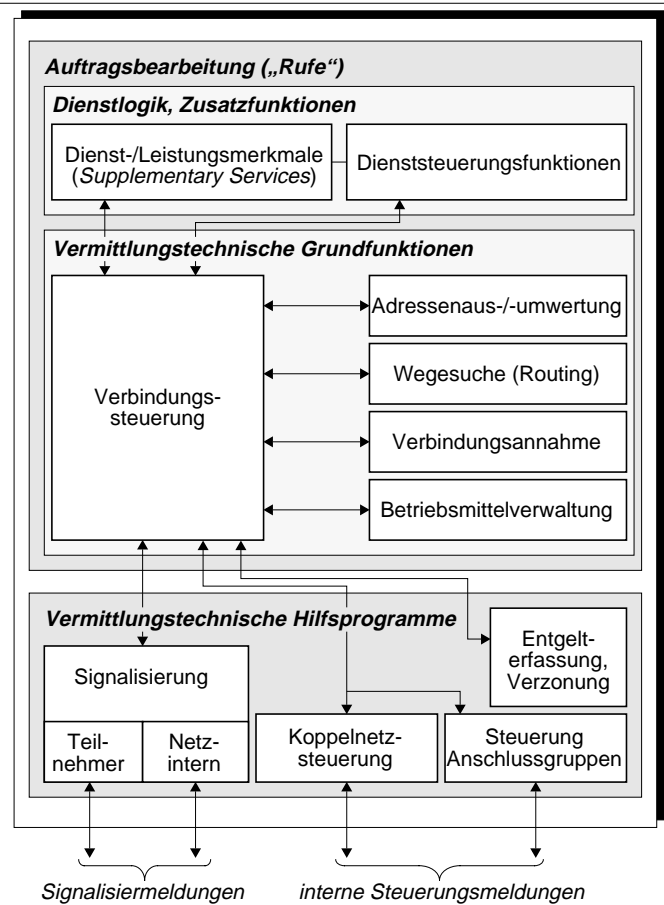
<sup>15</sup> Viele simultan ablaufende Einzelprozesse, die alle auf zentrale Komponenten (z.B. Ressourcen-Verwaltung) zugreifen, müssen koordiniert werden.

rung der Anwendungsprogramme), Speicherverwaltung, Zeitverwaltung<sup>16</sup>, Unterbrechungsbehandlung, Prioritätsebenen, Ablaufsteuerung (Berücksichtigung von Echtzeitaspekten) und interne Kommunikationsmechanismen (Interprozess- und Interprozessorkommunikation) gestellt. Die *Vermittlungssoftware* ist die hier interessierende Anwendung. Bild 2-5 zeigt eine exemplarische Strukturierung in wichtige Programmblöcke.<sup>17</sup> Für eine allgemeine Beschreibung siehe beispielsweise [5, 8, 560].

Vermittlungstechnische Hilfsprogramme dienen der Steuerung der Hardwarekomponenten einer Vermittlungsstelle sowie u. a. der Zeichengabe, Gebührenerfassung und Rufdatenaufzeichnung. Die Zeichengabeprogramme empfangen Signalisiermeldungen, ordnen diese Anreize einem vermittlungstechnischen Ereignis zu und melden dieses der Verbindungssteuerung. In der Regel werden mehrere verschiedene Zeichengabeverfahren unterstützt, z. B. verschiedene Impulswahlverfahren, Mehrfrequenzwahl, D-Kanal-Protokoll oder Zentralkanalzeichengabesysteme Nr. 5/Nr. 7.

Zu den wichtigsten Aufgaben der Bearbeitung vermittlungstechnischer Aufträge („Rufe“) gehören der Auf- und Abbau von Verbindungen. Die Verbindungssteuerung ist weiterhin für eine Koordinierung der Signalisierung und Koppelnetzsteuerung zuständig.

Alle Anwenderprogramme und selbst einige Programme des Betriebssystems werden als parallellaufende Prozesse in verschiedenen Dringlichkeitsstufen betrieben und vom Betriebssystemkern gesteuert. Von der Realisierung her wird nicht jede Steuerungsaufgabe als eigenständiger Prozess implementiert, da dies zu einer sehr großen Zahl von parallellaufenden Prozessen (u. U. mehrere zehntausend in großen Systemen [108]) mit beträchtlichem organisatorischem Aufwand und ungünstigem Reaktionsverhalten führen würde. Vielmehr werden mehrere Aufgaben, Prozeduren oder Anwenderprogramme, je nach Häufigkeit und gewünschten Reaktionszeiten, zu einzelnen Prozessen zusammengefasst.



**Bild 2-5:** Module der Vermittlungssoftware

### 2.3.3 Der Vermittlungsprozess

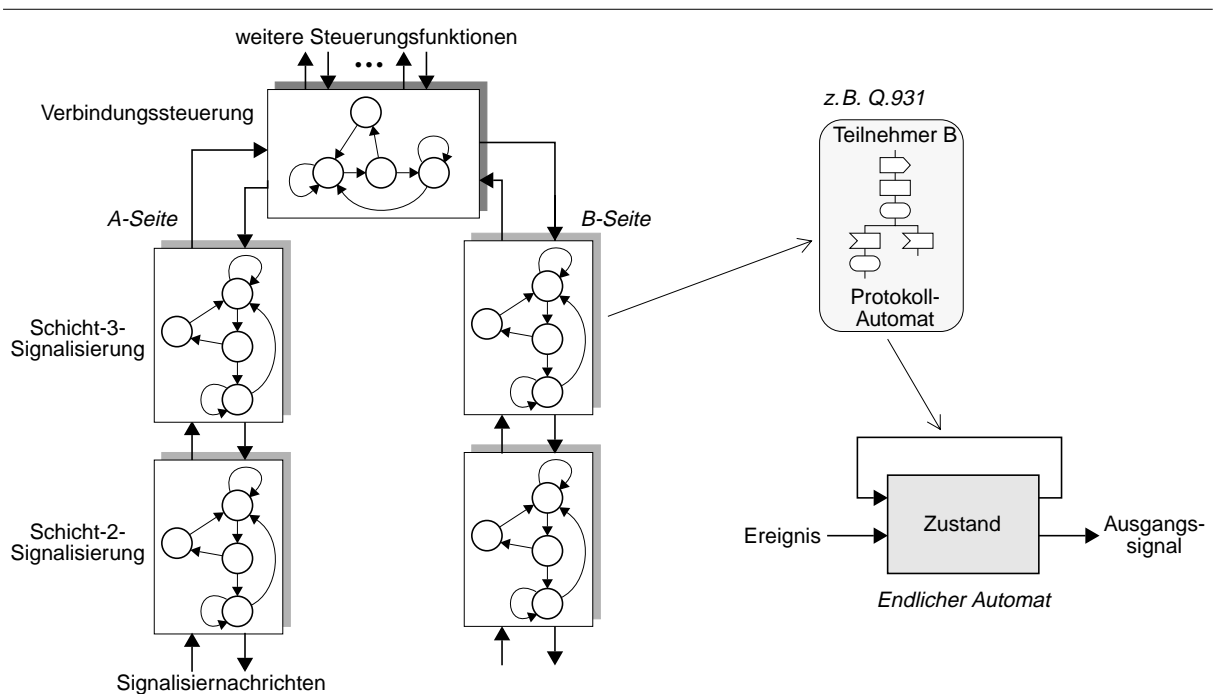
Die Vermittlungsaufgaben müssen simultan für sehr viele gleichzeitig bestehende, bzw. im Auf- oder Abbau befindliche, unabhängige Verbindungen durchgeführt werden. Die Anzahl von Verbindungen, die in Bearbeitung oder aktiv sind, kann zu einem Zeitpunkt mehrere zeh-

<sup>16</sup> Die Zeitverwaltung ermöglicht den Anwenderprogrammen, Zeitaufträge abzusetzen und zu löschen.

<sup>17</sup> Funktionalität des Blockes *Dienstlogik* kann auch durch andere Recheneinheit erbracht werden (z. B. Dienststeuerungsknoten (→ SCP) im Intelligenten Netz).

tausend Verbindungen in größeren Systemen betragen. Bei der Steuerung von Vermittlungssystemen handelt es sich im Prinzip um eine Prozesssteuerung, wenn man den Aufbau und Abbau von Verbindungen und die Vorgänge des Betriebes und Sicherns als Prozesse bezeichnet [107].

Steuerungsvorgänge werden immer aufgrund des Eintretens von Ereignissen ausgelöst. Bestimmte Funktionen der Vermittlungssteuerung werden durch diese Ereignisse aktiviert und müssen spätestens nach einer maximalen Antwortzeit reagieren. Zu den Ereignisquellen aus Sicht des Vermittlungsrechners zählen hauptsächlich *externe* (ankommende Signalisier Nachrichten von Endgeräten, Nachbarvermittlungsstellen, Dienststeuerungsknoten oder sonstigen Netzelementen; Befehle vom Bedienrechner) und *interne Ereignisse* (interne Steuerungsmeldungen anderer Systemkomponenten (z.B. periphere Gruppen), Meldungen anderer Prozessereinheiten in Mehrprozessor-Steuerungssystemen, interne Protokollereignisse (z.B. Timerabläufe), systeminterne Ereignisse (z.B. Alarmer, Fehlermeldungen)). Im weiteren werden nur noch Ereignisse betrachtet, die die vermittlungstechnischen Funktionen betreffen.



**Bild 2-6:** Vermittlungsvorgang (Ausschnitt) – Mehrere kooperierende Teilprozesse sind im Rahmen einer Kommunikationsbeziehung für anfallende Steuerungsvorgänge verantwortlich.

Der Signalisierverkehr, oder genauer das Eintreffen von Signalisierungsmeldungen, ist bekanntermaßen stochastischer Natur, d.h. diese Ereignisse treten rein zufällig auf. Aufgrund der Vielzahl von Signalisierungspunkten in einer Vermittlungsstelle stehen deshalb unter Normallast mehrere, zeitlich überlappende, jedoch voneinander unabhängige Ereignisse zur Bearbeitung durch die Software an. Für die Echtzeitbearbeitung von solchen „zeitgleichen“ Ereignissen bedient man sich des Prozesskonzeptes [40], das in der Lage ist, die verschiedenen parallel ablaufenden Programme zu koordinieren. Wie bereits angedeutet, müssen die hierzu notwendigen Mechanismen vom Betriebssystem bereitgestellt werden [159,167,293,345,424].

Bild 2-6 zeigt einen Ausschnitt eines beispielhaften Vermittlungsvorganges für eine Verbindung zwischen zwei Teilnehmern. Die einzelnen Teilprozesse repräsentieren endliche Automata-

ten. Jedes zufällig ankommende Teilnehmerereignis (hier: Schicht-2-Signalisiermeldung) stößt den Ablauf einer vom momentanen Zustand des betreffenden Automaten abhängigen Steuerungsfunktion an, die nach Ausführung bestimmter Aufgaben diesen Teilprozess in einen neuen Zustand versetzt. Als Reaktion auf dieses Ereignis werden in der Regel ein oder mehrere Ausgangssignale an andere Automaten weitergegeben. Jede neu ankommende Signalisierung löst somit in der Vermittlungssteuerung eine Folge von internen Signalen, die auch als Meldungskette bezeichnet wird, aus. Nach einem Zustandsübergang eines Automaten – d.h. die Befehlsfolge dieses Teilprogrammes ist abgearbeitet – wird auf neue Ereignisse gewartet. In der Zwischenzeit werden andere Ereignisse für quasi parallel laufende Vermittlungsprozesse oder nichtvermittlungstechnischen Steuerungsfunktionen abgearbeitet.

In der Vermittlungssteuerung wird der ankommende und der abgehende Teil einer Verbindung (sogenannte Halbverbindungen oder A-/B-Seite) in getrennten Softwaremodulen behandelt. Dieses Prinzip wurde ursprünglich eingeführt um Größe und Komplexität in Grenzen zu halten, und wird heute entsprechend den objektorientierten Ansätzen beibehalten [446].

## 2.3.4 Rechenzeitintensive Steuerungsfunktionen

Betrachtet man das steuerungstechnische Betriebsmittel *Prozessor*, dann lassen sich für die Komplettierung eines Rufes drei besonders rechenzeitintensive Steuerungsfunktionen identifizieren: *Signalisierung*, *Verkehrslenkung* und *Rufbearbeitung*.

### 2.3.4.1 Signalisierung

Die Signalisierung beinhaltet die Vorgänge, die für den Austausch vermittlungstechnischer Informationen zwischen den Endstellen und der zugehörigen Teilnehmervermittlungsstelle sowie zwischen Vermittlungsstellen notwendig sind. Über die Signalisierung gelangen Nachrichten zur Steuerung, die sie z.B. in Einstellbefehle für die Durchschaltung umsetzt, oder der Aktivierung von Betriebs- und Leistungsmerkmalen dienen. In Abschnitt 2.5 werden die für diese Arbeit relevanten Zeichengabe-Protokolle vorgestellt.

Eine Vermittlungssteuerung muss aus Flexibilitäts- und Kompatibilitätsgründen gleichzeitig verschiedene Signalisierungssysteme und -versionen unterstützen. In [121] wird angeführt, dass bereits in rechnergesteuerten Vermittlungssystemen der ersten Generation „*mehr als 20 Signalisierungssysteme*“ berücksichtigt wurden.<sup>18</sup>

### 2.3.4.2 Verkehrslenkung

Anhand der Zieladresse muss ein geeigneter Weg zum Zielort gefunden werden. Diese Steuerungsfunktion wird auch als Wegesuche oder Routing bezeichnet. Knotenintern ist beispielsweise ein Wegesuchverfahren bei ATM-Koppelnetzen notwendig, da hier immer mehrere Verbindungsmöglichkeiten zwischen einem Ein- und Ausgangsport bestehen. Knotenextern wird Routing für die netzweite Einrichtung von Verbindungen bzw. die netzweite Übermittlung von Paketen benötigt. Da ein Wegesuchverfahren neben Topologiedaten auch andere Informationen, beispielsweise aktuelle Verkehrswerte oder Kostenparameter, berücksichtigen muss, werden grundsätzlich sehr hohe Anforderungen an einen Routingalgorithmus gestellt.

---

<sup>18</sup> Vermittlungssystem *ITT 1220* von der *International Telephone & Telegraph Corporation*: dekadische Signalisierung, Mehrfrequenzsignalisierung, Direktwahl-Signalisierung, Zentralkanal-Zeichengabe).

### 2.3.4.3 Rufbearbeitung

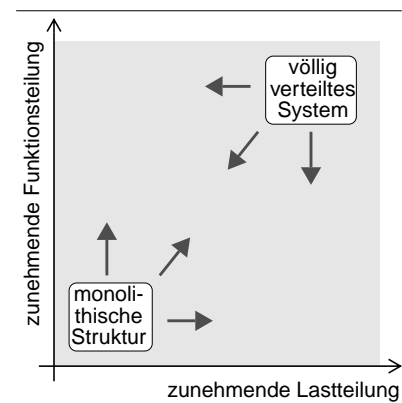
Hierunter werden alle übergeordneten Funktionen zur Dienstvermittlung und -steuerung verstanden. Während Signalisierung und Routing der Standardisierung unterworfen sind, werden diese Aufgaben herstellerspezifisch implementiert. Als Spezifikationsprache steht SDL zur Verfügung. Der Vermittlungsprozess ist grundsätzlich ereignisgesteuert (siehe Abschnitt 2.3.3). Die ATM-spezifischen Vermittlungsfunktionen werden in Abschnitt 2.6.2.2 vorgestellt.

### 2.3.5 Last- und Funktionsteilung

Die ersten rechnergesteuerten Vermittlungssysteme hatten eine monolithische Steuerungsstruktur [249], d.h. einen sehr hohen Zentralisierungsgrad. Dieser starke Zentralisierungsgrad hatte aus wirtschaftlichen Gründen durch den Einsatz der verfügbaren Rechner- und Speichertechnologie seine Berechtigung. Auf der anderen Seite bringt dieses Architekturkonzept aber auch eine Reihe von Problemen mit sich, beispielsweise hinsichtlich der Verfügbarkeit und Leistungsfähigkeit der Steuerung sowie der Softwarekomplexität. Die Zuverlässigkeit des Gesamtsystems wird in hohem Maße durch die Verfügbarkeit der zentralen Steuerung bestimmt. Ausfälle oder Fehler zeigen hier eine sehr große Wirkungsbreite. Aus Sicherheitsgründen sind deshalb Redundanzkonzepte vorzusehen [436].

Bereits frühzeitig, mit der Verfügbarkeit kostengünstiger hochintegrierter Schaltkreise, begann die Realisierung verteilter Steuerungsstrukturen mit geringem Zentralisierungsgrad [323]. Die Verteilung von Steuerungsaufgaben auf mehrere Teilsteuerungen kann nach dem Last- und Funktionsteilungsprinzip (*load* bzw. *function sharing*) erfolgen. Jede Steuerungsaufgabe kann in eine Folge von Teilfunktionen oder Operationen aufgeteilt werden, die durch ein oder mehrere Prozessoren bearbeitet werden können. Man unterscheidet nun zwei verschiedene Richtungen bezüglich der Dezentralisierung von Steuerungsaufgaben [7, 134, 146]:

- vertikale Dezentralisierung mit *funktionsorientierter* Aufteilung,
- horizontale Dezentralisierung mit *lastorientierter* Aufteilung.



**Bild 2-7:** Funktions- und Lastteilungsdiagramm [7]

Der Übergang von zentralisierten zu verteilten Steuerungsstrukturen ist fließend. Durch die Definition geeigneter Maßzahlen lässt sich der Grad an Last- und Funktionsteilung von Steuerungsstrukturen bestimmen (Bild 2-7, [7]). Da sich die Steuerungsaufgaben für unterschiedliche Netztechnologien unterscheiden, lassen sich schlecht Steuerungen für verschiedene Netzgenerationen miteinander vergleichen.

Die Dezentralisierung von Steuerungsaufgaben führt zur Verbesserung der Zuverlässigkeit, Verkehrsleistung, Modularität, Flexibilität und, was insbesondere für ATM-Systeme große Bedeutung hat, der Skalierbarkeit. Die Hauptnachteile dagegen sind zum einen der sogenannte Kommunikations-overhead für den notwendigen Informationsaustausch innerhalb eines Mehrprozessorsystems, zum anderen haben bestimmte Steuerungsaufgaben einen „stark zentralen Charakter“ (z.B. Ressourcenverwaltung) und ihre Dezentralisierung ist mit „hohen Kosten“ verbunden.



## 2.3.6 Architekturen und Strukturen von Steuerungseinheiten

Für die Konzeptionierung einer Überlastabwehr oder Leistungsregelung bildet die zugrundeliegende Steuerungsstruktur eine wesentliche Randbedingung. Dabei besitzen Steuerungssysteme für bestimmte Vermittlungsverfahren und Systemgenerationen ihre spezifischen Architekturmerkmale. In Anhang A werden im Zusammenhang mit dieser Arbeit relevante Steuerungsstrukturen kurz vorgestellt und diskutiert. Der Entwurf einer konkreten Regelkreisstruktur für die Leistungsoptimierung eines Vermittlungssystems ist ohne Berücksichtigung der Steuerungsstruktur nicht möglich. In Kapitel 4 erfolgt deshalb die Vorstellung eines generischen B-ISDN-Vermittlungsrechners basierend auf einer skalierbaren Multiprozessorstruktur.

## 2.4 Dienste und Protokolle des Breitband-ISDN

Der zu bearbeitende Verkehr eines Vermittlungsrechners in einem Netzknoten wird hauptsächlich durch Kommunikationsanwendungen und -dienste verursacht. In diesem Abschnitt werden für diese Arbeit relevante Begriffe und Aspekte näher erläutert. Der Begriff des Dienstes nimmt hierbei eine zentrale Stellung ein.

### 2.4.1 Der Dienstbegriff

Die Aufgabe eines Kommunikationsnetzes ist es, Nutzinformationen zwischen entfernten Teilnehmern mittels Übertragungs- und Vermittlungseinrichtungen auszutauschen. Die Nutzinformationen können dabei unterschiedlichster Ausprägung sein: Sprach-, Bewegtbild-, Dateninformation usw. oder auch Kombinationen einzelner Typen (Multimediatdaten). Die klassischen Netztechnologien und auch die heute noch vorwiegend betriebenen Netze sind dagegen für eine bestimmte Kommunikationsform spezialisiert.

Für den Begriff *Dienst* sind verschiedene Definitionen gebräuchlich. „Die Fähigkeit eines Netzes, Informationen einer bestimmten Art zu übertragen, wird als Dienst bezeichnet“ [57]. Eine engere Definition gibt [41]: Unter einem *Kommunikationsdienst* versteht man die „Eigenschaften eines Kommunikationsnetzes zur Unterstützung einer bestimmten Kommunikationsform einschließlich aller funktionellen, qualitativen und rechtlichen Aspekte“. Das in der Norm DIN ISO 7498 [618] (bzw. ITU X.200 [687]) beschriebene Basisreferenzmodell für die offene Kommunikation zwischen unterschiedlichsten Systemen, geht von einer abstrakteren Definition des Dienstkonzeptes aus. Bei diesem Schichtenmodell nutzt die Schicht  $N$  die Summe aller Funktionalitäten der darunterliegenden Schichten. Die Funktionalität von Schicht  $N-1$  (Dienstbringer) steht Schicht  $N$  (Dienstnutzer) als Dienst zur Verfügung. Der höchsten Schicht – der Verarbeitungs- oder Anwendungsschicht – sind die Anwendungsprozesse eines Endsystems zugeordnet. Eine Anwendung, z.B. eine verteilte Multimedia-Konferenz, nutzt deshalb Kommunikationsdienste für die Steuerung von Verbindungen und den Transport von Informationen.

In [215] werden als *Breitbanddienste* die Funktionen bezeichnet, die ein Netzbetreiber zur Unterstützung von *Breitbandanwendungen* über einen integrierten Teilnehmerzugang bereitstellt, d.h. über einen einzigen Netzanschluss, mit dem der Zugang zu allen Diensten möglich ist.<sup>19</sup> Von einem *Breitbandanschluss* spricht man üblicherweise, wenn dem Teilnehmer am Netzzu-

gang Übertragungsgeschwindigkeiten über 2 Mbit/s (oder 1,5 Mbit/s in den USA) zur Verfügung stehen.

## 2.4.2 Klassifizierungsmerkmale von Diensten

Ein Dienst lässt sich nach den unterschiedlichsten Gesichtspunkten klassifizieren. In [41] erfolgt eine Einteilung nach folgenden Kriterien:

- *Dienstmerkmale*: Bandbreite, Betriebsverhalten, Konfigurationen u. a.
- *Netzanforderungen* eines Dienstes an das Kommunikationsnetz: z. B. Übertragungs- und Vermittlungsgüte, Verfügbarkeit, Wirtschaftlichkeit.
- *Dienstqualität*, wobei hier wiederum zwischen Dienst- (*Quality of Service* [631]), Netz- und Verkehrsgüte unterschieden wird.<sup>20</sup>
- *Zusätzliche Dienst- und Leistungsmerkmale*, die über die „Basisfunktionen“ eines Dienstes hinausgehen, die Attraktivität des Dienstes erhöhen oder dessen Nutzung optimieren.

Gerade in diensteintegrierenden Netzen ist das Angebot an Dienstmerkmalen rapide gewachsen bzw. wird weiter ausgeweitet. Die Erbringung dieser zusätzlichen Funktionalität liegt vornehmlich im Bereich der Netzknoten bzw. speziellen Dienststeuerungsknoten bei Diensten des Intelligenten Netzes.

## 2.4.3 Diensteintegration

Wirtschaftliche Überlegungen bildeten mit die Motivation, alle heutigen und zukünftigen Dienste in einem Netz zu integrieren und nicht wie früher, oder zum Teil auch heute noch, eine Vielzahl dedizierter Spezialnetze zu betreiben. So wurde bereits vor ungefähr zehn Jahren mit der Einführung des diensteintegrierenden Digitalnetzes (ISDN) begonnen. Die Standardisierung für dieses Schmalband-ISDN ist weitestgehend abgeschlossen (siehe ITU-Normenreihe *I.xxx „Integrated Services Digital Network“*). Eine Übersicht bietet beispielsweise [632], eine gute Einführung in das ISDN ist in [9] zu finden. Ein wesentliches Merkmal von ISDN sind die 64-kbit/s-Transportkanäle, die durchgehend digital vermittelt werden. Pro Teilnehmer sind im einfachen Basisanschluss zwei 64-kbit/s-Transportkanäle (*Bearer Channel*, Nutz-Kanal) und ein 16-kbit/s-Signalisierkanal (*Data Channel*, Zeichengabe-Kanal) eingerichtet. Für Endgeräte mit größerem Bandbreitenbedarf ist der sogenannte Primärmultiplexanschluss vorgesehen. Diese Schnittstelle stellt am Netzzugang 30 Nutzkanäle mit 64 kbit/s und einen 64-kbit/s-Signalisierkanal zur Verfügung. Die Nutzbandbreite für eine Anwendung ist somit im Schmalband-ISDN auf maximal 1920 kbit/s begrenzt.

Die Integration von Breitbanddiensten (z. B. Abruf von Filmen oder Kopplung lokaler Netze über ein Breitbandnetz) erfordert derzeit Bitraten auf der Teilnehmeranschlussleitung, die teil-

---

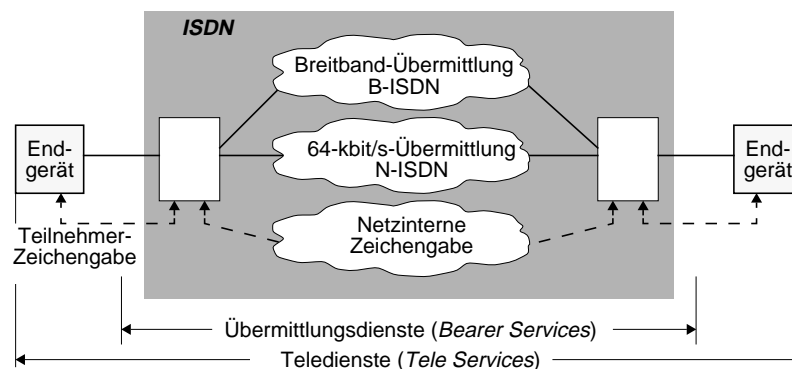
<sup>19</sup> Durch die fortschreitende Entwicklung auf dem Gebiet der Komprimieralgorithmen wird der tatsächlich benötigte Bandbreitenbedarf immer geringer. Der effektive Ressourcenbedarf eines „Breitbanddienstes“ wird deshalb vermutlich zukünftig in einer ähnlichen Größenordnung klassischer Schmalbanddienste liegen.

<sup>20</sup> *Dienstgüte* (QoS): Gesamtheit der Qualitätsmerkmale eines Kommunikationsdienstes aus der Sicht der Benutzer dieses Dienstes. *Netzgüte* (Network Performance, NP): Fähigkeit eines Kommunikationsnetzes oder eines Teils dieses Netzes, die Kommunikationsaufgaben zu erfüllen. *Verkehrsgüte* (Grade of Service, GoS): Teil der Netzgüte, der von der Bemessung der Anzahl von Netzbetriebsmitteln abhängt.

weise obige Grenze weit überschreiten. Dieser Bedarf war mit die Grundlage für die Konzeption des Breitband-ISDN (B-ISDN). Das Spektrum von Diensten und Dienstklassen, die das B-ISDN unterstützen soll, ist noch breiter und weiter gefasst als im ISDN. So sollen nun die verschiedensten Anwendungen mit den unterschiedlichsten Eigenschaften, z.B. konstante Bitrate oder büschelförmiger Verkehr, Bandbreitenbereiche von wenigen kbit/s bis zu mehreren Mbit/s, Punkt-zu-Punkt-Verkehrsbeziehungen bzw. Verteildienste, von dieser Netztechnologie unterstützt werden. Die hierzu notwendige Flexibilität lässt sich nur schwer mit den starren, synchronen Kanälen fester Bandbreite des ISDN erbringen. Deshalb wurde ATM als Transfermodus<sup>21</sup> für B-ISDN gewählt. Die sehr unterschiedlichen Anforderungen von Breitband-Verkehrsquellen werden durch die einheitliche Übertragung der Daten in Zellenform und dem virtuellen Verbindungsprinzip erfüllt. Eine durchgängige ATM-Infrastruktur bietet durch die bedarfsgerechte Zuordnung der Übermittlungsbandbreite eine sehr hohe Übermittlungsgeschwindigkeit und die Unterstützung von Echtzeitanwendungen [58]. Neben dem Vorteil der *effizienten Bandbreitennutzung* zeichnet sich ATM vor allem durch *Skalierbarkeit* gegenüber anderen Netztechnologien aus.

## 2.4.4 Tele- und Übermittlungsdienste

Als *Übermittlungsdienst* wird der Fernmeldedienst zur Nachrichtenübermittlung zwischen den Benutzer-Netzschnittstellen verstanden [613]. Ein Übermittlungsdienst ist nur in den unteren drei Schichten definiert.



**Bild 2-8:** Tele- und Übermittlungsdienste im ISDN<sup>22</sup>

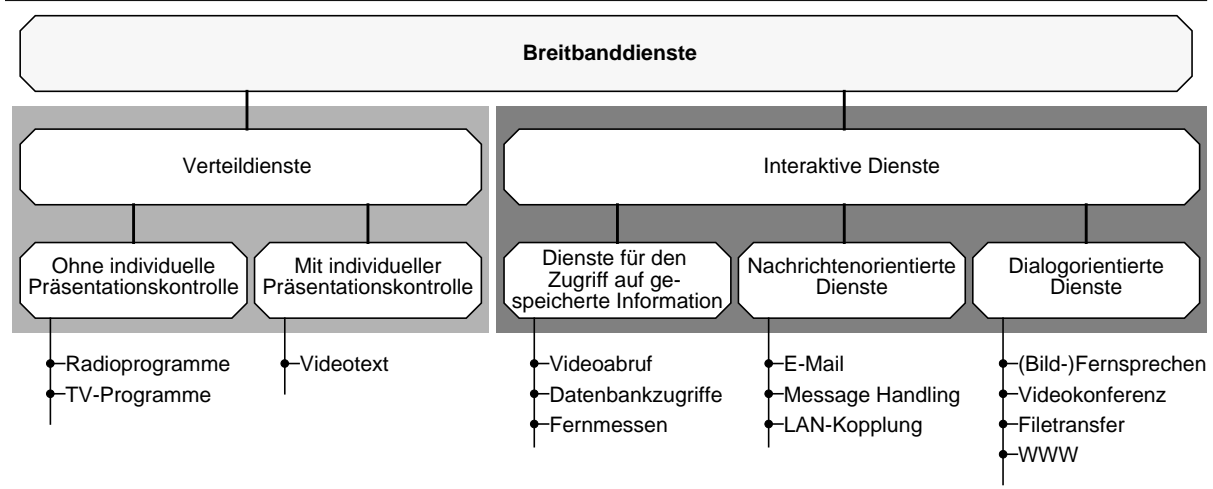
Die Ende-zu-Ende-Kommunikation zwischen den Benutzern wird von einem *Teledienst* erbracht (Bild 2-8). Ein Teledienst umfasst somit auch Funktionen in den Endeinrichtungen und deckt alle Schichten des Sieben-Schichten-Modells ab. Typische Teledienste des B-ISDN und ihre Charakteristiken sind beispielsweise in [90] beschrieben.

<sup>21</sup> Unter einem *Transfermodus* wird das Übermittlungsverfahren verstanden, welches einheitlich beim Übertragen, Multiplexen und Vermitteln von Nachrichten verwendet wird [41].

<sup>22</sup> Bild angelehnt an I.121 [633].

## 2.4.5 Neue und zukünftige Dienste

Die ITU-T-Empfehlung I.211 [636] nimmt folgende Einteilung der Dienste im B-ISDN aus Sicht des Netzes vor:



**Bild 2-9:** Klassifizierung von Breitband-Kommunikationsdiensten und typische Anwendungsbeispiele

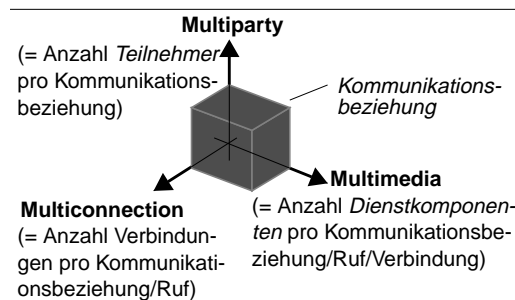
Es werden die beiden Hauptgruppen *interaktive Dienste* und *Verteildienste* unterschieden, die selbst wiederum in verschiedene Kategorien aufgeteilt werden (vgl. Bild 2-9). Für eine genauere Charakterisierung von Dienst- und Netzaspekten wurde die Methodik der Dienst- und Netzattribute entwickelt [634]. Solche *Dienstattribute* beschreiben beispielsweise folgende Eigenschaften:

- *Informationsart* (oder *Dienstkomponente*): Sprache, Daten, Fest- und Bewegtbilder
- *Verkehrseigenschaften*: Bitratenbedarf, zeitlicher Verlauf der Übertragungsrate (konstant, veränderlich, burstartig) usw.
- *Zeitbeziehung*: synchron bzw. asynchron mit/ohne Echtzeitanforderung
- *Dienstgüteanforderungen*: Toleranz gegenüber Verzögerungen, Verzögerungsschwankungen, Verlusten und Verfälschungen, usw.
- *Verbindungskonfiguration*: Punkt-zu-Punkt, Punkt-zu-Mehrpunkt (*Multicasting*), Mehrpunkt-zu-Mehrpunkt oder Verteilung (*Broadcasting*)
- *Symmetrie*: Unidirektional oder bidirektional (symmetrisch, asymmetrisch)
- *Kommunikationsprinzip*: Verbindungslos oder verbindungsorientiert
- *Einrichtung einer Verbindung*: Permanent, semi-permanent oder Wählverbindung

Ein Dienst lässt sich durch diese Attribute aus Teilnehmer- bzw. Netzsicht entsprechend charakterisieren. Bei den einzelnen Anwendungen stehen oft unterschiedliche Dienstgüteparameter im Vordergrund, die von den Vermittlungssystemen garantiert werden müssen [320, 374]. Diese Anforderungen müssen in der Überlastabwehr und Leistungsregelung Berücksichtigung finden. Dabei stellt sich die prinzipielle Frage, wie stark die Qualität unter hohen Lasten beeinträchtigt werden darf. Was wird überhaupt an Qualitätseinbuße vom Teilnehmer toleriert bzw. wahrgenommen? Diese Thematik und Problematik soll nachfolgend anhand ausgewählter Aspekte diskutiert werden.

### 2.4.5.1 „Gesprächs-“ bzw. Kommunikationsbeziehungen

Für jeden Informationstyp wurde in der Vergangenheit ein dedizierter Telekommunikationsdienst eingesetzt. Der zugehörige Übermittlungsdienst war für diesen Typ optimiert. Die Kommunikationsbeziehungen werden als Gespräch oder Ruf (*call*) in der Telekommunikation, als Dialog oder Sitzung (*session*) in der Datenkommunikation bezeichnet [441]. Ein Benutzer tritt mit einem oder mehreren Kommunikationspartnern in Beziehung. Klassisch war ein Ruf identisch mit einer durchgeschalteten Punkt-zu-Punkt-Verbindung, in der Rechnerwelt wurden vornehmlich verbindungslose Kanäle für den Datenaustausch eingesetzt. Bei einer zukünftigen Kommunikation hingegen werden ein oder mehrere Informationstypen in eine Kommunikationsbeziehung miteinbezogen. Bei diesen Anwendungen werden in der Regel mehrere, typspezifische ATM-Verbindungen parallel, oder auch zeitlich versetzt, im Kommunikationsnetz genutzt.



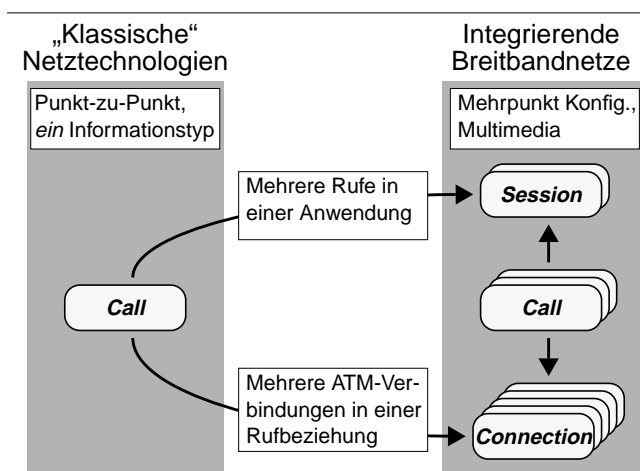
**Bild 2-10:** Mögliche Klassifizierung von Kommunikationsbeziehungen

Im B-ISDN wird deshalb zwischen einem Ruf (*call*) und einer Verbindung (*connection*) unterschieden. Eine Rufbeziehung umfasst ein oder mehrere Verbindungen (*multiconnection call*).<sup>23</sup> Die Eigenschaften jeder einzelnen Verbindung können dabei während des gesamten Rufes variiert werden. Für jeden Informationstyp wird in der Regel eine separate Verbindung eingerichtet. Ein Ruf mit mehreren Informationstypen wird häufig als „Multimedia-Gespräch“ (*multimedia call*) bezeichnet (siehe z. B. [54]).

Die Hierarchisierung bzgl. der verschiedenen Beziehungsebenen kann und wird auch in der Literatur noch weiter getrieben. Eine Anwendung selbst basiert im allgemeinen Fall wiederum auf der Nutzung von ein oder mehreren Rufen. Beispielsweise können in eine Anwendung mehr als zwei Teilnehmer involviert sein, zwischen denen jeweils einzelne Rufbeziehungen existieren. Diese Kommunikationsbeziehung wird oft auch als *Multiparty Call* bezeichnet (Bild 2-10). YUMIBA [570] gibt hierzu ein Beispiel (siehe Bild 2-11), wobei er die oberste Ebene als *Sessions* benennt.

Die gleiche Dreiebeneinteilung – *Session, Call, Connection* – wird in einem Breitbandkonzept für das Intelligente Netz verwendet [536]. Noch weiter geht KOTSIS u. a. [325]. Sie unterscheiden bei WWW-Anwendungen sechs Beziehungsebenen, wobei hier die Zeit zwischen dem ein- und ausloggen eines Benutzers auf dem Endgerät als *Session* bezeichnet wird. Es ist jedoch fraglich, ob die Einführung weiterer Ebenen oberhalb eines Rufes überhaupt generell noch Sinn macht.

Für die Lastregelung in einem Netzknoten stellt sich allerdings die Frage, inwieweit die Qualität für einen Dienst unter Überlast eingehalten werden kann bzw. notfalls be-



**Bild 2-11:** Kommunikationsbeziehungen – Ausprägung in Breitbandnetzen aus der Sicht von NTT [570].

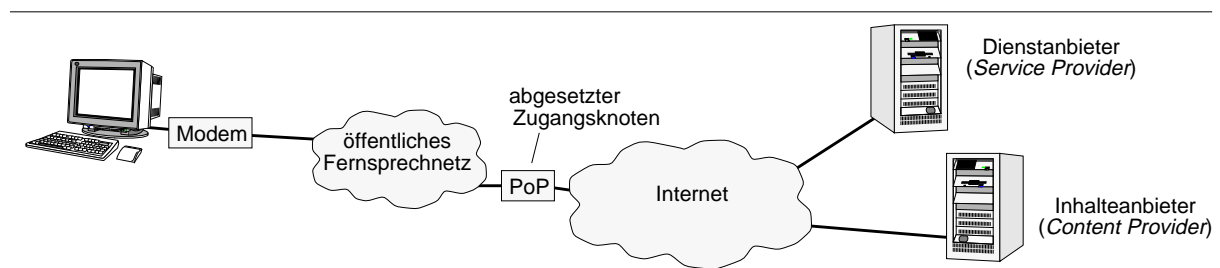
<sup>23</sup> Ein Ruf ohne Verbindung macht nur begrenzt Sinn.

schnitten werden muss. Werden neue Sitzungen bzw. neue Rufe angenommen oder abgelehnt, wie wird mit eingerichteten Rufbeziehungen verfahren, in denen beispielsweise neue Verbindungen angefordert oder bestehende Verbindungsparameter modifiziert werden sollen? Ein Beispiel einer komplexeren Anwendung soll diese Fragestellungen verdeutlichen:

Bei der *Telekooperation* (auch als *Joint Editing* oder *Collaborative Work* bezeichnet) arbeiten mehrere, räumlich entfernte Personen gleichzeitig an einer gemeinsamen Aufgabe, z.B. an einer Werbepräsentation. Jedem Teilnehmer sind die Daten gleichzeitig zugänglich und vorgenommene Änderungen werden unmittelbar auf dem Bildschirm angezeigt. Daneben seien Sprachverbindungen zwischen allen beteiligten Personen, die notwendig für die Koordinierung der gemeinsamen Arbeit sind, eingerichtet. Die Gruppe soll nun um einen weiteren Teilnehmer erweitert werden, d.h. zusätzliche Verbindungen sind einzurichten. Ein überlasteter Netzknoten muss nun entscheiden, ob diese weiteren Anfragen für den bereits eingerichteten Dienst bearbeitet oder ob andere, neue Rufanfragen, bedient werden.

### 2.4.5.2 Internet – Datendienste

Die Nutzung von Internet-Diensten zeichnet sich gegenwärtig durch ein rapides Wachstum aus. Damit wird zukünftig der in einem ATM-Netz vermittelte Verkehr mehrheitlich von IP-Anwendungen herrühren. Diese Tatsache hat entsprechende Konsequenzen für die Konzeption eines Breitbandknotens und die Belastung ihrer Vermittlungssteuerung [444], zumindest im Zugangsbereich. Gegenwärtig wird typischerweise ein Nutzer von Internet-Diensten über seine Ortsvermittlung mit dem *Point of Presence* (PoP) eines Diensteanbieters (ISP) verbunden (Bild 2-12). Ein *ISP PoP* befindet sich, zumindest derzeit, noch nicht an jeder Ortsvermittlung, so dass in diesen Fällen der Verkehr über Durchgangsvermittlungsstellen geleitet wird.



**Bild 2-12:** Nutzung von Netzressourcen bei Internetdiensten [232,463]

Für die Belegung von Betriebsmitteln in Nutzer- und auch Steuerungs-Ebene hat dieses Konzept große Auswirkungen die in Abschnitt 3.6.1 näher erläutert werden sollen. Zwei Internet-Dienste sollen stellvertretend kurz vorgestellt werden.

#### World-Wide Web (WWW)

Dieser Dienst basiert auf einem *Client-Server Modell*. Die Kommunikation verläuft in einem „Anfrage-Antwort“-Schema (*Request-Response-Szenario*), wobei die Anfrage jeweils vom Benutzer (*client*) ausgeht. Der Zugriff auf ein Dokument erfolgt mittels eines sogenannten *Web-Browsers*. Wählt ein Benutzer ein Dokument aus, dann setzt der Browser eine Anfrage an den zuständigen Server ab. Jedes Dokument kann sich aus mehreren Dateien zusammensetzen. Jede Datei wird gegenwärtig noch getrennt übermittelt, d.h. es wird jeweils eine TCP-Verbin-

dung (TCP: Transmission Control Protocol) auf- und wieder abgebaut. Messergebnisse und Verkehrscharakterisierung dieses Dienstes sind beispielsweise in [70] zu finden.<sup>24</sup>

### **Sprachtelefonie via Internet**

Das PSTN/ISDN bietet eine hochqualitativen, zuverlässigen und weltweiten Fernsprechdienst an. Fortschritte in der Sprachkomprimierung, die zunehmende Verbreitung multimedialfähiger PCs als Endgeräte und die Nachfrage nach billigen Ferngesprächen sind die Hauptgründe für die Entwicklung von *Voice over IP* (VoIP). Die größte Herausforderung ist dabei derzeit die schlechte Dienstgüte durch die Best-Effort-Eigenschaft der aktuellen Internet-Version. Trotzdem wird diesem echtzeitkritischen Dienst ein erhebliches IP-Verkehrspotential vorhergesagt, zumindest solange die gegenwärtigen Kostenstrukturen sich nicht ändern.

### **2.4.5.3 ATM-Übermittlungsdienst**

Jede Anwendung wird in einem ATM-Netz auf eine bestimmte Dienstklasse abgebildet. Hierzu wurden in ITU-T I.362 [640] mehrere allgemeine Dienstklassen definiert, die ein möglichst breites Anwendungsspektrum abdecken sollten. Eine Beschreibung dieser Klassen, Unterscheidungskriterien und typischer Anwendungsbeispiele sind beispielsweise in [36,54] zu finden. Aus Sicht der Lastregelung für einen B-ISDN-Vermittlungsrechner können diese AAL-Dienstklassen eine Entscheidungsgrundlage, wie eine neue Dienstanforderung zu behandeln ist, darstellen (siehe auch Abschnitt 3.6.2).

## **2.4.6 Breitband-ISDN im Kontext des Intelligenten Netzes**

Das bewährte Konzept des IN wird derzeit für Breitband-Netze weiterentwickelt (siehe z.B. [94,95,111,120,212,233,234,262,423,537]). Im B-ISDN betrifft dies hauptsächlich die Signalisierprotokolle. Durch die räumliche Trennung von Dienstvermittlung und -steuerung entstehen bei der Abwicklung eines IN-Dienstes zum Teil erhebliche Signalisieraktivitäten. Gerade für Breitband-IN-Dienste ist in dieser Richtung mit einer deutlichen Komplexitätszunahme zu rechnen [472,536]. Nach [413] soll sich diese Problematik durch den Funktionalitätszuwachs in Signalisierprotokollen zukünftig weiter verschärfen.<sup>25</sup> Die Konsequenzen für die Überlastabwehr eines B-SSP-Vermittlungsrechners werden in Abschnitt 3.6.1.1 diskutiert.

## **2.4.7 Charakterisierung von Diensten hinsichtlich ihres Steuerungsaufwandes**

Für die Last- und Leistungsregelung ist der Steuerungsaufwand, der sich hinter einem bestimmten Dienst verbirgt, das primär interessierende Merkmal. Gemeint ist damit das zeitliche Belegungsmuster von steuerungstechnischen Betriebsmitteln während der Lebensdauer eines Dienstes. Gerade während Hochlastphasen sind diese Ressourcen zu schützen. So dürfen beispielsweise für die Einhaltung von Dienstgüten durch das Netz bestimmte Verbindungspara-

---

<sup>24</sup> Z.B. wurden bereits 1995 an den Web-Server des *National Center for Supercomputing Applications* (NCSA) durchschnittlich pro Tag mehr als 350000 Anfragen abgesetzt.

<sup>25</sup> Beispielsweise sind im *Signalling Capability Set 2.2* u. a. folgende Aspekte angedacht: *IN Control*, *Session Control*, *Third Party Control*, *Sequential Call Establishment* oder *Call Transfer* [413].

meter nicht überschritten werden. Für diese Parameter werden in der ITU-T-Empfehlung I.352 [638] Obergrenzen für Rufauf- und -abbauzeiten im ISDN vorgegeben.

Ein anderes Beispiel betrifft den Aspekt des Verhältnisses von Schmal- zu Breitbanddiensten. So ist bei kompletter Nutzung eines ATM-Netzknotens für N-ISDN-Verbindungen eine theoretische Steuerungslast von 15 Millionen BHCA zu erwarten [533]. Diese Problematik rührt vom unterschiedlichen Ressourcenbedarf eines Dienstes in Nutzer- und Steuerungs-Ebene her. Für eine effiziente Überlastabwehr einer Breitbandvermittlungssteuerung ist es deshalb notwendig, Aussagen über den zu erwartenden Ressourcenbedarf eines Dienstes zu bekommen bzw. diesen abschätzen zu können. Die Problematik der Dienstcharakterisierung aus Sicht der Steuerungs-Ebene wird im Kontext zusätzlicher Breitbandaspekte in Abschnitt 3.6 weiter ausgeführt.

## 2.5 Zeichengabe im Breitband-ISDN

Aus Sicht der Überlastabwehr nimmt die Zeichengabe aus einer Vielzahl von Gründen eine zentrale Stellung ein. Wie bereits erwähnt wurde, bedingt die Bearbeitung der Signalisierung selbst einen nicht unwesentlichen Teil an Prozessorzeit [295]. Auf der anderen Seite bildet die Signalisierschicht die Schnittstelle zur Überlastabwehr, da eine neue Dienstanforderung sich erst anhand einer Signalisiernachricht identifizieren lässt. Möchte ein Netzknoten unter Überlast Verkehrsquellen drosseln, so ist dies auch nur im Rahmen der definierten Signalisierprotokolle möglich. Hier stehen je nach Protokoll oder Schicht verschiedene Möglichkeiten zur Verfügung. Weiterhin wird der Dienstablauf durch Signalisiernachrichten gesteuert. Durch die verschiedenen Typen, Ausprägungen oder auch potenziellen Fehlerfälle eines Dienstes sind nahezu unendlich viele Signalisierszenarien denkbar.

Durch B-ISDN und ATM kommen nun noch eine Fülle weiterer Anforderungen – z.B. weites Spektrum an Diensten und Anwendungen, verschiedene Dienstklassen und Dienstgüten, eine Vielzahl von Parametern zur Charakterisierung einer ATM-Verbindung, flexible Konfigurationen von Kommunikationsbeziehungen, Trennung von Ruf und Verbindung, u. v. a. m. – im Rahmen der Signalisierung hinzu. Aus den genannten Gründen wird nun im folgenden die Zeichengabe im B-ISDN und ihre neuen Möglichkeiten kurz vorgestellt.

### 2.5.1 Grundlegende Merkmale

Die Zeichengabe umfasst nicht nur den unmittelbar mit dem Auf- und Abbau einer Verbindung verknüpfte Steuerkommunikation, sondern darüberhinaus jeglichen Informationsaustausch zwischen den Vermittlungsstellen, zwischen Vermittlungsstellen und Netzzentren, zwischen Teilnehmer und Vermittlungsstelle und in diensteintegrierenden Netzen sogar den direkten Austausch von Steuerungsinformationen zwischen den Endgeräten.<sup>26</sup> Die durch die Zeichengabe zu übertragende Informationsmenge ist im wesentlichen von den vermittlungstechnischen Grundereignissen, von den zusätzlichen Dienstmerkmalen und von den technischen Merkmalen des Kommunikationsnetzes abhängig.

---

<sup>26</sup> Zeichengabe z.B. zur Kompatibilitätsprüfung der Endgeräte oder zum Dienstwechsel während der Verbindung. Diese Signalisiernachrichten werden nicht von den Netzknoten interpretiert sondern transparent durchgereicht (sogenanntes *User-to-User Signalling*).



## **Zeichengabe zwischen Vermittlungsstellen**

Für speicherprogrammgesteuerte Vermittlungssysteme ist die Zentralkanalzeichengabe prädestiniert. Im Gegensatz zur leitungsgebundenen Zeichengabe, wo Signalisierinformation auf dem für die Verbindung genutzten Nachrichtenkanal übertragen wird, erfolgt die Zentralkanalzeichengabe auf eigens zu diesem Zweck reservierten Zeichengabekanälen. Die Zuordnung der Zeichen zum jeweiligen Nachrichtenkanal erfolgt mittels einer Adressinformation. Dadurch ist die Ausnutzung eines zentralen Zeichengabekanals so gut, dass die Zeichengabe für mehrere tausend Nachrichtenkanäle auf einem einzigen Zeichenkanal erfolgen kann [6].<sup>27</sup>

Die Zentralkanalzeichengabe ist eine in hohem Maße zentralisierte Funktion [6]. Damit haben Ausfälle eine große Wirkbreite und könnten zu Verkehrseinschränkungen im Nachrichtenkanalnetz führen. Eine Reihe von Sicherheitsvorkehrungen muss deshalb getroffen werden.

Das in der heutigen Zeit bedeutendste Zeichengabesystem stellt das ITU-Zeichengabesystem Nr. 7 [191,650] dar. Es wurde für eine digitale Umgebung optimiert, protokollstrukturiert aufgebaut und für viele Anwendersysteme angelegt. Es ist sowohl für verschiedene Dienste, die in verschiedenen Netzen abgewickelt werden, als auch für integrierte Dienste in einem Netz (ISDN) einsetzbar. Ausführlichere Informationen zu diesem Signalisiersystem sind beispielsweise in [575] zu finden.

## **Zeichengabe an der Teilnehmerschnittstelle**

Sie stellt das Bindeglied zwischen den Funktionen des Endgerätes – und damit mittelbar den Maßnahmen des Teilnehmers – und den Funktionen der lokalen Vermittlungsstelle dar. Zeichengabesysteme für die Teilnehmeranschlussleitung sind auf die besonderen Rahmenbedingungen dieser Schnittstelle wie Informationsmenge, Übertragungstechnik und Intelligenzverteilung zwischen Endgerät und Vermittlungsstelle ausgelegt.

Neben den Grundereignissen – z.B. Verbindungsaufbau oder Alarmieren eines Teilnehmers – fallen zunehmend Steuerungsinformationen, z.B. zur Aktivierung und Deaktivierung von Dienstmerkmalen, für eine Nutzverbindung an. Man kann allgemein feststellen, dass das Verhältnis Steuerungs- zu Nutzinformation pro Verbindung im Zuge der fortschreitenden Fernsprechvermittlungstechnik, immer größer wurde.

Im Zuge des ISDN mit seinen komplexen Endgerätefunktionen sowie seinen verschiedenen, vielfältigen Diensten wurden die Anforderungen an die Zeichengabe noch mannigfaltiger. Für die Zeichengabe im ISDN wurde daher das D-Kanal-Protokoll entwickelt. Die Protokollschicht drei führt die eigentliche Verbindungsbehandlung für leitungsvermittelte Dienste (Fernsprechen, Daten), paketvermittelte Datendienste, Benutzer-Benutzer-Zeichengabeverbindungen und das Steuern von Dienstmerkmalen durch.

## **2.5.2 Anforderungen im Breitband-ISDN**

Die „klassische“ Signalisierung macht keine Unterscheidung zwischen Verbindungs-, Ruf- und Dienststeuerung, sondern behandelt diese Funktionen als „monolithischen Block“ [373]. Signalisierung im B-ISDN dagegen ist ein wesentlich komplexerer Vorgang, im Vergleich zu allen anderen Vorgängernetztechnologien [45,218]. Sie basiert auf den Prinzipien und Protokollen die für das Schmalband-ISDN entwickelt wurden [650-656]. Die offensichtlichste Änderung ist die Umstellung des Meldungstransports auf ATM-Zellen. In den Protokollschichten,

<sup>27</sup> Die zentralen Zeichengabekanäle bilden ein dem Nutzkanalnetz überlagertes Netz. Die Strukturen der beiden Netztypen können unterschiedlich sein. Die Vermittlungsstellen stellen Endpunkte im Zeichengabenetz dar.

die für die eigentliche Verbindungssteuerung zuständig sind, wurden zahlreiche Änderungen und Erweiterungen aufgrund einer Reihe neuer Anforderungen notwendig:

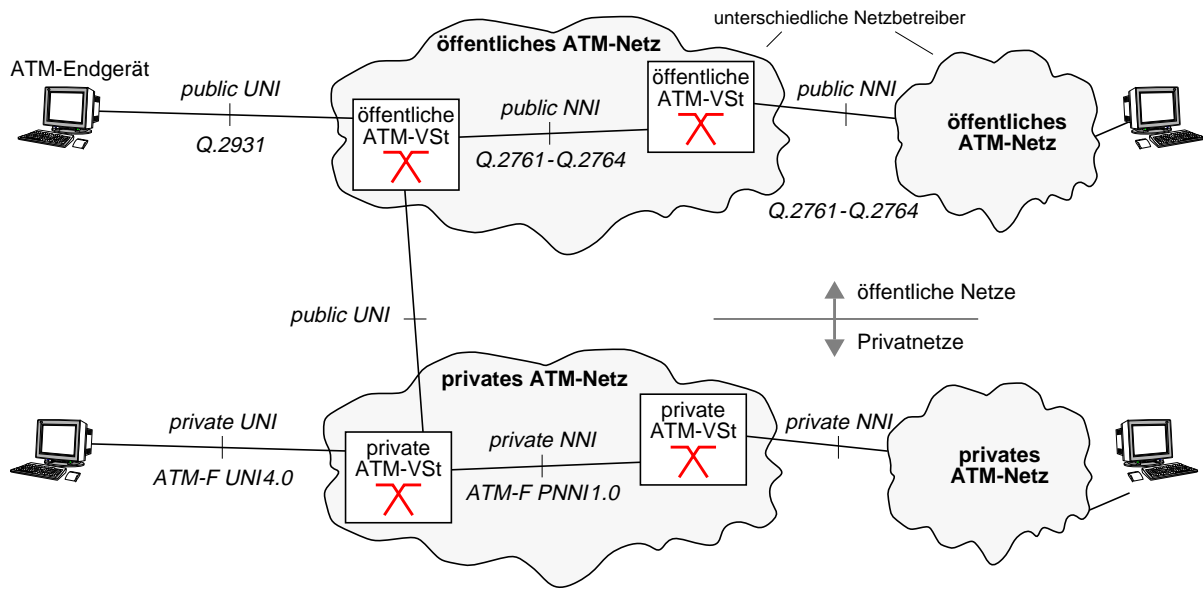
- Signalisierkanäle sind nicht starr festgelegt wie im N-ISDN. Signalisierverbindungen werden auf virtuelle Kanäle (VC) abgebildet und können u.a. durch die sogenannte Metasignalisierung [661] auf- und abgebaut werden.
- Ein Teilnehmer kann gleichzeitig mehrere Signalisierbeziehungen zu verschiedenen Kommunikationspartnern unterhalten.
- Einrichtung und Auslösung von virtuellen Kanälen und Pfaden beinhaltet eine Reihe von Funktionen:
  - Auf- und Abbau von VCs/VPs als Wählverbindungen oder auf semi-permanenter bzw. permanenter Basis mit allen angeforderten Verbindungsmerkmalen.
  - Aushandeln von Verbindungsparametern.
  - Veränderung von Verbindungsparametern einer aktiven Verbindung auf Anforderung.
- Unterstützung verschiedenster Typen von Ruf- und Verbindungskonfigurationen (Kommunikationsbeziehungen) wie beispielsweise Punkt-zu-Punkt-, Punkt-zu-Mehrpunkt- (für *Multicasting* [530]) und *Broadcast*-Verbindungen.
- Neben „klassischen“ Rufen und Schmalband-Diensten, Unterstützung verschiedener neuer Rufotypen wie z.B. *Multiconnection Calls* (für z.B. Multimedia Dienste) oder *Multiparty Calls*.
- Notwendige Steuerungsfunktionen für diese neuen Rufotypen wie beispielsweise simultaner Auf- oder Abbau mehrerer rufbezogener Verbindungen, Hinzufügen bzw. Terminieren von Verbindungen zu oder von einer Rufbeziehung, Hinzufügen bzw. Entfernen eines Teilnehmers (*Party*) zu oder von einem Mehrteilnehmerruf (*Multiparty Call*), Behandlung korrelierender Verbindungen als *Multiconnection Calls*, Rekonfiguration oder Aufteilung eines *Multiparty Calls* in mehrere einfache Rufbeziehungen.
- Mobilfunknetze bedingen einen relativ großen Anteil nichtrufbezogener Steuerungsfunktionen (*non-call related control functions*), z.B. Handy-Registrierung durch Einschalten („Einbuchen im Netz), *Location Update*, Abfrage der Sprachbox oder *Short Message Service* als Paketdienst.

Gerade die Steuerung von Mehrpunktverbindungen (siehe Q.2971 [684]) und Multimedia-Kommunikationsbeziehungen bedingte die einschneidendsten Veränderungen für die Signalisierprotokolle bzw. die Ruf- und Verbindungssteuerung. Mit aus diesen Gründen wird bei der Signalisierung im B-ISDN zwischen Ruf- und Verbindungssteuerung unterschieden („*Separation between Call Control and Connection/Bearer Control*“).

### 2.5.3 Schnittstellen und Protokollübersicht

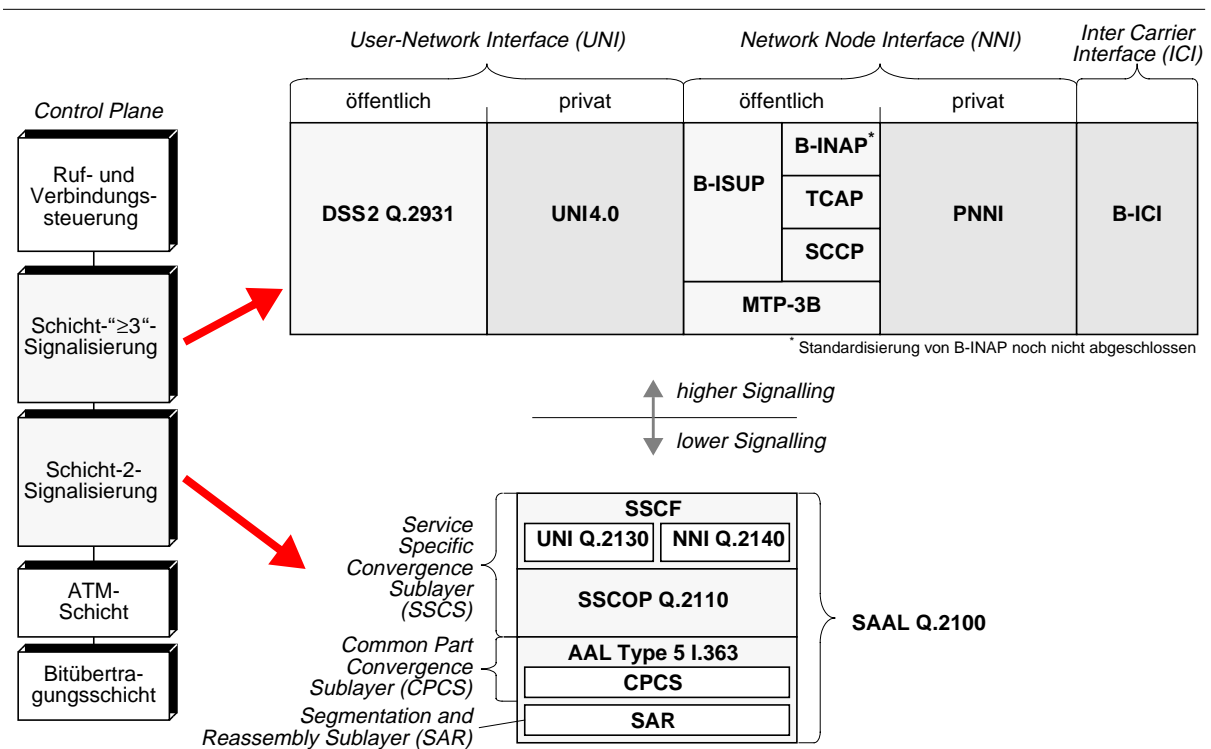
Bei der ATM-Signalisierung werden folgende Schnittstellenaspekte unterschieden (Bild 2-13):

- Signalisierung zwischen Endgerät und Vermittlungsstelle (*User-Network Interface*),
- Signalisierung zwischen Vermittlungsstellen eines Netzbetreibers (*Network Node Interface* bzw. *Network-Network Interface*),
- Signalisierung zwischen Vermittlungsstellen unterschiedlicher Netzbetreiber (*Inter Carrier Interface*)<sup>28</sup>,
- Signalisierung zwischen öffentlichen und Privatnetzen.



**Bild 2-13:** Signalisierschnittstellen in öffentlichen und privaten ATM-Netzen (Stand der Standardisierung: 1998)

Die wichtigsten ATM-Signalisierprotokolle für die unterschiedlichsten Schnittstellenkombinationen sind in Bild 2-14 zusammengestellt. Der aktuelle Stand der jeweiligen Protokollspezifikationen ist den Referenzen *ATM-Forum* [608-611] und *ITU-T* [659-684] zu entnehmen.<sup>29</sup>



**Bild 2-14:** Übersicht wichtiger ATM-Signalisierprotokolle

<sup>28</sup> In dieser vom ATM-Forum erarbeiteten Spezifikation sind insbesondere die spezifischen Management-Aspekte abgedeckt, die beim Übergang zwischen den Netzen unterschiedlicher Betreiber auftreten.

<sup>29</sup> Entsprechende Empfehlungen anderer Standardisierungsgremien, wie z.B. *ETSI*, *ANSI* oder *IETF*, wurden nicht referenziert.

Daneben müssen weiterhin alle Schnittstellen zwischen ATM- und Nicht-ATM-Netzen, z.B. Übergang zum N-ISDN, zu Signalisiernetzen oder SMDS, unterstützt werden [415].

## 2.5.4 ATM-Anpassungsschicht für die Signalisierung (SAAL)

Die *Signalling ATM Adaptation Layer* (SAAL, Q.2100 [659]) hat die Aufgabe, der über ihr liegenden eigentlichen Signalisierungsschicht einen zuverlässigen Übertragungsdienst anzubieten, da die Schicht-3-Signalisierungsprotokolle in der Regel keine Fehlerkompensationsmechanismen besitzen. Für den Transport der Signalisierungsinformation wird AAL Typ 5 verwendet (I.363.5 [642]).<sup>30</sup>

### 2.5.4.1 Wichtige Funktionen des *Service Specific Connection Oriented Protocol*

Das *Service Specific Connection Oriented Protocol* (SSCOP, Q.2110 [660]) stellt einen gesicherten, verbindungsorientierten Übermittlungsdienst für die Schicht-3-Zeichengabe bereit. Dieses Sicherungsprotokoll geht auf HDLC zurück und verwendet typische Merkmale dieser Protokolle, wie z.B. Fehlererkennung und -behebung oder Flusssteuerung [240,287].

Durch die *Service Specific Coordination Function* (SSCF, Q.2130 oder Q.2140 [662,663]) erfolgt die Anpassung an die Anforderungen der darüberliegenden Zeichengabeprotokolle.

Neben der Nutzdatenübertragung sind im Rahmen dieser Arbeit vor allem folgende Steuerungsfunktionen des SSCOP von Bedeutung:

- *Reihenfolgesicherung*  
Bei der gesicherten Übertragung werden alle Dateneinheiten zyklisch durchnummeriert (Modulo  $2^{24}$ ). Dadurch kann auf korrekte Reihenfolge geprüft werden.
- *Fehlerkorrektur durch wiederholte Übertragung*  
Übertragungsfehler können nicht festgestellt werden. Vielmehr können anhand der Folgenummern der Verlust von Dateneinheiten erkannt, und durch selektive Anforderung (*Selective Repeat*) zur wiederholten Übertragung diese Verluste korrigiert werden.
- *Flusssteuerung*  
Das SSCOP erlaubt eine Flusssteuerung auf Basis eines Fenstermechanismus (*Window Flow Control*), bei dem nur für eine bestimmte Anzahl von Dateneinheiten Quittierungen ausstehen dürfen. Möglicherweise interessant für Überlastabwehrverfahren ist die Tatsache, dass bei diesem Protokoll die Sendefenstergröße dynamisch vom Empfänger eingestellt und modifiziert werden kann [244]. Die Fenstergröße beträgt maximal  $2^{24} - 1$  Sendeberechtigungen (sogenannte *Credits*).

### 2.5.4.2 Überwachung der Signalisierstrecken

Durch die zentrale Bedeutung eines Signalisiernetzes sind die Anforderungen an die Verfügbarkeit sehr hoch. Werden zwischen zwei SSCOP-Instanzen über einen längeren Zeitraum keine Daten übertragen, dann wird durch Timerabläufe (*Keep-Alive*- bzw. *Idle*-Timer)<sup>31</sup> der

<sup>30</sup> Derzeit wird diskutiert, alternativ AAL Typ 2 (I.363.2 [641]) auch für die Signalisierung zu verwenden.

<sup>31</sup> Einstellwerte am UNI bzw. NNI: *Poll*-Timer 750ms/100ms, *Keep-Alive*-Timer 2s/100ms, *Idle*-Timer 15s/100ms und *No-Response*-Timer 7s/1,5s.

Austausch von sogenannten *Poll*-Steuereinheiten provoziert, um damit die Durchgängigkeit des Signalisierabschnittes nachzuweisen. Durch die *Poll*-Zyklen ergibt sich eine erhebliche *dy-namische Grundlast* für die Protokollinstanzen, auch wenn keine oder nur wenig Signalisierungsinformation ausgetauscht werden [54].

### 2.5.4.3 Realisierung der Schicht-2-Signalisierung in Netzelementen

Im Schmalband-ISDN erfolgt die Trennung zwischen Hard- und Softwarerealisation des Protokollstacks in der Steuerungs-Ebene zwischen Schicht zwei und drei. Dies ist hauptsächlich auf den Standardisierungsgrad, der naturgemäß mit wachsender Schichtnummer abnimmt, und der Komplexität von Funktionen in höheren Schichten zurückzuführen. Es existieren speziell für die Signalisierung eine Vielfalt an dedizierten Hardwarelösungen (z.B. *LAPD-Controller*) oder Mikrocontroller mit unterstützenden Hard- und Firmwarefunktionen (z.B. *MC68360* mit Signalisierung Nr. 7).

Für das B-ISDN sieht die aktuelle Marktsituation noch anders aus. Da derzeit noch keine hochintegrierten Schaltkreise auf dem Markt sind, die eine vollständige Realisierung in Hardware der Schicht-2-ATM-Signalisierung unterstützen, wird diese bei heutigen ATM-Systemen in Software implementiert. Diese Situation hat mit entscheidend die Steuerungsarchitektur aktueller Systeme geprägt. Bei ATM-Vermittlungen im Kernnetz können mehrere hunderttausend Signalisierkanäle gleichzeitig aufgebaut sein. Bei diesen Systemen erfolgt die Schicht-2-Signalisierung durch mehrere Mikroprozessoren die nach dem Prinzip der Lastteilung arbeiten. Jeder einzelne Prozessor ist dabei für eine konfigurierbare Anzahl von Signalisierungsverbindungen verantwortlich.

In Zukunft ist zu erwarten, dass auch für die Schicht-2-Signalisierung entsprechende Chips für eine Hardwarelösung verfügbar sind.<sup>32</sup>

## 2.5.5 Zeichengabe am Netzzugang: Teilnehmersignalisierung

Bei der ITU-T wird die Teilnehmersignalisierung (*access signalling*) für das B-ISDN unter der Bezeichnung *Digital Subscriber Signalling System No. 2* (DSS2, [670]) spezifiziert. Neben der öffentlichen Zeichengabeschnittstelle existiert die Empfehlung UNI 4.0 [611] des ATM Forums für Privatnetze. Stellvertretend sollen hier wichtige Eigenschaften des DSS2 vorgestellt werden.

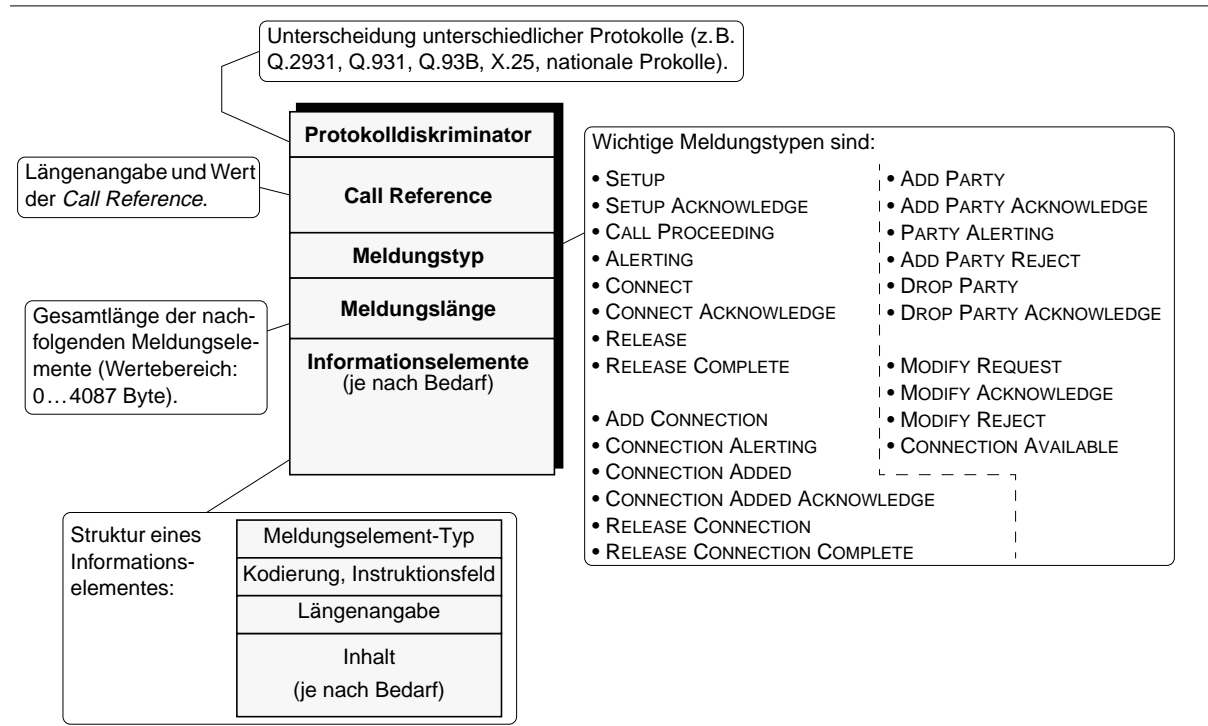
### 2.5.5.1 Meldungstypen und -format

Das Format einer Q.2931-Meldung umfasst fünf globale Felder (siehe Bild 2-15). Von einem einzigen Endgerät können gleichzeitig sehr viele Nutzverbindungen ausgehen. Zu jeder Verbindung existiert während ihrer gesamten „Lebensdauer“ eine zugehörige Signalisierungsbeziehung, die durch die *Call Reference* gekennzeichnet wird.<sup>33</sup> Der Wert für eine neue Verbindung wird von der initiiierenden Seite aus vergeben. Aufgrund der *Call Reference* können die einzel-

---

<sup>32</sup> Was eine Hardwarelösung derzeit so erschwert, ist weniger der SSCOP-Protokollautomat, als vielmehr die theoretisch maximale Länge einer Nachricht (65524 Byte), der ggf. notwendige Puffer aufgrund der Fenstergrößen, Mechanismen wie *Selective Repeat* oder *Local Data Retrieval*, Zeitgeberfunktionen oder die Vielzahl von Signalisierkanälen, die parallel unterstützt werden sollten. Vermutlich wird sich hier eine Lösung auf Basis eines dedizierten Mikrocontrollers (z.B. *SAAL*- oder *SSCOP*-Controller), mit ATM-Ports als Ein-/Ausgabeschnittstellen, einer schnellen CPU mit Firmware und möglichst viel Puffer für Signalisierungsdaten, abzeichnen.

nen Signalisierungsmeldungen ihren jeweiligen Protokollinstanzen im Endgerät und in der Vermittlungsstelle zugeordnet werden.



**Bild 2-15:** Format einer Schicht-3-Nachricht am UNI

Wichtige *Meldungstypen* im Rahmen dieser Arbeit sind vor allem die *SETUP*- und die *RELEASE*-Signalisierungsnachricht für eine neue Verbindungsanfrage und die Auslösung einer Verbindung. Eine Signalisierungsnachricht kann theoretisch mehrere tausend Byte lang sein. In der Praxis können jedoch die meisten Meldungen in eine oder wenige ATM-Zellen abgebildet werden. Die *SETUP*-Meldungen haben dabei den größten Umfang [509].

Die weiteren Meldungsfelder (*Information Element*) enthalten die eigentlichen Signalisierungsinformationen und Parameter wie z. B. Verbindungskennung (*VPCI/VCI*), Protokollanforderungen, Adress-, Zustands- und sonstige Steuerungsinformationen.

### 2.5.5.2 Anforderung von neuen Rufen und Verbindungen

Die Anforderung einer neuen Verbindung wird mittels einer *SETUP*-Nachricht von Sendeseite beantragt. Ein Ruf repräsentiert eine Beziehung zwischen Kommunikationspartnern, die sich in einer gewissen Anzahl von Verbindungen mit bestimmten Parametern zwischen ihnen äußert.<sup>34</sup> Ein Ruf kann eine, mehrere oder sogar keine Verbindung umfassen. Ein Ruf ohne zugehörige Verbindung macht beispielsweise dann Sinn, wenn vorab zwischen zwei Instanzen eine Kompatibilitätsüberprüfung stattfindet oder bestimmte Parameter ausgehandelt, die eigentliche(n) Verbindung(en) aber erst zu einem späteren Zeitpunkt aufgebaut werden sollen.

<sup>33</sup> Bei Mehrpunktverbindungen und/oder Rufen mit mehreren Verbindungen werden zusätzlich folgende logische Referenzen unterschieden [373]: zur Rufidentifizierung die *Call Reference*, die *Connection Reference* zur Kennzeichnung einer bestimmten Verbindung, für die Teilnehmer jeweils eine *Endpoint Reference* u. a.

<sup>34</sup> Jeder Ruf hat nur einen einzigen Teilnehmer als Eigentümer (sogenannter *Call-Owner-Party*) [373]. Bei traditionellen Rufen ist dies immer der rufende Teilnehmer (*Calling Party*).

Eine SETUP-Meldung enthält vornehmlich Informationen zum angeforderten Dienst.<sup>35</sup> Für diese Angaben stehen 17 Informationselemente zur Verfügung, die teilweise vorgeschrieben (*mandatory*) oder optional sind. Um aus diesen Angaben in einer Vermittlungssteuerung Rückschlüsse auf die dahinterliegende Anwendung, d.h. für die Überlastabwehr, machen zu können, sind eventuell folgende Informationselemente interessant:

- AAL Typ (Feld *AAL parameters*),
- Bandbreiteangaben in Zellen pro Sekunde (Feld *ATM traffic descriptor*),
- Dienstgüteeanforderung für diese Verbindung (Feld *QOS parameter*),
- Breitband-Übermittlungsdienst (Feld *Broadband bearer capability*),
- Schichtinformationen (Felder *Broadband high layer information* und *Broadband low layer information*) und
- Teiladressen (Felder *Calling party sub-address* und *Called party sub-address*) da diese u. a. wie beim TCP gewisse „Dienstzugangspunkte“ repräsentieren könnten.

Die Standardisierung ist derzeit allerdings noch nicht soweit, dass die Kodierung aller Informationselemente vollständig festliegt. Viele Parameter und Angaben sind gegenwärtig noch nicht definiert oder „reserviert“.

Handelt es sich bei einer Anfrage um eine Punkt-zu-Mehrpunkt-Verbindung, dann wird dies im Informationselement „*Broadband bearer capability*“ angegeben. Momentan wird nur eine unidirektionale Informationsübermittlung vom rufenden Teilnehmer (*Root*) zu mehreren gerufenen Teilnehmern (*Leafs*) unterstützt [611,664,684]. Teilnehmer können dynamisch durch die Meldungen ADD PARTY und DROP PARTY zu- und weggeschaltet werden. Für diese zusätzlichen Verbindungen wird die ursprüngliche *Call Reference* weiterverwendet.

Ein weiteres Informationselement für die SETUP-Nachricht wurde in Q.2959 [674] (und Q.2726.2 [665]) spezifiziert, das dem rufenden Teilnehmer ermöglicht, einer neuen Verbindung eine gewisse Priorität (*Call Priority*) zuzuordnen. Fünf Prioritätsstufen (*Level 1... 5*) wurden definiert. Diese Angabe ist optional und kann von Vermittlungsstellen während der Aufbauphase ausgewertet und als Steuerungsparameter genutzt werden. Ein gerufener Teilnehmer kann bei einem Ressourcenengpass bereits bestehende niederprioritäre Rufe unterbrechen („*pre-emption*“) oder abbauen, um auf die Anfrage des höher priorisierten Rufes zu antworten. Dieser Parameter ist für eine Überlaststeuerung nicht uninteressant.

Soll eine neue Verbindung angefordert werden, dann müssen nicht alle Parameter bereits der ersten SETUP-Meldung mitgegeben werden (*Blockverfahren*). Vielmehr können noch fehlende Informationselemente durch mehrere nachfolgende SETUP-Meldungen nachgeliefert werden (*Overlap-Verfahren*). Weiterhin bietet Q.2962 [678] die Möglichkeit, ATM-Schicht-Parameter während des Verbindungsaufbaus auszuhandeln. Dabei können neben den gewünschten Verbindungsparametern, alternative Angaben oder minimal akzeptable Werte angegeben werden.

Wie bereits oben erwähnt, kann der rufende Teilnehmer Kompatibilität und Zustand des gerufenen Teilnehmers vor dem Aufbau einer Nutzverbindung überprüfen (sogenanntes *Look Ahead*, Q.2964.1 [683]). Dieser Mechanismus ist sinnvoll um Blindbelegungen von angeforderten Netzressourcen zu verhindern.

---

<sup>35</sup> Momentan ist relativ unklar, inwieweit diese Information der Überlastabwehrsteuerung einer Vermittlungsstelle zugänglich gemacht werden soll/kann und es ist vor allem fraglich, ob diese Information überhaupt einen relevanten Nutzen (z. B. für Diensterkennung) besitzt (siehe auch Abschnitt 3.6.2.1). Interessant wäre beispielsweise die Frage, ob ein Netzknoten aus einer SETUP-Meldung erkennen kann, welche Anwendung, z. B. ein Spiel oder dringender Notruf, hinter dieser Verbindungsanfrage steht.

### 2.5.5.3 Modifikation von Verbindungsparametern

Während der aktiven Phase einer Verbindung sollen bestimmte Parameter dynamisch verändert werden können (Q.2963 [679]). Dafür sind die Meldungstypen MODIFY REQUEST, MODIFY ACKNOWLEDGE, MODIFY REJECT und CONNECTION AVAILABLE vorgesehen. Momentan wird von Q.2963.1 [680] eine Modifikation der Spitzenzellrate für Punkt-zu-Punkt-Verbindungen unterstützt. Nur der rufende Teilnehmer kann neue Bandbreiteeinstellungen anfordern. Während der Änderung bleibt die Verbindung weiterhin aktiv.

### 2.5.5.4 Modifikation der Verbindungszahl

Zur Unterstützung von Multimedia-Gesprächen soll es möglich sein, mehrere Nutzverbindungen über eine gemeinsame Signalisierungsbeziehung zu koordinieren und bei Bedarf einzelne Verbindungen hinzuzufügen (Meldung ADD CONNECTION) oder auszulösen (Meldung RELEASE CONNECTION). Diese Gesprächsbeziehung wird als *Multiconnection Call* bezeichnet. Der zugehörige Standard Q.298x [685] ist allerdings noch nicht stabil.

### 2.5.5.5 Zustandsautomat und Protokolltimer

Für jede Teilnehmersignalisierungsbeziehung<sup>36</sup> wird ein Schicht-3-Protokollautomat instanziiert. Aus Sicht der Überlastabwehr interessiert hier vornehmlich, wie ein Endgerät auf unbeantwortete oder abgewiesene Dienstanforderungen reagiert.

Die Antworten nahezu aller vom Endgerät gesendeten Signalisiernachrichten werden auf Teilnehmerseite zeitüberwacht. Besonders kritisch ist der Q.2931-Timer *T303* für die Kontrolle einer SETUP-Nachricht. Läuft dieser Zeitgeber nach vier Sekunden ab wird erneut eine Anforderung gesendet. Beim zweiten Timerablauf wird die Anwendung informiert. Weiterhin wird durch eine RELEASE-Nachricht ein SETUP zurückgewiesen. Wiederum wird dies der Anwendung mitgeteilt. In beiden Fällen wird der Zustandsautomat deinstanziiert.

In beiden Fällen hängt es von der Realisierung und Implementierung der Applikation ab, welches Rufwiederholverhalten durch diese nicht erfolgreich bearbeiteten Anforderungen erzeugt wird.

## 2.5.6 Zeichengabe im Netz: Zwischenamts-Signalisierung

Aus Sicht einer lokalen Last- und Leistungsregelung in einem B-ISDN-Vermittlungsrechner ist die Signalisierung am NNI weniger interessant, da neue Dienstanforderungen aus dem Netzinneren grundsätzlich mit höherer Priorität bearbeitet werden sollten als UNI-Anfragen. Aus diesem Grund sollen hier nur kurz die wichtigsten Protokolle für die netzinternen Signalisierschnittstellen in Breitbandnetzen vorgestellt werden.

### 2.5.6.1 Öffentliche Schnittstellen

Das Protokoll B-ISUP [666] ist der B-ISDN-Anwenderteil im Signalisiersystem Nr. 7 und umfasst die Funktionen, die für die Unterstützung der Verbindungssteuerung für die einfachsten

---

<sup>36</sup> Im Kontext von IN wird hier von einem *Half-Call* [234] gesprochen.



Übermittlungsdienste sowie für die zusätzlichen Dienstmerkmale benötigt werden. Initiiert durch eine SETUP-Meldung des rufenden Teilnehmers wird bei einem abgehenden Ruf in der Teilnehmer-Vermittlungsstelle eine IAM-Meldung (INITIAL ADDRESS MESSAGE) erzeugt, wobei u. a. die in der SETUP-Meldung enthaltenen verbindungs- und dienstbezogenen Informationen übernommen werden. Diese wird durch das Netz weitergeleitet, wobei jeweils die entsprechenden VPCI/VCI-Werte und Bandbreiten für die Nutzkanalverbindung abschnittsweise reserviert werden.

Für die Überlastabwehr repräsentiert eine IAM somit eine neue Dienstanforderung eines ankommenden Rufes oder eines Transitrufes. Entsprechend der Weiterentwicklung der Teilnehmersignalisierung wird auch der B-ISUP um die notwendigen Dienstmerkmale ergänzt.<sup>37</sup>

### 2.5.6.2 Private Schnittstellen

Vom ATM Forum wurde PNNI [610] zur Zeichengabe zwischen privaten ATM-Netzknoten als auch zur Verschaltung von Privatnetzen unterschiedlicher Betreiber spezifiziert. Neben der Signalisierung umfasst PNNI ein sehr leistungsfähiges Routing-Protokoll, das es erlaubt, die Wegesuche dynamisch an die sich ändernde Topologie der Netze sowie an die sich ändernden Lastverhältnisse anzupassen. Die Verbindungssteuerung des PNNI wurde so definiert, dass möglichst viel von der UNI-Signalisierung des ATM Forums übernommen werden konnte.

### 2.5.7 Signalisierszenarien

Signalisiernachrichten stellen den Großteil des Steuerungsverkehrs einer Vermittlungssteuerung dar. Jeder Dienst lässt sich durch eine bestimmte Menge charakteristischer Signalisierszenarien kennzeichnen. Beispielsweise ist der Signalisierablauf des Basisrufes – im Kontext des B-ISDN entspricht dieser Dienst eines Punkt-zu-Punkt-Rufes mit einer Verbindung – in jedem Buch über Signalisierung zu finden. Ein exemplarisches Signalisierszenario für einen komplexeren Dienst wird in Anhang C.3.2.2 näher erläutert.

## 2.6 B-ISDN/ATM-Vermittlungssysteme

ATM wurde als Transporttechnologie und Vermittlungsverfahren für das B-ISDN in I.121 [633] festgelegt. Auf dem Markt erhältliche ATM-Vermittlungssysteme besitzen deshalb als kleinsten gemeinsamen Nenner eine Koppereinrichtung basierend auf ATM-Zellen. Die Funktionalität eines B-ISDN/ATM-Knotens umfasst aber mehr, als Zellen von Eingängen zu Ausgangsports zu vermitteln. Unter einem B-ISDN/ATM-Vermittlungssystem wird ein Netzknoten verstanden, der die volle Funktionalität und alle Protokolle gemäß den B-ISDN-Standards unterstützt.<sup>38</sup>

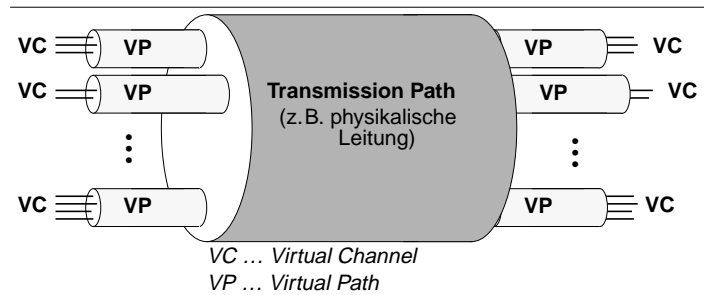
---

<sup>37</sup> Hierzu wurden zwei weitere Stufen definiert, die als *Capability Set 2.1* [664] bzw. *2.2* bezeichnet werden.

<sup>38</sup> Wenn im folgenden von einem *B-ISDN-System* gesprochen wird, dann ist ein Netzknoten mit voller Funktionalität und Protokollunterstützung gemeint. *ATM-Vermittlungssysteme* dagegen haben mehr rudimentären Charakter, da sie nur eine Untermenge der B-ISDN-Systeme darstellen. Beispielsweise besitzen die meisten der derzeit leistungsfähigsten *Router* von IP-Paketen als Vermittlungskern ein ATM-Koppelnetz.

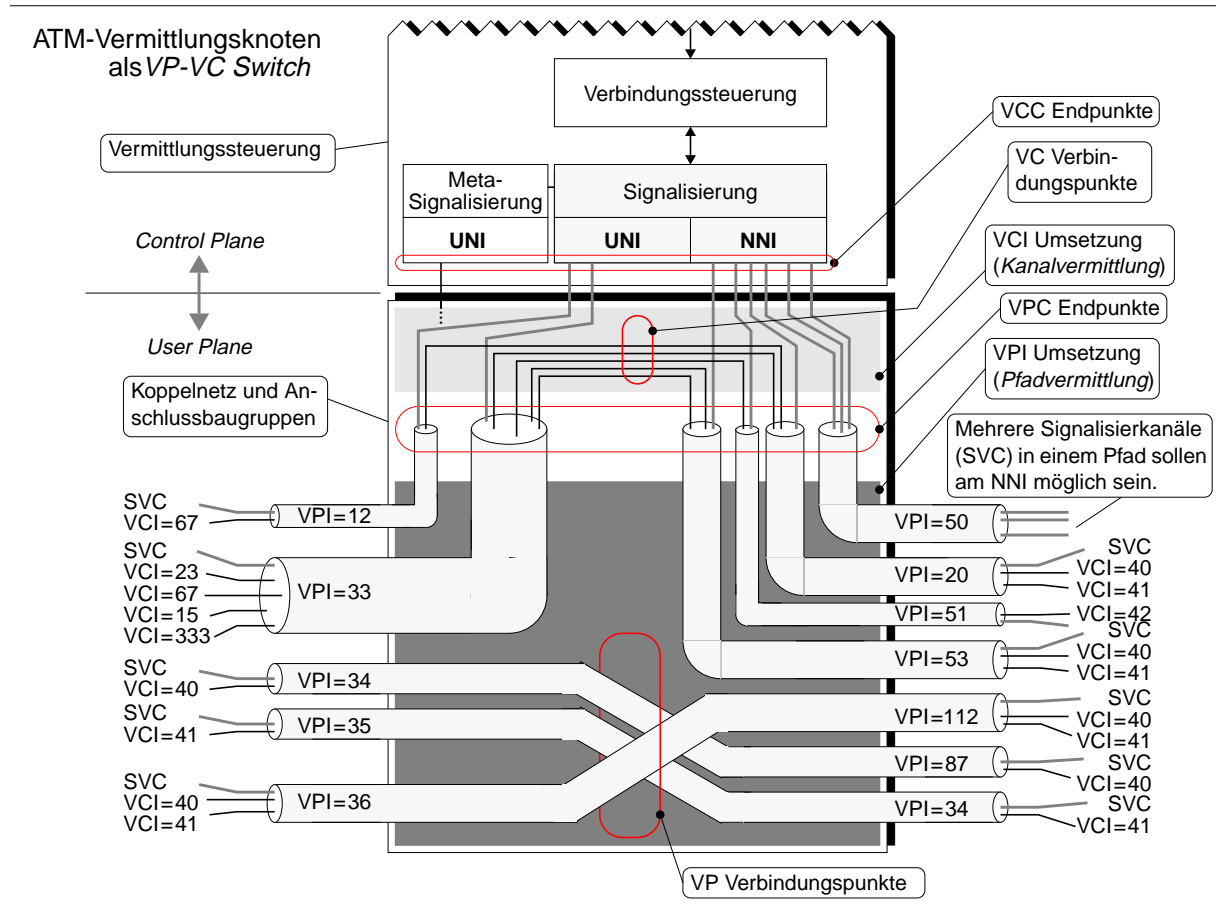
## 2.6.1 ATM-Vermittlungstechnik

Die eigentliche Vermittlung von ATM-Zellen geschieht in der Koppelanordnung (*Cell Switching Network*) eines ATM-Knotens. THEIMER [515] gibt eine ausführliche Darstellung über Funktionsumfang, Architekturen und ihre Leistungsfähigkeit von ATM-Koppelnetzstrukturen. Bei Punkt-zu-Mehrpunkt-Verbindungen werden im Koppelnetz die betreffenden Zellen vervielfacht und zu verschiedenen Ausgängen weitergeleitet (*Multicasting*). Durch das *VPI/VCI*-Feld wird jede ATM-Zelle einem virtuellen Kanal innerhalb eines virtuellen Pfades zugeordnet (Bild 2-16). Alle Zellen einer ATM-Verbindung besitzen dieselbe Kennung. Die Verbindungskennung ist jeweils nur für einen Übertragungsabschnitt gültig und wird deshalb in jedem Netzknoten umgewertet.



**Bild 2-16:** Beziehung zwischen Kanälen und Pfaden

die betreffenden Zellen vervielfacht und zu verschiedenen Ausgängen weitergeleitet (*Multicasting*). Durch das *VPI/VCI*-Feld wird jede ATM-Zelle einem virtuellen Kanal innerhalb eines virtuellen Pfades zugeordnet (Bild 2-16). Alle Zellen einer ATM-Verbindung besitzen dieselbe Kennung. Die Verbindungskennung ist jeweils nur für einen Übertragungsabschnitt gültig und wird deshalb in jedem Netzknoten umgewertet.



**Bild 2-17:** Beispiel für die Vermittlung von Kanälen und Pfaden (*VP-VC Switch*). Meta-Signalisierkanäle wurden nicht eingezeichnet. Nach ITU-T I.311 [637] werden Verbindungspunkte und Verbindungs-Endpunkte unterschieden.

ATM-Netzknoten, die eine Zellvermittlung nur auf Basis der Pfadkennung realisieren, werden als ATM-Pfadvermittlung bezeichnet, während die ATM-Kanalvermittlungen die vollständige Verbindungskennung auswerten.<sup>39</sup>

Die Signalisierinformationen werden getrennt von den eigentlichen Nutzdaten in speziellen Signalisierkanälen (*Signalling Virtual Channel*) transportiert (Außerband-Signalisierung, vgl. Bild 2-17). Für die Signalisierverbindungen sind auf ATM-Schicht bestimmte VPI/VCI-Kombinationen reserviert. Für Punkt-zu-Punkt-Konfigurationen wird beispielsweise die Kodierung  $VPI=x/VCI=5$  (mit  $x \neq 0$ ) verwendet [639], d.h. für jeden Pfad ist ein eigener Signalisierkanal vorgesehen. Ein Pfad kann aber auch mehrere Signalisierkanäle transportieren (unterschiedliche *VCIs*).

## 2.6.2 Generische Architektur einer B-ISDN/ATM-Vermittlungsstelle

### 2.6.2.1 Spezifische Funktionen

Neben den allgemeineren Funktionen eines Vermittlungssystems und seiner Steuerung (siehe Abschnitte 2.1.1 und 2.3.1) sind spezifische Aufgaben von einem B-ISDN/ATM-Vermittlungsknoten zu erfüllen. Im Rahmen dieser Arbeit stehen die Aufgaben der Steuerungs-Ebene (*Control Plane*) im Vordergrund. Die meisten Aufgaben der Nutzer-Ebene (*User Plane*) in einem Netzknoten haben aber natürlich einen indirekten Einfluss auf die Steuerungs-Ebene, da diese Funktionen von der Vermittlungssteuerung aktiviert, konfiguriert, überwacht usw. werden müssen.

Für jede Verbindung wird zwischen einem Teilnehmer und dem Netz ein Verkehrsvertrag (*Traffic Contract*) abgeschlossen, der die Grundlage der Verkehrssteuerung bildet. Dieser Vertrag umfasst Verkehrsparameter und Anforderungen an die Dienstgüte für die betrachtete Verbindung. Für diese Verkehrssteuerung wurde eine ganze Reihe von sich einander ergänzenden Einzelmechanismen spezifiziert (siehe z.B. [362,383,393,611,644]).

#### 2.6.2.1.1 Nutzer-Ebene

Betrachtet werden nur Funktionen oberhalb der Bitübertragungsschicht. Die Nutzinformation von ATM-Zellen ist transparent für ein ATM-Vermittlungssystem. Näher betrachtet wird nur der ATM-Zellkopf. Folgende Standardfunktionen sind zu bearbeiten:

- Filterfunktion für ungültige und fehlgeleitete Zellen
- Überprüfung und Umsetzung der Verbindungskennung (VPI/VCI)<sup>40</sup>
- Überwachung von ATM-Zellströmen bezüglich ihres Verkehrsvertrages (Policing-Algorithmen), namentlich die *Usage Parameter Control* (UPC) und *Network Parameter Control* (NPC) [126]
- Verkehrsformung (*Traffic Shaping*)

<sup>39</sup> Für die Bezeichnung eines Netzknotens unterscheidet I.311 [637], welche Verbindungskennung umgesetzt wird (Pfadkennung *VPI*, Kanalkennung *VCI* oder beide *VPI/VCI*), und ob die Einstellinformation von der Management-Ebene für eine Festverbindung (Netzknoten = *Cross-Connect*), oder von der Steuerungs-Ebene für eine Wählverbindung (Netzknoten = *Switch*) kommt. Die Begriffe reichen somit vom *VP Cross-Connect* bis zum *VP-VC Switch*.

<sup>40</sup> Um Suchverfahren (Tabellengröße, notwendiger Speicherbedarf) zu vereinfachen oder bei selbststroutenden Koppelnetzen wird in der Regel knotenintern ein herstellerepezifisches, verkleinertes Adressformat eingesetzt.

- Prioritätensteuerung (z.B. Unterscheidung hoch- und niederpriore Zellen innerhalb einer Verbindung, Zeit- bzw. Verlustprioritäten [219] oder verbindungs-spezifische Prioritäten wie beispielsweise *Weighted Fair Queueing*)
- Unterstützung von Flusskontrollverfahren [285]
- Schnelle Ressourcenzuteilung
- Ein-/Auskopplung von Signalisier- und Managementzellen
- Überlastbehandlung auf ATM-Schicht [442,443]
- ATM-Schicht-Verkehrsmessungen

Die Verkehrssteuerungs-Mechanismen können aufgrund ihres Reaktionsverhaltens verschiedenen zeitlichen Ebenen – Zellübertragungszeit, Signallaufzeit im Netz oder Verbindungsdauer – zugeordnet werden.

Wichtige Funktionen der ATM-Anpassungsschicht sind die Unterstützung der verschiedenen Verbindungstypen (z.B. ITU I.371 [644] *ATM Transfer Capabilities* wie *DBR, SBR, ABR, ABT* oder ATM-Forum [612] *Service Categories* wie *CBR, rt-VBR, nrt-VBR, ABR, UBR*) und Dienstklassen (z.B. *Klasse A, B, C* oder *D* nach I.362 [640]).

Für diese Arbeit ist das Verständnis der einzelnen Funktionen weniger wichtig. Die meisten Mechanismen werden in der Aufbauphase für eine Verbindung konfiguriert. Viele Bausteine für eine schnelle Bearbeitung der ATM-/AAL-Schichtfunktion in Hardware sind bereits auf dem Markt. Bestimmte Funktionen erfordern aber während der aktiven Verbindungsphase das Eingreifen der Vermittlungssteuerung. Entweder kommen die Anreize hierfür durch die Signalisierung (z.B. Modifikation von Verbindungsparametern) oder wie beispielsweise beim ABR-Dienst von bestimmten Nutzzellen (sogenannte *Resource Management Cells*).

#### 2.6.2.1.2 Steuerungs-Ebene

Die meisten allgemeinen Aufgaben der Steuerung unterscheiden sich wenig von denen eines Schmalband-ISDN-Knotens. Die vermittlungsspezifischen Steuerungsfunktionen wurden jedoch aufgrund der ATM-Technik und den B-ISDN-Protokollen umfangreicher und komplexer (siehe 2.5.1). Es wird zwischen Rufen und Verbindungen unterschieden, es werden mehr Ruf-typen und Verbindungskonfigurationen unterstützt, und es muss ein weites Spektrum unterschiedlichster Anwendungen und Dienste abgedeckt werden. Allein die Entscheidung der Annahme eines neuen Rufes oder einer neuen Verbindung (*Call/Connection Admission Control*), ist, bei einer effizienten Unterstützung des statistischen Multiplexens, eine aufwendige und rechenzeitintensive Funktion.

Die verbindungsbezogenen Aufgaben der zentralen Steuerung lassen sich bei einem ATM-System in folgende Funktionsblöcke unterteilen:

- *Auswertung der Verbindungskennung*: Bestimmung des Zielknotens des gerufenen Teilnehmers sowie Behandlung ungültiger Adressen.
- *Verkehrslenkung (Routing)*: Auf Basis der Rufnummer (Adresse) des gerufenen Teilnehmers muss ein Weg, d.h. eine Folge von Teilnetzen und Vermittlungsstellen bestimmt werden, über die eine Verbindung geführt wird. Routingalgorithmen werden in der Regel nicht von Standardisierungsgremien festgelegt. In [532] werden Vorschläge von Routingprotokollen für öffentliche ATM-Netze diskutiert.
- *Verbindungsannahme (Connection Admission Control, CAC)*: Eine Verbindung wird durch ihre Verkehrsparameter und Dienstgüteanforderungen charakterisiert. Durch das

statistische Multiplexen bei ATM ist eine gegenseitige Beeinflussung der Verbindungen unvermeidlich. Unter Berücksichtigung dieser Kennwerte muss die CAC überprüfen, welcher Ressourcenbedarf benötigt wird und ob die Verbindung ohne Verletzung der Dienstgüte-Kriterien bereits eingerichteter Pfade und Verbindungen angenommen werden kann. Die Gesamtentscheidung für die Annahme oder Ablehnung einer Verbindung beruht auf den Einzelentscheidungen für die betreffenden Verbindungsabschnitte. Bei Ablehnung besteht immer noch die Möglichkeit der Nachverhandlung mit modifizierten Verbindungsparametern.

Für eine angenommene Verbindung müssen weiterhin die *Policing*-Parameter bestimmt und in den Leitungsmodulen eingestellt werden. Die CAC ist weiterhin bei Anforderungen bzgl. der Modifikation von Verbindungsparametern involviert. In [328] werden zahlreiche CAC-Verfahren vorgestellt und bewertet. Die CAC hat einen ausgeprägten zentralen Charakter, da sie sehr eng mit der Ressourcenverwaltung zusammenarbeitet. Sie stellt deshalb einen potentiellen Systemengpass dar. Es existieren Ansätze, diese Funktion auf mehrere Recheneinheiten zu verteilen [14], was allerdings erhöhte Anforderungen hinsichtlich Datenkonsistenz und Interprozessor-Kommunikation aufwirft.

- *Ressourcenverwaltung*: Die Verwaltung von vermittlungstechnischen Betriebsmitteln erfolgt in zentralen Datenbanken. VCI/VPI-Kennungen, Bandbreite und Zellpuffer sind die wichtigsten Ressourcen, die beim Aufbau, Modifikation und Auslösen einer Verbindung belegt und wieder freigegeben werden müssen.
- *Interne Wegesuche*: Ein Resultat der netzweiten Wegesuche ist der zu verwendende Ausgangsport des Koppelnetzes. Für diesen Port muss die interne Zieladresse ermittelt werden. Bei bestimmten Koppelnetztypen kann ein Ausgang über mehrere interne Wege erreicht werden. Hierzu muss mittels eines internen Wegesuchalgorithmus vor dem eigentlichen Verbindungsaufbau ein Weg ausgewählt werden [515].
- *Einstellung der Koppelanordnung*: Bestimmte Einstellinformationen, wie beispielsweise Zieladressen oder Pufferstrategien, sind bei bestimmten Koppelarchitekturen notwendig. Mit einem Minimum an externer Steuerinformation kommen die sogenannten selbststeuernden Koppelnetze (*selfrouting networks*) aus.
- *Einstellung der verbindungspezifischen Parameter*: In den Leitungsmodulen sind für jede Nutzverbindung Policing-Parameter und Adressenumwertinformationen einzustellen.

Die Ressourcenverwaltung der Steuerungs-Ebene selbst werden im nachfolgenden Kapitel 3 im Zusammenhang mit der Überlastabwehr diskutiert.

### 2.6.2.1.3 Management-Ebene

Die dritte Ebene im B-ISDN Referenzmodell repräsentiert die Managementaufgaben. Für das Systemmanagement eines ATM-Netzknotens umfasst dies die Funktionsblöcke *Fault* -, *Performance* -, *Configuration* -, *Security* -, *Accounting* - und *Traffic Management*. Für die Überwachung der ATM-Schicht wurden umfangreiche OAM-Prozeduren spezifiziert, die vollständig in Hardware realisiert werden. Ein Großteil der Managementaufgaben wird in der Regel nicht im eigentlichen Vermittlungsrechner, sondern in einem eigenen, oft abgesetzten Bedienrechner bearbeitet. Da diese Aufgaben zusätzlich in einer höheren Zeitebene ablaufen („weniger echtzeitkritisch“) als die eigentlichen Vermittlungsaufgaben, werden sie im Rahmen dieser Arbeit nicht weiter mehr betrachtet.

### 2.6.2.1.4 Schnittstellenunterstützung

B-ISDN/ATM-Systeme müssen die zugrundeliegenden Protokolle unterstützen. Daneben steht aber auch immer die Forderung, Schnittstellenfunktionen – vertikal als auch horizontal – für andere Dienste und Technologien zur Verfügung zu stellen. Diese Forderung betrifft das Interworking mit anderen Netztechnologien wie beispielsweise *Schmalband-ISDN*, *Frame Relay*, *Ethernet*, *FDDI* oder *SMDS* und die Unterstützung bestehender Dienste. Andere Beispiele sind Kopplungen an reine Signalisier- (z.B. SS Nr. 7 mit MTP 1-3) und Netzverwaltungsnetze (z.B. *TMN* [645]) sowie im *Intelligenten Netz* die Schnittstelle zwischen *SSP* und *SCP* bzw. *SDP*.

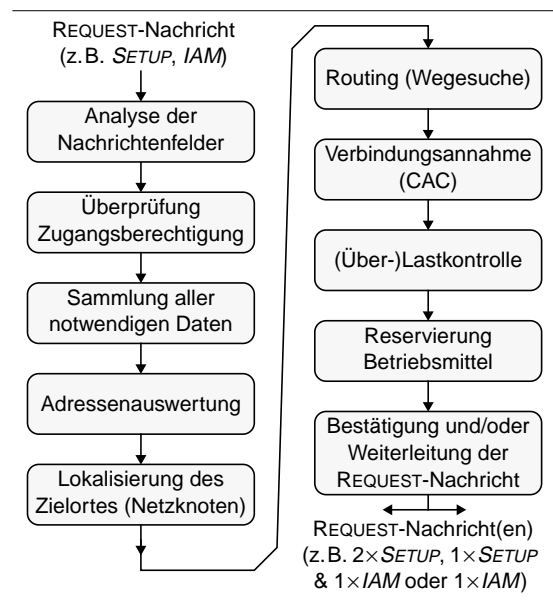
Für Nicht-ATM-Dienste sind dienstspezifische Anpassungs- und Umsetzungsfunktionen (*Interworking Function*, *IWF*) notwendig. Wenn lediglich die unteren, transportorientierten Protokollschichten ineinander umgesetzt werden, spricht man vom *Netz-Interworking*. Werden auch die höheren Protokollschichten, die in den Endgeräten angesiedelt sind, in die Umsetzung miteinbezogen, spricht man vom *Dienst-Interworking* [54].

### 2.6.2.2 Bearbeitungsschritte beim Ruf- und Verbindungsaufbau

Jede neue Dienstanforderung durchläuft eine charakteristische Folge von Bearbeitungsschritten. Diese Sequenz (Bild 2-18, auch kanonische Reihenfolge genannt) ist auch für Schmalband-Systeme gültig. Im B-ISDN besitzen die einzelnen Funktionsblöcke aber weitestgehend eine andere inhaltliche Ausprägung.

Zunächst wird für jede neue Rufanforderung die Zugangsberechtigung und die Dienstnutzungsrechte überprüft.<sup>41</sup> Weiterhin müssen alle notwendigen Daten gesammelt werden, da die erste *SETUP*-Nachricht nicht die komplette Rufinformation liefern muss. Bei vollständig vorliegender Information werden dann die Adressen ausgewertet und ggf. umgesetzt. Beispielsweise wird auch festgestellt, ob es sich um einen IN- oder Nicht-IN-Ruf handelt. Anhand der Adresse des gerufenen Teilnehmers wird dann ermittelt, über welchen Vermittlungsknoten das entsprechende Endgerät erreicht werden kann. Handelt es sich nicht um einen lokalen Ruf, dann ist eine Wegesuche notwendig. Bei erfolgreichem Routing liegen nun Ein- und Ausgangsport (siehe auch Bild 2-19) dieser Verbindung am Netzknoten fest und die *CAC* kann überprüfen, ob die angeforderten Ressourcen lokal verfügbar sind. Bei erfolgreicher Verbindungsannahme werden die entsprechenden Betriebsmittel reserviert und anschließend diese Anforderungsmeldung bestätigt und/oder weitergeleitet.

Für ein optimales Leistungsverhalten eines Vermittlungsrechners ist eine Lastkontrolle unerlässlich. Je nach Zielsetzung kann dieser Funktionsblock an unterschiedlichen Punkten in die Bearbeitungskette eingefügt werden (z.B. nach *CAC* in Bild 2-18).



**Bild 2-18:** Sequenz notwendiger Bearbeitungsschritte bei erfolgr. Rufannahme

<sup>41</sup> Auch Teilnehmerprofil genannt. Z.B. wird geprüft, ob Teilnehmer nur lokale oder auch Ferngespräche führen darf.

### 2.6.2.3 Funktionsblöcke und generische Architektur

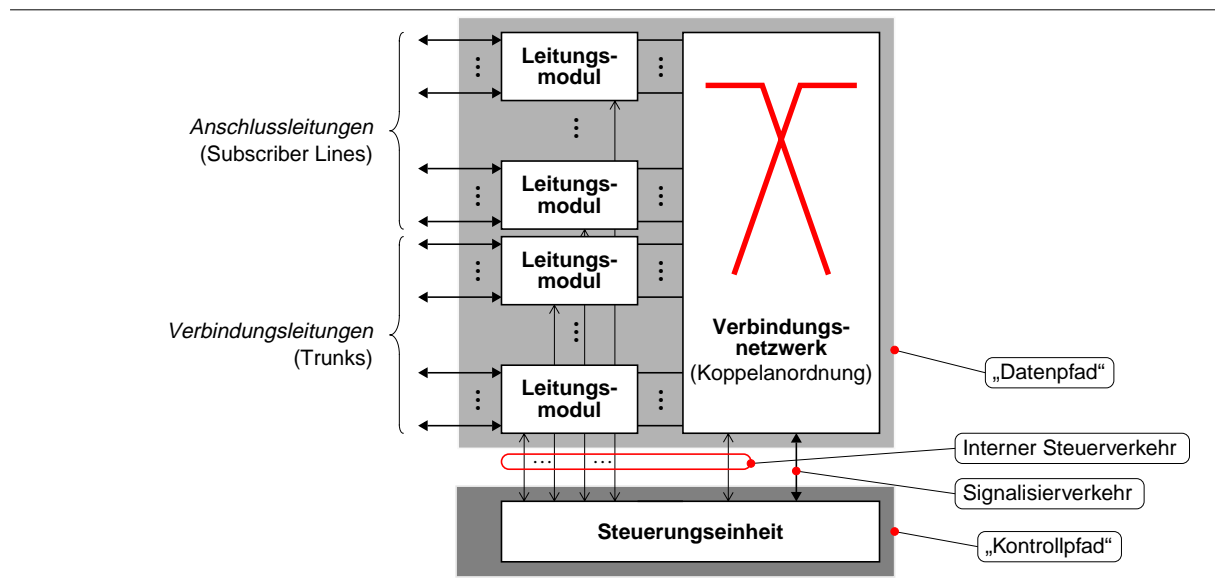
Aufgrund der unterschiedlichen Anforderungen hinsichtlich Schichtzugehörigkeit, Bearbeitungsgeschwindigkeit und Komplexität an die einzelnen Funktionen lassen sich drei Hauptfunktionsblöcke unterscheiden (Bild 2-19, [14]):

- *Koppelnetz* oder *Verbindungsnetzwerk (Cell Switching Network/Fabric)*:

Diese Anordnungen für die eigentliche Zellvermittlung waren und sind Gegenstand zahlreicher Untersuchungen (z.B. [171,393,515,571]). Wichtige Leistungswerte des Koppelfeldes sind die Art und Anzahl der Anschlüsse (z.B. STM-1 155.52-Mbit/s-Port), der Gesamtdurchsatz (z.B. 10 Gbit/s [695]), Verlustwahrscheinlichkeit und Zell-Verzögerungszeit (z.B. 10  $\mu$ s [698,700]). Einstellvorgänge und Überwachungsaufgaben werden von einem oder mehreren lokalen Prozessoreinheiten wahrgenommen.

- *Leitungs-Anschlussgruppen* oder *-module (Input & Output Modules)*

Der Aufgabenbereich dieser Einheiten umfasst die Bearbeitung der Protokolle der Bitübertragungsschicht, der ATM- und der ATM-Anpassungsschicht. Weiterhin erfolgt hier die Anpassung an interne Schnittstellenformate und lokale Betriebs- und Wartungsfunktionen. Wie im Koppelnetz erfolgt auch hier die Bearbeitung der Dateneinheiten (Zellen) vollständig in Hardware. Für die Einstellung, Konfiguration und Überwachung des Zellpfades besitzt jedes Leitungsmodul ein oder mehrere lokale Prozessoren. Diese, aus Sicht der Zentralsteuerung, peripheren Prozessoren sind z.B. für Einträge in die Adressenumsetzungstabelle, Einstellung der Policing-Parameter oder Pufferstrategien verantwortlich.



**Bild 2-19:** Generische Architektur eines ATM-Netzknottens

- *Vermittlungsrechner* als zentrale Steuerungseinheit (*Switching* oder *Control Processor*)

Diese Komponente ist vornehmlich für die Abwicklung des Aufgabenbereichs der Steuerungs- und der Management-Ebene verantwortlich. Darunter fallen alle Funktionen im Zusammenhang mit der Ruf- und Verbindungssteuerung, der Betriebs- und Wartungsaufgaben, Funktionen des Netzmanagement und der Sicherungstechnik (siehe Abschnitte 2.6.2.1.2, 2.6.2.1.3). Die Funktionen sind vollständig in Software realisiert. Als Hard-

wareplattform deutet sich der Trend an, vermehrt Standard-(Industrie-)Rechnerbaugruppen mit kommerziellen Prozessortypen einzusetzen.<sup>42</sup> Aufgrund der Anforderung nach Skalierbarkeit reichen die Steuerungsstrukturen von Einzelrechner- bis zu sehr aufwendigen Mehrprozessorsystemen (siehe Anhang A.2).

Weitere typische Funktionsblöcke sind beispielsweise Anschlussgruppen für Zellvervielfachung (z.B. *Cell Copy Unit* [223]) zur Unterstützung der Multicastfähigkeit für Punkt-zu-Mehrpunktverbindungen<sup>43</sup> oder Einheiten zur Echounterdrückung. Leitungsmodule und Koppelnetz arbeiten auf Zeitebene von Zellen. Die Steuerungseinheit auf Ebene von Verbindungen.

### 2.6.3 Kennwerte von ATM-Vermittlungssystemen

Eine umfassende Bewertung der *Gesamtleistung eines ATM-Vermittlungssystems* muss die unterschiedlichen Aspekte von Teilnehmer-, Netz- und Betreibersicht berücksichtigen. Die Schwerpunkte liegen dabei auf folgenden Eigenschaften:

- *Teilnehmersicht*

Im Vordergrund stehen die Leistungsdaten und verfügbaren Dienste der Nutzer-Ebene wie beispielsweise unterstütztes Dienstespektrum und Dienstparameter, nutzbare Bandbreite (angebotene Schnittstellen z.B. PDH E3, SDH STM-1 oder Sonet STS-3c), Antwortzeitverhalten für Signalisier Nachrichten [539] oder Zelltransportcharakteristik (Zellverzögerung, Verzögerungsschwankung, Verlust- und Fehlleitungswahrscheinlichkeit).

- *Netz- und Betreibersicht*

Vermittlungskapazität in Zellen pro Sekunde, Verarbeitungs- oder Steuerungskapazität in "BHCA" ("Rufleistung"), Koppelnetzdurchsatz (Systeme mit *Tbit/s*-Werten sind bereits angekündigt [202]), Anzahl von Anschlüssen (z.B. *Abacus Switch* [123] mit mehreren tausend Ports), maximale Anzahl gleichzeitiger Verbindungen, verfügbare Protokolle (z.B. für Wählverbindungen) und Protokollversionen, Verbindungsannahmemechanismus und Routingverfahren, Verbindungsüberwachung (Policing) und Verkehrssteuerung, Lastmanagement und Überlastabwehr, verfügbare Schnittstellen, Netzübergänge, Erweiterbarkeit, Verfügbarkeit, Redundanzkonzepte oder Skalierbarkeit.

Für die Bewertung der *Leistungsfähigkeit einer Vermittlungssteuerung* sind spezielle Kenngrößen von Interesse. Geeignete Leistungsmaße werden im Zusammenhang mit der Überlastproblematik in Abschnitt 3.1.3 diskutiert.

### 2.6.4 Aufbau von Steuerungseinheiten

Die Regelkreisstruktur einer Überlastabwehr in einem Vermittlungsrechner orientiert sich primär an dessen Hard- und Softwarearchitektur. Aus diesem Grund werden nachfolgend spezifi-

---

<sup>42</sup> Aus wirtschaftlichen Gründen geht man davon ab, eigenentwickelte Spezialprozessoren, Rechnerkarten oder Schnittstellenformate (z.B. Busse) einzusetzen. Heutige Vermittlungssteuerungen verwenden dieselben Prozessor- und Controllertypen wie sie auch im PC- und Arbeitsplatzrechnerbereich eingesetzt werden. Dieser Trend setzt sich in der Softwareentwicklung fort, wo der Trend weg von eigenentwickelten Sprachen oder ITU-Standards (z.B. *CHILL* [688]) hin zu industrieeüblichen Hochsprachen wie *C* oder *C++* geht.

<sup>43</sup> Modul ist dann notwendig, wenn diese Funktion nicht bereits durch das Koppelnetz erbracht werden kann.



sche Implementierungskonzepte und Architekturmerkmale von ATM-Vermittlungssteuerungen vorgestellt.

### 2.6.4.1 Implementierungskonzepte

Die Steuerungsstruktur für einen B-ISDN/ATM-Netzknoten ist, im Vergleich zu den Vermittlungssteuerungen für das analoge Telefonnetz oder Schmalband-ISDN, weitergehenden Anforderungen unterworfen. In [54] werden wichtige Kriterien diskutiert:

- Aufwand für die Steuerung einer einzelnen Verbindung wird zunehmend größer. Ursachen hierfür sind die Bereitstellung von Dienstmerkmalen im Weitverkehrsbereich, die bisher nur in Privatnetzen angeboten wurden, oder neue Dienste, die im Rahmen des Intelligenzen Netzes eingeführt wurden.
- Trend zum Einsatz größerer Vermittlungsknoten im Kernnetzbereich. Die zugrundeliegende Motivation für einen Netzbetreiber ist die Anzahl von Netzelementen und auch der Netzebenen zu reduzieren.
- Zeichengabeprotokolle werden immer komplexer.
- Steigende Mobilität der Teilnehmer.
- ATM als gemeinsame Basis führt zu einem Verschwinden der Grenzen zwischen den bisher streng getrennten Welten der privaten und öffentlichen Netze, der Übertragungs- und Vermittlungstechnik sowie der Daten-, Sprach- und Verteilkommunikation. Jede dieser Welten hat ihre gewachsenen Steuerungskonzepte, die nun in eine Kontrollplattform integriert werden müssen.
- Entkopplung des Nutzdatendurchsatzes von der erforderlichen Steuerungsleistung. Im Gegensatz zu den durchschaltvermittelnden Netzen, wo die Anzahl der verfügbaren Nutzkanäle eine relativ genaue Abschätzung der Steuerungsleistung erlaubt, lässt bei ATM-Systemen die Leistungsfähigkeit des Koppelnetzes keine Rückschlüsse auf die erforderliche Kapazität der zentralen Steuerung zu. Der Steuerungsaufwand für eine Verbindung korreliert mit der zugrundeliegenden Anwendung. Aufgrund der Vielzahl von Diensten, die unterstützt werden sollen, ist nun hier ein breites Spektrum an Szenarien denkbar.<sup>44</sup>

In [54] werden diese Überlegungen in der Forderung zusammengefasst, dass die Steuerungskapazität unabhängig von der reinen Zellvermittlungskapazität flexibel skalierbar sein muss und an die jeweiligen Erfordernisse angepasst werden kann.

### 2.6.4.2 Hardwareaspekte der Steuerung

Bei Vermittlungssystemen zeichnet sich der Trend ab, von eigenentwickelten Spezialprozessoren mit speziell für die Vermittlungstechnik optimierten Merkmalen wegzugehen und allgemein verfügbare Standardprozessoren einzusetzen. Diese Richtung begründet sich aus der sich

---

<sup>44</sup> Angenommen, durch den Dienst *Video on Demand (VoD)* werde eine mittlere Bandbreite  $C_{VoD}=4$  Mbit/s für ungefähr 150 min in Anspruch genommen. Auf der anderen Seite ließen sich mit demselben Ressourcenbedarf circa 3000 Fernsprechverbindungen, bei einer mittleren Rufdauer von 3 min, vermitteln. Betrachtet man den *Steuerungsaufwand*, dann liegt hier der Ressourcenbedarf von *VoD* bei wenigen Fernsprechrufen. D.h., in diesem Beispiel liegt das Verhältnis der unterschiedlichen Steuerungsbelastung ungefähr bei 1:1000!

Noch krasser ist das Verhältnis bei einem *ATM-Cross-Connect*. Dieser besitzt dieselben Komponenten in der Nutzer-Ebene wie die vollwertigen *VP-VC Switches*, ist auf der anderen Seite aber nur mit einer sehr einfachen Steuerung ausgestattet sind, da nur durch das Netzmanagement gesteuerte Festverbindungen unterstützt werden müssen.

laufenden Leistungssteigerung kommerzieller Mikroprozessoren und wirtschaftlichen Gesichtspunkten. Zusätzlich reduziert sich der Softwareentwicklungsaufwand, da beispielsweise eine breite Palette verfügbarer Echtzeit-Betriebssysteme eingesetzt werden kann.

Für kleine ATM-Vermittlungsanlagen (z.B. im Privatnetzbereich) kann bereits ein Prozessor-knoten (Mikrorechner) als Steuerungseinheit ausreichend sein. Für die größeren, universell einsetzbaren Knoten im Weitverkehrsnetz liegen die Anforderungen an die Steuerung in Bereichen, welche die Leistungsfähigkeit einzelner Mikroprozessoren bei weitem überschreiten.<sup>45</sup> Daraus ergibt sich, dass für größere Vermittlungssysteme nur Mehrprozessor-Architekturen in Frage kommen [479].

### **2.6.4.3 Softwareaufbau**

Wie bei den Hardware-Komponenten zeichnet sich auch hier der Trend ab, kommerziell verfügbare Software-Produkte einzusetzen. Durch den Einsatz marktüblicher Betriebssysteme wird eine weitgehende Unabhängigkeit von der zugrundeliegenden Hardware-Plattform erreicht. Weiterhin werden auch vermehrt Softwarekomponenten, z.B. Signalisierprotokoll-stacks oder Managementsoftware, für diese Betriebssysteme angeboten, so dass sich prinzipiell der Entwicklungsaufwand reduzieren lässt. Auf der anderen Seite sind gerade in der öffentlichen Technik im Lauf einer Produktfamilie tausende von Mannjahren in die Vermittlungsprogramme eingeflossen. Diese Investitionen will und kann man verständlicherweise nicht so ohne weiteres aufgeben. Die grundsätzlichen Anforderungen an die Steuerungssoftware eines B-ISDN-Vermittlungssystems unterscheiden sich nicht von den grundlegenden Konzepten wie sie bereits in Abschnitt 2.3.2 beschrieben wurden.

### **2.6.4.4 Verteilung von Aufgaben: Last- und Funktionsteilung**

Die leistungsfähigsten Vermittlungsrechner basieren auf Multiprozessorarchitekturen mit einer verteilten Steuerung. Aus den oben erwähnten Aspekten gilt dies umso mehr für ATM-Systeme. Beispielarchitekturen mit verteilten Steuerungsstrukturen werden in Anhang A.2.2 vorgestellt. Wie bereits in Abschnitt 2.3.5 erwähnt wurde, wird zur Leistungssteigerung das Last- und Funktionsteilungsprinzip angewendet. Im Zusammenhang mit Regelkreisstrukturen zur Last- und Leistungsregelung in einem B-ISDN-Vermittlungsrechner soll dieses Prinzip nachfolgend (Abschnitt 2.6.5) als Klassifizierungsschema von Steuerungseinheiten angewendet werden.

### **2.6.4.5 Interne Steuerungskanäle**

Jede größere Hardwaregruppe (z.B. Koppelnetz oder Leitungsmodule) der Nutzer-Ebene besitzt für lokale Steuerungsaufgaben einen eigenen Mikroprozessor. Diese werden aus Sicht des zentralen Vermittlungsrechners als *periphere Prozessoren* bezeichnet. Zwischen Nutzer- und Steuerungs-Ebene existieren zwei verschiedene Arten von Steuerungssignalflüssen: Nachrichten der Signalisierprotokolle und knoteninterne Steuerungsnachrichten.

Für die Kommunikation zwischen Nutzer- und Steuerungs-Ebene bestehen grundsätzlich zwei verschiedene Möglichkeiten. Entweder wird eine dedizierte Infrastruktur (z.B. leistungsfähiger Standardbus oder LAN-Protokoll) vorgesehen oder es wird die Kommunikationsinfrastruktur

---

<sup>45</sup> In [54] werden 50000 BHCA als maximale Kapazität für einen modernen Einzelprozessor (als optimistische Abschätzung) angegeben.

tur der Nutzzellen mitverwendet (siehe auch Bild 2-19). In [54] werden die Vor- und Nachteile beider Varianten ausführlich diskutiert.

Im Gegensatz zu Schmalband-ISDN macht es bei ATM Sinn, die komplette Signalisierbearbeitung im zentralen Vermittlungsrechner zusammenzufassen. Ein Signalisierkanal wird dann nicht bereits im Leitungsmodul terminiert, sondern über das Koppelfeld direkt bis an die zentrale Steuerung geführt. Mit aus diesem Grund wird mehrheitlich die zweite Variante, die ausschließliche Kommunikation über das Koppelnetz, in den meisten ATM-Systemen eingesetzt. Der Austausch interner Steuerungsnachrichten erfolgt dann ebenfalls mittels ATM-Zellen. Als Protokoll wird i.d.R. eine vereinfachte Schicht-2-Signalisierung (z.B. *SSCOP*-Variante) vorgesehen.

Von dieser Systemvariante wird auch im Rahmen dieser Arbeit ausgegangen, d.h. der Protokollstack der *Control Plane* wird bis zur ATM-Schicht in der Nutzer-Ebene und oberhalb dieser Schicht komplett im zentralen Vermittlungsrechner bearbeitet. Die in bestimmten Systemarchitekturen durchaus sinnvolle Auslagerung von Signalisierungsfunktionalität in die Leitungsmodule (dezentraler Ansatz) wird beispielsweise in [128] diskutiert.

## 2.6.5 Klassifizierung von Steuerungseinheiten

Zur Einteilung von Steuerungen rechnergesteuerter Vermittlungssysteme wurde von DENZEL [146] ein Schema eingeführt, das sich am strukturellen Aufbau orientiert. Die Klassifizierung beruht dabei auf einer Menge von Grundelementen<sup>46</sup>, wobei die jeweilige Realisierungsform ohne Belang ist. Da in dieser Arbeit primär die Verarbeitungseinrichtungen von Interesse sind, wird dieses Strukturierungsschema nicht weiter betrachtet.

Die nachfolgende Klassifizierung orientiert sich deshalb in erster Linie an der Anordnung und Anzahl von Steuerungsprozessoren. Im zweiten Schritt werden dann weitere Aspekte, beispielsweise welche Funktionsaufteilung vorgenommen wird, berücksichtigt. Die Kommunikationsbeziehungen zwischen verschiedenen Verarbeitungseinheiten und die Infrastruktur zum Steuerungsdatenaustausch werden nicht als Ordnungsmerkmale für die Klassifizierung verwendet.

Betrachtet man den räumlichen Wirkungskreis einer Lastregelung, dann sei die unterste Ebene auf einen Prozessor begrenzt. Das zentrale Grundelement sei deshalb der *Prozessorknoten* (PN), worunter hier ein Mikroprozessor mit eigenem Speicher verstanden wird.<sup>47</sup> Damit kann hier ein Vermittlungsrechner durch folgende einfache *Notation* beschrieben werden:

Ein (*A, B, C, ...*)-*Vermittlungsrechner* besitzt mehrere Funktionsebenen die mit *A, B* usw. bezeichnet werden sollen. Jede Funktionsebene besitzt dabei mindestens einen oder mehrere gleichberechtigte Prozessorknoten.<sup>48</sup>

Typischerweise bewegt sich die Anzahl an Funktionsebenen derzeit zwischen eins und drei bei kommerziellen Systemen. Bei einem *System mit drei Funktionsebenen* wird oft folgende Aufgabenteilung gemacht:

- *A-Ebene*: niederere Signalisierung, d.h. Schicht-2-Protokolle (z.B. *SAAL*),

---

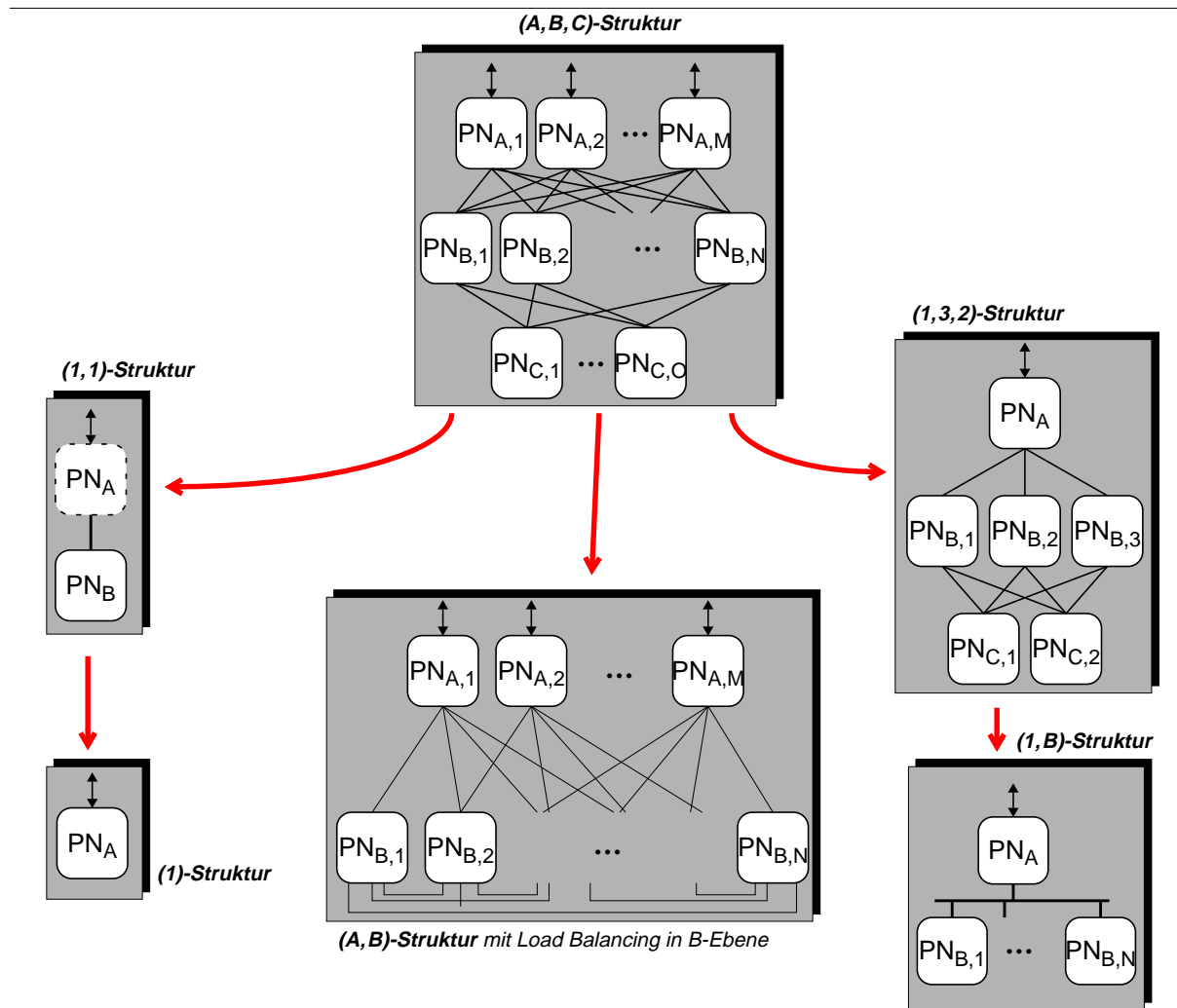
<sup>46</sup> Dieses Schema basiert auf folgenden Grundelementen: KN (Koppelnetzmodul), PE (periphere Einrichtung), ÜE (Übertragungseinrichtung) und VE (Verarbeitungseinrichtung).

<sup>47</sup> Dies entspricht auch üblichen Realisierungsformen, wobei ein Prozessorknoten als eigenständige Einschubkarte (sogenannter Mikrorechner) realisiert wird. Mit zunehmender Integrationsdichte werden aber auch mehrere PN auf einer Karte zusammengefasst.

<sup>48</sup> Die Tatsache, dass eine Funktionsebene keinen PN besitzt, ergibt in der Praxis keinen Sinn.

- *B-Ebene*: höhere Signalisierung (z.B. *Q.2931*, *B-ISUP*, *INAP*), Dienstvermittlungs- und Dienststeuerungsfunktionen<sup>49</sup>,
- *C-Ebene*: Funktionen mit stark zentralem Charakter (z.B. Verbindungsannahmesteuerung oder zentrale Ressourcenverwaltung).

In Bild 2-20 sind bekannte Strukturen (siehe Anhang A.2) zusammengestellt, wobei nun hier auch mögliche Kommunikationsbeziehungen sichtbar werden.



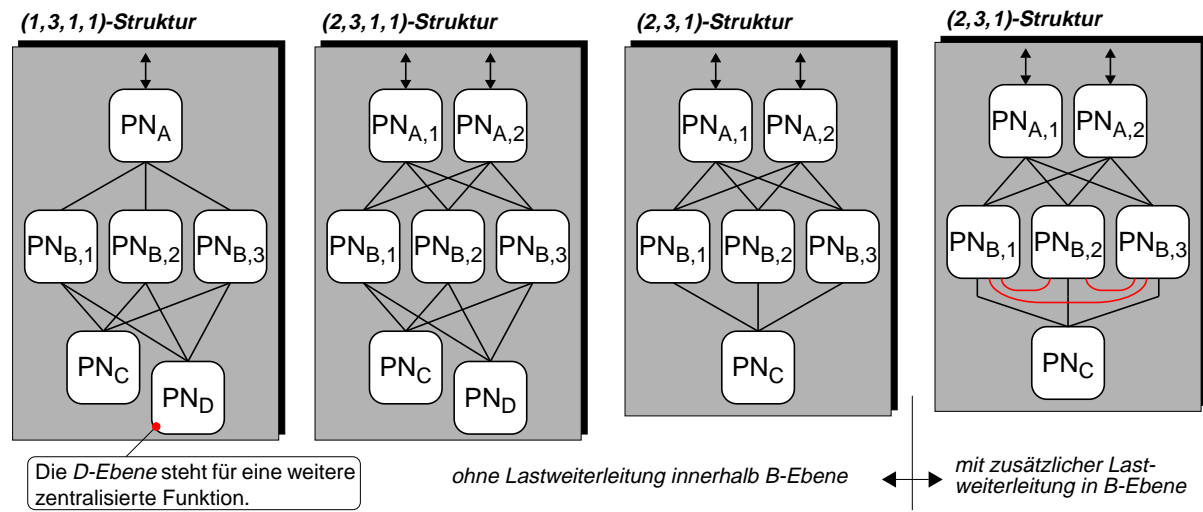
**Bild 2-20:** Varianten der *(A, B, C)*-Steuerungsstruktur, die auch in realen ATM-Systemen eingesetzt werden oder Gegenstand von Leistungsuntersuchungen waren.

In Bild 2-21 ist eine Familie von Steuerungsstrukturen abgebildet, die im Rahmen dieser Arbeit durch Leistungsuntersuchungen näher analysiert wurden. Die Motivation für diese Auswahl erfolgt in Abschnitt 4.1. An dieser Stelle sollen nur die unterschiedlichen Merkmale der vier Varianten kurz erläutert werden:

- Besitzt die *A*-Ebene genau einen Prozessorknoten, d.h. eine  $(1, \dots)$ -Struktur, dann nimmt dieser die Aufgabe der Lastverteilung an die *B*-Prozessoren ( $PN_{B,j}$ ) vor. Durch den  $PN_A$  als zentrale Instanz ist eine relativ einfache globale Lastverteilung realisierbar.

<sup>49</sup> In einem IN sind diese in die Dienststeuerungsknoten (SCP) ausgelagert.

- Bei mehreren  $PN_{A,i}$  sind verschiedene Ansätze der Lastweiterleitung von einem A- an einen B-Prozessorknoten denkbar. Diese Thematik wird in Abschnitt 4.6 erläutert.
- Ist neben einem  $PN_C$  noch ein  $PN_D$  vorhanden, dann ist zwar die mittlere Belastung für diese Einheiten geringer, auf der anderen Seite wächst aber die Interprozessorkommunikation an.
- Innerhalb der B-Ebene kann eine zusätzliche Lastweiterleitung Sinn machen [603].



**Bild 2-21:** Beispiele für Steuerungsarchitekturen mit drei bzw. vier Funktionsebenen aber *unterschiedlichen Lastteilungsmerkmalen* (z.B. Lastzuteilung durch Scheduling zentral/verteilt, Lastweiterleitung<sup>50</sup> ja/nein)

## 2.7 ATM-Knoten für heterogene Netzumgebungen

Die in den vorangegangenen Abschnitten gemachten Erläuterungen bezogen sich ausschließlich auf ein reines B-ISDN-Vermittlungssystem. Darunter werden ATM-Transportplattformen mit ATM-Signalisierung und anderen B-ISDN-spezifischen Steuerungsfunktionen verstanden. Da gegenwärtig, und vermutlich auch zukünftig, heterogene Netzumgebungen existieren, sollte ein ATM-Netzknoten so flexibel konzipiert sein, auch für Nicht-B-ISDN-Umgebungen konfiguriert werden zu können. Typische Anforderungen sind LAN-WAN-Kopplungen oder die Unterstützung einer Vielzahl etablierter Datendienste (z.B. *IP*, *SMDS*, *Frame Relay*).

Solche hybriden Netzknoten werden häufig als *ATM Multimedia Multiservice Switch*<sup>51</sup> bezeichnet [563]. Darunter werden Systeme mit einer ATM-Zellvermittlung (bis Schicht zwei, sogenanntes *Native ATM*) und ab Schicht drei Unterstützung (zusätzlich) weiterer Dienste neben B-ISDN verstanden. Neben anderen Protokollfamilien sind dies Adaptions- oder dienstspezifische Funktionen höherer Schichten. Diese Aufgaben werden entweder in die Vermittlungssoftware eines Netzknotens integriert oder in Server ausgelagert.

<sup>50</sup> Auch *Load Balancing* genannt (Begriffserläuterung siehe Abschnitt 3.1.1, S. 48).

<sup>51</sup> Service bezieht sich hier auf unterstütztes Protokoll. Um die Kompatibilität solcher hybrider ATM-Vermittlungseinrichtungen zu gewährleisten, haben im November 1998 zahlreiche Hersteller und Netzbetreiber das *Multiservice Switching Forum* (MSF) gegründet (siehe auch <http://www.msforum.org/>).

Durch den gegenwärtigen rasanten Anstieg der Datenkommunikation, ausgelöst durch die zunehmende Nutzung des Internet, ist die Vermittlung von IP-Verkehr über ATM-Netze die derzeit wichtigste Herausforderung [254]. Dabei steht das Internet-Protokoll mit seiner verbindungslosen Kommunikation und seinem Best-Effort-Dienst [485] dem verbindungsorientierten und ressourcenreservierenden ATM gegenüber. Die Abbildung von IP-Paketen auf ATM-Zellen ist prinzipiell weniger problematisch. Die Schwierigkeiten kommen mit der Anforderung, dass IP-Anwendungen auch die inhärenten Vorteile von ATM, verschiedene Dienstklassen, garantierte Dienstgüten, große Übertragungsgeschwindigkeit, usw., für eine effiziente Kommunikation nutzen sollten.<sup>52</sup>

Zu dieser Problematik sind eine Vielzahl von Lösungsansätzen von der ATM- und Internet-Welt, insbesondere von ihren Standardisierungsgremien *ATM-Forum* und *IETF*, vorgestellt worden (siehe auch z.B. [127, 158, 175, 221, 398, 413, 415, 417, 463, 468, 505]). Diese Entwicklung hält gegenwärtig noch an und soll hier nicht weiter vertieft werden, da sie nicht Gegenstand dieser Arbeit ist.

Für die Last- und Leistungsregelung bestehen aber folgende Berührungspunkte. In der Nutzer-Ebene sind ggf. Änderungen in der Hardware, für die Umwandlung von ATM-Zellen in IP-Pakete, notwendig [157]. Die Aktivierung und Konfiguration der Verkehrssteuerungsfunktionen (z.B. Policing) oder die Unterstützung von Multicasting obliegt der Vermittlungssoftware. Dagegen sind in der Steuerungs-Ebene die größten Änderungen und Erweiterungen vorzunehmen, was wiederum die Vermittlungssoftware betrifft. Weiterhin gibt es Unterschiede, ob der Vermittlungsknoten im Inneren oder am Rand einer Netz-Domain lokalisiert ist.

Entscheidend für eine Lastregelung in der Vermittlungssteuerung ist, welche Form der Ressourcenreservierung und Schicht-3-Signalisierung von einer Anwendung genutzt wird. Eine neue Dienstanforderung kann sich beispielsweise auf eine Verbindung, einen temporären Paketstrom (*Data Flow* oder *Shortcut* genannt) oder ein einzelnes Schicht-3-Paket beziehen.<sup>53</sup> Dies wiederum wirft unterschiedliche Anforderungen an die Überlastabwehr auf.

---

<sup>52</sup> Eine weitere Motivation für die Einbindung der Schicht-2-ATM-Vermittlung bildet der Aspekt, dass das Wachstum von IP-Verkehr den Anstieg der Leistungsfähigkeit von IP-Routern übersteigt [491].

<sup>53</sup> Allgemein erhebt sich immer die Frage, ob eine Kommunikation verbindungslos oder verbindungsorientiert sein sollte. Betrachtet man nur den Signalisieraspekt, dann sollten zumindest die Vorteile die Kosten überwiegen. Eine Signalisierung macht prinzipiell dann Sinn, wenn die Verbindungsdauer wesentlich größer ist als die Verbindungsaufbauzeit. Messungen weisen nun auf relativ große Antwortzeiten der ATM-Signalisierung bei der gegenwärtigen Systemgeneration hin. Die Identifizierung von „IP-Verbindungen“ erfordert auch einen größeren Aufwand. Weiterhin ist der IP-Verkehr durch folgende ungefähre Aufteilung charakterisiert: 80% der Beziehungen stellen kurze Datenflüsse dar die 20% des Datenvolumens ausmachen [535].

Aus diesem Grund wurden Verfahren zur sogenannten Kurzsignalisierung wie *Tag Switching* oder *Multiprotocol Label Switching (MPLS)* [229] vorgeschlagen. Hierbei werden für einen kurzen Datenfluss Ressourcen reserviert. Diese *Shortcuts* können deshalb eine erhebliche Belastung für eine Vermittlungssteuerung darstellen.

# Kapitel 3

## Problematik der Lastregelung in B-ISDN-Vermittlungsrechnern

Die Problematik der Systemüberlastung, der Überlasterkennung und -abwehr begleitet die Entwicklung aller Generationen von Vermittlungssystemen. Bei der konventionellen Fernsprechvermittlungstechnik war die Dringlichkeit der Entwicklung, Erprobung und Implementierung von Überlastabwehrmaßnahmen noch nicht gegeben, da die herkömmlichen, nicht-rechnergesteuerten Vermittlungssysteme häufig eine für Überlast relativ unempfindliche dezentrale Steuerungsstruktur aufweisen [522]. Mit der Einführung rechnerbasierter Vermittlungssteuerungen wurde die Überlastabwehr eine wichtige Funktion für den Schutz dieses steuerungs-technischen Betriebsmittels. Im Zuge der Diensteintegration wurden weitergehende Anforderungen an diese Last- und Leistungsregelkreise gestellt. Durch das B-ISDN kommen weitere Aspekte hinzu. Diese sollen in diesem Kapitel erläutert und diskutiert werden.

Da das B-ISDN auf bereits etablierten Netztechnologien aufsetzt, müssen Erfahrungen dieser Netz- und Systemgenerationen in das Konzept einer B-ISDN-Lastregelung einfließen. Aus diesem Grund sollen weiterhin relevante Forschungs- und Entwicklungsaspekte aufgearbeitet und dargestellt werden.

### 3.1 Einführung

Jedes System<sup>1</sup> wird aus realisierungstechnischen und ökonomischen Gründen mit begrenzten Ressourcen ausgestattet. In Einrichtungen dieser Art werden diese Ressourcen für die Auftragsbearbeitung (Dienstleistungserbringung) angefordert, belegt und nach einer gewissen Zeit wieder freigegeben. Solche Systeme stehen in Wechselwirkung mit ihrer Umwelt. In der Rechner-technik werden sie als Antwortsysteme bezeichnet. Typische Beispiele sind Stand-Alone-Computer, Terminal-Host- oder allgemein Client-Server-Architekturen. In der Kommunikationstechnik gehören beispielsweise alle Netzelemente zu dieser Systemkategorie: Übertragungssysteme, Vermittlungsstellen, Verbindungsabschnitte, gesamte Netzbereiche usw. Ein Lastregelmechanismus sollte aus einer Vielzahl von Gründen, die später erläutert werden, in diesen Systemen implementiert sein.

Generell gibt es keine Lösung die allen Aspekten, die bei dieser Problematik auftreten, gerecht wird und werden kann! Aus diesem Grund wird die Überlastabwehr auch zukünftig, vor allem in der Entwurfsphase eines Systems, immer ein wichtiges Thema sein. Diese Aussage gilt für ressourcenreservierende Kommunikationsnetze, als auch für solche, in welchen Ressourcen gemeinsam konkurrierend genutzt werden.

---

<sup>1</sup> Neben rein technischen Systemen (Maschinen) existiert die Überlastproblematik auch in vielen anderen Bereichen: Wirtschaftssysteme [87,481], biologische Systeme (z.B. Immunabwehr im menschlichen Körper) usw.

### 3.1.1 Wichtige Begriffe

Als generelle Zielsetzung der Überlastabwehr wird vorgegeben, ein System vor Durchsatzeinbruch, Absinken der Dienstqualität, Instabilität, Verfügbarkeit usw. unter „hohen Lasten“ zu schützen. Diese allgemeine Aussage lässt sich für die Kategorie von Systemen, die in dieser Arbeit betrachtet werden, präzisieren. Dabei wird von folgenden Begriffsdefinitionen ausgegangen:

- *Last* bezeichnet das Verkehrsangebot für eine Prozessoreinheit oder den gesamten Vermittlungsrechner.
- *Überlast* ist der Anteil des Verkehrsangebotes, der die Nennkapazität übersteigt. Dieser Anteil wird üblicherweise in Prozent bezogen auf die Nennkapazität angegeben.
- *Überlastung* (NTG 0903 [617]): „Über die Nennkapazität hinausgehende Inanspruchnahme einer vermittlungstechnischen Einrichtung (...).“
- *Betriebsmittel*: Aus der Menge aller vermittlungstechnischen Einrichtungen erfolgt bei der Überlastabwehr eine Beschränkung auf die Einrichtungen der Steuerungs-Ebene. Beispiele steuerungstechnischer Betriebsmittel sind: Prozessor(zeit) oder Pufferplätze aber auch Register, Busbandbreite, Sätze, Tongeneratoren.<sup>2</sup> Neben physikalischen Betriebsmitteln werden von Überlastabwehren auch logische Betriebsmittel (z.B. Adressen, Rufreferenzen, Nummern) überwacht.
- *Nennkapazität* (nach ITU Q.543 ‚*Engineered Capacity*‘ [648]): Bezeichnet das mittlere Verkehrsangebot, das unter Einhaltung aller Dienst- und Netzgüteanforderungen gerade noch bearbeitet werden kann.<sup>3</sup>
- *Durchsatz* (ITG 5.2-03 [616]): „Quotient aus der Menge der innerhalb eines Zeitabschnittes erfolgreich bearbeiteten Anforderungen und der Länge dieses Zeitabschnittes.“
- *Überlastindikator* (ITG 5.2-01 [614]): „*Verkehrsmess- oder Zustandsgröße*, die auf eine Überlastsituation schließen lässt.“
- *Überlastabwehr* bzw. *-steuerung* (ITG 5.2-01 [614]): „Maßnahmen einer Einrichtung des Netzes als Reaktion auf Überlast, mit dem Ziel, auch bei Überlast einen hohen Nutzdurchsatz zu gewährleisten.“
- *Überlastabwehrstrategie* (NTG 0903 [617]): „Strategie für die Reaktion einer Vermittlungseinrichtung auf eine Überlastung mit dem Ziel, die Folgen der Überlastung abzuschwächen.“

Im englischen Sprachgebrauch werden *Overload* und *Congestion* oft synonym für den Zustand der *Überlastung* verwendet [55]. Die Definitionen für diese Begriffe unterscheiden sich inhaltlich auch nicht. Die oft unterschiedliche Verwendung von *Overload Control* und *Congestion Control* scheint mehr historische Gründe zu haben.<sup>4</sup> Für Signalisiernetze wird in ITU E.744 [630] eine entsprechende Begriffsunterscheidung definiert, die aber letztlich auch keine saubere Trennung darstellt. Auf dem Gebiet der ATM-Vermittlungstechnik hat sich folgende Begriffsverwendung eingebürgert: In der heutigen Literatur wird als *Overload Control* nach wie

<sup>2</sup> Die Objekte der Überlastabwehr („Regelstrecken“) sind eine Untermenge aller steuerungstechnischen Betriebsmittel. Und zwar diejenige Menge, die unter Überlastung kritisch sind und Systemengpässe darstellen.

<sup>3</sup> Lässt man die Netzgüteanforderungen außer acht, dann ist die tatsächlich erreichbare Bearbeitungskapazität höher. Diese Größe für den maximal möglichen Durchsatz wird als *Verkehrskapazität* (*traffic capacity*, [616]) bezeichnet.

<sup>4</sup> Der Begriff *Overload Control* wird seit je her für die Überlaststeuerung für Steuerungseinheiten rechnergesteuerter Vermittlungssysteme verwendet. Auf der anderen Seite wurde (und wird noch) als *Congestion Theory* („Stautheorie“) allgemein das Gebiet der Verkehrstheorie bezeichnet.



vor die Überlastabwehr in der Steuerungs-Ebene (*Control Plane*) bezeichnet, während *Congestion Control* sich mehr auf die Verkehrssteuerungsverfahren für die Nutzer-Ebene (*User Plane*) bezieht [154,190,192,206,226,228,434].<sup>5</sup> Ein weiterer Zusammenhang bildet die Tatsache, dass *Overload* eine mögliche Ursache von *Congestion* ist.<sup>6</sup>

Bei Mehrprozessorsystemen, die nach dem *Lastteilungsprinzip* (*Load Sharing*) arbeiten, sind übergeordnete Lastregelverfahren notwendig. Folgende Begriffsdefinitionen sollen hier verwendet werden:

- *Scheduling* bezeichnet die Lastzuteilung durch eine zentrale Instanz (*Scheduler*) in einem verteilten System.
- *Load Balancing* bezeichnet Lastausgleichsverfahren in verteilten Systemen mit global verfügbarer Information über den Belastungszustand der einzelnen Prozessorknoten.
- *Load Forwarding* bezeichnet allgemein die Lastweiterleitung zwischen verschiedenen Prozessorknoten.

Gerade diese Begriffe werden äußerst uneinheitlich in der Literatur verwendet.

## 3.1.2 Das Überlastphänomen

### 3.1.2.1 Leistungseinbruch bei hohen Lasten

Der Aufbau einer Vermittlungssteuerung basiert allgemein auf einem Rechnersystem. Dessen Verarbeitungskapazität ist prinzipiell begrenzt. Der Grenzwert wird als Nennkapazität bezeichnet. Bei Lastwerten unterhalb dieses Arbeitspunktes ist die Leistung (z.B. Nutzdurchsatz) annähernd proportional zur Last. Übersteigt das Angebot diesen Punkt, dann bricht die Leistung typischerweise ein, da die Abwehr der die Nennkapazität übersteigenden Last auch Betriebsmittelbelegungen verursacht. Dieser nicht zum wirtschaftlichen Gewinn beitragende Verkehrsanteil wird als Blindlast bezeichnet.

Dieser typische Funktionsverlauf der Leistung in Abhängigkeit der Last ist seit Beginn der rechnergesteuerten Vermittlungstechnik bekannt [116,255] und wird auch in der Standardisierung zur Veranschaulichung der Überlastabwehr verwendet (siehe ITU Q.543 [648]). Dieser Effekt wurde auch durch eine Vielzahl von Messungen an realen Systemen oder Simulationsmodellen festgestellt. In Anhang B.3.2 wird ein einfaches Warteschlangenmodell präsentiert, das eine analytische Lösung für dieses Leistungsverhalten erlaubt.

Neben diesem typischen Verhalten realer Systeme sind theoretisch auch andere Leistungsfunktionen denkbar:

- *Ideale Systeme*

Als ideale System sollen solche verstanden werden, die zwar eine begrenzte Verarbei-

---

<sup>5</sup> Beispielsweise wird in I.370 [643] *Congestion Control* folgendermaßen definiert: „Refers to real-time mechanisms to prevent and recover from congestion during periods of coincidental peak traffic demands or network overload (e.g., resource failures). Congestion Control includes both congestion avoidance and congestion recovery mechanisms.“

<sup>6</sup> *Congestion* → Blockierung (von Betriebsmitteln).

In einem amerikanischen Wörterbuch wird folgende Unterscheidung gemacht:

- Congestion: „In a communication system, a state or a condition that occurs when more user attempts to use a service than the system is capable of handling.“
- Overload: „A load, placed on a device, that is greater than the device is capable of handling, so that the device can't perform the functions for which it was defined. E.g., Traffic entered into a communication system that is greater than the traffic capacity of the system.“

tungskapazität besitzen, alle ankommenden Rufe aber erfolgreich bearbeiten und demzufolge keine Blindlast verursachen. Hierbei wächst die Leistung linear bis zu einem Maximalwert an und bleibt dann bei weiter ansteigender Last konstant maximal.

- *Omnipotentes System* [209]

Die Leistung verhält sich proportional zur Last. Solch ein Systemverhalten lässt sich nur durch vielfache Ressourcenüberdimensionierung für einen begrenzten Arbeitsbereich erreichen. Aus wirtschaftlichen Gründen sind solche Systeme nicht praxistauglich.

Da Vermittlungssteuerungen der Klasse von Echtzeit-Systemen angehören, ist eine Überlastabwehr absolut notwendig. Ein Ziel jeder Überlastabwehr ist, das Leistungsverhalten eines realen Systems dem einem Idealen anzunähern.

### 3.1.2.2 Wechselwirkung zwischen Teilnehmer und Vermittlungssystem

Der Teilnehmer als Dienstanwender und das Kommunikationsnetz mit seinen Netzelementen als Dienstbringer stehen in enger Wechselwirkung. Die subjektiv empfundene Dienstgüte kann entscheidend das Verhalten eines Teilnehmers beeinflussen. Eine Änderung des Teilnehmerverhaltens spiegelt sich in Laständerungen wider, was wiederum einen Einfluss auf die Leistung und damit die Dienstgüte hat.

In Fernsprechnetzen reagiert typischerweise ein Teilnehmer auf einen nicht erfolgreichen Ruf mit einem Wiederholversuch. Dieses Phänomen der Rufwiederholungen (*Repeated Call Attempts* oder *Follow On Calls*) wurde bereits 1947 von KOSTEN [324] und 1957 von COHEN [133] erwähnt. Jede Rufwiederholung bedeutet einen Lastanstieg und kann gerade unter Überlast zu einem Leistungseinbruch führen. Wenn diese Rückwirkung einer *Mitkopplung* entspricht, dann können die Verkehrswerte lawinenartig ansteigen [73], was natürlich unter allen Umständen vermieden werden sollte.

Bei der Auslegung einer Überlastabwehr ist deshalb dieser Effekt zu berücksichtigen. Aus diesem Grund wurden bereits in den 1960er und 70er Jahren, aber auch noch heute, hierzu sehr viel an Forschungsarbeit geleistet (z.B. [112,278,355,394,426,462,556]). Dabei entstanden eine Vielzahl von Modellen (z.B. [27]), die zum Teil sehr detailliert das Verhalten eines menschlichen Teilnehmers oder auch Endgerätes (z.B. automatische Wahlwiederholungen) nachbilden [352].

In Rechnernetzen wurden ähnliche Rückwirkungen beobachtet. Bei Terminal-Host-Systemen ist das Verkehrsangebot eine Funktion des Dienstgüteparameters Antwortzeit. Bei kurzer Antwortzeit sind viele Benutzer bereit, Aufträge abzusetzen, bei schlechtem Antwortverhalten nur noch wenige [243]. Dieser *Marktmechanismus* wird auch als *Teilnehmerentmutigung* bezeichnet. Es liegt hier eine *Gegenkopplung* vor, so dass überhaupt keine Überlast entstehen kann (siehe auch Anhang B.2.2).

### 3.1.2.3 Weitere Effekte

Das Wählen oder Aufhängen bevor der Wählton aufgeschaltet wird, sind Beispiele für Protokollverletzungen des Teilnehmers.<sup>7</sup> Solche Verhaltensweisen, auch Ungeduld des Teilnehmers genannt, führt zu abgebrochenen Rufen und erhöht damit die Blindlast. Diese Effekte lassen sich in der Teilnehmermodellierung berücksichtigen. Wiederum entsteht eine Rückkopplung und damit ein Regelkreis zwischen Teilnehmer und Vermittlungssystem.

---

<sup>7</sup> Der Wähltonverzög ist ein Maß für die Dienstgüte und kann unter hohen Lasten sehr groß werden.

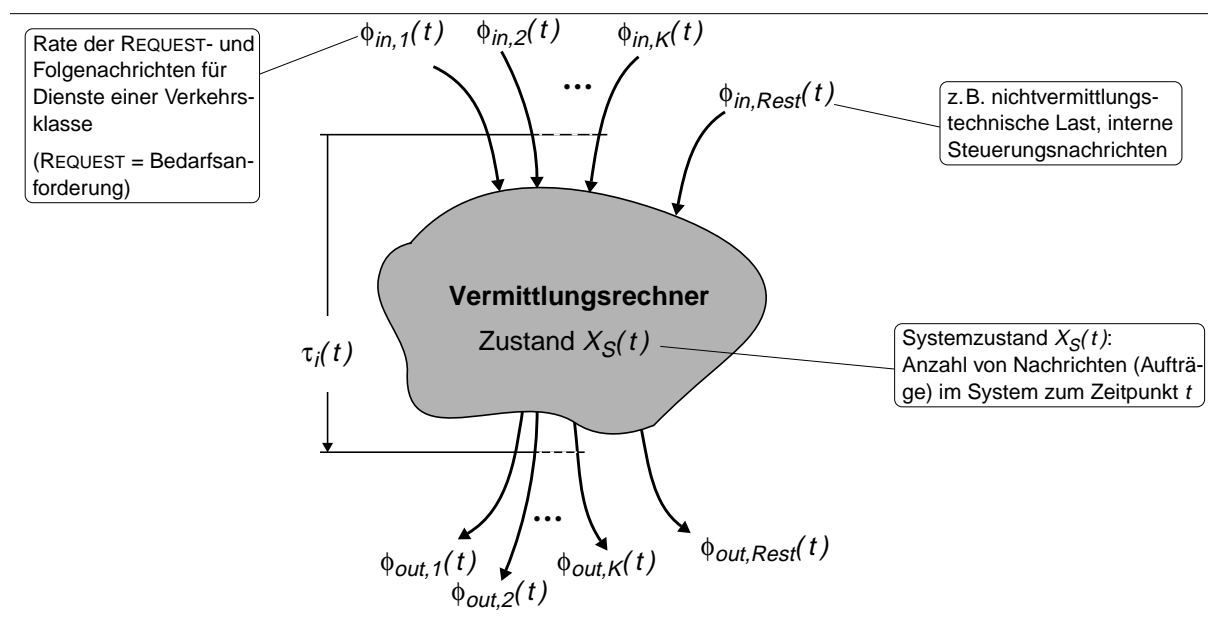
Unter dem sogenannten *Speichereffekt* wird folgendes Verhalten verstanden. Jeder Vermittlungsrechner besitzt einen Puffer für neu ankommende Rufe. Unter Überlast können sich bei großem Wartespeicher sehr viele Anforderungen ansammeln. Geht nun das Angebot wieder zurück, dann wird trotz des externen Verkehrsabfalls die interne Last zunächst nicht kleiner. Dieser dynamische Effekt führt zu Trajektorien im Leistungs-Last-Diagramm, da bei kleiner werdender Last ein (kurzzeitiger) Leistungsanstieg zu beobachten ist [208,523]. Eine Überlastabwehr, die beispielsweise als Lastindikator die Rufankunftsrate verwendet, muss diesen internen Effekt berücksichtigen.

### 3.1.3 Leistungsmaße

Für die qualitative und quantitative Bewertung der Systemleistung ist man an möglichst aussagekräftigen Kenngrößen interessiert. Dies betrifft die Beurteilung von Optimierungsmaßnahmen oder den Vergleich unterschiedlicher Systemarchitekturen bzw. Algorithmen. Speziell für die Last- und Leistungsregelung ist die Auswahl eines geeigneten Leistungsmaßes von zentraler Bedeutung. In der Literatur (z.B. [225]) wurde hierzu bereits eine Vielzahl von Kenngrößen vorgeschlagen. Üblicherweise sind diese als Zeitmittelwerte für eingeschwungene Systeme definiert. Für dynamische Systembetrachtungen – bei der Leistungsregelung sind Leistungskenngrößen direkt in den Regelkreis involviert – sind ggf. Anpassungen oder Erweiterungen zu berücksichtigen. Im folgenden werden für die Steuerungs-Ebene in Frage kommende Kenngrößen zusammengestellt.

#### 3.1.3.1 Elementare Kenngrößen

Für ein bestimmtes System lassen sich Kenngrößen für die unterschiedlichsten Betrachtungsebenen definieren. In dieser Arbeit wird die Steuerungsleistung am Bearbeitungsverhalten von Schicht-3-Nachrichten gemessen. Entsprechend der zugrundeliegenden Dienstmodellierung (siehe Abschnitt 5.2.6) werden hier  $K$  Verkehrsklassen unterschieden. Bild 3-1 zeigt die verschiedenen Verkehrsflüsse, die der Last des Vermittlungsrechners entsprechen.



**Bild 3-1:** Relevante Kenngrößen für die Definition von Leistungsmaßen

Alle *Leistungsmaße*  $\Lambda$  beziehen sich auf äquidistante *Zeitintervalle* der Länge  $T$  (z.B. Mess- oder Regelintervall). Die *Durchsatzkenngröße*  $\Lambda_\phi$  gibt allgemein die *Anzahl erfolgreich bearbeiteter Aufträge pro Zeiteinheit* an. Im Kontext dieser Arbeit wird diese für zwei verschiedene Nachrichtentypen<sup>8</sup> verwendet:

- **Durchsatz von Dienstanforderungen** (*Request Messages*) für Verkehrsklasse  $v$

$$\Lambda_{\phi, RM, v}(k) = \int_{(k-1)T}^{kT} \phi_{out, RM, v}(t) dt \quad (3-1)$$

bzw. für Gesamtverkehr

$$\Lambda_{\phi, RM}(k) = \int_{(k-1)T}^{kT} \phi_{out, RM}(t) dt = \sum_{v=1}^K \Lambda_{\phi, RM, v}(k). \quad (3-2)$$

- **Durchsatz von komplettierten Dienstanforderungen** (*Completed Service Requests*) für Verkehrsklasse  $v$

$$\Lambda_{\phi, CS, v}(k) = \int_{(k-1)T}^{kT} \phi_{out, CS, v}(t) dt \quad (3-3)$$

bzw. gesamter Nutzdurchsatz:

$$\Lambda_{\phi, CS}(k) = \int_{(k-1)T}^{kT} \phi_{out, CS}(t) dt = \sum_{v=1}^K \Lambda_{\phi, CS, v}(k). \quad (3-4)$$

Da der Betriebsmittelbedarf für die einzelnen Verkehrsklassen verschieden ist, unterscheiden sich selbst bei gleichen Ankunftsdaten die jeweiligen Durchsätze. Der Durchsatzkennwert ist ein gutes Maß für die Systemleistung insgesamt, da die einzelnen Auftraggeber nicht eingehen. Ein einzelner Teilnehmer hingegen beurteilt die Leistung des Vermittlungssystems meist anhand der Zeitspanne, bis seine individuelle Anforderung „beantwortet“ ist. Das Antwortzeitverhalten ist deshalb ein wesentliches Dienstgütekriterium. Folgende *Antwortzeitkenngrößen*<sup>9</sup> werden hier verwendet:

- **Antwortzeit der Dienstanforderungen** für Verkehrsklasse  $v$ :

$$\Lambda_{\tau, RM, v}(k) = \frac{1}{N_v(k)} \cdot \sum_{i=1}^{N_v(k)} \tau_{RM, v, i}(t) \text{ für } (k-1)T \leq t < kT. \quad (3-5)$$

Es wird der Mittelwert über die Antwortzeit aller im betrachteten Intervall erfolgreich bearbeiteten REQUEST-Nachrichten (Anzahl  $N_v(k)$ ) berechnet. Verlässt keine Meldung das System ( $N_v(k) = 0$ ), dann ist die Maßzahl für das entsprechende Intervall nicht definiert. Falls diese Kenngröße intern als Eingangsvariable in einem Regelkreis verwendet wird, dann muss in solchen Fällen ein Wert geschätzt werden.

Für den Gesamtverkehr ergibt sich entsprechend

$$\Lambda_{\tau, RM}(k) = \frac{1}{K} \cdot \sum_{v=1}^K \Lambda_{\tau, RM, v}(k). \quad (3-6)$$

<sup>8</sup> Erste bzw. letzte Nachricht (REQUEST bzw. RELEASE) eines Dienstszenarios (siehe auch Bild 4-3, S. 95).

<sup>9</sup> Oft auch als Verweil-, Verzögerungs-, Durchlauf- oder Systemzeit bezeichnet. In [354] wird von der mittleren Antwortgeschwindigkeit (*Average Speed of Answer*) gesprochen.

Die Antwortzeit wird mit Hilfe von Zeitstempeln ermittelt. Kann dieses Messverfahren nicht angewendet werden, dann können Antwortzeiten prinzipiell mit Hilfe des LITTLE'schen Gesetzes [211,359,433] geschätzt werden. In [363] werden entsprechende Definitionen für instationäre Systeme vorgeschlagen. Aufgrund der hier vorliegenden Systemcharakteristik (z.B. klassenindividuelle Bedienzeiten, Prioritäten, Regelstrategien, Meldungsverluste) und einem zeitabhängigen Verkehrsprofil (d.h. der Erhaltungssatz ist ungültig:  $\phi_{in}(t) \neq \phi_{out}(t)$ ) ist die Anwendung von *Little* kritisch (siehe auch [474]). Für die Leistungsregelung ist aber u.U. bereits ein grober Näherungswert für eine Tendenzangabe ausreichend, so dass folgende Approximation ihre Berechtigung besitzt:

- **Geschätzte Antwortzeit für eine Nachricht:**

$$\Lambda_{\tau}(k) = \frac{\bar{X}_S(k)}{\bar{\phi}_{out}(k)} \text{ oder } \tilde{\Lambda}_{\tau}(k) = \frac{\bar{X}_S(k)}{\bar{\phi}_{in}(k)} \text{ für } N(k) > 0, \quad (3-7)$$

wobei  $X_s(t)$  der Systemzustand bedeutet (vergl. Bild 3-1). Welche Variante verwendet werden soll, hängt von den verfügbaren Messgrößen ab. Wie gesagt, dies kann nur eine Faustformel sein und kann, gerade während hochgradig instationärer Phasen, auch ungenaue Ergebnisse liefern.

### 3.1.3.2 Zusammengesetzte Kenngrößen

Die Maße *Durchsatz* und *Antwortzeit* sind weit verbreitete Leistungskenngrößen, reflektieren aber eine etwas zu einseitige Sicht, wenn es um eine mehr ganzheitliche Optimierungsfunktion für eine Leistungsregelung geht. Für die Bewertung von Optimierungsmaßnahmen wurden deshalb kombinierte Leistungsmaße definiert. Aus dem Bereich der Unternehmensforschung (*Operations Research*) sind zahlreiche Gewinn-Kosten-Funktionen bekannt [500]. Für Warteschlangensysteme stammt beispielsweise von TIPPER [518] folgende Definition:

- **Gewinn-Kosten-Maß** (*Net Benefit* bzw. *Cost of Operation*):

$$\Lambda_{BC}(k) = \int_{(k-1)T}^{kT} [\alpha \cdot \bar{X}_S(k, t) - \beta \cdot \bar{\phi}_{out}(k, t)] dt. \quad (3-8)$$

Der erste Term soll die Kosten, der zweite Term den Gewinn repräsentieren. Deren relative Bedeutung wird mit den Gewichtungskonstanten  $\alpha$  und  $\beta$  eingestellt.

Eine Optimierung würde hier eine Minimierung von  $\Lambda_{BC}$  bedeuten. Man versucht hier somit gleichzeitig das *Antwortzeit- und Durchsatzverhalten* zu optimieren. Aus Sicht des Dienstnutzers (Teilnehmer) steht die Antwortzeit, aus Sicht des Dienstbringers (Netz) der Nutzdurchsatz im Vordergrund. Aus globaler Sicht sollte der gegensätzliche Charakter dieser Kenngrößen deshalb in einem gemeinsamen Leistungsmaß vereinigt werden. Die bekannteste Kompromisslösung hierzu ist die Größe *Power*, die von GIESSLER ET AL. [208] in den 1970er Jahren für Rechnernetze definiert wurde, und vor allem durch KLEINROCK eine weite Verbreitung fand [310,312].<sup>10</sup> *Power* wird durch den Quotienten von Durchsatz und Antwortzeit definiert:

- **Power:**

$$\Lambda_P(k) = \frac{\bar{\phi}_{out}(k)}{\bar{\tau}(k)} \text{ für } N(k) > 0. \quad (3-9)$$

<sup>10</sup> In sehr vielen Veröffentlichungen wird fälschlicherweise KLEINROCK als Urheber von *Power* angegeben. KLEINROCK selbst weist aber in [310] deutlich auf die Definition in [208] hin.

Betrachtet man nicht das Gesamtsystem, dann ist vornehmlich die klassenindividuelle *Power*<sup>11</sup> der REQUEST-Nachrichten von Interesse:

$$\Lambda_{P, RM, v}(k) = \frac{\Lambda_{\phi, RM, v}(k)}{\Lambda_{\tau, RM, v}(k)}. \quad (3-10)$$

Für viele Systeme zeigt *Power* in ihrem funktionalen Verlauf ein globales Maximum (siehe auch Anhang B), so dass diese Kenngröße tatsächlich ideal zur quantitativen Bewertung einer Optimierung von Antwort- und Durchsatzverhalten verwendet werden kann.<sup>12</sup> Für spezielle Systeme wurden Erweiterungen von *Power* definiert (z. B. [93,311]). Für verteilte Systeme hat sich obige *Power*-Definition als unbrauchbar erwiesen [478] und es wurde das *Power-Produkt* (*Product of Power*) als erweitertes Leistungsmaß eingeführt [267,358,377]. Diese Kenngröße berücksichtigt die Überlagerung von Verkehrsströmen als auch Fairnessaspekte.

Im Rahmen der Leistungsregelung für einen B-ISDN-Vermittlungsrechner lässt sich eine entsprechende Kenngröße wie folgt definieren.

- **Power-Produkt:**

$$\Lambda_{P_{RM}^{\Pi}}(k) = \prod_{v|N_v(k) > 0} \frac{\Lambda_{\phi, RM, v}(k)}{\Lambda_{\tau, RM, v}(k)} = \prod_{v|N_v(k) > 0} \Lambda_{P, RM, v}(k). \quad (3-11)$$

Problematisch bei dieser Definition sind Zeitintervalle, in welchen keine REQUEST-Meldungen einer Verkehrsklasse erfolgreich bearbeitet werden (d. h.  $N_v(k) = 0$ ), und dass die Relevanz unterschiedlicher Verkehrsklassen nicht einfach durch Gewichtungsfaktoren berücksichtigt werden kann. Durch den Produktansatz werden aber solche Lösungen vermieden, die versuchen, das *Power*-Maß für eine bestimmte Verkehrsklasse – auf Kosten der anderen – zu maximieren. Das *Power-Produkt* repräsentiert deshalb einen akzeptablen Kompromiss, zumal diese Funktion auch ein globales Maximum besitzen soll [358]<sup>13</sup>, und stellt ein globales Optimierungsziel für eine Leistungsregelung dar.

Neben dem Produkt ist auch die Summe aller klassenindividuellen *Power*-Werte von Interesse.

- **Power-Summe:**

$$\Lambda_{P_{RM}^{\Sigma}}(k) = \sum_v g_v \cdot \frac{\Lambda_{\phi, RM, v}(k)}{\Lambda_{\tau, RM, v}(k)} = \sum_v g_v \cdot \Lambda_{P, RM, v}(k). \quad (3-12)$$

Dieses Leistungsmaß besitzt die Vorteile, dass zum einen Gewichtungen möglich sind, und zum anderen, da Summanden auch null sein dürfen, seine Berechnung einfacher wird. Für die Lösung eines Optimierungsproblems besitzt diese Funktion aber den gravierenden Nachteil, dass mehrere lokale Maxima existieren können (siehe auch Anhang B.4).

<sup>11</sup> Die Bezeichnung *Power* (dt. Kraft, Leistung) hat sich etabliert, ist aber m. E. vom Begriff her irreführend. In [208] wird ausdrücklich auf den Zusammenhang mit der physikalischen Leistung – Durchsatz soll der Energie, Delay der Zeit entsprechen – hingewiesen und als Einheit [*Packets/s<sup>2</sup>*] oder [*Packets/s*] angegeben! Der Durchsatz ist aber bereits ein Maß für die Leistung, so dass *Power* eher einer ‚zeitgewichteten Leistung‘ entspricht. Eine Nachfrage 1998 bei GIESSLER ergab, dass die damalige Verwendung des Begriffs *Power* keine tiefere Bedeutung hatte.

<sup>12</sup> Beim *M/M/1*-System hat *Power* sein Maximum bei  $A = 0.5$  Erl. In diesem Arbeitspunkt beträgt der mittlere Durchsatz die Hälfte des maximal Möglichen und das mittlere Delay entspricht der doppelten Minimumantwortzeit.

<sup>13</sup> Bis jetzt ist kein strenger Beweis für die Existenz eines globalen Maximums einer *Power-Produkt-Funktion* bekannt. In der Literatur wird vielmehr argumentiert, dass dieses Extremum plausibel ist oder durch zahlreiche Messungen belegt wird [358].

## 3.2 Ressourcenbedarf

### 3.2.1 Mittel zur Charakterisierung von Rufbeziehungen

Maßgeblich für die Verkehrscharakterisierung ist der notwendige Ressourcenbedarf einer Kommunikationsbeziehung. Ressourcen werden sowohl in der Nutzer- als auch in der Steuerungs-Ebene in Anspruch genommen (siehe Tabelle 3-1). In [332] erfolgt eine zusätzliche Unterteilung der *steuerungstechnischen Betriebsmittel* in

- *aktive Betriebsmittel (active resources)*: Belegung ist unmittelbar durch die auszuführende Funktion bestimmt (z.B. Prozessor, Koppelement)
- *passive Betriebsmittel (passive resources)*: Belegung ist im allgemeinen an aktive Betriebsmittel gekoppelt und hängt vom momentanen Belegungszustand ab (z.B. Verbindungsspeicher)

	<b>Nutzer-Ebene</b> (vermittlungstechnische Betriebsmittel)	<b>Steuerungsebene</b> (steuerungstechnische Betriebsmittel)
physikalisch	<ul style="list-style-type: none"> <li>• <i>Port</i> (externe Netzknotenschnittstelle)</li> <li>• <i>Link</i> (Bandbreite)</li> <li>• <i>Zellpuffer</i> (im Koppelnetz und Leitungsmodulen)</li> <li>• Zellkopiermodule</li> <li>• Echokompensationseinheiten</li> </ul>	<ul style="list-style-type: none"> <li>• Prozessor</li> <li>• Puffer für Signalisier- und interne Steuerungsnachrichten</li> <li>• Zeitgeber</li> <li>• Speicher für rufbezogene Daten</li> <li>• Stapelspeicher (Stack)</li> </ul>
logisch	<ul style="list-style-type: none"> <li>• Betriebssoftware für Zellverzögerungs- und -verluststeuerung (nach E.735 [627])</li> <li>• ATM-Adressen</li> <li>• VPI/VCI-Kennungen</li> </ul>	<ul style="list-style-type: none"> <li>• Rufreferenzen (<i>Call Reference</i>)</li> <li>• Verbindungsreferenzen (<i>Connection Reference</i>)</li> <li>• aktive Prozesse</li> </ul>

**Tabelle 3-1:** Typische Betriebsmittel in einem ATM-Netzknoten

Jede erfolgreiche oder nicht erfolgreiche Rufbeziehung belegt während ihrer „Lebensdauer“ Betriebsmittel in einem Netzknoten. Der Bedarf kann dabei zeitlichen Schwankungen unterliegen (z.B. Dienstklassen mit variabler Bitrate). Gerade in der Steuerungs-Ebene weist die temporäre Intensität einer Signalisierbeziehung üblicherweise einen hochgradig büschelförmigen Charakter auf.<sup>14</sup>

Die verfügbaren Betriebsmittel bilden die natürliche Leistungsgrenze eines Vermittlungssystems. Die in Anspruch genommenen Betriebsmittel stellen die Kosten für einen Dienst aus Sicht eines Netzknotens dar. Die Entgeltberechnung für die Teilnehmer basiert in bekannten Netztechnologien hauptsächlich auf der Betriebsmittelnutzung in der Nutzer-Ebene. Dieser Ansatz scheint diskussionswürdig im B-ISDN, da die enge Korrelation zwischen den Kosten in Nutzer- und Steuerungsebene, aufgrund der unterschiedlichsten Dienste, nicht mehr gegeben ist. In der Steuerungsebene werden die Kosten vornehmlich durch die Signalisierbearbeitung verursacht. Die Signalisieraktivität und -intensität kann aber nun gravierend unterschiedlich für verschiedene Dienste sein. So gibt es immer verschiedene Realisierungsvarianten der Dien-

<sup>14</sup> Dies lässt sich beispielsweise am Dienst einer Videokonferenz verdeutlichen. In Anhang C.3.2.2 (S. 260) ist hierzu ein exemplarisches Signalisierzenario dieser *Multiparty*-Konfiguration mit *Multiconnection Calls* näher ausgeführt. Zusätzliche Verbindungen werden hier zu- und wieder weggeschaltet oder Verbindungsparameter werden temporär modifiziert, z.B. um die Bildauflösung kurzzeitig zu erhöhen.

sterbringung durch ein Netz, z.B. IN-Konzept oder Nutzung eines Internet-Dienstes<sup>15</sup>, sehr große Unterschiede hinsichtlich der zeitlichen Verkehrscharakteristik und des Ressourcenbedarfs, z.B. Fernsehverteildienst gegenüber E-Mail, oder die unterschiedliche Repräsentation eines Dienstes, z.B. Sprachtelefonie im Fest- und Mobilnetz.

### 3.2.2 Zustandsraum

Die Belastung eines Vermittlungssystems lässt sich allgemein durch folgendes Zustandsmodell beschreiben. Alle betrachteten Betriebsmitteltypen werden in zwei Vektoren, die gemeinsam den Zustandsraum  $\mathfrak{R}$  aufspannen, zusammengefasst:

$$\mathfrak{R}(t) = \{\vec{R}_C(t), \vec{R}_U(t)\} \quad (3-13)$$

mit

- $\vec{R}_U(t) = \{R_{U,1}(t), R_{U,2}(t), \dots, R_{U,m}(t)\}$  für Betriebsmittel der Nutzer-Ebene und
- $\vec{R}_C(t) = \{R_{C,1}(t), R_{C,2}(t), \dots, R_{C,n}(t)\}$  für die Steuerungs-Ebene.

Die einzelnen Vektorelemente  $R_{U,i}$  bzw.  $R_{C,j}$  repräsentieren dabei eine Maßzahl von den in Tabelle 3-1 aufgelisteten Betriebsmitteltypen, beispielsweise die *effektive Bandbreite* [328] einer Nutzverbindung. Durch die begrenzte Verfügbarkeit der Betriebsmittel weist der Zustandsraum eine endliche Ausbreitung in jeder Dimension auf. Die Grenzfläche stellt die maximale Belastung dar. Die Lastregelung muss für einen stabilen Arbeitspunkt innerhalb dieses Bereiches sorgen. Aufgrund des Entwurfes eines Systems, Schnittstellenspezifikationen und sonstigen Randbedingungen konzentriert sich die Zustandsbeschreibung in der Praxis auf wenige Parameter kritischer Betriebsmittel in einem Netzknoten. In der Nutzer-Ebene sind dies üblicherweise *Bandbreite* oder *Zellpufferkapazität*, in der Steuerungs-Ebene die verfügbare *Rechenleistung*. Somit kann die Beschreibung des Systemzustandsmodells wesentlich einfacher erfolgen. Im folgenden (siehe Abschnitt 3.6) wird auf die hier eingeführten Ressourcenvektoren zurückgegriffen.

## 3.3 Überlastursachen und -situationen

In [353] wird berichtet, dass 1997 in einigen dicht besiedelten Gebieten in den USA der Internetverkehr so intensiv war, dass ein normaler Fernsprechteilnehmer nicht einmal mehr einen Wählton bekam, als er den Hörer abnahm. Die Ursache dieses Ressourcenengpasses ist die Nutzung des Netzes durch einen Dienst, dessen spezielle Charakteristik nicht die Grundlage der Netzplanung bildete.

Eine kurze Übersicht weiterer Beispiele von Überlastursachen und -situationen soll in diesem Unterkapitel gegeben werden.

---

<sup>15</sup> Beim WWW wird durch die Anwahl (Mausklick) eines *Uniform Resource Locators* eine sequentielle Folge von TCP-Verbindungen, für den Transport von Text-, Bild- oder Audioinformation, auf- und wieder abgebaut.

*Realisierung I:* Die Verbindungssteuerung erfolgt im Internet. Im ATM-Netz existiert eine transparente Verbindung vom Endgerät zu einem Internet-Provider für die Dauer einer WWW-Sitzung. Während dieser Sitzung gibt es keine Signalisieraktivität im B-ISDN.

*Realisierung II:* Die Dienst- und damit die Verbindungssteuerung erfolgt im B-ISDN (d.h. Endgerät und Netzknoten). Damit würde jeder Mausclick eine Folge von Verbindungsauf- und -abbauten auslösen.



### 3.3.1 Einführung

Die Entwicklung und Auslegung eines Systems steht in engem Zusammenhang mit dem Bedarf an Betriebsmitteln innerhalb des spezifizierten Arbeitsbereiches. Wird dieser Lastbereich verlassen, ist mit Leistungseinbußen oder Fehlverhalten zu rechnen. Potenzielle Überlastursachen sollen hier nach verschiedenen Gesichtspunkten strukturiert und exemplarisch erläutert werden, da diese Aspekte auch für Vermittlungssteuerungen von B-ISDN-Knoten gültig sind.

### 3.3.2 Überlast aus Sicht des Architekturmodells

Betrachtet man das Architekturmodell, dann sind sowohl *vertikal* und als auch *horizontal* Überlastursachen denkbar. Beispielsweise findet zwischen übereinanderliegenden Protokollschichten (vertikal) üblicherweise eine Datenflusssteuerung und damit eine Lastregelung statt. Durch Rückstaumechanismen können sich lokale Überlastungen auf höherliegende Schichten ausbreiten.

Bei der *Peer-to-Peer*-Kommunikation (horizontal), d.h. zwischen räumlich oder logisch getrennten Protokollinstanzen, ist in vielen Protokollfamilien eine Art Laststeuerung vorgesehen. Im Signalisiersystem Nr. 7 sind beispielsweise auf den Ebenen zwei (MTP2), drei (MTP3), vier (SCCP) und der Anwenderebene (z.B. ISUP oder transaktionsorientierte Anwenderteile wie INAP) Verfahren wie Flusskontrolle oder Überlastabwehrstrategien spezifiziert [379,575]. Als Beispiele aus der B-ISDN-Teilnehmersignalisierung wäre die Schicht-2-Flusskontrolle (SSCOP) oder die Antwortzeitüberwachung von Schicht-3-Nachrichten (Q.2931) zu nennen.

Durch diese Protokollmechanismen werden Lastregelkreise definiert. Findet dabei eine Kommunikation über Netzgrenzen oder zwischen verschiedenen Protokollfamilien statt (sogenanntes *Interworking*), dann müssen die jeweiligen Regelkreisstrukturen besonders sorgfältig konzipiert werden.

### 3.3.3 Überlast in verschiedenen Netztypen

Jede Netztechnologie hat ihre spezifischen Überlastprobleme. Gerade in einem integrierenden Netz sind deshalb die verschiedenen netz- und diensttypischen Einflüsse zu beachten.

#### 3.3.3.1 Paket- bzw. Rechnernetze, Internet

Bereits im ersten paketvermittelnden Netz, dem ARPANET [350], wurden in der Anfangsphase mit vier Netzknoten (Router) 1970 Überlastuntersuchungen angestellt. Da offensichtlich noch kein Bedarf für eine Modellierung bestand, wurden diese Untersuchungen in Form von Feldtests, also im Wirkbetrieb, im Netz durchgeführt. Die notwendige künstliche Überlast wurde durch Softwaremanipulationen erzeugt. Ende 1973 trat der erste (ungewollte) größere Zusammenbruch eines Knotens auf [26].<sup>16</sup> Zu Beginn waren die Knoten noch ohne eine Überlastab-

---

<sup>16</sup> Durch einen Hardwarefehler wurden Routing-Tabellen falsch aktualisiert. Hierauf funktionierte die Verkehrslenkung im Netz nicht mehr korrekt und eine lawinenartige Sturzflut von Paketen ergab sich in Richtung eines bestimmten Netzknotens.

wehr ausgestattet. Störungen eines einzelnen Knotens konnten sogenannte Dominoeffekte auslösen, die sich letztlich über das gesamte Netz ausbreiteten.

Eine große Übersicht zur Überlastproblematik in paketvermittelnden Netzen ist in [60] zu finden. Durch das gegenwärtige rapide Wachstum an Internet-Verkehr gewinnt diese Problematik (z.B. Anfragen an zentrale Server, Gigarouter) wieder an Bedeutung [376].

### 3.3.3.2 Fernsprech- bzw. Telekommunikationsnetze

Aufgrund der sehr hohen Anforderungen an Verfügbarkeit, Zuverlässigkeit und Dienstgüte öffentlicher Telekommunikationsnetze hat hier die Überlastabwehr schon immer einen zentralen Charakter. Die meisten Impulse zu dieser Thematik stammen auch aus dem Gebiet der Analyse rechnergesteuerter Fernsprechvermittlungssysteme.

Typische Problematiken sind Teilnehmerrückwirkungen (siehe Abschnitt 3.1.2.2) oder der Einbruch der Rufkomplettierungsrate unter starker Überlast. Frühzeitig wurde erkannt, dass der Ressourcenverbrauch blindlastverursachender Rufe nicht vernachlässigbar ist. In [106] wird berichtet, dass ein sogenannter „False Start Call“<sup>17</sup> 30% und „Partial Calls“ bereits 35 % bis 85 % der vollen Bedienzeit beanspruchen.

### 3.3.3.3 Signalisiernetze

Signalisiernetze stellen das Rückgrat eines jeden (verbindungsorientierten) Nutzdatennetzes dar und müssen deshalb besonders robust gegen Überlast sein. Besonders problematisch sind hier lokale Überlastungen, die sich sehr schnell auf große Netzbereiche ausbreiten können [132,256,493].

### 3.3.3.4 Intelligentes Netz

Im Intelligenten Netz erfolgt die Dienststeuerung in dedizierten Rechnern (SCP). Aufgrund dieses Architekturkonzeptes bilden diese zentralen Netzelemente prinzipielle Systemengpässe unter hohen Verkehrslasten. Die Dienstbearbeitung erfolgt nach dem Client-Server-Prinzip zwischen SSP und SCP [119]. Ein SSP sendet eine Anfrage (*Request*, TCAP) an den SCP, der nach Bearbeitung eine Antwort (*Response*) zurückschickt. Hinter jedem IN-Dienst verbirgt sich ein Szenario dieser Meldungspaare (Dialog). Bei Überlastabwehruntersuchungen für SCPs wurde nun erkannt, dass sich hinter einem „Call“ die unterschiedlichsten Ressourcenanforderungen verbergen können. Diese Tatsache wurde bei einfachen und auch bei den ersten SCP-Überlastabwehren bewusst oder unwissentlich außer acht gelassen. Die Problematik der Laststeuerung, die aus der Unterstützung eines breiten Spektrums unterschiedlichster Dienste herrührt, hat deshalb ähnliche Aspekte im IN und ISDN (insbesondere B-ISDN).

Typische Überlastsituationen sind in [318,338,404] beschrieben.

## 3.3.4 Überlast im Normalbetrieb

Hierunter soll die Problematik verstanden werden, wenn Planungsparameter für die Ressourcenbedarfsabschätzung von Netzelementen nicht mehr gültig sind, wie in dem eingangs er-

<sup>17</sup> Rufender Teilnehmer erhält Wählton und beendet den Anrufversuch ohne zu wählen.

wähnten Beispiel durch Internetverkehr. Dieser Punkt betrifft hauptsächlich die zunehmende Erweiterung des Dienstespektrums und die damit verbundene Einführung von neuen Diensten in bestehende Netze. Dies betrifft besonders Dienste aus den Bereichen Mobilkommunikation, ISDN, IN und Internet. Diese spezielle Thematik wird in [575] diskutiert.

### 3.3.5 Überlast durch spezielles Teilnehmerverhalten

Der Bedarf an geschäftlicher und privater Kommunikation eines Teilnehmers ist zeitlichen Schwankungen unterworfen. Ein weiterer entscheidender Einfluss auf das Teilnehmerverhalten bildet die Tarifierung. Die Verkehrsprofile der tages- und jahreszeitlichen Schwankungen sind bekannt und werden vom Netzmanagement berücksichtigt.<sup>18</sup> Dabei sind große Unterschiede im tageszeitlichen Lastprofil zwischen verschiedenen Diensten feststellbar. Aus wirtschaftlichen Gründen werden Vermittlungssysteme nicht für zu erwartende Spitzenwerte dimensioniert, sondern die Grundlage bildet der Verkehrswert der mittleren Hauptverkehrsstunden-Belegungsversuche [616]. Aufgrund des Benutzerprofils und der stochastischen Verkehrseigenschaften ist deshalb immer auch mit Überlastphasen zu rechnen.

Ein zweites Beispiel für spezielles Teilnehmerverhalten sind im Fernsehen eingeblendete Rufnummern oder die Aufforderung, sich an TED-Umfragen (Televotum) zu beteiligen [528].<sup>19</sup> Man spricht hier von getriggertem Verkehr oder medienstimulierter Überlast [117], die zu sehr hohen und fokussierten Lastspitzen führen kann.

Den dritten Aspekt stellen die Wechselwirkungen zwischen Netz und Teilnehmer dar, wobei in bestimmten Fällen Mitkopplungseffekte und damit Lastanstiege entstehen können (siehe Abschnitt 3.1.2.2).

### 3.3.6 Überlast aufgrund Systemausfällen und Naturkatastrophen

Der Ausfall einer Sendeeinrichtung in Utah 1961 war ein Ereignis, dass die gezielte Untersuchung von Überlastphänomenen in rechnergesteuerten Kommunikationsnetzen auslöste [227]. Die einschneidendste Erfahrung dabei war, dass der Ausfall einer Netzkomponente eine Überlast erzeugen kann, die sich sehr schnell über weite Netzbereiche ausbreiten kann.

Erdbeben in dicht besiedelten Gebieten sind diejenigen Naturkatastrophen, die eine der größten Herausforderungen an die Verfügbarkeit eines Kommunikationsnetzes stellen. Folgende drei Ursachen begründen dies: Systemausfälle durch Beschädigung, Systemausfälle aufgrund unzureichender Energieversorgung<sup>20</sup> und das Zusammenbrechen der Systemleistung aufgrund Überlastung.<sup>21</sup> Gerade die rapide Zunahme des Fernsprechverkehrs von und in das betroffene Gebiet induzieren sehr große Verkehrsmengen. Bereits 1971 wurde nach einem Erdbeben in der Region von Los Angeles eine 35fache Folgelast [391] gemessen.

---

<sup>18</sup> Beispiele saisonaler Überlastspitzen sind Weihnachten, Silvester oder speziell das chinesische Neujahrsfest in den USA [74].

<sup>19</sup> Das jüngste Beispiel dieser Art war der telefonische Verkauf von Tickets zur Fußball-WM'98. Durch den europaweiten Aufruf waren tagelang viele Netzknoten einer unverhältnismäßig hohen Überlastung ausgesetzt.

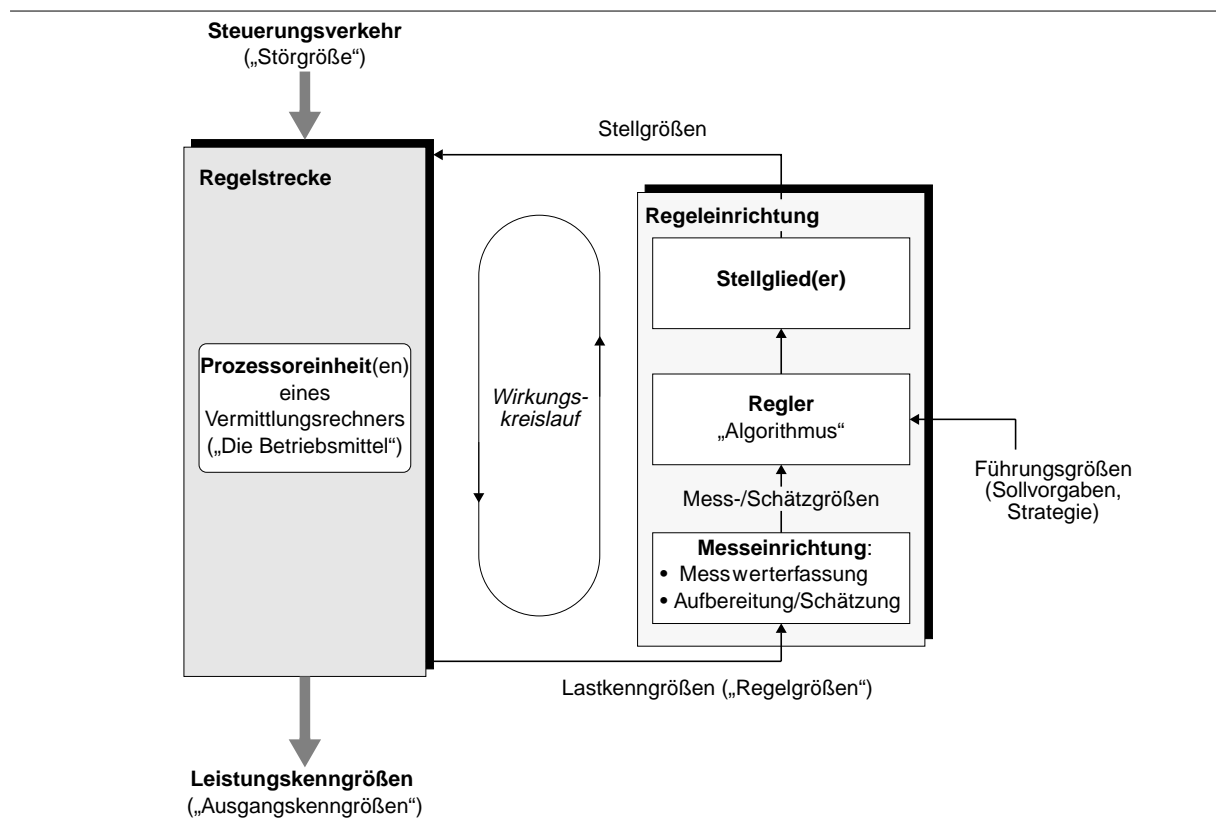
<sup>20</sup> Falls die Notstromversorgung durch Batterien erbracht werden soll, so ist deren Leistungsfähigkeit oft nur auf wenige Stunden beschränkt. In [513] wird über den Ausfall nahezu aller Telefonzellen berichtet: öffentliche Kartentelefone aufgrund fehlender Stromversorgung und die Münztelefone aufgrund voller Münzsammelboxen!

<sup>21</sup> Die Auswirkungen sind nicht nur auf das Festnetz begrenzt. Gerade der enorme Verkehrsanstieg stellt auch für Mobilfunk und Satellitenkommunikation die Ursache eines möglichen Leistungseinbruches dar.

Die schlimmste Katastrophe in einer Industrienation bildete das große *Hanshin/Awaji Erdbeben* 1995 in Japan. Die Auswirkungen auf das öffentliche Kommunikationsnetz wurden in [513,568] ausführlich dokumentiert. Bemerkenswert war die sehr große und relativ lang andauernde Überlastung.<sup>22</sup>

### 3.4 Überlasterkennung und Regeleinrichtungen

Jede Überlastabwehr versucht lastregulierend, im Sinne einer Leistungsoptimierung, zu wirken. Die Beeinflussung basiert auf einer ermittelnden Lastinformation. Durch die Eingriffe der Überlastabwehr entsteht somit eine Rückwirkung und damit ein geschlossener Wirkungskreislauf. Somit stellt diese Funktion einer Vermittlungssteuerung einen Regelkreis dar. Problematik, Eigenschaften und spezielle Struktur eines Regelkreises zur Last- und Leistungsregelung in einem Vermittlungsrechner sollen in diesem Abschnitt aufgezeigt werden.



**Bild 3-2:** Blockschaltbild der Überlastabwehr als Regelkreis. Die Regelstrecke repräsentiert den Daten-, die Regeleinrichtung den Kontrollpfad. Die Regelung wird als Softwarefunktion realisiert, die Regelstrecke sind vornehmlich Hardwarebetriebsmittel des Vermittlungsrechners.

<sup>22</sup> Es wurde eine 50fache Verkehrsintensität festgestellt. Dieser unglaubliche Verkehrsanstieg entspräche einer plötzlichen Verdopplung des Verkehrs auf ganz Japan bezogen ([568]: „The congestion that occurred was on a huge scale never seen before!“). Überlast: 1. Tag → 50fach, 2. Tag → 20fach, 3. Tag → 6fach; insgesamt dauerte die Überlastung sechs Tage

### 3.4.1 Regelkreiskomponenten

Jede Regeleinrichtung (Regler) lässt sich prinzipiell durch ihre Übertragungsfunktion charakterisieren. Aufgrund von Zufallsprozessen handelt es sich hier um stochastische Reglertypen. Die Wirkungskette eines Reglers in einem Vermittlungsrechner ist in Bild 3-2 dargestellt. Die einzelnen Regelkreiskomponenten werden in den nachfolgenden Abschnitten vorgestellt und diskutiert.

Weiterhin ist der direkte Wirkungskreis eines Reglers immer auf räumliche und/oder zeitliche Bereiche (engl. *control horizon* [430,558]) begrenzt. Dieser Aspekt wird für die in dieser Arbeit vorgestellten Regelkreisstruktur in Abschnitt 4.3 behandelt.

### 3.4.2 Allgemeine Regelziele

Die allgemeinen Zielsetzungen für eine Regeleinrichtung [29] gelten auch für den speziellen Typus einer Vermittlungssteuerung. Wichtig sind hier vor allem die dynamischen Eigenschaften wie Übergangsverhalten und Stabilitätskriterien. Weitere Regelziele lassen sich aus den spezifischen Anforderungen an Überlastabwehreinrichtungen ableiten, die ausführlich in Abschnitt 3.4.8 erläutert sind.

### 3.4.3 (Über-)Lastindikatoren

Entscheidenden Einfluss auf die Güte einer Überlastabwehr bilden Umfang, Genauigkeit und Aktualität der Informationen über den Belastungszustand der Regelstrecke. Dieses Wissen wird über Sensoren, die damit hier sogenannten Lastindikatoren entsprechen, erfasst. Allgemein lassen sich diese in *Last-* und *Leistungskenngrößen* unterteilen. Beispielsweise stellt die CPU-Auslastung eine last-, der Durchsatz eine leistungscharakterisierende Größe dar.

Eine weitere Klassifizierung lässt sich in *zeitbehaftete Größen*<sup>23</sup> und *Zustandsgrößen* (auch Momentangrößen genannt) vornehmen. Verzögerungszeiten und Flussraten sind Beispiele zeitbehafteter Größen. Die Warteschlangenbelegung dagegen ist eine Zustandsgröße.

Aus einem Indikatorwert lässt sich entweder direkt eine Last- oder Leistungsinformation gewinnen, oder man verknüpft diese mit einer anderen Größe zu einer zusammengesetzten Kenngröße. Die Anzahl allozierter Protokollzeitgeber ist ein Beispiel einer *direkten* Lastkenngröße, die nach *Little* berechnete Durchlaufzeit, das *Power-Produkt* oder der Gradient einer Warteschlangenbelegung (auch Leer- bzw. Füllgeschwindigkeit eines Puffers) sind Beispiele für *indirekte, abgeleitete* Kenngrößen.

Weitere Aspekte betreffen die prinzipielle Eignung, die Verfügbarkeit oder den Messaufwand einer Systemzustandskenngröße, die insbesondere bei der Auslegung einer Überlastabwehr wichtige Randbedingungen darstellen. Neben den bereits oben erwähnten Beispielen für Sensortypen gibt es noch eine Reihe weiterer bekannter Indikatoren, wie z.B. die Blockierwahrscheinlichkeit oder Verlustrate, bestimmte Betriebssystemressourcen, Ausführungszeiten bestimmter Programmmodule<sup>24</sup>, die *Recovery Time*<sup>25</sup> [504], Aufenthaltszeiten [68,348], die *Packet Retransmission Ratio* (TCP) oder Ankunftsdaten bestimmter Meldungstypen.

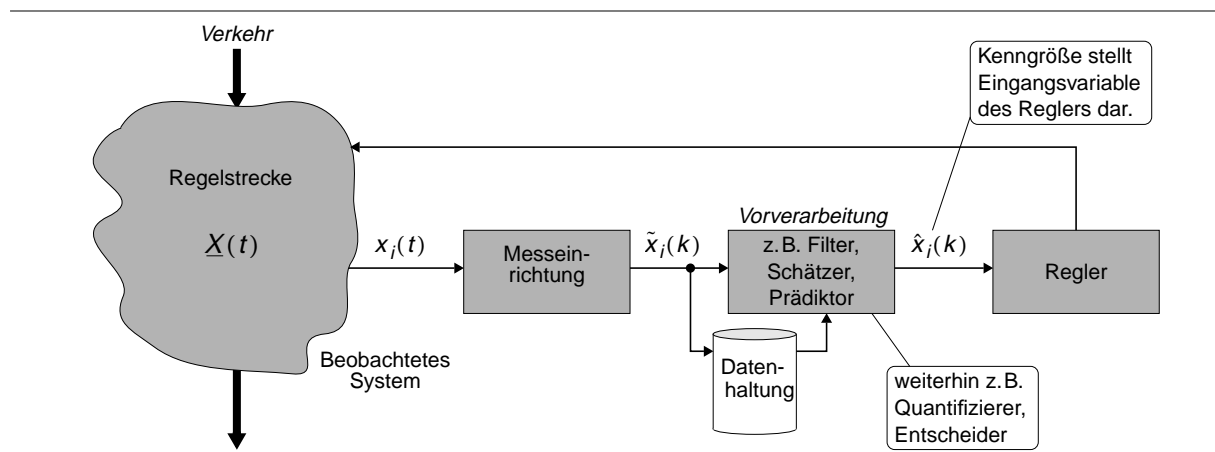
<sup>23</sup> Gleich Zeitgrößen (z.B. Wartezeit) sowie Größen die nur durch eine Zeitmittelung gewonnen werden können (z.B. Blockierrate).

### 3.4.4 Beobachtung und Messung von Kenngrößen

Gerade in einem Vermittlungsrechner stellen Beobachtung und Messung von Systemzustandskenngrößen eine nicht einfache Aufgabe dar [246]. Weiterhin ist eine Aufbereitung der Messdaten notwendig, um zwischen normalen stochastischen Lastschwankungen und tatsächlichen Überlastsituationen unterscheiden zu können.

#### 3.4.4.1 Problematik der Messdatenerfassung und -aufbereitung

Bild 3-3 zeigt die prinzipiellen Bearbeitungsschritte einer Last- oder Leistungskenngröße.



**Bild 3-3:** Erfassung und Aufbereitung einer Kenngröße

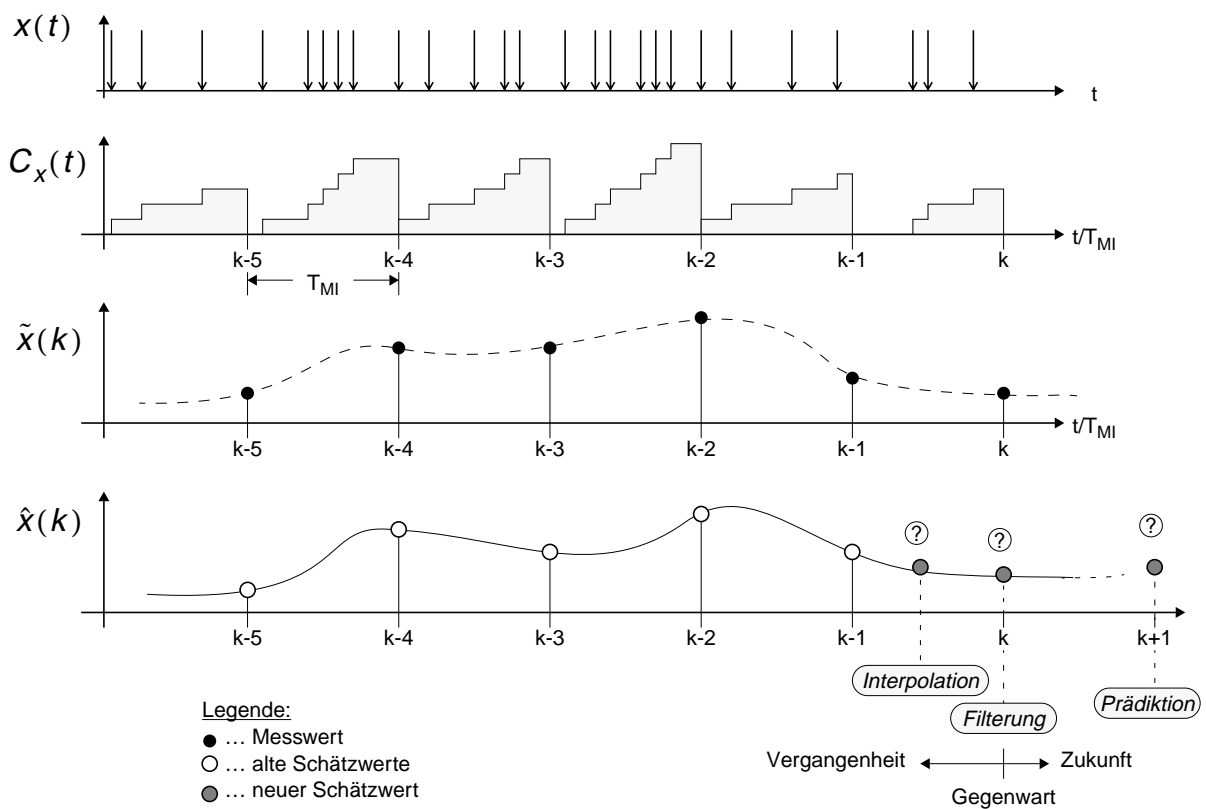
Aus Aufwandsgründen kann die Beobachtung nicht zeitkontinuierlich erfolgen, sondern die Verarbeitung neuer Schätzwerte erfolgt zu diskreten Zeitpunkten. Auch die Zeitpunkte, an denen ein Regler eingreifen kann, müssen somit diskrete Zeitpunkte sein. Sinnvollerweise entspricht die Regelintervalldauer einem ganzzahligen Vielfachen der Messintervalldauer.

Es stellt sich häufig das Problem, dass sich diejenige Größen, an deren Kenntnis man interessiert ist, nicht direkt messen lassen, weil die Messung entweder durch unvermeidbare Störgrößen beeinträchtigt wird oder das Messsignal nur eine abgeleitete Sekundärinformation von den Größen ist, die primär interessieren. Es besteht dann die Aufgabe, das gemessene Signal nachträglich in einem sogenannten *Filter* aufzubereiten, wobei ihm unter Verwendung vorausgegangener Messwerte der relevante Informationsgehalt entnommen wird.

Bild 3-4 soll dies anhand eines Beispiels verdeutlichen. Die Größe  $x(t)$  repräsentiere hier einen *Meldungsfluss* im betrachteten System. Das Messgerät verwendet einen Zähler  $C_x(t)$  als Hilfsgröße für eine Mittelwertbildung pro Messintervall ( $T_{MI} = konst.$ ). Mittels der Messreihe  $\{\tilde{x}(k), \tilde{x}(k-1), \dots\}$  wird anschließend ein Schätzwert  $\hat{x}$  berechnet.

<sup>24</sup> Beispielsweise wird im No. 2 ESS [439] als Lastindikator die Zeit verwendet, die für die Programmausführung der Hauptschleife benötigt wird (Schwellwert  $T_{Main\ Cycle} = 175ms$ ).

<sup>25</sup> Es wird der Systemzustand (z.B. Warteschlangenlängen, Prozessorauslastung) zum Zeitpunkt  $t_0$  festgehalten. Als *Recovery Time* wird die Zeit definiert, bis derselbe Systemzustand erreicht wird.



**Bild 3-4:** Beispiel für die Aufbereitung einer Kenngröße für einen Regler. Gegenüberstellung der verschiedenen Schätzverfahren.

Ein *zeitdiskretes Filter* berechnet allgemein aus den anliegenden Messwerten  $\tilde{x}$  einen Schätzwert  $\hat{x}$  nach folgender Funktion:

$$\hat{x}(k + \zeta) = f(\tilde{x}(k), \tilde{x}(k - 1), \tilde{x}(k - 2), \dots) \quad (3-14)$$

In der *Zeitreihenanalyse*<sup>26</sup> unterscheidet man folgende Fälle (Tabelle 3-2):

Parameter $\zeta$	Bezeichnung	Bedeutung von Schätzwert $\hat{X}$ (Zeitbereich)
• $\zeta < 0$ (z.B. $\zeta = -0.5$ )	<i>Interpolation</i>	Schätzwert $\hat{X}$ bezieht sich auf den tatsächlichen Wert $x$ zu einem früheren Zeitpunkt
• $\zeta = 0$	<i>Filterung</i>	Schätzwert $\hat{X}$ bezieht sich auf den aktuellen Signalwert
• $\zeta > 0$ (z.B. $\zeta = 1$ )	<i>Prädiktion, Extrapolation, Prognose</i>	Schätzwert $\hat{X}$ am Filterausgang bezieht sich auf einen zukünftigen Wert des Signals $x$

**Tabelle 3-2:** Prinzipielle Zeitbereiche für die Vorhersage eines Schätzwertes  $\hat{x}(k + \zeta)$

<sup>26</sup> Die allgemeine Begriffsverwendung ist nicht einheitlich was *Filterung*, *Schätzung* usw. betrifft. Die hier verwendeten Begriffe orientieren sich an [17].

Das hier betrachtete technische System – die Steuerungseinheit („Kontrollpfad“) einer rechnergesteuerten Vermittlungsstelle – ist aus Sicht der Lastregelung durch eine Menge endlich verfügbarer Betriebsmittel gekennzeichnet. Typische Betriebsmittel sind beispielsweise Prozessoren, Speicher für Nachrichten (z.B. Signalisierungsmeldungen), Speicher für Verbindungsdaten, Speicher für Teilnehmerdaten oder Betriebssystemressourcen wie z.B. Zeitgeber. Die einzelnen Ressourcen lassen sich selbst hinsichtlich ihres Werte- und Zeitbereiches, die jeweils diskret oder kontinuierlich sein können, unterscheiden.<sup>27</sup>

Der Belegungszustand dieser Betriebsmittel selbst stellt ein Maß für die aktuelle Systembelastung dar. Neben diesen Kenngrößen existieren eine Reihe abgeleiteter Größen die ebenfalls Anwendung als Überlastindikatoren finden (siehe Abschnitt 3.4.3).

### 3.4.4.2 Eignung und Einsatzbereiche bestimmter Kenngrößen

Die Eignung einer Kenngröße als Lastindikator hängt eng damit zusammen, welchen Informationswert sie besitzt und welcher funktionale Zusammenhang, z.B. ein linearer oder nichtlinearer, zwischen Kenn- und tatsächlicher Lastgröße besteht. So ist beispielsweise der *Ruf-* oder *Wähltonverzögerung* weniger als Lastindikator für eine Regelung der *Prozessorauslastung* geeignet. Anforderung und Zielsetzung einer Regelung bestimmen deshalb den Kreis geeigneter Sensortypen.

Diese Auswahl wird oft in einem Vermittlungsrechner durch das Problem der Beobacht- oder Observierbarkeit gewisser Kenngrößen eingeschränkt (siehe auch Abschnitt 3.4.4.4).

### 3.4.4.3 Allgemeine Definition der Systembelastung

Alle relevanten Lastkenngrößen lassen sich zu einem Vektor  $\underline{X}(t)$  zusammenfassen. Jede Vektorzeile repräsentiert eine bestimmte Kenngröße. *Dieser Lastvektor ist ein Maß für den aktuellen Systemzustand.* Für die Lastregelung werden in der Regel die einzelnen Kenngrößen quantifiziert, d.h. ihr Wertebereich wird in wenige Bereiche (z.B. 2, 3 oder 64 [259]) unterteilt. Diese Abbildung erleichtert die Handhabung und ist auch ausreichend für den praktischen Einsatz. Diese Quantifizierung wird immer vorgenommen, wenn ein Regelalgorithmus tabellengestützt realisiert wird. Natürlich müssen die Lastkenngrößen nicht immer quantifiziert werden, zum Beispiel wenn der Regelalgorithmus durch eine (abschnittsweise) kontinuierliche Funktion repräsentiert wird und die Lastkenngröße direkt eine Eingangsvariable dieser Funktion darstellt.<sup>28</sup>

Jeder Bereich entspricht einem bestimmten Lastzustand. Durch diese Abbildung erhält somit jede Kenngröße eine endliche Zustandskette als Repräsentanten. Der Lastvektor entspricht einem Ausschnitt aus dem kompletten mehrdimensionalen, diskreten Zustandsraum.

### 3.4.4.4 Messproblematik von Systemzustandsgrößen

Die aktuelle Systembelastung  $\underline{X}(t)$  ist durch die tatsächlichen Werte  $x_i(t)$  der einzelnen Lastkenngrößen zum Zeitpunkt  $t$  gekennzeichnet. Für die Überlastabwehr wäre es ideal, wenn die-

---

<sup>27</sup> Änderungszeitpunkte (Freiwerden bzw. Belegung) eines Betriebsmittels

<sup>28</sup> Durch die endliche Anzahl von Stellen in einem Digitalrechner erfolgt durch die Rundung der kontinuierlichen Lastkenngröße jedoch wieder prinzipiell eine Quantisierung. Der resultierende Zustandsraum ist jedoch ein Vielfaches größer.



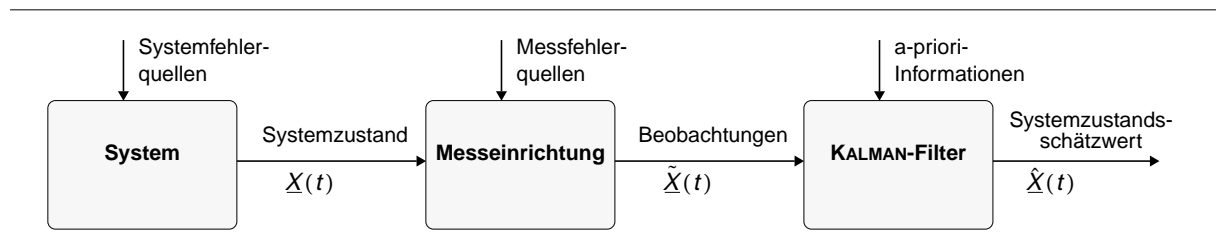
se Größen unverfälscht und ohne Zeitverzug dem Regler zur Verfügung gestellt werden könnten. In der Realität gibt es jedoch folgende Einschränkungen:

- Bestimmte Typen von Kenngrößen, z.B. alle zeitbezogenen Größen, müssen über eine Zeitspanne aufgenommen werden, was einen Zeitverzug zur Folge hat.
- Jede Messung erfordert allgemein Zeit.
- Die Messungen erfolgen in der Regel unterstützt durch sogenannte Softwaremonitore. Jede Messung benötigt somit selbst Ressourcen, stellt einen Eingriff in das beobachtete System dar und verfälscht unter Umständen somit indirekt die zu erfassenden Größen.
- Von bestimmten Kenngrößen möchte man oft nicht den tatsächlichen Wert, sondern beispielsweise Mittelwert, Standardabweichung oder Gradient.
- Vorverarbeitung durch Filter z.B. zur Unterdrückung von Rauschen.
- Notwendiger Kompromiss zwischen Reagibilität<sup>29</sup> und Stabilität einer Regelung [569].

Aufgrund dieser Problematik wird die gemessene bzw. geschätzte Belastung  $\hat{X}(t + \delta t)$  nur näherungsweise dem gesuchten tatsächlichen Systemzustand  $X(t)$  entsprechen.

### 3.4.5 Zustandsgrößenschätzung

Die Eingabevariablen eines Last- bzw. Leistungsreglers sind bestimmte Zustandsinformationen der Regelstrecke und werden durch Messungen gewonnen (Bild 3-5). Dem interessierenden Nutzsignal sind aufgrund des Zufallsverkehrs stochastische Störsignale überlagert, die durch geeignete Filter wieder getrennt werden müssen. Für zeitdiskrete Systeme haben sich die sogenannten *Kalman-Filter* [11] aufgrund einer Vielzahl von Vorteilen gegenüber anderen Filterverfahren (z.B. *Wiener-Filter*) in vielen technischen Disziplinen bewährt [56].<sup>30</sup>



**Bild 3-5:** Zustandsschätzung eines gestörten dynamischen Systems mittels eines Kalman-Filters

Mit Hilfe dieser Filtertechnik kann der Zustand eines dynamischen Systems aus verrauschten Messungen optimal geschätzt werden. Die optimale Zustandsschätzung nutzt alle verfügbaren Informationen, um die bestmögliche Aussage über den Zustand des Systems zu einem beliebigen Zeitpunkt zu machen. Die nötigen Informationen stammen aus der Differentialgleichung des Systems, aus allen verfügbaren Messwerten und aus den statistischen Eigenschaften von Messunsicherheiten und Systemstörungen.

<sup>29</sup> Unter *Reagibilität* versteht man die *Fähigkeit*, sehr sensibel (und damit schnell) zu reagieren. *Reaktivität* dagegen ist eher ein Maß für die Rückwirkung (bezogen auf einen bestimmten Normwert).

<sup>30</sup> Das sogenannte *Kalman-Bucy-Filter* ist die erweiterte Version für *zeitkontinuierliche* Signale.

Auf dem Gebiet der Kommunikationsnetze lassen sich viele Anwendungen, die für dieses Schätzverfahren prädestiniert sind, identifizieren. So gibt es hierzu auch beispielsweise Vorschläge für *CACs* [164], die ATM-Messtechnik [286], die Verkehrsvorhersage [620] oder aus dem Bereich des Netzmanagement [125,416]. Speziell für Lastregelverfahren haben sich *Kalman-Filter* jedoch nicht durchgesetzt. Dies hat im wesentlichen folgende Gründe<sup>31</sup>:

- In jedem Regelintervall sind *aufwendige Rechnungen* notwendig (z.B. Matrixinversionen). Die verfügbare Rechenzeit ist aber gerade für die Lastregelung in Vermittlungsrechnern streng begrenzt (siehe auch Abschnitt 3.4.8), so dass der Kosten-Nutzen-Aufwand hier nicht zu rechtfertigen ist.
- Für eine Kalmanfilterung werden die *Varianzen der Fehlerquellen* benötigt. Diese sind aber in der Regel hier unbekannt.
- Bei einer Kalmanfilterung wird davon ausgegangen, dass Nutz- und Störsignal im gleichen Frequenzbereich liegen, so dass die Trennung nicht durch einfache *Bandfilterung* möglich ist, sondern zur Trennung *Schätzmethoden* angewendet werden müssen. Gerade für Überlastabwehralgorithmen sind aber oft einfache *Tiefpässe* ausreichend (siehe z.B. [199,471]).
- *Kalman-Filter* werden hauptsächlich für sehr sensible Regelstrecken eingesetzt, wo es tatsächlich auf eine optimale Zustandsschätzung ankommt. Dies trifft hier aber weniger zu. Im Gegenteil wird oft – wie bereits erläutert wurde – der Zustandsraum in wenige (Last-)Bereiche quantifiziert.

Damit reduziert sich das Problem der Zustandsgrößenschätzung hier auf praktikable Messverfahren und einfache, geeignete Filter. Diese Aspekte sind Gegenstand des nachfolgenden Abschnittes.

### 3.4.6 Störsignalfilterung

Bei vielen Kenngrößen muss aus dem gemessenen Signal  $x(t)$  ein Nutzsinal  $y(t)$  bestimmt werden, das durch ein Störsignal  $n(t)$  beeinflusst wird. Bei einer additiven Beeinflussung, d.h.  $y(t) = x(t) + n(t)$ , kommen für die Trennung beider Signale prinzipiell folgende Verfahren zur Anwendung [30]:

- *Bandpassfilter* (Frequenzspektren liegen in verschiedenen Bereichen) bzw.
- *Schätzverfahren* (Frequenzspektren liegen im gleichen Bereich), wobei hier allerdings der Einfluss des Störsignals nicht vollständig eliminiert, sondern lediglich minimiert werden kann.

In der hier nun vorliegenden Anwendung ist man vornehmlich nur an den Signalmittelwerten, also den „sehr niederfrequenten“ Anteilen interessiert. Hierzu sind bereits einfache, diskrete Filter ausreichend. Der wichtigste Algorithmus stellt dabei die *rekursive Mittelwertbildung* dar, wobei sich vor allem folgende Varianten in der Praxis bewährt haben (Notation entsprechend Bild 3-4):

- *Mittelwertbildung mit begrenztem Gedächtnis*<sup>32</sup>, d.h. die letzten  $N$  Messwerte werden in die Berechnung mit einbezogen:

---

<sup>31</sup> In [302] werden beispielsweise einige Nachteile bei der Anwendung von *Kalman-Filtern* für *Flusskontrollverfahren* aufgeführt.

<sup>32</sup> Auch *gleitende Durchschnitte* oder *gleitendes Fenster* genannt.

- Gleichgewichtung (gleitender Durchschnitt 1. Ordnung):

$$\hat{x}(k) = \frac{1}{N} \cdot \sum_{i=k-N+1}^k \tilde{x}(i) . \quad (3-15)$$

- Gewogene gleitende Durchschnitte:

$$\hat{x}(k) = \frac{1}{N} \cdot \sum_{i=k-N+1}^k a_{i-k+N} \cdot \tilde{x}(i) \quad \text{mit} \quad \sum_{j=1}^N a_j = 1 . \quad (3-16)$$

- *Mittelwertbildung mit unendlichem, nachlassendem Gedächtnis*
  - Exponentielles Glätten (*exponential smoothing*) als wichtigster Vertreter dieser Klasse:

$$\hat{x}(k) = \xi \cdot \tilde{x}(i) + (1 - \xi) \cdot \hat{x}(k - 1) \quad \text{mit} \quad \xi \text{ als Glättungsfaktor}^{33}. \quad (3-17)$$

Eine ausführliche Diskussion der verschiedenen Verfahren ist beispielsweise in [12,467] zu finden. Aufgrund einer Vielzahl von Vorteilen, z.B. kleiner Rechen- und Speicheraufwand, gute Genauigkeit, Reagibilität und auch Stabilität, wird in praktischen Anwendungen (z.B. [79,246,269,382]) fast ausschließlich das Prinzip des *exponentiellen Glättens* eingesetzt. So wird auch in dieser Arbeit für die *schnelle Überlastabwehr* (siehe Abschnitt 4.3.2) dieses einfache Schätzverfahren verwendet.<sup>34</sup> Für Regelkreise, die auf *länger andauernde Überlasten* reagieren sollen, werden dagegen eher *gleitende Fenster* (z.B. mit  $N=1000$ ) zur Mittelwertbildung eingesetzt.

### 3.4.7 Festlegung der Zeitkonstanten – Mess-, Schätz- und Regelintervalle

Gerade bei einer stochastischen Regelung muss ein Kompromiss zwischen den konträren Anforderungen Stabilität und Reagibilität gefunden werden [569]. Zum einen sollte das Ergebnis einer Messung möglichst frühzeitig dem Regler verfügbar sein, andererseits wird eine Aussage mit einer gewissen statistischen Sicherheit erwartet. Beide Kriterien stehen in engem Zusammenhang mit der Länge des Messzeitraumes. Bei der Last- und Leistungsregelung in einem Vermittlungsrechner sind folgende Zeitebenen zu unterscheiden: Elementarereignisse (*elementary interval*, z.B. Ankunft einer Meldung, Freiwerdezeitpunkt eines Betriebsmittels), Messfenster (*measurement interval*), Schätzintervall (*estimation interval*) bzw. Vorhersagezeitraum (*prediction interval*) und Regelintervall (*control interval*).<sup>35</sup> Für jede Zeitebene ist eine separate Intervallabschätzung notwendig.

Für Verfahren zur Abschätzung von Messintervalllängen sei auf [569] verwiesen. Speziell für die Kenngrößen Auslastung und Durchsatz werden hierzu in [243] Algorithmen angegeben.

<sup>33</sup> Eine wichtige Bedeutung kommt der Festlegung des *Glättungsfaktors*  $\xi$  hinzu. Die Wahl von  $\xi$  stellt eine Kompromissfindung zwischen Reagibilität und Stabilität dar. Erfahrungsgemäß wird  $\xi$  zwischen 0.01 und 0.3 gewählt [12]. Weiterhin kann durch geeignete Parameterwahl die unterschiedlichen Verfahren zu den gleichen Schätzwerten führen: ein Glättungsfaktor von  $\xi = 2/(N+1)$  entspricht beispielsweise einer Gleichgewichtung mit Fenstergröße  $N$ .

<sup>34</sup> In [199,471] wurden die aufgeführten Verfahren, und auch weitergehende Filtertechniken, für die Anwendung in einem Vermittlungsrechner in zahlreichen Parameterstudien getestet.

Für die Überlastabwehr in Vermittlungsrechnern werden eher längere Messzeiträume verwendet. Im Rahmen dieser Arbeit wird die Messfenstergröße auch vom jeweiligen Regelkreistyp (siehe Abschnitt 4.3.2) beeinflusst. Die Länge der Schätz- und Regelintervalle ist üblicherweise ein ganzzahliges Vielfaches des Messintervalls.

### 3.4.8 Anforderungen und Kriterien an Überlastabwehreinrichtungen

An die Funktion der Überlastabwehr werden folgende, teilweise sich widersprechende, Anforderungen gestellt:

- *Echtzeitfähigkeit*  
Das Antwortzeitverhalten, insbesondere erfolgreich bearbeiteter REQUESTS, unter Überlast sollte nicht merklich schlechter sein als unter normalen Lastbedingungen.
- *Simplität, Komplexität (oder Implementation Complexity)*  
Z.B. wenige Parameter, einfache Algorithmen oder kurzer Programmcode.
- *Blindlastanteil*  
Eine Minimierung dieses Anteiles bedeutet eine Leistungsmaximierung.
- *Overhead (Rechenzeit- und Speicherplatzbedarf)*  
Der Rechenzeitbedarf für die Überlastabwehr sollte möglichst klein sein.<sup>36</sup>
- *Reagibilität (oder Responsiveness)*  
Steht allgemein für die Fähigkeit eines Regelkreises „sehr sensibel zu reagieren“. Als quantitatives Maß bietet sich die Einschwingzeit auf Lastsprünge an. Dabei werden folgende Teilgrößen mit berücksichtigt:
  - *Erkennungszeit (oder Detection Time)*  
Zeit von Überlastbeginn bis zum Erkennen durch die Überlastabwehr.
  - *Reaktionszeit (oder Reaction Time)*  
Zeit vom Erkennen bis zum Eingreifen (d.h. Einstellen der Aktoren) durch die Überlastabwehr. Diese Zeit entspricht häufig der Regelintervalldauer.
  - *Einschwingzeit (oder Operation Time)*  
Zeit vom Eingreifen durch die Überlastabwehr bis zum Einstellen eines stabilen Systemzustandes.
- *Stabilität*  
Ein instabiles Regelsystem ist ineffizient [178]. Für einen Stabilitätsnachweis existieren eine Vielzahl von Verfahren, für die allerdings die Übertragungsfunktion benötigt wird [29]. Diese ist für Lastregelkreise von Vermittlungsrechner meistens aber nicht bekannt. Häufig ist sie auch in diesem Anwendungskreis eine stark nichtlineare Funktion.
- *Trägheit*  
Im Gegensatz zu einem schnellen Reaktionsvermögen sollte aber eine Lastregelung tole-

---

<sup>35</sup> Die Beobachtung von Größen erfolgt entweder durch Abtastung (*Polling*) zu äquidistanten Zeitpunkten („aktive Messung“, synchron) oder ist an bestimmte Ereignisse geknüpft („passive Messung“, asynchron). Diese Prinzipien lassen sich auf allen Zeitebenen anwenden. Die kleinste Zeiteinheit („Elementarintervall“) ist die Zeit zwischen zwei Ereignissen. Diese kann in der Modellwelt prinzipiell auch beliebig kleine Zeiten umfassen (z.B. Messung der Warteschlangenbelegung einer Queue, die den Verkehr einer Quelle mit Poisson-Charakteristik puffert). In realen System gibt es hier jedoch Einschränkungen.

<sup>36</sup> Für kommerzielle Systeme gibt es entsprechende Vorgaben. Beispielsweise steht dem Prozess für die Lastregelung nicht mehr als ungefähr ein Prozent Prozessorzeit zur Verfügung.

rant gegenüber natürlichen stochastischen Verkehrsschwankungen sein. Eine Überlastabwehr muss deshalb gewisse Filterfunktionen für ihre gemessenen Lastsignale vorsehen.

- *Fairness*

Die unter Überlast notwendigen Regeleinriffe sollten sich gerecht auf die „betroffenen Objekte“ auswirken. Ein absolutes Fairnesskriterium gibt es nicht, demzufolge gibt es auch im Umfeld der Überlastabwehr sehr viele Definitionen bezüglich dieses Aspekts. Der Fairnessbegriff unterscheidet sich beispielsweise für folgende Zielsetzungen:

- Bezüglich *Verkehrszusammensetzungen* (Dienste- oder Rufmix): In [296] wird gefordert, dass das Verhältnis angenommener zu abgelehnter Rufe für jeden Ruftyp gleich sein sollte.<sup>37</sup>
- Bezüglich *Ruf-/Verbindungstyp* (z.B. Multicasting [490]).
- Bezüglich *Dienstauswahl*: geschäftliche Nutzern wird eine höhere Kapazität zugeteilt als privaten Nutzern ([290]: „*Business Services before Domestic Services*“, z.B. eine Videokonferenz hat Priorität vor dem Dienst Filmabruf).
- Bezüglich *Tageszeit*: Beispielsweise wird dem Dienst Filmabruf in den Nachmittag- und Abendstunden mehr Systemkapazität zur Verfügung gestellt als in den Spitzenzeiten des Vormittags.
- Bezüglich *Kundentyp*: Gute Kunden bekommen eine bessere Dienstgüte als Nutzer, die weniger oft einen Dienst in Anspruch nehmen.

- *Robustheit*<sup>38</sup>

Selbst unter extremen Überlastsituationen sollte der maximale Durchsatz noch nahe der Nennkapazität sein.<sup>39</sup>

- *Nichtvermittlungstechnische Prozesse*

Unter Überlast dürfen nichtechtzeitkritische Aufträge für eine gewisse Zeit zurückgestellt werden. Allerdings muss bestimmten Aufgaben (z.B. Sicherheitstechnik) noch ein gewisser CPU-Anteil garantiert werden.

- *Wirkungskreis* (oder *Control Horizon*)

Dieser muss auf den für eine Überlastabwehr spezifizierten Bereich beschränkt bleiben. So darf beispielsweise ein lokaler Lastregelkreis nicht in benachbarte Netzknoten eingreifen und deren Leistungsparameter bewusst (nachteilig) beeinflussen.<sup>40</sup>

Betrachtet man bekannte Verfahren aus der Vermittlungstechnik, dann stand bei der Überlastabwehr nicht unbedingt die *Leistungsoptimierung* für bestimmte Lastbereiche im Vordergrund. Die Gewichtung lag vielmehr bei einfachen, stabilen und vor allem robusten Mechanismen. Gerade bei B-ISDN-Systemen wird durch das Kriterium „maximal zur Verfügung stehende Prozessorzeit“ eine wesentliche Vorgabe an die Komplexität eines Lastregelverfahrens gemacht.

---

<sup>37</sup> Davon ausgenommen werden allerdings Ruftypen mit einer „relativ niedrigen“ Ruftrate, die von der Überlastabwehr nicht reguliert werden sollten.

<sup>38</sup> In der Regelungstheorie wird ein Regler dann als *robust* bezeichnet, wenn er in einem gewissen Umfang *unempfindlich* gegenüber dem zugrundeliegenden, zwangsläufig ungenauen Prozessmodell, und dem sich dynamisch veränderlichen Prozessverhalten ist [29].

<sup>39</sup> Ein Maß für Robustheit wird in ITU Q.543 [648] wie folgt vorgegeben: „Bei 50%iger Überlast sollte der Durchsatz nicht unter 90% der Nennkapazität fallen.“ Im Zusammenhang der Verkehrssteuerung im Signalisiersystem Nr. 7 ist in ITU E.744 [630] eine weitere Definition dieses Kriteriums zu finden.

<sup>40</sup> Dies wäre hier Aufgabe einer übergeordneten Regelung.

## 3.4.9 Lastregel- und Überlastabwehrverfahren

### 3.4.9.1 Klassifizierungsschema für Wirkungsbereich

Ein Verfahren zur Last- oder Leistungsregelung lässt sich hinsichtlich seines räumlichen und zeitlichen Wirkungskreises klassifizieren. Betrachtet man die *Zeitebene*, in welcher ein Regelverfahren wirksam wird, dann lassen sich auf oberster Ebene zwei Zeitbereiche unterscheiden (siehe auch Abschnitt 4.3.2):

- Lastregelung ist bereits im sogenannten Normallastzustand aktiv, d.h. *vor* dem Auftreten einer Überlastung. Man spricht in diesem Zusammenhang von *präventiven* oder *proaktiven* Verfahren (*preventive/proactive controls*), da diese bereits im Vorfeld wirksam sind.
- Überlastabwehr, d.h. *nach* dem Erkennen einer Überlastung wird die Regelung aktiv. Diese Mechanismen werden als *reagierende* Verfahren (*reactive controls*) bezeichnet.<sup>41</sup>

Bei den reagierenden Verfahren werden oft noch weitere *relative* Zeitbereiche – z.B. kurz-, mittel- oder langfristig (*short-term, medium-term, long-term*) – unterschieden.

### 3.4.9.2 Konventionelle Methoden

Seit Beginn der rechnergesteuerten Vermittlungstechnik wurden zwischenzeitlich eine Vielzahl an Verfahren vorgeschlagen. Nachfolgend eine Reihe von Beispielen die bezüglich ihres *räumlichen Wirkungskreises* strukturiert sind.

- *Lokal*, d.h. für einen Netzknoten:
  - *Rufblockierung* oder *Zurückweisung* von Anforderungen
  - *Verzögerung* von Quittungen oder Rufablehnungen
  - *Aktive Verschlechterung* der Dienstgüte, um das Teilnehmerverhalten zu beeinflussen, z.B. durch Anlegen des Gassenbesetzttons, Wegfallen des Wähltons oder verzögerte Signalisierung des Freizeichens.
  - *Verwerfen* von Datenpaketen bzw. Rufanforderungen
  - Bearbeitungsstrategien (*Queueing Disciplines*) für den *Puffer* von *neuen Rufanforderungen* (siehe auch [152]): a) *Abfertigungsdisziplinen*, z.B. FIFO- oder LIFO-Strategie<sup>42</sup> während Überlastphasen; b) *Verdrängungsstrategien*, z.B. bei voller Warteschlange eignet sich *Blocking* (neuer Ruf geht verloren) oder *Push-Out* (ältester Ruf wird gelöscht). Weiterhin bietet *Time-Out* (überschreitet die Wartezeit einen Schwellwert, dann wird der jeweilige Ruf gelöscht) eine Begrenzung der maximalen Durchlaufzeit.
  - *Lastvertagung*, d.h. Zurückstellung niederpriorer Aufträge (z.B. Entgelterfassung)
  - *Lastteilung* oder *-verschiebung* in Mehrprozessorsystemen [134]
  - Änderung von Systemparametern
- *Global*, d.h. für einen Netzbereich:

---

<sup>41</sup> Gelegentlich werden in der Literatur auch die Begriffe *passive* für *präventive* bzw. *aktive* für *reagierende* Verfahren gebraucht (z.B. [414]).

<sup>42</sup> Hintergrund dieser Technik aus PSTN-Netzen ist die Beobachtung, dass bei größer werdendem Wähltonverzug die Wahrscheinlichkeit zunimmt, dass der rufende Teilnehmer vorzeitig seinen Rufversuch abbricht, bzw. ohne den Wählton abzuwarten sofort einen erneuten Rufversuch startet [106]. Durch die ansteigenden Wartezeiten bei FIFO-bedienten Queues wirkt sich diese Wahrscheinlichkeit für alle ankommenden Rufe aus. Wird jedoch während einer Überlastphase nach der LIFO-Strategie gearbeitet, dann kann zumindest ein Bruchteil (Nutzdurchsatz) der Teilnehmer mit einer Mindestgüte (hier: minimaler Wähltonverzug) bedient werden (siehe auch [27, 151, 187, 241]). Das AXD 301 von Ericsson ist ein ATM-Vermittlungssystem, das dieses Laststeuerungsprinzip verwendet [96].

- Laststeuerung durch Tarifierung
- Netzzugangskontrolle [596]
- Adaptive Verkehrslenkung, Rerouting
- Speichermanagement oder allgemein Ressourcenreservierung
- Flusskontrolle bzw. zeitweise Inaktivierung bestehender virtueller Verbindungen
- Priorisierung ankommender bzw. endender Verkehrsströme [169]

Die aufgeführten Mechanismen decken den Großteil bekannter Verfahren ab.

### 3.5 Beispiele aus etablierten Netztechnologien

In verschiedenen Netztechnologien wurden im Lauf der Zeit verschiedene, bewährte Verfahren zur Überlastabwehr entwickelt. Die Anforderungen an die Komponenten dieser Lastregelkreise resultieren aus den jeweiligen spezifischen Randbedingungen. Für die Konzeption einer Last- und Leistungsregelungsstruktur für eine B-ISDN-Vermittlungssteuerung ist es notwendig, diese Strategien und Algorithmen zu kennen, da gewisse Elemente übernommen oder in abgewandelter Form eingesetzt werden können. Aus diesem Grund sollen nachfolgend verschiedene ausgewählte Beispiele kurz erläutert werden.

- *Fernsprechnetze*

Hinweise zur Überlastabwehr in den ersten rechnergesteuerten Vermittlungssystemen sind in [32] enthalten. In SPC-Systemen der ersten Generation bildete nicht nur der Prozessor den Engpass unter Überlast, sondern oft hatte die Überlastabwehr andere Ressourcen als Regelobjekte. In [464] wird z.B. ein Algorithmus auf Basis der Rufregisterbelegung vorgestellt. Weiterhin lassen sich generelle Unterschiede durch die Evolution in der Signalisierungstechnik festmachen.<sup>43</sup> Typische Verfahren sind in [82,112,187,235,237,246,248,282,292,294,465,480,496,497,519,525,541,545,566,553,598,600] zu finden.

Vermittlungssteuerungen von PSTN- und ISDN-Systemen besitzen in der Regel eine hierarchische Steuerungsstruktur. Dabei ist die Signalisierung auf die Leitungsmodule verteilt und die Rufbearbeitung in einem Vermittlungsrechner zentralisiert.<sup>44</sup> Die entworfenen Überlastabwehrkonzepte basieren auf dieser Hardwarearchitektur sowie insbesondere den spezifischen Merkmalen des Fernsprechdienstes.

- *Schmalband-ISDN*

Relativ aufwendige Verfahren wurden beispielsweise für die ISDN-Systeme von ALCA-TEL [67,387,543] (z.B. unterschiedliche Prioritäten für verschiedene Rufstypen), ERICSSON [252] (z.B. Klassifizierung einer neuen Dienstanforderung, Abschätzung der zu erwartenden, dienstspezifischen Bearbeitungszeit), SIEMENS [137,138,231] (z.B. zwei kooperierende Regelkreise, Blindlastvermeidung in zentralem Prozessor, adaptive Parameternachführung), NOKIA [348] (z.B. interne Flusskontrolle zwischen Peripherie- und Zentralprozessor) entwickelt. Weitere interessante Varianten und Überlastabwehrstrategien von N-ISDN-Vermittlungssteuerungen sind u. a. in [89,168,317,497,504] zu finden.

- *Nebenstellentechnik, TK-Anlagen*

---

<sup>43</sup> Für die Überlastabwehr macht es beispielsweise einen Unterschied, ob die gesamte Rufinformation in einem Datenpaket (z.B. DSS1) ankommt, oder ob über einen Zeitraum von mehreren Sekunden ankommende Wählimpulse bearbeitet werden müssen.

<sup>44</sup> Einem zentralen Rufprozessor können somit mehrere tausend Signalisierprozessoren untergeordnet sein.

Für TK-Anlagen hatte die Überlastabwehr nie die Bedeutung wie für die öffentliche Technik. Hier standen andere Anforderungen (z.B. Vielfalt an Leistungsmerkmalen) im Vordergrund. Sehr leistungsstarke Systeme sind aber natürlich mit einer Lastregelung ausgestattet (z.B. [259]). Es wird erwartet, dass die Trennung in öffentliche und private Technik zukünftig verschwinden wird.

- *Rechner- bzw. Paketnetze*

Lastregelverfahren für das Internet wurden bereits anfangs der 1980er Jahre publiziert [142,475]. Allgemein werden in Paketnetzen lokale (z.B. Paketweiterleitungsstrategien, interne Lastverteilung [593]), globale (z.B. Routing [380]) oder Ende-zu-Ende-Mechanismen (z.B. Flusskontrolle [204,266,310]) unterschieden. Ein weit verbreitetes Verfahren ist die Überwachung von Paketpuffern. In Abhängigkeit vom Füllgrad werden dann Pakete verworfen [339]. Je nach Lastzustand kann sich die Verwerfung auf unterschiedliche Pakettypen beziehen [143]. In einem Verfahren [419], das sich auf gesamte Paketnetzbereiche bezieht, wird der Schätzwert für die Netzauslastung indirekt aus der Auswertung von Paketzeitstempeln errechnet. Eine gute Übersicht von Verfahren zur Überlastabwehr in paketvermittelnden Netzen ist in [73,141,222,316,561,573] zu finden.

- *Signalisiernetze*

Für die Verkehrssteuerung und Überlastabwehr im Signalisiersystem Nr. 7 wird in ITU E.744 [630] ein Rahmenwerk vorgegeben. Beispiele zu Lastregelverfahren für diesen Netztyp sind in [194,371,449,450,451,572,575] zu finden. Viele Laststeuerungsstrategien für Signalisiernetze kombinieren Elemente von Überlastabwehrverfahren paket- (z.B. Flusskontrolle, *Throttling* [492]) und durchschalteter Netze (z.B. *Call Gapping*) [71].

- *Intelligentes Netz*<sup>45</sup>

Der Systemengpass im *IN* wird durch die zentralen Rechner für die Dienststeuerung (*SCP*) gebildet. Die Kommunikation ist dialogorientiert und findet auf der Anwenderebene (*INAP*) statt. Weiterhin wird im *IN* ein breites Dienstespektrum unterstützt. Diese drei Hauptaspekte bilden das Rahmenwerk bekannter Überlastabwehrverfahren im *IN*.

Am wichtigsten ist der Schutz der *SCPs* vor Überlast. Dies wird durch eine Zugangskontrolle in den *SSPs* bzw. allgemein durch eine Verkehrsrosselung zum *SCP* erreicht (z.B. [72,196,239,273,308,337,346,403,405,409,428,429,494,529,574]). An zweiter Stelle kommt die lokale Lastregelung in einem *SCP*-Rechner (z.B. [262]). Komplexere Verfahren basieren auf der Berücksichtigung eines breiten Dienstespektrums (z.B. [296,297,307,514]).<sup>46</sup>

Neben dem *SCP* werden auch dedizierte Verfahren für den Schutz eines *SSP* [361], *IP* [305,306] oder *SDP* entwickelt [177,183,184].

- *Mobilkommunikationsnetze*

Mobilfunkdienste besitzen in der Regel aufwendigere Signalisierszenarien als reine Festnetzdienste. Daneben sind in zellularen Netzen die Mobilitätsverwaltung oder Hand-Over-Entscheidungen besonders lastintensive Prozeduren für die Mobilitätsvermittlung-

---

<sup>45</sup> Im Intelligenten Netz wird eine Verteilung der Dienstvermittlungs- und der Dienststeuerfunktionen auf verschiedene Netzelemente vorgenommen. An der Zieladresse erkennt ein Vermittlungsknoten (*SSP*), ob es sich um einen *IN*-Dienst handelt. Vor allem während der Aufbauphase eines *IN*-Dienstes findet ein Dialog zwischen *SSP* und *SCP* statt. Globale *IN*-Lastregelmechanismen stehen in engem Zusammenhang mit Verfahren des *SS* Nr. 7.

<sup>46</sup> Im *IN* kann eine Klassifizierung in Gruppen ähnlicher Dienste anhand der Zieladresse erfolgen. In [349] werden hierzu Messverfahren beschrieben, deren Ergebnisse (z.B. die mittlere Dauer eines *IN*-Rufes) u.a. die Parametrisierung der *SCP*-Überlastabwehr beeinflussen.



steuerungen. Die Überlastabwehrproblematik steht in engem Zusammenhang zum IN bzw. Signalisiernetzen [333]. In [197,420,586] werden Algorithmen für Mobilfunkvermittlungsstellen präsentiert.

- *Netzmanagement*

Eine Funktion des Netzmanagements stellt die Laststeuerung für einen begrenzten Netzbereich auf den oberen Zeitebenen dar [77]. Damit kann die räumliche Ausdehnung des Wirkungskreises relativ groß werden. Die zeitlichen Aspekte betreffen vornehmlich präventive aber auch reaktive Maßnahmen [531].

Allgemeinere Ansätze beschäftigen sich mit der Verwendung von *Fuzzy Logic* und *Neuronalen Netzen* für die Überlastabwehr von Vermittlungssteuerungen. Gerade im B-ISDN mit seinen „ungewissen“ Verkehrsmustern soll die *Fuzzy Logic* prädestiniert sein [510].

## 3.6 Neue Aspekte und Herausforderungen durch die Dienstintegration in Breitbandnetzen

Durch die Dienstintegration im ISDN und den relativ großen Steuerverkehrsanteil im IN und in Mobilfunknetzen, haben sich die Anforderungen an die Vermittlungssteuerungen in den letzten Jahren vergrößert. Mit der Einführung des B-ISDN und dem Transport von IP-Verkehr über ATM-Netze erweitert sich die Problematik der Last- und Leistungsregelung für ATM-Vermittlungsrechner. Insbesondere durch die Aspekte Verkehrs- und Dienstmix, Teilnehmer- und Endgeräteverhalten, Problematik der Planung und Vergabe von Betriebsmitteln, Wechselwirkung zwischen Nutzer- und Steuerungs-Ebene sowie die spezifische Steuerungsstruktur von ATM-Vermittlungsrechnern. Diese Fragestellungen sollen nachfolgend erläutert und diskutiert werden.

### 3.6.1 Dienstintegration und Verkehrscharakteristik

#### 3.6.1.1 Allgemeine Aspekte

Die bekannte mittlere *Tagesverkehrskurve* basiert vorwiegend auf dem Fernsprehdienst und kann deshalb zukünftig nicht mehr für eine Lastabschätzung herangezogen werden. Es wird vielmehr eine gleichmäßigere Verteilung des Verkehrs erwartet. Derzeit liegt die *Hauptverkehrsstunde* noch am späten Vormittag, dieser Zeitraum wird sich zukünftig vermutlich nicht mehr so eindeutig festlegen lassen. Die Angabe der Steuervermittlungsleistung auf Basis von *Hauptverkehrsstunden-Belegungsversuchen* (BHCA) kann nicht ohne weiteres auf B-ISDN-Systeme übertragen werden, sondern muss meines Erachtens neu definiert werden (siehe auch Abschnitt 3.6.2.2).

Die *Charakteristik des Rufankunftsprozesses* wird für unterschiedliche Dienste verschieden sein. Die Annahme einer Poisson-Verteilung im PSTN (siehe auch Abschnitt 5.2.2) wird nicht für alle Dienste haltbar sein. Bestimmte Dienste können stoßartige Lasten (auch *mass calling situations* genannt) hervorrufen.

Die Dimensionierung und Auslegung durchschalteter Netze basierte immer auf mittleren Rufankunftsdaten und mittleren Belegungszeiten. Durch den unterschiedlichen Grad

an Komplexität verschiedener Dienste ist hier nun eine gewisse Sorgfalt geboten. Da der sogenannte *BHCA-Wert* auch eine wichtige Führungsgröße für die Überlastabwehr war und auch zukünftig sein kann, wird diese Problematik in Abschnitt 3.6.2 ausführlicher diskutiert.

Das *IN-Konzept* wird auch im B-ISDN weiterhin bestehen. Vermutlich werden sogar fast alle Netzdienste als IN-Dienste angeboten [19]. Diese Aussage wird allerdings durch das gegenwärtige immense Wachstum an Internet-Verkehr relativiert. Es scheint sich deshalb eher der Trend in Richtung vermehrter Nutzung von Internet-Diensten abzuzeichnen. Aufgrund dessen wird die Bearbeitung eines Dienstes durch eine verteilte Dienstvermittlung und -steuerung realisiert. Dies bedeutet, dass mehrere Steuerungsrechner während der Dauer einer Kommunikationsbeziehung in die Bereitstellung eines Dienstes involviert sind. Der Austausch der notwendigen Steuerungsinformation erfolgt durch Signalisierung, was wiederum eine erhöhtes Verkehrsaufkommen in der Steuerungs-Ebene zur Konsequenz hat.<sup>47</sup>

### 3.6.1.2 Konsequenzen für eine Vermittlungssteuerung

Die Vermittlung eines Spektrums verschiedenster Dienste im B-ISDN mit unterschiedlichsten Eigenschaften erfordert den Einsatz vielfältiger Verkehrssteuerungsverfahren, deren Hauptziele darin bestehen, die angeforderte Dienstgüten einzuhalten und die Auslastung von Netzressourcen („Gewinn“) zu optimieren. Diese Problematik der Diensteintegration ist nicht B-ISDN-spezifisch, sondern war auch bereits bei Zeitmultiplexsystemen eine Herausforderung. Mit Entwicklung des N-ISDN wurden dann erweiterte Verfahren der Bandbreitenzuweisung und der Rufannahmekontrolle erforderlich (z.B. [198,326,447]). Diese Steuerungsmechanismen sind nur auf der Rufebene tätig. Eine N-ISDN-Überlastabwehr kontrolliert deshalb beispielsweise nur die Rufannahmerate.<sup>48</sup>

Im B-ISDN ist nun eine getrennte Behandlung von Ruf- und Verbindungsebene notwendig. Ein weiterer Faktor stellt die stochastische Überlagerung von Verkehrsflüssen auf Zellebene im ATM dar. Diese Anforderungen bedingen erheblich leistungsfähigere Ressourcenverwaltungen als in früheren Netztechnologien. Deutlich wird dies u. a. durch die Vielzahl von CAC-Veröffentlichungen (siehe z.B. [99]).

In [201,260,386] wird angedeutet, dass aufgrund der sehr rechenzeitintensiven Funktionen im B-ISDN, wie z.B. Routing, CAC oder Verbindungsaufbau, der Engpass in einem Netz vermutlich nicht mehr die Übertragungspfade, sondern vielmehr die Vermittlungssteuerungen sein werden.

### 3.6.1.3 Weitere Dienstaspekte

Die Vielfalt an potenziellen Signalisierszenarien spiegelt sich in der prognostizierten Vielzahl unterschiedlichster Anwendungen und Dienste des B-ISDN wider. Das Verhältnis des Ressourcenbedarfs in Nutzer- und Steuerungs-Ebene für die unterschiedlichen Dienste wird zusätzlich großen Schwankungen unterworfen sein (in Abschnitt 3.6.3 wird dieser spezielle Aspekt weiter vertieft).

---

<sup>47</sup> KANT ET AL. [288] bietet eine gute Übersicht zur Signalisierproblematik im Zusammenhang des Zusammenwachsens von Telekommunikations- und Rechnernetzen. Die Aussage, dass die Charakteristik des Signalisierverkehrs zukünftig immer komplexer wird, und damit auch die Anforderungen an die Vermittlungssteuerungen, wird durch eine Vielzahl von Argumenten belegt.

<sup>48</sup> Die Entscheidung für eine Verbindungsannahme (CAC) in einem durchschaltvermittelnden Netz beschränkt sich, im Vergleich zu B-ISDN, auf eine relativ einfache Ressourcenüberprüfung (siehe z.B. ITU-T E.731 [625]).

Auch wenn die Steuerungs-Ebene noch Lastreserven besitzt, kann es zu einer Ablehnung einer neuen Dienstanforderung aufgrund Überlast in der Nutzer-Ebene kommen, da beispielsweise die CAC gewissen Bandbreiteforderungen nicht nachkommen kann. Mit der Problematik der Ressourcenverwaltung und der Frage, ob eine neue Dienstanforderung zuerst von der Überlastabwehr als Ressourcenverwalter der Steuerungs-Ebene, oder zuerst von der CAC als Vertreter der Nutzer-Ebene analysiert wird, beschäftigt sich Abschnitt 3.6.4.1.

Weiterhin unterscheiden sich die Anteile an Vermittlungs- und Dienststeuerfunktionen für die unterschiedlichen Dienstypen. Diese Funktionalitäten benötigen zwar für ihre Bearbeitung ausschließlich steuerungstechnische Betriebsmittel, wobei allerdings der spezifische Bedarf stark variieren kann.

Ein weiterer Aspekt stellt die Tatsache dar, dass vermehrt eine Vielzahl von vermittlungstechnischen Steueraktivitäten nicht unmittelbar durch Teilnehmeranreize ausgelöst werden (z.B. *Handover*) oder sogar gänzlich *rufunabhängig* sind. Dieser vermutlich wachsende Steuerlastanteil muss bei der Überlastabwehrkonzeption berücksichtigt werden. So war beispielsweise die *Rufannahmerate* ein beliebter Lastindikator in PSTN- und ISDN-Systemen. Im B-ISDN mit seinen speziellen Verkehrsstrukturen ist dieser aus den genannten Gründen jedoch nur noch bedingt nutzbar, da sich beispielsweise aus der Lastinformation dieses speziellen Indikators keine Rückschlüsse auf die tatsächliche Belastung des Betriebsmittels Prozessor ableiten lassen.

#### **3.6.1.4 Auswirkung Internet**

Es ist bekannt, dass eine durchschnittliche Internetverbindung wesentlich länger dauert und auch sonst grundsätzlich andere statistische Eigenschaften hat als ein Standardtelefongespräch [75,353,357]. Auch die Tages- und Wochenzeitlastprofile haben eine andere Charakteristik als das klassische Benutzerverhalten [517].

### **3.6.2 Abschätzung des Ressourcenbedarfs**

Ein Regelkreis zur Last- und Leistungsregelung muss aus den regelmäßig ermittelten Messdaten der Sensoren Rückschlüsse auf die aktuelle und die zur erwartende Belastung gewinnen. Ohne eine ungefähre Vorstellung des benötigten Bedarfs an Betriebsmitteln ist eine sinnvolle Lastkontrolle nicht möglich. Diese Problematik betrifft vor allem die beiden Aspekte der Verkehrsklassifizierung und der Definition eines geeigneten Maßes für die Steuerungsleistung.

#### **3.6.2.1 Identifizierung eines bestimmten Dienstes**

Für eine genaue Lastregelung ist eine differenzierte, dienstspezifische Ressourcenbedarfsabschätzung notwendig. Aufgrund der Dienstvielfalt ist aus wirtschaftlichen und Komplexitätsgründen eine Abstrahierung in Dienstgruppen bzw. Verkehrsklassen unerlässlich. Aus diesem Grund ist eine in der Vermittlungssteuerung ankommende neue Dienstanforderung zunächst zu klassifizieren. Diese Aufgabe ist durchaus nicht trivial oder sogar teilweise unmöglich, da sie *im Netz* vorgenommen werden muss. Während im Endgerät die zugrundeliegende Anwendung sichtbar wird, werden im Netzzinneren nur die Informationen der unteren Protokollschichten ausgewertet und damit zugänglich.

Es erhebt sich also die Frage, ob aus dem Signalisierverkehr (z.B. einer SETUP-Meldung) Informationen über den dahinterstehenden Dienst bzw. dessen auslösende Anwendung extrahiert werden können. Ein erster Ansatz können die Schicht-3-Informationselemente, wie z.B. *AAL-Parameters*, *ATM Traffic Descriptor*, *QoS Parameter*, *Broadband Low Layer Information*<sup>49</sup> oder *Broadband High Layer Information*<sup>50</sup> bieten. Mittels der AAL-Parameter lassen sich die angeforderte ATM-Transportklasse<sup>51</sup> bzw. der Verbindungstyp (CBR, VBR usw.) identifizieren. Diese Teilinformation zu den Qualitätsanforderungen für die Nutzer-Ebene kann eventuell auch für eine grobe Dienstklassifizierung in der Steuerungs-Ebene herangezogen werden. Der Zugriff auf diese Informationselemente durch die Überlaststeuerung sowie eine anschließende Dienstidentifizierung und -gruppierung bedingt vermutlich nicht unerhebliche Kosten an Rechenzeit.

Anders sieht es bei einem IN-Dienst aus. Durch die Auswertung der Rufnummer ist bekannt, ob es sich um einen IN- oder Nicht-IN-Dienst handelt. Diese Information kann auch den Ausgangspunkt einer Klassifizierung bilden.

### 3.6.2.2 Nennkapazität als BHCA-Wert

In der Überlastabwehr eines SPC- bzw. allgemeiner eines PSTN-Systems wurde häufig die Rufankunftsrate  $\lambda_{CA}(t)$  als Lastindikator verwendet. Auf der anderen Seite war der BHCA-Wert  $\lambda_{BHCA}$  eine bekannte Leistungsgröße für die Nennkapazität dieses Systems. Damit war ein direkter Vergleich zwischen gemessenem und zulässigem Belastungsgrad möglich. Im einfachsten Fall bildete der BHCA-Wert den Schwellwert für die Unterteilung des Arbeitsbereiches in Normal- und Überlast.

Für das PSTN wurde ein Rahmenwerk für die Bestimmung des Nennkapazitätswertes von der ITU definiert (siehe Abschnitt 2.2.3).<sup>52</sup> Dieser Standard lässt sich nicht unmittelbar auf B-ISDN-Systeme übertragen. Grundsätzlich erhebt sich die Frage, ob die BHCA-Angabe als Leistungskennwert für einen Vermittlungsrechner überhaupt noch sinnvoll ist. Dies kann bejaht werden, wenn es um den objektiven Vergleich der Vermittlungssteuerleistung verschiedener Systeme geht. Hierbei müsste sich dann die Nennkapazitätsangabe entweder auf einen bestimmten Referenzdienst (z.B. klassische Festnetztelefonie), oder auf einen definierten Dienst- und Verkehrsmix (z.B. Erweiterung von Q.543 [648]) beziehen.

Auf der anderen Seite ist eine gewisse Vorsicht angebracht, wenn auf Basis dieses Kennwertes eine Lastregelung in einem B-ISDN-Vermittlungsrechner realisiert wird. Aufgrund der Dienstvielfalt kann aus der gemessenen Summenankunftsrate nicht unmittelbar ein Belastungswert abgeleitet werden. Soll dieser Indikatortyp verwendet werden, dann sollten im B-ISDN eine Dienstunterscheidung vorgenommen und entsprechende klassenspezifische Ankunftsrate  $\lambda_{CA,v}(t)$  ermittelt werden.

---

<sup>49</sup> Beispielsweise wird bei der Vermittlung von IP-Verkehr über ATM in dieses Feld eine entsprechende Eintragung gemacht [689]. Die Kennzeichnung eines Protokollendpunktes ist im UNI 3.1 nur in diesem Element möglich.

<sup>50</sup> Die genaue Spezifikation dieses Feldes ist derzeit noch nicht festgelegt. Vermutlich wird eine bestimmte Nummer einen bestimmten Dienst kennzeichnen, mittels derer die Endgeräte über eine Tabelle die zugrundeliegende Anwendung identifizieren. Diese Vorgehensweise wird auch bei den Dienstkennungen der Internet-Protokolle verwendet. Hier gibt es einen Satz von Nummern, die jede Anwendung – ursprünglich aus historischen Gründen, mittlerweile durch die *Internet Assigned Numbers Authority* (IANA) vergeben – so verwendet. Im Falle neuer Dienste müssen die neu vergebenen Nummern in die entsprechenden Tabellen gleich eingetragen werden – ansonsten kann nicht kommuniziert werden.

<sup>51</sup> Beispiele für die vier definierten Dienstklassen: *A* – Isochrone Dienste, Video mit konst. Bitrate; *B* – Audio, Video mit variabler Bitrate; *C* – verbindungsorientierte Datenkommunikation; *D* – verbindungslose Datenkommunikation

<sup>52</sup> Weitere Vorschläge für die Nennkapazitätsermittlung stammen z.B. von *Bellcore* [289] oder *DUTKIEWICZ* [163].

### 3.6.2.3 Ansätze für äquivalente „BHCA“-Definitionen in Breitbandnetzen

Für das ISDN war und ist die Charakterisierung der Verkehrseigenschaften für einen Teilnehmer bzw. ein Endgerät aufgrund des Dienstemixes schwierig (sogenannte „*Multiservice Traffic Characterization*“). Für Vermittlungsstellen in diensteintegrierten Netzen (sogenannte *Multiservice Nodes*) gibt es Ansätze in der Literatur für modifizierte BHCA-Definitionen. Zwei Beispiele sollen dies belegen.

GONZÁLEZ SOTO [214] macht den Vorschlag des Konzeptes der *Integrated Call Attempts* (auch „*Reference Equivalent User*“ genannt). Hierbei wird ein Dienstemix betrachtet, wobei für jeden Dienst optional selbst wiederum verschiedene Verkehrsklassen unterschieden werden können. Je nach Durchdringung (*penetration*) werden die Dienste unterschiedlich gewichtet und über alle Dienste Durchschnittsparameter ermittelt.<sup>53</sup>

MEDINA ET AL. [381] macht den Vorschlag, verschiedene, ausgewählte durchschalte- und paketvermittelte Schmalbanddienste als Basis für den BHICA-Wert<sup>54</sup> zu nehmen. Dabei werden zusätzlich die Anteile aus den beiden Teilnehmerklassen – privater oder geschäftlicher Nutzer – gewichtet einbezogen. Dieses Leistungsmaß berücksichtigt bereits die neue Dienstvielfalt.

### 3.6.3 Vergabe und Verwaltung von Ressourcen

Die Vergabe und Überwachung vermittlungstechnischer Betriebsmittel erfolgt durch die Ruf- und Verbindungsannahmesteuerung (*CAC*) und die sogenannte Ressourcenverwaltung und -allokation (siehe Abschnitt 2.6.2.1.2). Diese Funktionen basieren durch Auswertung von  $\vec{R}_U$ . In der Steuerungs-Ebene (hier  $\vec{R}_C$ ) sind für diese Aufgaben vornehmlich das Betriebssystem sowie als aufgesetzte Anwendung (siehe Bild 2-5) die Überlastabwehr verantwortlich.

Ein *neuer Ruf- oder Verbindungswunsch* (z.B. ADD CONNECTION) oder die *Modifikation* von Parametern einer bestehenden Verbindung (z.B. MODIFY REQUEST) sei durch die Anfrage  $\{\vec{r}_C^\kappa(t), \vec{r}_U^\kappa(t)\}$  bzw.  $\{\Delta\vec{r}_C^\kappa(t), \Delta\vec{r}_U^\kappa(t)\}$  charakterisiert. Parameter  $\kappa$  repräsentiere dabei eine systeminterne Variable, mit welcher alle akzeptierten Dienstanforderungen durchnummeriert werden.<sup>55</sup> Für eine optimale Dienstgüte müsste nun in jedem beteiligten Netzelement dieser Ressourcenbedarf, ohne Beeinträchtigung bestehender Kommunikationsbeziehungen, bereitgestellt werden. Beispielsweise sollte die Annahmesteuerung idealerweise überprüfen, ob folgende Bedingungen erfüllt werden können:

$$r_{C,j}^\kappa(t) + \sum_{\text{alle } i} r_{C,j}^i(t) \leq R_{C,j,max} \quad \text{für alle } j = 1, 2, \dots, m \quad \text{und} \quad (3-18)$$

$$r_{U,j}^\kappa(t) + \sum_{\text{alle } i} r_{U,j}^i(t) \leq R_{U,j,max} \quad \text{für alle } j = 1, 2, \dots, n. \quad (3-19)$$

<sup>53</sup> In einem Beispiel werden elf verschiedene ISDN-Dienste mit spezifischen Gewichtungsfaktoren betrachtet. Bestimmte Dienste werden in drei Anwendungsfällen gruppiert und beispielhafte Referenzparameter angegeben. Für die *Control Plane* und einen Vermittlungsrechner sind folgende Werte interessant: mittlere „Ruftrate“  $\lambda \approx 0.6 \dots 1.9$  BHCA und mittlere „Rufdauer“  $h \approx 120 \dots 140$ s. Dabei werden diverse Faktoren für *circuit switching* und *packet switching* unterschieden. Bestimmte Dienste (z. B. Teletex) können über beide Vermittlungstechniken abgewickelt werden.

<sup>54</sup> Der BHICA-Wert (*Busy Hour Integrated Call Attempts*) stellt einen Vektor, mit verschiedenen Komponenten wie z. B. Ruf- und Paketraten, Informationsvolumen eines Rufes bzw. Paketes u.a., dar. Die einzelnen Komponenten stellen gewichtete Mittelwerte über die verschiedenen Dienst- und Teilnehmerklassen dar.

<sup>55</sup> Diese Variable kann beispielsweise die *Call Reference* (siehe Abschnitt 2.5.5.1) sein.

Der Parameter  $i$  indiziert alle zum Zeitpunkt  $t$  existierenden Kommunikationsbeziehungen (auch *Active Calls* genannt). Der Arbeitsbereich ist durch die Grenzfläche  $\{\overrightarrow{R_{C,max}}, \overrightarrow{R_{U,max}}\}$  vorgegeben. Gleichungen (3-18, 3-19) beschreiben im Prinzip das Rahmenwerk für eine Lastplanung, da diese Bedingungen für die gesamte Dauer einer Kommunikationsbeziehung garantiert werden sollten. In der Realität lässt sich diese Zielsetzung, aufgrund vieler unbekannter Verkehrseigenschaften (alle Parameter besitzen oft stochastischen Charakter) eines Dienstes, oft nicht umsetzen. Vielmehr wird dann durch Abstrahierung der betrachteten Dienstszenarien versucht, diese durch wenige aussagekräftige Merkmale (z.B. mittlere Rufdauer) zu beschreiben.

Für die Zusage vermittlungstechnischer Betriebsmittel ist in ATM-Netzen die CAC hauptverantwortlich. Diese stellt aufgrund des statistischen Multiplexens von Verbindungen verschiedener Dienstklassen eine besondere Herausforderung dar. Viel Forschungsarbeit wurde in die Entwicklung und Untersuchung praktikabler Algorithmen in den vergangenen Jahren investiert (z.B. [98,200,207,328,425]). Einen vielversprechenden Ansatz stellen die indirekten Verfahren dar, bei welchen eine Verbindung durch ihre *effektive Bitrate*<sup>56</sup> charakterisiert wird (z.B. [279]). Auf Basis der momentanen Verkehrszusammensetzung und der zulässigen Verlustwahrscheinlichkeit muss nun die CAC eine Entscheidung, für die Annahme oder Ablehnung der neuen Ressourcenanfrage, treffen. Wesentlich dabei ist, es findet keine in die Zukunft gerichtete Lastplanung statt („die Verbindungsdauer geht nicht ein“) und die „einfache“ Charakterisierung des Ressourcenbedarfs durch nur zwei Parameter  $\vec{r}_U^K = \{r_{U,1}, r_{U,2}\}$ , mit  $r_{U,1}$  als *effektive Bitrate* und  $r_{U,2}$  als „indirektes“ Maß für Zellpufferbedarf, vorgegeben durch die garantierte Zellverlustwahrscheinlichkeit.

Um den unterschiedlichen Belastungsgrad der beiden Ebenen vergleichen zu können, beschränkt man sich sinnvollerweise auf die kritischste Ressource in jeder Ebene. Diese seien durch  $R_{U,1}(t)$  (z.B. *Bandbreite* des Ports mit der höchsten Auslastung oder *Zellpuffer* eines Koppellementes mit Zentralspeicher mit dem momentan größten Angebot) und  $R_{C,v}(t)$  (z.B. verfügbare Rechenzeit des Steuerungsprozessors mit dem momentan größten Angebot) gekennzeichnet. Als *Belastungsgrad*  $\Theta$  kann dann

$$\Theta_U(t) = \frac{R_{U,1}(t)}{R_{U,1}^{max}} \quad \text{für die Nutzer- und} \quad \Theta_C(t) = \frac{R_{C,v}(t)}{R_{C,v}^{max}} \quad (3-20)$$

für die *Steuerungsebene* definiert werden.

Auf der anderen Seite gibt es nun bezüglich der Ressourcenvergabe von steuerungstechnischen Betriebsmitteln vergleichsweise wenig an Forschertätigkeit. Dies ist hauptsächlich durch folgende Punkte begründet:

- Im normalen Lastbereich kann bei Nicht-ATM-Vermittlungssystemen üblicherweise davon ausgegangen werden, dass der potenzielle Engpass in der Nutzer-Ebene liegt (d.h.  $\Theta_C < \Theta_U$ ). Bevor es zu Betriebsmittelengpässen in der Steuerungsebene kommt, wird die Kapazitätsgrenze im Datenpfad erreicht.<sup>57</sup>
- Üblicherweise wird nur für hohe Lastzustände durch die Überlastabwehr ein „Sicherheitsventil“ vorgesehen. Dieses stellt jedoch bezüglich der Ressourcenvergabe nur eine sehr einfache und weniger planerische Funktion dar.
- Die Entwicklung von ATM-Vermittlungssystemen hat gegenwärtig die Prototypenphase verlassen. Bei dieser Systemgeneration liegt primär das Augenmerk auf einer hohen

<sup>56</sup> Auch *äquivalente* oder *virtuelle* Bitrate genannt.

<sup>57</sup> Natürlich gibt es aber auch präventive „Überlastabwehrverfahren“, die bereits im Niederlastbereich aktiv werden.

Zellvermittlungsleistung und weniger auf einem optimierten Vermittlungsrechner mit breiter Unterstützung aller Protokolltypen und -versionen bzw. voller Softwarefunktionalität.

Mittlerweile wurde erkannt, dass im B-ISDN von der Bedingung  $\Theta_C < \Theta_U$  nicht grundsätzlich ausgegangen werden kann. Durch die variierenden Verkehrszusammensetzungen aus unterschiedlichsten Diensten können sich theoretisch in einem Vermittlungssystem die Extremfälle  $\Theta_C(t) \ll \Theta_U(t)$  und  $\Theta_C(t) \gg \Theta_U(t)$  sogar permanent abwechseln. Durch eine dem Einsatzzweck angepassten Ausbau bezüglich der Leistungsfähigkeit eines Vermittlungssystems wird dem zu begegnen versucht. Trotz sorgfältigster Planung können aber diese Zustandswechsel nicht verhindert werden.

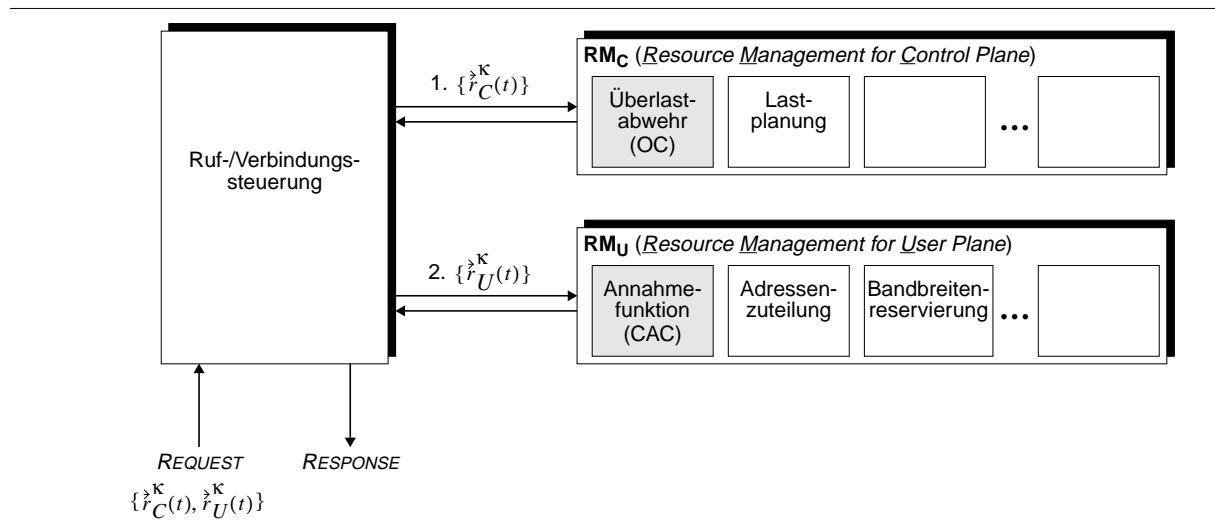
In Abschnitt 4.4 wird deshalb versucht, eine einfache Ressourcenplanung für die Steuerungsebene mit in die Lastregelung zu integrieren. Ein weiterer Aspekt – die Wechselwirkung zwischen Nutzer- und Steuerungsebene – wird im nachfolgenden Abschnitt 3.6.4 diskutiert.

### 3.6.4 Wechselwirkungen zwischen Nutzer- und Steuerungsebene

Durch verschiedene Ressourcenverwaltungen für Nutzer- und Steuerungsebene, und des unterschiedlichen Ressourcenbedarfs eines Dienstes in beiden Ebenen, können Wechselwirkungen auftreten. Diese Interaktionen sind kontraproduktiv und mindern die Vermittlungsleistung. Folgende Beispiele sollen dies dokumentieren.

#### 3.6.4.1 Ressourcenverwaltungen

Die Ressourcenverwaltungen für Nutzer- und Steuerungsebene sind in Bild 3-6 durch  $RM_U$  bzw.  $RM_C$  bezeichnet. Die CAC und die Überlastabwehr sind dabei nur stellvertretende, aber wichtige Teilfunktionen. Ein neuer Verbindungswunsch durchläuft üblicherweise zunächst die  $RM_C$ . Bei erfolgreicher Bearbeitung wird dieser dann an die  $RM_U$  weitergeleitet. Eine Anfrage zur Modifikation von Verbindungsparametern wird direkt an  $RM_U$  gesendet.



**Bild 3-6:** Ressourcenverwaltungen eines B-ISDN-Vermittlungssystems (diese Funktionsblöcke stellen nur einen Ausschnitt des Vermittlungsrechners dar).

Ein neuer, durch die Überlastabwehr bereits akzeptierter Verbindungswunsch kann immer noch nachträglich von der CAC abgelehnt werden. Dabei gilt grundsätzlich: je mehr Ressourcen angefordert werden, umso größer ist die Wahrscheinlichkeit einer Ablehnung. Diese Korrelation wird in [508] als Ansatz für einen speziellen Annahmealgorithmus verwendet. Ein Verbindungswunsch mit hohem Bandbreitenbedarf wird – wenn Bandbreite momentan nicht verfügbar – nicht sofort abgelehnt, sondern für eine gewisse Zeit gepuffert.<sup>58</sup> Dieser Vorschlag unterstützt die Überlastabwehr, indem er u. a. Rufwiederholungen entgegenwirkt.

Weiterhin ist zu überlegen, ob nicht bereits die  $RM_C$  gewisse Informationen von  $\dot{r}_U^K(t)$  in ihre Entscheidungsfindung miteinbeziehen sollte. Konkret würde dies die Auswertung von Parametern des Verkehrsvertrages bedeuten. Für diesen Ansatz spricht, dass die rechenzeitintensive Bearbeitung in der CAC bereits im Vorfeld vermieden werden kann, oder dass Rückschlüsse auf die Belastung der Nutzer-Ebene möglich sind. Auf der anderen Seite ist die Analyse von Nachrichtefeldern „problematisch“ (Rechenzeit, event. Schichtverletzung), und es ist fragwürdig, ob auf die gewünschte Information direkt zugegriffen werden kann.<sup>59</sup>

Das Kosten- zu-Nutzen-Verhältnis wird entscheiden, ob diese Problematik zukünftig in das Konzept einer Betriebsmittelverwaltung eingehen wird. Allgemein kann aber festgehalten werden, dass die Ressourcenvergabe weit komplexer ist als in Nicht-ATM-Netzen.

### 3.6.4.2 Durch aktive Nutzverbindungen induzierte Steuerungslast

Aktive Nutzverbindungen bestimmter Dienstklassen (z.B. ABR) erfordern eine permanente Parametersteuerung durch die Ressourcenverwaltung in einem Vermittlungssystem. Es wird eine lokale Last im Vermittlungsrechner durch interne Steuerungsnachrichten, ohne explizite Signalisierung, erzeugt. Folgendes Beispiel soll dies dokumentieren:

Es macht einen Unterschied für die Vermittlungssteuerung, ob z.B. für eine Datenkommunikation eine VBR- oder eine ABR-Verbindung angefordert wird. Bei einer VBR-Verbindung ist nur beim Verbindungsaufbau eine Einstellung der UPC-Parameter notwendig. Bei einer ABR-Verbindung [457,458] dagegen ist im schlimmsten Falle für jede *RM-Zelle* eine Reaktion der Vermittlungssteuerung notwendig. Der theoretisch induzierte Steuerungsverkehr kann dabei eine erhebliche Belastung für den Vermittlungsrechner darstellen.<sup>60</sup>

Die Standards [612] machen bezüglich dieser Problematik keine Vorgabe. Dieser Punkt wird vielmehr je nach Hersteller unterschiedlich gehandhabt. In den Leitungsmodulen muss auf jeden Fall jede *RM-Zelle* analysiert werden. Hierzu wird der Verkehr zyklisch (z.B. pro ms) gemessen und anschließend eine entsprechende Verbindungstabelle aktualisiert. Die zentrale Ressourcenverwaltung in der Steuerungs-Ebene wird dagegen in der Praxis vermutlich nur nach jeder *k. RM-Zelle* („*k sehr groß*“) in die Ressourcenüberprüfung involviert.<sup>61</sup> Auch in den Veröffentlichungen zu ABR wird diese Wechselwirkung zwischen Nutzer- und Steuerungs-Ebene nicht berücksichtigt und geht in die Bewertung eines Verfahrens auch nicht mit ein.

<sup>58</sup> Im Detail geht der Vorschlag auf einen erweiterten Q.2931-Protokollautomaten hinaus. Neben einem zusätzlichen Zustand, einem neuen Timer und Informationselement, wird die Nachricht CALL QUEUED eingeführt. Statt einem RELEASE wird der rufenden Instanz zunächst mit dieser Nachricht geantwortet. Für eine gewisse Zeit wird anschließend zyklisch versucht, den Verbindungswunsch erfolgreich zu bearbeiten.

<sup>59</sup> Oder muss zunächst eine aufwendige Berechnung, wie z.B. bei der effektiven Bitrate, gemacht werden.

<sup>60</sup> Z.B. ist bei ABR mit der „Explicit Rate“-Variante für jede *RM-Zelle* ein Zugriff auf die Ressourcenverwaltung notwendig. Dabei repräsentiert jede 32. Zelle eine *RM-Zelle* [236,445]. Dies bedeutet beispielsweise bei einer Verbindung mit 10 Mbit/s, dass im schlimmsten Falle grob auf 1000 *RM-Zellen* pro Sekunde reagiert werden muss. D.h. der Nutzkanal verursacht einen erheblichen knoteninternen Steuerungsverkehr, ohne dass irgendwelche Signalisierungen für diese Verbindung bearbeitet werden müssen!

<sup>61</sup> Aus den Datenblättern sind hierzu z.Zt. noch keine Angaben zu entnehmen. Dieser Parameter beeinflusst aber auf jeden Fall das ABR-Regelkreisverhalten und damit die Güte und Interoperabilität.



### 3.6.5 Steuerungsarchitektur

Wie bereits in Abschnitt 2.3.5 (bzw. 2.4.3) erwähnt wurde, zeichnet sich die ATM-Technologie durch eine durchgängige Skalierbarkeit des Netzes aus.<sup>62</sup> Die Anforderungen an die Nutzer-Ebene eines Netzknotens ergeben sich weitestgehend aus dem Einsatzumfeld, der Ausbaustufe und der Netzplanung. Die entsprechend notwendig bereitzustellende Leistung der Steuerungsebene muss auf diese Erfordernisse abgestimmt sein. Gefragt sind also Steuerungsarchitekturen, die ein hohes Maß an Flexibilität hinsichtlich ihrer Vermittlungsleistung bieten.

Steuerungsarchitekturen für leistungsfähige ATM-Systeme werden als verteilte Mehrprozessorsysteme mit einer Vielzahl lose gekoppelter Prozessoreinheiten realisiert werden [318].<sup>63</sup> Dies steht im Gegensatz zu den SPC-Steuerungen mit hohem Zentralisierungsgrad. In Anhang A.2 sind typische Steuerungsstrukturen repräsentativer ATM-Systeme dargestellt.

Dieses Architekturmerkmal erlaubt eine flexible Funktionsverteilung auf die einzelnen Prozessorknoten. Besonders stark belastete Funktionsebenen werden zusätzlich in Lastteilung betrieben. Durch das Lastteilungsprinzip kommen Lastverteil- (*Load Balancing*) bzw. Lastweiterleitungsalgorithmen (*Load Forwarding*) eine sehr wichtige Bedeutung zu. Diese Regelkreise stellen für den Vermittlungsrechner eine globale Lastregelung dar. Ergänzt wird diese Regelebene durch Überlastabwehrfunktionen, deren Wirkungskreis lokal auf einen Prozessorknoten begrenzt sind.

## 3.7 Vorschläge und bekannte Lösungsansätze

Im Zuge der Einführung diensteintegrierender Netze und der Unterstützung bestimmter Qualitätsmerkmale für verschiedene Dienstklassen sind weitergehende Ansätze notwendig. Für die Last- und Leistungsregelung betrifft dies vornehmlich folgende Aspekte:

- Die Unterscheidung in wichtige und weniger wichtige Dienste (z.B. Fernsprechen gegenüber Filmabruf oder einer Spielanwendung).<sup>64</sup> Diese Differenzierung bedingt eine unterschiedliche Behandlung unter Überlast.
- Einschränkung der Modifikation von Dienstparametern (z.B. Änderung der Bitrate beim Bildfernsprechen) unter Überlast. Die Überlastabwehr muss hier auf eine Folgemeldung in einem Szenario reagieren können.
- Einschränkung der Dienstqualität aktiver Rufe. Beispielsweise werden bestimmte Leistungsmerkmale temporär nicht mehr unterstützt.

Diese Punkte lassen sich in der Forderung zusammenfassen, dass die Dienstgütekriterien, mit welchen ein Dienst für die Nutzer-Ebene angefordert wurde (und für die auch unterschiedlich bezahlt werden muss), sich in einer differenzierten Dienstbearbeitung in der Steuerungsebene niederschlagen muss. Dies führt zu einer Priorisierung von Diensten sowie einer Priorisierung

---

<sup>62</sup> Nach [290] umfasst Skalierbarkeit folgende Netzparameter: Endgeräteanzahl pro Netzknoten, Anzahl von Netzknoten und Verbindungsabschnitten pro Netz, geographische Ausdehnung, Anzahl von Diensten, die unterstützt werden, und Größe von Datenobjekten.

<sup>63</sup> Dieselben Steuerungsarchitekturmerkmale werden in SS#7-STPs und in IN-SCPs bzw. -SDPs eingesetzt werden (müssen). Allen ist gemeinsam, dass die maximale Kapazität sich aus den Echtzeitanforderungen ergibt.

<sup>64</sup> Dabei ist zu diskutieren, wie detailliert diese Dienstdifferenzierung erfolgen soll. Ein erster Ansatz wäre beispielsweise die grobe Unterteilung in Schmal- und Breitbanddienste, IN- und Nicht-IN-Dienste oder eine Klassifizierung bezüglich der Echtzeitanforderungen.

verschiedener Meldungstypen innerhalb eines Signalisierszenarios in der Vermittlungssteuerung.<sup>65</sup>

Neben dieser mehr grundsätzlichen Anforderung soll im nächsten Kapitel ein Konzept einer Last- und Leistungsregelung für einen generischen B-ISDN-Vermittlungsrechner präsentiert werden. Aus der Literatur sind hierzu noch keine Verfahren bekannt.

Neben Vorschlägen mehr allgemeiner Natur (z.B. Überlaststeuerung durch „virtuelles Netzgeld“ von SAITO [459]<sup>66</sup>) wurden bereits Verfahren zu speziellen ATM-Systemen publiziert. Zwei Beispiele werden hierzu in Anhang A.3 näher erläutert. Bei diesen Verfahren steht aber die spezielle Steuerungsstruktur im Vordergrund – die Dienstproblematik geht nicht in diese Algorithmen ein.

---

<sup>65</sup> Daneben existiert weiterhin die „klassische“ Priorisierung verschiedener Rufstypen (z.B. hohe Priorität ankommender Rufe, Rufe mit Kurzwahlnummer (z.B. '110') oder Rufe von besonderen Anschlüssen), wie sie von verschiedenen nationalen und internationalen Standards gefordert wird [453].

<sup>66</sup> Vorgeschlagenes Verfahren zielt daraufhin ab, das Verhalten eines Teilnehmers aktiv, und somit direkt die Verkehrsquellen, zu beeinflussen. Durch dynamische Zuteilung bzw. Entzug von sogenanntem Netzgeld soll versucht werden, dem typischen Teilnehmerverhalten während einer Überlastphase – beispielsweise Wiederholungen abgelehnter Rufe – entgegenzusteuern. Das Netzgeld kann beispielsweise eine Gutschrift für die zukünftige Belegung von Netzressourcen sein. Bei erhöhter Last werden diese Ressourcen für den Benutzer scheinbar teurer. Somit findet hier eine netzweite Regelung mit direkter Rückwirkung auf die potenziellen Überlastverursacher statt.

# Kapitel 4

## Neues Konzept einer mehrstufigen hierarchischen Regelkreisstruktur zur Lastkontrolle und Leistungsoptimierung

In den beiden vorangegangenen Kapiteln wurden netz- und systemspezifische Herausforderungen an die Last- und Leistungsregelung für die Steuerungseinheit für B-ISDN-Vermittlungssysteme herausgearbeitet und diskutiert. Daraus ergeben sich vielfältige Randbedingungen für den Entwurf einer entsprechenden Regelkreisstruktur, die je nach Anwendungsumfeld mehr oder weniger stark gewichtet in den Anforderungskatalog einfließen müssen. Aufgrund dieser spezifischen Anforderungen wurde ein neues Konzept einer Überlastabwehrstrategie entwickelt, das in diesem Kapitel vorgestellt wird.

Zunächst werden alle relevanten Randbedingungen diskutiert und die betrachtete generische Steuerungsarchitektur eines B-ISDN-Vermittlungsrechners vorgestellt. Darauf setzt eine mehrstufige Regelkreisstruktur auf. Eine Reihe kooperierender Regelschleifen, die sich hinsichtlich ihres räumlichen und zeitlichen Wirkungsbereiches unterscheiden, bilden den Rahmen für diese übergeordnete Aufgabe in der Vermittlungssteuerung.

Aus dem Gesamtkonzept werden drei neuartige Ansätze ausgewählt und in den weiteren Unterkapiteln näher erläutert. Dabei handelt es sich um ein Verfahren zur präventiven Lastkontrolle, die Optimierung einer bestimmten Leistungskenngröße in einem individuellen Prozessknoten sowie die globale Lastverteilung in einer verteilten Steuerungsstruktur.

### 4.1 Randbedingungen und Merkmale

#### 4.1.1 Verkehr

In Abschnitt 3.6 wurde die Problematik der Laststeuerung für B-ISDN-Systeme detailliert ausgeführt. Die größte Herausforderung bildet dabei die Tatsache, dass der Ressourcenbedarf einer neuen Dienstanforderung weitestgehend unbekannt ist. Vorgabeinformationen für die Lastregelung sind nicht wie bei den meisten konventionellen Überlastabwehrverfahren (z.B. BHCA-Rahmenwerte) gegeben, sondern müssen aus Lastmessungen gewonnen und an die sich zeitlich ändernden Verkehrsangebote und Dienstmischungen angepasst werden. Man spricht in diesem Zusammenhang von einem adaptiven Verfahren, da sich die Führungsgrößen des Regelkreises laufend ändern. Eine effiziente Lastregelung muss den *unterschiedlichen Ressourcenverbrauch* im Vermittlungsrechner für die unterschiedlichen Rufotypen und Dienste berücksichtigen.

## 4.1.2 Vermittlungsrechner

Auf vielversprechende Konzepte für Steuerungsstrukturen für B-ISDN-Vermittlungsrechner wurde bereits in Abschnitt 3.6.5 eingegangen. Architekturmerkmale einer betrachteten Steuerungseinheit müssen in einem Lastregelkonzept berücksichtigt werden, da die Realisierung eines Lastregelkreises sehr stark von der zugrundeliegenden Steuerungsstruktur abhängt. In dieser Arbeit soll von einem generischen Vermittlungsrechner mit folgenden Merkmalen ausgegangen werden:

- Multiprozessorarchitektur als lose gekoppeltes, verteiltes System
- mehrere Funktionsebenen
- Lastteilungsprinzip unter den Prozessorknoten innerhalb einer Funktionsebene

Im Detail wird diese hierarchische Steuerungsstruktur in Abschnitt 4.2 vorgestellt. Das entsprechende Simulationsmodell mit seinen Teilkomponenten ist Gegenstand von Abschnitt 5.4.

## 4.1.3 Regelkreisstruktur

Die Struktur des Regelkreises muss auf die Systemarchitektur abgestimmt sein. Aus der hier angenommenen Steuerungsstruktur ist zunächst ein mehrstufiger Regelkreis hinsichtlich räumlicher Lokalität, d.h. räumlich begrenzter Wirkungsbereiche (*control horizon*) sinnvoll. Für bestimmte Prozessorknoten<sup>1</sup> wird deshalb auf unterster Ebene eine lokale Lastregelung vorgesehen. Eine zweite Ebene berücksichtigt den Vermittlungsrechner in seiner Gesamtheit.

Die härtesten Anforderungen hinsichtlich einer Lastregelung werden an die Zugangsvermittlungsstellen gestellt. Dieser Knotentyp befindet sich am Netzrand, d.h. zwischen Teilnehmern und Durchgangsvermittlungsstellen. Diese Art von Vermittlungsstelle soll in dieser Arbeit betrachtet werden.

Teilnehmer bzw. Endgeräte sind die originären Verkehrsquellen und -senken. Eine Überlastabwehr sollte sinnvollerweise, soweit dies protokolltechnisch überhaupt möglich ist, direkt auf die Quellen rückwirken. Wie bereits erwähnt wurde, ist die Gefahr von Rufwiederholungen bei der Auslegung einer Überlastabwehr zu berücksichtigen. Die Behandlung dieses Mitkopplungseffektes im Rahmen dieser Arbeit wird in Abschnitt 5.3 näher erläutert.

Neben der räumlichen Lokalität sollten Lastregelungen mehrere Zeitebenen unterscheiden. Der gesamte Aufbau der hier betrachteten Last- und Leistungsregelung wird in Abschnitt 4.3 vorgestellt.

## 4.1.4 Bewertungskriterien

In Abschnitt 3.4.8 wurden Eigenschaften von Überlastabwehrverfahren diskutiert. Die Vielfalt der dort angesprochenen Anforderungen und Vergleichskriterien zeigt, dass eine objektive und umfassende Bewertung verschiedener Lastregelverfahren außerordentlich schwierig ist. Gerade beim Reglerentwurf widersprechen sich darüberhinaus oft manche der gestellten Anforderungen, so dass im Einzelfall geeignete Kompromisse gefunden werden müssen. So wird ein

---

<sup>1</sup> Diejenigen, die potenzielle Systemengpässe darstellen.

komplexer Algorithmus, der auf einer genauen Dienstcharakterisierung basiert, einen höheren Nutzdurchsatz erzielen als ein einfaches Verfahren, das nur wenige Parameter berücksichtigt.

Weitere Probleme beim Vergleich mit bekannten Verfahren aus der Literatur ergeben sich daraus, dass diese oft andere Strategien zum Ziel haben und teilweise auch von anderen Randbedingungen ausgehen. Beispielsweise wird in den meisten Untersuchungen von Überlastabwehrverfahren für B-ISDN-Systeme noch von einem klassischen Rufverkehr ausgegangen. Die Dienstvielfalt mit ihren Auswirkungen in der Steuerungs-Ebene wird dort vernachlässigt.

Ein weiterer Punkt für die Bewertung unterschiedlicher Verfahren ist, dass notwendigerweise für einen quantitativen Vergleich alle relevanten Systemparameter vorliegen müssen. Bei Leistungsuntersuchungen von Algorithmen für die Nutzer-Ebene sind diese üblicherweise vollständig bekannt (z.B. Linkgeschwindigkeit, Abfertigungsdisziplinen oder Prioritätsmechanismen). Im Gegensatz hierzu werden absolute Leistungsangaben für die Steuerungs-Ebene in Untersuchungen so gut wie nie angegeben. Gerade Bedienzeiten für Signalisierverkehr oder der Rufdurchsatz werden fast immer normiert aufgeführt.<sup>2</sup> Die Bearbeitungszeiten haben aber einen nicht unwesentlichen Einfluss auf die Parameter eines Lastregelkreises (z.B. Messintervalldauer oder Schwellwerte für Warteschlangen).

Aus den genannten Gründen ist ein quantitativer Vergleich des hier vorgestellten Rahmenwerkes für eine Lastregelung mit bekannten Verfahren schwierig bzw. nur in wenigen Punkten überhaupt möglich. Die hier verwendeten Bewertungskriterien werden deshalb im Einzelfall für die jeweiligen Messungen und Leistungsuntersuchungen in den entsprechenden Kapiteln angesprochen.

## 4.2 Generische Steuerungsarchitektur

Aus Sicht eines gesamten Vermittlungssystems (im Sinne von Bild 2-19) hat der betrachtete Vermittlungsrechner zentralen Charakter. Dessen Systemstruktur selbst besteht aus einem skalierbaren Multiprozessorsystem, wobei sich sowohl der Funktions- als auch der Lastteilungsgrad flexibel variieren lassen sollen. Die Konfiguration eines Prozessorknotens – und damit die Zuordnung zu einer bestimmten Funktionsebene – erfolgt durch Laden der entsprechenden Software und liegt somit statisch fest.

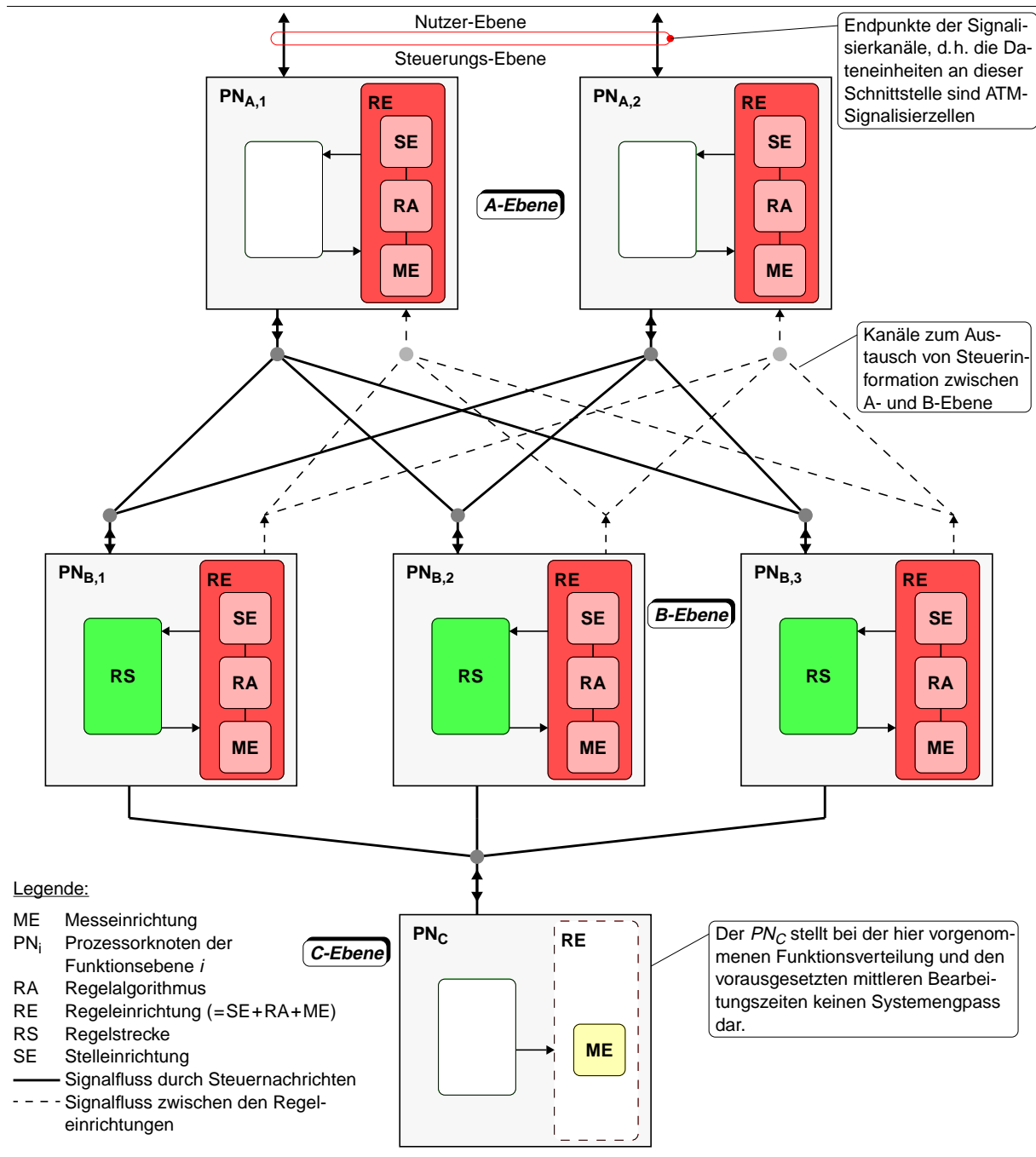
Durch diese prinzipiellen Merkmale lassen sich viele verschiedene Steuerungsstrukturen konfigurieren (generische Architektur).

### 4.2.1 Übersicht

Eine konkrete Steuerungsarchitektur mit sechs Prozessorknoten als Verarbeitungseinheiten ist in Bild 4-1 dargestellt. Jeder Knoten besitzt prinzipiell einen lokalen Regelkreis, welcher auf oberster Ebene aus der zu regelnden Strecke und der zugehörigen Regeleinrichtung besteht. In einem Prozessorknoten können lokal mehrere, parallel wirkende Regelkreise implementiert sein, wobei sinnvollerweise die zugehörigen Regelobjekte disjunkt sein sollten. Eine Regeleinrichtung selbst enthält allgemein eine Mess- und Stelleinrichtung, die beide durch einen Regelalgorithmus verknüpft sind.

---

<sup>2</sup> Dies hat herstellerepezifische Gründe. In einem Datenblatt eines B-ISDN-Knotens fehlen entweder diese Angaben oder die aufgeführten Zahlen gelten für eine zukünftige Produktversion.



**Bild 4-1:** Generische Steuerungsarchitektur des zentralen Vermittlungsrechners –  $(2, 3, 1)$ -Struktur mit Lastzuteilung durch die A- an die B-Prozessoren

Verschiedene Regeleinrichtungen können miteinander kommunizieren, d.h. Last- und Leistungsinformationen können prinzipiell lokal, zwischen verschiedenen PN einer Ebene oder zwischen verschiedenen PN unterschiedlicher Ebenen ausgetauscht werden. Somit lassen sich auch übergeordnete bzw. hierarchische Regelkreise definieren. Ein anderer Aspekt bildet die Tatsache, dass sich die verschiedenen Komponenten einer Regeleinrichtung räumlich verteilen lassen. Beispielsweise ist das der Fall bei Verwendung eines Lastindikators der B-Ebene und

der Realisierung der Aktorik in einem  $A$ -Prozessor. In diesem Beispiel befindet sich die Messeinrichtung in einem  $PN_{B,j}$  und die Stelleinrichtung in einem  $PN_{A,i}$ .

Die Grauschattierung der Regelstrecken in Bild 4-1 soll andeuten, dass die PN der  $B$ -Ebene hier das kritischste Betriebsmittel darstellen. Dadurch sind diese das Zielobjekt der nachfolgend beschriebenen Regelkreise.

## 4.2.2 Gewählte Funktionsverteilung

Die im Rahmen dieser Arbeit betrachtete Steuerungsarchitektur besitzt drei Funktionsebenen. Nach der in Abschnitt 2.6.5 eingeführten Notation liegt damit ein  $(A, B, C)$ -Vermittlungsrechner vor. Die Aufgabenzuordnung auf die einzelnen Funktionsebenen soll ebenfalls der dortigen Aufteilung entsprechen, d.h. die sogenannte niederere Signalisierung wird durch die  $PN_{A,i}$  bearbeitet; höhere Signalisierung, Dienstvermittlungs- und Dienststeuerfunktionen sind Aufgaben der  $PN_{B,j}$ ; Funktionen mit stark zentralem Charakter (z.B. Verbindungsannahmesteuerung, zentrale Ressourcenverwaltung oder Datenbanksystem) sind im  $C$ -Knoten lokalisiert.

Stellt der  $PN_C$  einen Systemengpass dar, dann müssen dessen Aufgaben ebenfalls auf mehrere Verarbeitungseinheiten verteilt werden. Da diese Funktionen auf zentrale Daten zugreifen, sind entsprechend leistungsfähige, verteilte Algorithmen notwendig (z.B. verteilte CAC). Diese Problematik ist aber nicht Gegenstand dieser Arbeit und wird deshalb nicht weiter vertieft.<sup>3</sup>

## 4.2.3 Merkmale der Lastverteilung

ATM-Zellen mit Signalisierinformation werden aufgrund ihrer speziellen VCI/VPI-Kennung durch das Koppelnetz an die  $A$ -Ebene geroutet. An den  $PN_{A,i}$  enden deshalb die Signalisierungskanäle des  $UNI$  und  $NNI$ . Da mehrere  $A$ -Prozessoren nach dem Lastteilungsprinzip arbeiten, muss definiert werden, wie der ankommende Signalisierverkehr auf die verschiedenen  $PN_{A,i}$  verteilt wird. Die einfachste Möglichkeit ist eine *statische Zuordnung* eines bestimmten  $PN_{A,i}$  zu einer Gruppe von Leitungsmodulen. Durch Variation der Gruppengröße kann beispielsweise der Teilnehmersignalisierung weniger  $A$ -Prozessorkapazität als der  $NNI$ -Signalisierung zur Verfügung gestellt und damit ein erster Priorisierungsmechanismus für eine globale Lastregelung installiert werden. Weitere Strategien sind beispielsweise *Round-Robin* oder eine zufällige Zuteilung.<sup>4</sup> In [410] wird hierzu ein hybrider Ansatz beschrieben.

Aufgrund der Funktionsverteilung wird davon ausgegangen, dass die  $A$ -Ebene hier keinen Systemengpass bildet. Deshalb wird auch die Lastverteilung an dieser Zell-Schnittstelle im Rahmen dieser Arbeit nicht weiter betrachtet. Auch wird eine Lastweiterleitung innerhalb dieser Ebene ausgeschlossen.

Die nächste Regelkreisebene bezieht sich auf die Lastweiterleitung von einem  $A$ - an einen bestimmten  $B$ -Prozessorknoten. Nach einer bestimmten Strategie muss die  $A$ -Ebene neue Anforderungen auf die  $PN_{B,j}$  verteilen. Entsprechend der Funktionszuordnung gibt ein  $PN_{A,i}$  eine Schicht-2-Signalisierdateneinheit einer neuen Dienstanforderung an die  $B$ -Ebene weiter, wo dann die Schicht-3-Signalisierung stattfindet bzw. das anwendungsorientierte Protokoll (z.B.

---

<sup>3</sup> Alternativ können die Aufgaben der  $C$ -Ebene auch von den  $PN_{B,j}$  erbracht werden. Damit entfällt eine Funktionsebene (siehe auch Anhang A.2.1.5 mit einer entsprechenden Steuerungsarchitektur).

<sup>4</sup> Die zufällige Verteilung wird dabei bereits von modernen Mehrpfad-Koppelnetzen unterstützt. Dadurch wird eine homogene Belastung des Koppelnetzes, gerade bei inhomogenen Verkehrsangeboten, angestrebt.

B-ISUP) terminiert wird. Diese Art von Lastweiterleitung ist Gegenstand dieser Arbeit (siehe Abschnitt 4.6).

Nicht betrachtet wird dagegen eine Lastweiterleitung innerhalb der *B*-Ebene. Solch ein Algorithmus des *Load Balancing* wurde für das spezielle Anwendungsumfeld eines Einklassensystems in [603] patentiert.

## 4.3 Regelungstechnischer Ansatz – Mehrstufige Regelung

Die Notwendigkeit eines regelungstechnischen Ansatzes ergibt sich aus der Tatsache, dass Vermittlungsrechner hochgradig dynamische Systeme repräsentieren und deren Leistungsverhalten durch das Zusammenwirken von Konfiguration (Steuerungsstruktur, Funktionsaufteilung usw.) und aktueller Last bestimmt wird. Während die Konfiguration weitgehend statisch ist<sup>5</sup>, treten bei der Last, aufgrund der stochastischen Verkehrseigenschaften, meist relativ starke zeitliche Schwankungen auf. Regelkreise bieten den Vorteil, dass sie für die Fälle, in denen das Verhalten des Rechnersystems im Voraus nicht bekannt ist, mit ihren rückgekoppelten Wirkungskreisläufen auf die dynamischen Schwankungen Einfluss nehmen können [225].

Wie bereits erwähnt, lassen sich Regelkreise hinsichtlich verschiedener Merkmale klassifizieren. Für die Last- und Leistungsregelung in einem Vermittlungsrechner sind dabei der räumliche und zeitliche Wirkungskreis die wichtigsten Klassifizierungsaspekte. Beide werden in den nächsten Unterkapiteln weiter diskutiert. Als drittes Merkmal kann der Informationstyp, auf den der Regelkreis wirkt, herangezogen werden.<sup>6</sup>

### 4.3.1 Räumliche Ebenen

Mindestens zwei sinnvolle räumliche Ebenen lassen sich für eine mehrstufige Regelkreisstruktur einer Mehrprozessorarchitektur identifizieren, die folgendermaßen bezeichnet werden sollen:

- *Lokale* Regelung – Wirkungskreis ist auf einen Prozessorknoten begrenzt
- *Global* Regelung – Wirkungskreis ist auf den zentralen Vermittlungsrechner begrenzt

Die lokalen Regelkreise werden in den am stärksten belasteten Recheneinheiten installiert. Im vorliegenden Fall sind dies die *B*-Prozessoren. Für die Realisierung einer globalen Regelung gibt es verschiedene Möglichkeiten. Für die betrachtete Steuerungsstruktur werden Regelstrecke und Regelungseinrichtungen auf verschiedene Funktionsebenen verteilt. Dabei stellt die *B*-Ebene das Regelobjekt dar. Die Regeleinrichtung, d.h. Sensorik, Aktorik und Regelfunktion, wird räumlich verteilt durch die *A*-Prozessoren erbracht. In Abschnitt 4.6 wird diese globale Regelkreisstruktur näher ausgeführt.

---

<sup>5</sup> Skalierbare B-ISDN-Systeme erlauben eine flexible Konfiguration (z.B. durch zusätzliche Einschubkarten) bezüglich der bereitgestellten Steuerleistung. Dadurch hat auch die Steuerungsstruktur prinzipiell dynamischen Charakter. Dieser Aspekt wird im Rahmen dieser Arbeit nicht weiter verfolgt.

<sup>6</sup> Ein Regelkreis reagiert beispielsweise nur auf einen bestimmten *Meldungstyp* (z.B. auf SETUPS und nicht auf IAMs), bestimmte Dienste oder bestimmte *Verkehrsklassen*.



### 4.3.2 Berücksichtigte Zeitebenen

Hinsichtlich *zeitlicher Lastschwankungen* lassen sich verschiedene Bereiche unterscheiden (Tabelle 4-1). Während mittel- und langfristige Lastschwankungen in den Bereich des Netzmanagements fallen, sind kurzfristige Schwankungen Angriffspunkt lokaler Lastregelungen.

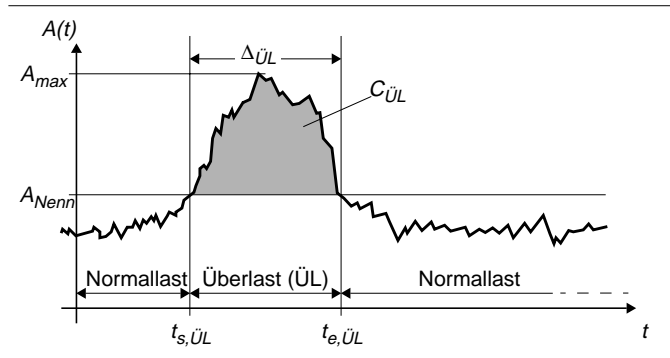
Typus	Zeitbereich	Ursachen, sonstige Bemerkungen
kurzfristig	ms ... min	Zufallscharakter des Verkehrs (→ stochastische Prozesse), nicht voraussagbar
mittelfristig	min ... h	„Tageszeitkurve“, eher voraussagbar und deshalb auch planbar
langfristig	> d	z.B. Einführung neuer Dienste, Tarifpolitik, geändertes Nutzerverhalten

**Tabelle 4-1:** Typische Zeitbereiche bei Lastschwankungen

Speziell aus Sicht der Überlastabwehr ist eine weitere Verfeinerung des Zeitbereiches für kurzfristige Lastschwankungen sinnvoll. Überlastsituationen lassen sich hinsichtlich ihrer „Intensität“ und Dauer unterscheiden (Bild 4-2).

Eine Vermittlungssteuerung (als Monoprocessorsystem) befindet sich in Überlast wenn das aktuelle *Angebot*  $A(t)$  größer als die *Nennkapazität*  $A_{Nenn}$  ist, d.h.  $A(t) > A_{Nenn}$  wobei  $A_{Nenn} < 1$  ist.

Beispielsweise beträgt die angegebene Nennkapazität oft nur 80% der maximalen Systemkapazität. Der *Überlastanteil*  $A_{ÜL}(t)$  ergibt sich dann zu



**Bild 4-2:** Überlastimpuls  $A_{ÜL}(t)$  der Dauer  $\Delta_{ÜL}$  und der Verkehrsmenge  $C_{ÜL}$

$$A_{ÜL}(t) = A(t) - A_{Nenn} \quad (4-1)$$

Ein *Überlastimpuls* habe nun die *Dauer*  $\Delta_{ÜL}$

$$\Delta_{ÜL} = t_{e,ÜL} - t_{s,ÜL} \quad (4-2)$$

und besitze den *Überlastfaktor*  $\zeta_{ÜL}$

$$\zeta_{ÜL} = \frac{A_{ÜL, max}}{A_{Nenn}} = \frac{A_{max} - A_{Nenn}}{A_{Nenn}} = \frac{A_{max}}{A_{Nenn}} - 1 \quad (4-3)$$

Ohne Überlastabwehr würde sich in einem reinen Wartesystem die Verkehrsmenge  $C_{ÜL}$

$$C_{ÜL} = \int_{t_{s,ÜL}}^{t_{e,ÜL}} A_{ÜL}(t) dt \text{ mit Einheit [Erlh]} \quad (4-4)$$

anhäufen. Prinzipiell kann eine Überlast in folgenden beiden Extremformen, bei gleicher Verkehrsmenge  $C_{\ddot{U}L}$ , auftreten:

- (1) Schwache, aber relativ lang andauernde Überlast ( $\zeta_{\ddot{U}L}$  sehr klein,  $\Delta_{\ddot{U}L}$  sehr groß),
- (2) Starke, kurze Lastspitze ( $\zeta_{\ddot{U}L}$  sehr groß,  $\Delta_{\ddot{U}L}$  sehr klein).

Beide Varianten sollen dabei Ausprägungen einer *kurzfristigen Lastschwankung* (z.B.  $\Delta_{\ddot{U}L,(1)}$  mehrere Minuten,  $\Delta_{\ddot{U}L,(2)}$  wenige Sekunden) sein und fallen deshalb in dieselbe, oben definierte Kategorie. Während auf Überlastimpuls (2) sehr schnell reagiert werden muss, sind die Zeitanforderungen an Typ (1) weniger hart. Aufgrund dieser Randbedingung werden hier zwei getrennte Regelkreise für die Überlastabwehr eingeführt, die im folgenden als *schnelle* bzw. *retardierende* Überlastabwehr bezeichnet werden.<sup>7</sup>

Damit erfolgt in dieser Arbeit eine *Beschränkung auf drei Zeitebenen*, die nachfolgend noch einmal zusammengestellt sind:

- präventiv wirkend
  - *präventive* Lastregelung (*preventive control*)
- reaktiv wirkend
  - *schnelle* Überlastabwehr (*immediate reactive control*) für stoßartige Lasten
  - *retardierende* Überlastabwehr (*retard reactive control*)

Da sich die einzelnen Zeitebenen nicht strikt voneinander trennen lassen, überlappen sie sich teilweise. Praktisch bedeutet dies, dass gleichzeitig mehrere Regelkreise aktiv sein können. Die oben aufgeführten Regelkreise werden in nachfolgendem Abschnitt konkretisiert.

### 4.3.3 Lastebenen

Betrachtet man den Verlauf einer Leistungskenngröße als Funktion einer bestimmten Lastkenngröße, dann ist es aus Sicht einer Regelung sinnvoll, den Wertebereich der Lastgröße in verschiedene Arbeitsbereiche, auch Lastebenen genannt, aufzuteilen. Im Rahmen dieser Arbeit erfolgt eine allgemeine Aufteilung in die Bereiche *Normal-*, *Hoch-* und *Überlast*.<sup>8</sup> Die konkreten Bereichsgrenzwerte hängen dabei vom jeweiligen Regelkreistyp ab.

### 4.3.4 Verfahren zur Last- und Leistungsregelung

Die Motivation für ein mehrstufiges Regelkreiskonzept wurde bereits in den vorangegangenen Abschnitten erläutert. Im Folgenden soll nun eine Übersicht über die einzelnen Regelkreise für eine zugrunde liegende generische Steuerungsstruktur gegeben werden. Eine ausführlichere Begründung erfolgt bei der Diskussion der Algorithmen in den jeweiligen Unterkapiteln.

In Tabelle 4-2 sind die einzelnen Regelkreise für die unterschiedlichen Zeit- und Raumebenen zusammengestellt. Die Aufstellung gibt nur eine kleine, begrenzte Auswahl von Algorithmen wieder und kann bei weitem nicht vollständig sein. So wäre eine Vielzahl weiterer Verfahren denkbar. Aus den beschriebenen Randbedingungen und Anforderungen kommt aber den folgenden Mechanismen (grau schattiert in Tabelle 4-2) hier eine besondere Bedeutung zu:

<sup>7</sup> In realen System gibt es Implementierungen, die eine noch feinere Aufteilung vornehmen (z.B. [470]).

<sup>8</sup> Gegebenenfalls fallen Hoch- und Überlastbereich zusammen, so dass in diesem einfachen Fall nur zwei Arbeitsbereiche unterschieden werden.

- *Annahmeverfahren für neue Dienstanforderungen*<sup>9</sup>

Dieses Annahmeverfahren für neue Dienstanforderungen (*NRAC, New Request Admission Control*) basiert auf einer Bedarfsabschätzung steuerungstechnischer Betriebsmittel. Während eine *CAC* nur Ressourcen der Nutzer-Ebene berücksichtigt, hat diese neue Methode ihren Anwendungskreis in der Steuerungs-Ebene. Da aus der Literatur hierzu nichts Vergleichbares bekannt ist, wird dieses vielversprechende Verfahren am detailliertesten untersucht (siehe Abschnitt 4.4).

		Raumebene ( <i>spatial control horizon</i> )	
		Lokale Regelung	Globale Regelung
Zeitebene ( <i>time control horizon</i> )	präventive Lastregelung	<b>Annahmekontrolle für neue Dienstanforderungen</b> <i>(NRAC, New Request Admission Control)</i>	<b>Lastzuteilung</b> <i>(LFAB, Load Forwarding from A- to B-Level)</i>
		<i>Verbindungsannahmesteuerung</i> <i>(CAC, Connection Admission Control)</i>	
	schnelle Überlastabwehr	<b>Behandlung neuer Dienstanforderungen</b> <i>(HNR, Handling of New Requests)</i>	<b>Lastweiterleitung</b> <i>(BLB, B-Level Load Balancing)</i>
		<b>Lastvertagung nichtvermittlungstechnischer Aufträge</b> <i>(DNSJ, Deferment of Non-Switching Jobs)</i>	<b>Dynamische Flusskontrolle der Schicht-2-Signalisierung</b> <i>(FCL2S, Flow Control of Layer-2-Signalling)</i>
	retardierende Überlastabwehr	<b>Deaktivierung von Dienst- und Leistungsmerkmalen</b> <i>(DSS, Deactivation of Supplementary Services)</i>	<b>Interne Routingstrategie für neue Signalisierendpunkte</b> <i>(RSISE, Routing Strategy for Internal Signalling Endpoints)</i>

Prozessorknoten
zentraler Vermittlungsrechner

**Tabelle 4-2:** Mehrstufiges Regelkreiskonzept (Regelkreise mit „grauer Schattierung“ bilden Schwerpunkt dieser Arbeit; Rest: typische Beispiele von Lastregelverfahren für entsprechenden Anwendungskreis<sup>10</sup>)

- *Behandlung neuer Dienstanforderungen* (siehe Abschnitt 4.5)

Diese Funktion ist für einen schnell wirksamen Überlastschutz eines Prozessorknotens notwendig. Neben der Ablehnung oder Verwerfung einer Signalisiernachricht kommt speziell im B-ISDN der Dienstbehandlung eine wichtige Bedeutung zu. Hierunter wird verstanden, dass der Signalisierverkehr im Vermittlungsrechner bewusst mit *unterschiedlicher Dienstgüte* bearbeitet wird. Folgende *Eingriffsmöglichkeiten* (Stellgrößen) ver-

<sup>9</sup> Dienstanforderungen (REQUESTS) können sich allgemein auf eine *Verbindung*, einen *Ruf* oder eine *Sitzung* beziehen.

<sup>10</sup> Anmerkung zur *Verbindungsannahmesteuerung*: Dieser Mechanismus basiert als einziger nicht auf Ressourcen-Maßzahlen der *Control Plane* (aus diesem Grund wurde der Eintrag gestrichelt umrahmt). Die *CAC* wird hier aufgeführt, da zum einen sich ihre Entscheidungen indirekt auf den Bedarf steuerungstechnischer Betriebsmittel auswirkt und zum anderen sie mit das wichtigste präventive, lokale Lastkontrollverfahren ist.

deutlichen dies:

- Unterschiedliche (statische und dynamische) Prioritäten für verschiedene Verkehrsklassen
  - Bevorzugung „älterer Dienste“<sup>11</sup>
  - bewusstes Verzögern (innerhalb der durch die Signalisierprotokolle vorgegebenen Zeitschranken) von Antwortmeldungen
  - Optimierungsfunktion basierend auf einem speziellen Leistungsmaß (z.B. *Power-Produkt*), das bereits aufgrund seiner Definition den Dienstaspekt unterstützt
- *Lastzuteilung* (siehe Abschnitt 4.6)  
Hierunter wird die Strategie für die Zuteilung einer neuen Dienstanforderung von einem A- an einen B-Prozessorknoten verstanden. Die Strategie orientiert sich dabei an gewissen Kriterien. Ein besonderes Merkmal stellt der hier völlig verteilte Ansatz dar, d.h. es gibt keine übergeordnete, zentrale Instanz, die die Lastverteilung koordiniert.

Diese Auswahl an Lastregelverfahren bildet den Schwerpunkt dieser Arbeit. Folgende, eventuell ergänzende Verfahren werden hier nicht näher untersucht, da sie keine grundsätzlichen Neuerungen darstellen:

- *Lastvertagung nichtvermittlungstechnischer Prozesse*  
Dieser Mechanismus ist ein bewährtes Überlastabwehrverfahren („*Deferable Work*“) und wird auch weiterhin eingesetzt. Die zugrundeliegende Softwarearchitektur und das Betriebssystem haben einen wesentlichen Einfluss auf das Potenzial dieses Verfahrens.
- *Lastweiterleitung*  
Ein eigenständiges *Load Balancing* zwischen den Prozessorknoten der B-Ebene kann sich bei sorgfältiger Auslegung positiv auf das Leistungsverhalten auswirken, gerade auch bei den Unwägbarkeiten im B-ISDN-Steuerungsverkehr. Aufgrund des oben vorgesehenen Lastzuteilungsverfahrens wäre hier aber eine zusätzliche Lastweiterleitung aus wirtschaftlichen Gründen nicht vertretbar.
- *Temporäre Deaktivierung von Dienst- und Leistungsmerkmalen*  
Unter extremen Lastbedingungen wird nur noch eine gewisse Basisfunktionalität bereitgestellt. Aufgrund der zunehmenden Dienstkomplexität und des immer breiter werdenden Dienstespektrums gewinnt dieser Mechanismus, gerade auch im B-ISDN-Umfeld, immer mehr an Bedeutung.
- *Dynamische Flusskontrolle der Schicht-2-Signalisierung*  
Die Fenstergröße einer SSCOP-Verbindung kann dynamisch verändert werden (siehe Abschnitt 2.5.4.1). Durch diese Maßnahme ließe sich prinzipiell der Signalisierverkehr an der Teilnehmer-Netz-Schnittstelle (UNI) durch Drosselung der Endsysteme reduzieren.
- *Interne Routingstrategie für neue Signalisierendpunkte*  
Alle Nachrichten einer Signalisierverbindung werden im betrachteten Vermittlungssystem zu gleichen A-Prozessorknoten geroutet. Wie bereits in Abschnitt 4.2.3 diskutiert wurde, gibt es bei dieser Steuerungsarchitektur mehrere Möglichkeiten der Vergabe der Signalisierendpunkte. Ein sinnvolles Routingverfahren könnte beispielsweise nach fol-

---

<sup>11</sup> An einem Prozessorknoten kommen die REQUEST-Meldungen sequentiell, d.h. zu unterschiedlichen Zeitpunkten an. Für jeden akzeptierten Dienst wird ein Datensatz angelegt und eine Referenznummer ( $\rightarrow$  *Call Reference*, siehe auch 2.5.5.1 bzw. 2.5.5.2, S. 29ff) vergeben. Diese wird üblicherweise sequentiell hochgezählt (Modulo-Operation). Es sei nun angenommen, dass die beiden (in Bearbeitung befindlichen) Dienste die Referenzen  $CR_i$  und  $CR_j$  besitzen, wobei  $CR_i < CR_j$  sei. Für die Ankunftszeitpunkte der entsprechenden Dienstanforderungen gilt dann  $t_{A,i} < t_{A,j}$ , d.h. Dienst Nr.  $i$  ist hier der „Ältere“. Die Realisierung dieser Bedienstrategie bedingt eine relativ aufwendige Pufferverwaltung aufgrund rechenzeitintensiver Listenoperationen.

gender Strategie verfahren: Von den  $K$  Prozessorknoten der A-Ebene seien  $\nu$  für UNI- und  $K-\nu$  für NNI-Signalisierendpunkte reserviert. Durch dynamische Veränderung des Parameters  $\nu$  kann die verfügbare Bearbeitungskapazität für Schicht-2-Signalisierung variiert und damit indirekt eine Last- und Leistungsregelung vorgenommen werden.<sup>12</sup>

Wie bereits oben erwähnt, ist noch eine Reihe weiterer, vielversprechender Ansätze denkbar. Die getroffene Auswahl stellt deshalb keine grundsätzliche Wertung für diesen Kreis von Lastregelverfahren dar. In den folgenden Abschnitten werden die im Rahmen dieser Arbeit untersuchten Algorithmen näher erläutert.

Durch die getroffene Auswahl liegt der Schwerpunkt auf der präventiven Last- und Leistungsregelung, da gerade hier der Dienstaspekt besonders zum Tragen kommt. Bei der schnellen Überlastabwehr als Reaktion auf stoßartige Lasten kommt den konventionellen „Sicherheitsventilen“ aber immer noch eine wichtige Bedeutung zu. Damit rückt auch der Aspekt bezüglich *Repeated Call Attempts* im Rahmen dieser Arbeit in den Hintergrund.

## 4.4 Annahmekontrolle für neue Dienstanforderungen

Im folgenden wird ein neues Konzept für eine präventive Lastregelung vorgestellt. Diese sogenannte *Annahmekontrolle für neue Dienstanforderungen (NRAC, New Request Admission Control)* ist Teil der Ressourcenverwaltung für steuerungstechnische Betriebsmittel und wird als parameteradaptive Regelung<sup>13</sup> realisiert.

### 4.4.1 Problematik und Anforderungen

Vor der Annahme einer neuen Dienstanforderung findet auf Basis des Verkehrsvertrages eine Überprüfung der vermittlungstechnischen Betriebsmittel durch die CAC statt. Ein vergleichbarer Mechanismus für die Steuerungs-Ebene existiert derzeit noch nicht, obwohl Verfahren dieser Art speziell für B-ISDN-Systeme durchaus ihre Berechtigung hätten. Während die CAC vornehmlich das Leistungsverhalten des Datenpfades positiv beeinflusst, kann durch ein adäquates Verfahren für den Kontrollpfad die Systemleistung zusätzlich gesteigert werden.

Die CAC für die Nutzer-Ebene ist in ITU-T I.371 [644] definiert. Es wird aber kein bestimmter Algorithmus spezifiziert. Vielmehr sind viele Lösungen denkbar und die Entscheidung, welches Verfahren implementiert wird, bleibt letztlich dem Hersteller oder Netzbetreiber überlassen.<sup>14</sup> Die Grundlage jeder CAC-Entscheidung ist der Verkehrsvertrag, mit dem eine Quelle ihren Verbindungswunsch charakterisieren muss. Je nach Algorithmus gehen verschiedene Parameter des *Traffic Contracts* in die Berechnung mit ein. Dies ist der entscheidende Unterschied. Einem Annahmeverfahren für die *Control Plane* fehlt diese Informationsbasis! Ein Algorithmus für die Steuerungs-Ebene kann deshalb zwangsläufig nicht so genau eine Lastabschätzung, und damit eine Planung des Ressourcenbedarfs, vornehmen wie eine CAC. Deshalb

<sup>12</sup> Eine Verkleinerung von  $\nu$  unter Überlast ist beispielsweise eine denkbare Strategie. Damit wird den vom Netz kommenden Signalisiernachrichten (auf Kosten der UNI-Meldungen) mehr Rechenzeit zur Verfügung gestellt. Der Parameter  $\nu$  repräsentiert somit ein grobes Prioritätsmaß für die beiden Verkehrsklassen.

<sup>13</sup> *Adaptive* Regelung: Regler, die ihr Verhalten dynamisch den sich ändernden Eigenschaften der zu regelnden Prozesse und deren Signale anpassen [29].

<sup>14</sup> In ITU E.736 [628] werden Vorschläge für eine Reihe praktikabler CAC-Verfahren gegeben. Drei prinzipielle Betriebsarten werden unterschieden: (1) *Multiplexing of constant bit-rate streams*, (2) *Rate envelope multiplexing* und (3) *Rate sharing statistical multiplexing*.

stellt hier das für die Steuerungs-Ebene konzipierte Annahmeverfahren einen *präventiven Lastregelmechanismus* dar.

Allgemein ergeben sich die Anforderungen für ein Annahmeverfahren – ob nun für die Nutzer- oder die Steuerungs-Ebene – aus folgenden Randbedingungen:

- Welche *Ressourcen* eines Vermittlungssystems gehen in die Berechnung ein?
- Welche *Verkehrsparameter* für die Charakterisierung eines Dienstes sind verfügbar bzw. müssen abgeschätzt werden?
- Welche *Dienstgüteparameter* gehen in die Berechnung ein und welche sollen garantiert werden?

Die *Ressourcenproblematik* wurde bereits in den Abschnitten 3.2 und 3.6.4.1 ausführlich erläutert. Für einen ersten Ansatz ist es ausreichend, sich auf wenige Betriebsmitteltypen zu beschränken und beispielsweise nur die (vermutlich) kritischste Ressource zu betrachten. Die wichtigste Ressource der Steuerungs-Ebene ist in der Regel die einem Prozess zugewiesene *Rechenzeit*. Die CPU stellt somit das zentrale Betriebsmittel eines Prozessorknotens dar. Die Charakterisierung eines Dienstes soll deshalb auf dieser Komponente basieren. Damit entspricht der Ressourcenbedarf eines Dienstes der Summenbelegungszeit aller zugehörigen Nachrichten während der gesamten Dienstdauer. Das nachfolgend vorgestellte Verfahren berücksichtigt deshalb den CPU-Bedarf eines Dienstes, der sich in der zeitabhängigen Auslastung  $\rho(t)$  niederschlägt.

Weitaus problematischer ist die *Dienstcharakterisierung*. Der Ressourcenbedarf eines durchschnittlichen Dienstes bzw. Rufes ist bei weitem nicht mehr so gut abschätzbar wie im analogen Fernsprechnetze oder ISDN (siehe beispielsweise Abschnitt 5.2.2). Damit entfällt im B-ISDN die wichtigste Planungsgrundlage, da die meisten bekannten Nicht-B-ISDN-Verfahren<sup>15</sup> diesen Parameter mit als Entscheidungsgrundlage heranziehen. Wünschenswert ist eine Aussage über das zu erwartende Lastprofil für jede neue Dienstanforderung. Allein aus der Vielzahl möglicher Diensttypen und ihren jeweiligen stochastischen Ausprägungen kann solch eine Aussage nur hypothetischer Natur sein. Es ist deshalb erforderlich, die verschiedenen Dienste auf wenige Verkehrsklassen zu abstrahieren und bestimmte Parameter vorauszusetzen, d.h. Annahmen zu treffen. Ein erster Ansatz zur Abschätzung von B-ISDN-Steuerverkehrsparameter wird in Abschnitt 4.4.2.1 vorgestellt.

Während CAC-Verfahren für ihre Berechnungen hauptsächlich die Dienstgüteparameter Zellverlustwahrscheinlichkeit, Zellverzögerung und Zellverzögerungs-Schwankung verwenden, müssen für Annahmeverfahren der Steuerungs-Ebene andere Qualitätsparameter berücksichtigt werden. Die Maßzahlen Blockierwahrscheinlichkeit oder garantierte Antwortzeiten sind hier beispielsweise weniger geeignet. Einem bereits akzeptierten Dienst muss aber garantiert werden, dass mindestens die wichtigsten Nachfolgemeldungen<sup>16</sup> bearbeitet werden und es nicht zu einem Dienstabbruch aufgrund von Fehlerfällen kommt. Höchste Priorität hat deshalb die Komplettierung eines Dienstes.

Da der Signalisierverkehr für die Zeitdauer einer abgeschlossenen Dienstbearbeitung extrem büschelförmigen Charakter besitzt, sollte aufgrund des erzielbaren Multiplexgewinns das Prinzip des *statistischen Multiplexens* angewandt werden. Diese Eigenschaft muss in dem Annahmealgorithmus berücksichtigt werden. Unter starker Belastung sind kurzzeitige Dienstgüteeinbrüche für aktive Kommunikationsbeziehungen unvermeidlich. Dies obliegt der reaktiven

---

<sup>15</sup> Z.B. als Parameter zur Berechnung der Nennkapazität eines Steuerungsrechners

<sup>16</sup> Eine „weniger wichtige“ Nachfolgemeldung ist beispielsweise der Wunsch nach einer Parametermodifikation einer bereits existenten Verbindung unter Hochlast.

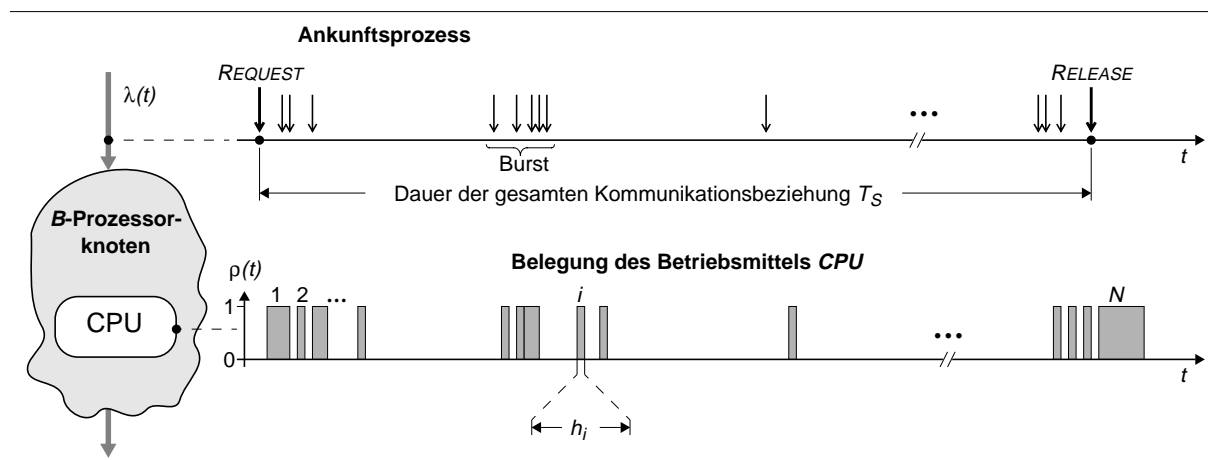
Lastregelung (siehe Dienstbehandlung Abschnitt 4.5). Im folgenden wird die *NRAC* (*New Request Admission Control*) näher vorgestellt.

## 4.4.2 Zentrale Ressource

### 4.4.2.1 Lastcharakterisierung

Die Lastcharakterisierung hängt eng mit der Dienstmodellierung (siehe Abschnitt 5.2) zusammen. Je realistischer die Verkehrscharakteristik einer Dienstbeschreibung sein soll, umso detaillierter müssen Informationen über Quellen oder sonstige Merkmale vorliegen. Wie bereits mehrfach ausgeführt wurde, ist dies gerade im B-ISDN-Umfeld außerordentlich problematisch. Die *NRAC* ist als lokales Lastregelverfahren für einen *B*-Prozessorknoten konzipiert. An der Schnittstelle dieser Funktionsebene werden Schicht-3-Nachrichten mit den  $PN_A$  ausgetauscht. Der an einem  $PN_B$  ankommende Verkehrsstrom setzt sich deshalb überwiegend aus Schicht-3-Signalschicht-Nachrichten zusammen. Die unterste Modellierungsebene für eine Charakterisierung des Steuerungsverkehrs ist deshalb hier die Signalschicht.

Bild 4-3 zeigt den Ankunftsprozess und die resultierende Inanspruchnahme des Betriebsmittels *CPU* am Beispiel der kompletten Bearbeitung eines Dienstes. Das vollständige Szenario umfasse  $N$  Nachrichten. Der Dienst beginnt hier mit einer *REQUEST*- und endet mit einer *RELEASE*-Meldung. Die mittlere Bedienzeit einer Nachricht sei mit  $h_i$  bezeichnet.



**Bild 4-3:** Typischer Dienstverlauf auf Signalisiererebene. Die Zeitdiagramme zeigen den Ankunftsprozess und die resultierende Ressourcenbelegung.

Dieses *Lastprofil* einer Kommunikationsbeziehung lässt sich in erster Näherung durch die *stationäre Prozessorauslastung*  $\bar{\rho}_S$  beschreiben. Für eine Lastplanung wäre als zweiter Parameter die *mittlere Gesamtdauer*  $T_S$  einer Kommunikationsbeziehung<sup>17</sup> wünschenswert. Dieser Ansatz soll im folgenden als *Zwei-Parameter-Modell* bezeichnet werden.

<sup>17</sup> Bei Internetanwendungen wird in diesem Zusammenhang von einer Sitzung (*Session*), in der Telekommunikationstechnik von einem Ruf (*Call*) gesprochen. In [54] wird eine *ganze Kommunikationsbeziehung* als „*Gespräch*“ (*Call*) bezeichnet.

Mittels der *Verkehrsmenge*  $C_S$  lässt sich der Zeitmittelwert der Prozessorauslastung berechnen. Die Steuerverkehrsmenge<sup>18</sup> (*control traffic volume*) entspricht hierbei der Summe der Belegungsauern aller zu einer Kommunikationsbeziehung gehörenden Signalisier- und internen Steuerungsnachrichten:

$$T_H = \sum_{i=1}^N h_i \quad \text{in [s] oder } C_S = \frac{T_H}{3600} \quad \text{in [Erlh].} \quad (4-5)$$

Damit ergibt sich eine *mittlere Prozessorauslastung*  $\bar{\rho}_S$  bzw. ein *Verkehrswert*  $Y_S$  von

$$\bar{\rho}_S = \frac{T_H}{T_S} \quad \text{in [%] bzw. } Y_S = \frac{C_S}{T_S} \quad \text{in [Erl].} \quad (4-6)$$

Es kann somit nur ein stationärer Mittelwert für die Ressourcenbelegung angegeben werden. Da aber der aus einer Kommunikationsbeziehung rührende Signalisierverkehr stark *büschelartigen Charakter* (u.a. diensteabhängig) hat, wäre ein weiterer Parameter zur Kennzeichnung der Dynamikeigenschaft wünschenswert.<sup>19,20</sup> In der Nutzer-Ebene wird auf Büschelebene hierfür die Maßzahl *Burstiness* verwendet [328].<sup>21</sup> Im Zusammenhang mit ATM-Zellströmen entspricht diesem Büschelfaktor das Verhältnis aus maximaler und mittlerer Zellrate.

Für das betrachtete steuerungstechnische Betriebsmittel könnte als entsprechendes Aktivitätsmaß das Verhältnis von maximaler zu mittlerer Prozessorauslastung definiert werden. Da der Maximalwert  $\rho_{max}$  aber gleich eins ist, wäre der „Büschelfaktor“ hier gleich dem Kehrwert des Mittelwertes und würde somit keine weitere Information einbringen. Dagegen wäre die Ankunftsrate  $\lambda(t)$  (mit  $\lambda_{max}$  und  $\bar{\lambda}$ ) oder, speziell aus Sicht der Lastregelung, das Angebot  $A(t)$  (mit  $A_{max}$  und  $\bar{A}$ ) für die Definition eines „*Burstiness*“-Maßes für die Steuerungs-Ebene geeigneter.

Allgemein erhebt sich natürlich die Frage, ob weitere und wieviele Parameter in die Dienstcharakterisierung miteinbezogen werden sollten. Die Kenntnis des exakten *Zeitverlaufes* wäre für eine Ressourcen- und Lastplanung verständlicherweise wünschenswert.<sup>22</sup> In [325] wird beispielsweise ein sehr detailliertes, generisches Arbeitslastmodell für Internet-Dienste vorgeschlagen. Da nun aber im allgemeinen über die Charakteristik des Steuerungsverkehrs wesentlich weniger bekannt ist als über den Nutzverkehr, soll hier nur eine einfache Maßzahl verwendet werden.

<sup>18</sup> Verkehrsmenge (*traffic volume*) ist in [616] wie folgt definiert: Summe der Belegungsauern eines vermittlungstechnischen Betriebsmittel innerhalb eines beliebig wählbaren Zeitabschnitts.

<sup>19</sup> Am Beispiel des klassischen Fernsprechverkehrs soll diese Dynamik im Signalisierverkehr aufgezeigt werden. Die mittlere Rufdauer sei  $T_S = 180$ s und ein Vermittlungsknoten besitze eine Nennkapazität von beispielsweise 500000 BHCA. Daraus ergibt sich für die durchschnittliche Summenbelegungsauer eines Rufes  $T_H = 7.2$ ms CPU-Zeit, was einer Verkehrsmenge von  $C_S = 2 \cdot 10^{-6}$  Erlh entspricht. Die *mittlere Belastung* des Prozessors für einen Ruf, und damit der Verkehrswert, beträgt  $\rho_S = 0.004\%$  oder  $Y_S = 4 \cdot 10^{-5}$  Erl. Der hier vorliegende Steuerungsverkehr besitzt somit für das Verhältnis von maximaler zu mittlerer Ressourcenbelegung einen Wert von  $1 : \bar{\rho}_S = 25000$ .

<sup>20</sup> In [135] wird die *Signalisierintensität* eines Dienstes als Leistungskriterium für die Bewertung von Ressourcenverwaltungsstrategien verwendet. Die *Signaling Intensity* ist dort definiert als „*the total amount of atomic signaling messages performed by a switching node divided by the total number of connection requests*“.

<sup>21</sup> Der *Büschelfaktor* kennzeichnet das Verhältnis aus mittlerem und minimalem Ereignisabstand [616].

<sup>22</sup> Bestrebungen in dieser Richtung werden in [567] skizziert. Es wird ein weiteres Informationselement für die SETUP-Nachricht vorgeschlagen, in welchem Beginn und Ende der angeforderten Kommunikationsbeziehung eingetragen wird. Motiviert wird dieser Vorschlag durch eine zeit- und bedarfsgerechtere Ressourcenplanung für die Nutzer-Ebene. Diese Information ließe sich aber auch für eine zeitliche Betriebsmittelplanung für die Steuerungs-Ebene nutzen.



#### 4.4.2.2 Effektive Prozessorbelastung

Für eine effiziente Nutzung von Betriebsmitteln der Nutzer-Ebene (z.B. Linkkapazität, Zellpuffer) wird beim ATM das Prinzip des statistischen Multiplexens angewendet. Für Verbindungen mit variablem, im Zeitverlauf schwankendem Bitratenbedarf, findet hierbei eine bewusste „Überbuchung“ von Ressourcen statt. Der sogenannte Multiplexgewinn<sup>23</sup> ist dabei umso größer, je deutlicher der Mittelwert vom Maximalwert abweicht. Damit ist dieses Multiplexverfahren für zeitlich variierende Lastprofile, wie ihn beispielsweise der Signalisierverkehr innerhalb einer Kommunikationsbeziehung aufweist, geradezu prädestiniert.

Bei CAC-Algorithmen wird als Verkehrscharakterisierung oft als Ersatzparameter für eine Nutzverbindung die sogenannte *äquivalente Zellrate* oder *effektive Bandbreite* verwendet [328] (siehe auch Abschnitte 3.2.2 und 3.6.3). Diese Maßzahl stellt eine Verknüpfung von Verkehrs- und Netzgüteparametern dar, da z.B. auch die Verlustwahrscheinlichkeit in die Berechnung eingehen kann. Aus Sicht der Ressourcenverwaltung bzw. CAC ist die effektive Bandbreite auch deshalb besonders attraktiv, da dieser Ersatzparameter eine lineare Kapazitätszuweisung erlaubt.

Für die Bestimmung des Effektivwertes einer neuen Verbindung müssen allerdings prinzipiell neben den Verkehrsvertragsangaben auch die Charakteristiken aller bereits bestehenden Verbindungen sowie Systemparameter (z.B. Puffergrößen) berücksichtigt werden. Weitere Berechnungskriterien ergeben sich aus den allgemeinen Anforderungen, dass ausreichend Ressourcen reserviert werden und keine Überlastungen auftreten. Andererseits sollte der Effektivwert aber möglichst klein gewählt werden, um den Multiplexgewinn zu maximieren [532]. Die Berechnung solch einer Maßzahl ist deshalb alles andere als trivial, zumal diese auch in Echtzeit ausgeführt werden muss.

Für eine ATM-Nutzverbindung liegt damit der zugehörige Effektivwert  $r_{U, eff}$  allgemein zwischen Mittel- und Maximalwert:

$$\bar{r}_U \leq r_{U, eff} \leq r_{U, max}. \quad (4-7)$$

Für die Ressourcenverwaltung wäre die Verwendung des Maximalwertes die einfachste Möglichkeit. Das andere Extrem für die Zulassung von Verbindungen bildet die Beurteilung auf Grundlage des Mittelwertes. Für die Steuerungs-Ebene lässt sich nun, mit der im vorherigen Abschnitt vorgestellten Lastcharakterisierung, die *effektive Prozessorauslastung*  $\hat{r}_p$  als entsprechende Maßzahl einführen. Da eine Quelle keine Steuerverkehrsparameter für eine neue Kommunikationsbeziehung spezifiziert, kann aber prinzipiell auch keine effektive Prozessorbelastung vorab berechnet werden.<sup>24</sup>

Aufgrund dieser Tatsache sollen – im Rahmen dieser Arbeit – durch fortlaufende Messungen Effektivwerte geschätzt und damit adaptiv nachgeführt werden. Die eigentliche Ressourcenverwaltung für die Steuerungs-Ebene arbeitet somit tabellengestützt. Es sei an dieser Stelle angemerkt, dass – im Gegensatz zur Nutzer-Ebene – diese stationäre Maßzahl *alle Verbindungen* einer Kommunikationsbeziehung einbezieht.<sup>25</sup>

---

<sup>23</sup> Effizienzzunahme, welche durch asynchrones Multiplexen erzielt werden kann, wenn die zur Bereitstellung einer gegebenen Dienstgüte insgesamt benötigte Bitrate geringer ist als die Summe der Spitzenbitraten der einzelnen (virtuellen) Verbindungen [616].

<sup>24</sup> Das Circumflex-Zeichen „^“ soll deshalb andeuten, dass es sich bei dieser Größe nur um einen Schätzwert handeln kann.

<sup>25</sup> Denkbar wären auch *verbindungs-spezifische* Parameter. Eine *Verbindung* könnte beispielsweise durch eine *effektive Bandbreite* und eine *effektive Prozessorbelastung* (gemeinsam) charakterisiert werden.

Für eine Realisierung werden zweckmäßigerweise verschiedene Diensttypen gruppiert und in Verkehrsklassen zusammengefasst. Dieses Vorgehen ist aufgrund der Komplexität des Problems notwendig und wird auch üblicherweise bei Lastmodellen angewandt. Der Parameter  $\hat{r}_{p,v}$  soll deshalb die effektive Belastung durch einen Dienst der *Verkehrsklasse*  $v$  bezeichnen. Ein wesentliches Problem stellt die Ermittlung brauchbarer Werte dar. Da für B-ISDN in der Literatur hierzu noch keine Angaben als Ausgangsgrundlage verfügbar sind, müssen diese Parameter geschätzt werden.

Solche Schätzwerte sind zum jetzigen Zeitpunkt zwangsläufig sehr ungenau und können nur durch fortlaufende Messungen an realen Netzumgebungen konkretisiert werden.<sup>26</sup> Bis eine stabile Datenbasis vorliegt, vergehen in der Regel sehr viele Jahre des Netzbetriebs. In den meisten Fällen sind genaue Aussagen über Quellparameterschätzungen sowieso nur schwer möglich [131]. In solch einem Fall müssen die Parameter dynamisch nachgeführt werden. Als Techniken können Lernalgorithmen, indirekte Verfahren<sup>27</sup> oder adaptive Anpassungen aufgrund On-Line-Messungen eingesetzt werden. Diese Art von Parameteradaption soll hier eingesetzt werden.

In der Nutzer-Ebene sind bereits ruf- und verbindungsindividuelle Lastmessungen durch die Policing-Funktionen oder die Datenerfassung für Entgeltabrechnungen integriert. Der tatsächliche Ressourcenverbrauch kann hier mit dem durch den Verkehrsvertrag angeforderten verglichen werden. Werkzeuge für *dienstindividuelle* Messungen<sup>28</sup> wären theoretisch auch im Vermittlungsrechner einsetzbar. Aus praktischen Gründen – der Aufwand wäre unverhältnismäßig groß und die Messkosten zu teuer – muss darauf verzichtet werden. Somit werden nicht für jede Verkehrsklasse alle Betriebsmittelbelegungen überwacht und ausgewertet. Da als Sensor für die Lastregelung sowieso die Prozessorauslastung gemessen werden soll, kann diese Größe für eine Parameteradaption herangezogen werden. Es ist klar, dass mit dieser „reduzierten“ Informationsmenge eine Nachführung nicht so leistungsfähig sein kann, wie wenn klassenindividuelle Messdaten einbezogen werden könnten. Dieser Algorithmus wird im Unterabschnitt 4.4.3.3 vorgestellt.

Abschliessend soll noch einmal darauf hingewiesen werden, dass diese Ressourcenmaßzahl bewusst als Effektivwert bezeichnet wird, da zum einen der stationäre Mittelwert unbekannt ist und zum anderen das Prinzip des statistischen Multiplexens angewendet werden soll. Durch die adaptive Nachführung der Schätzparameter ist die effektive Prozessorauslastung auch keine Konstante, sondern hängt neben der Verkehrsklasse von den aktuellen "Umgebungsbedingungen" - d.h. Anzahl und Typ aktiver Kommunikationsbeziehungen - sowie, je nach angewandtem Filter- und Schätzverfahren, gegebenenfalls von der Vergangenheit ab.

---

<sup>26</sup> Solche Messungen, um an Daten für Arbeitslastmodelle heranzukommen, sind ein Problem für sich. Heutige Endsysteme sind Rechnerplattformen mit vielschichtiger Software. Das Messwerkzeug wird in der Regel als Softwaremonitor realisiert. Jede Messung hat somit eine Rückwirkung auf das beobachtete System und führt zwangsläufig zu einer Verfälschung der Messdaten. Die Abweichungen können teilweise gravierend sein.

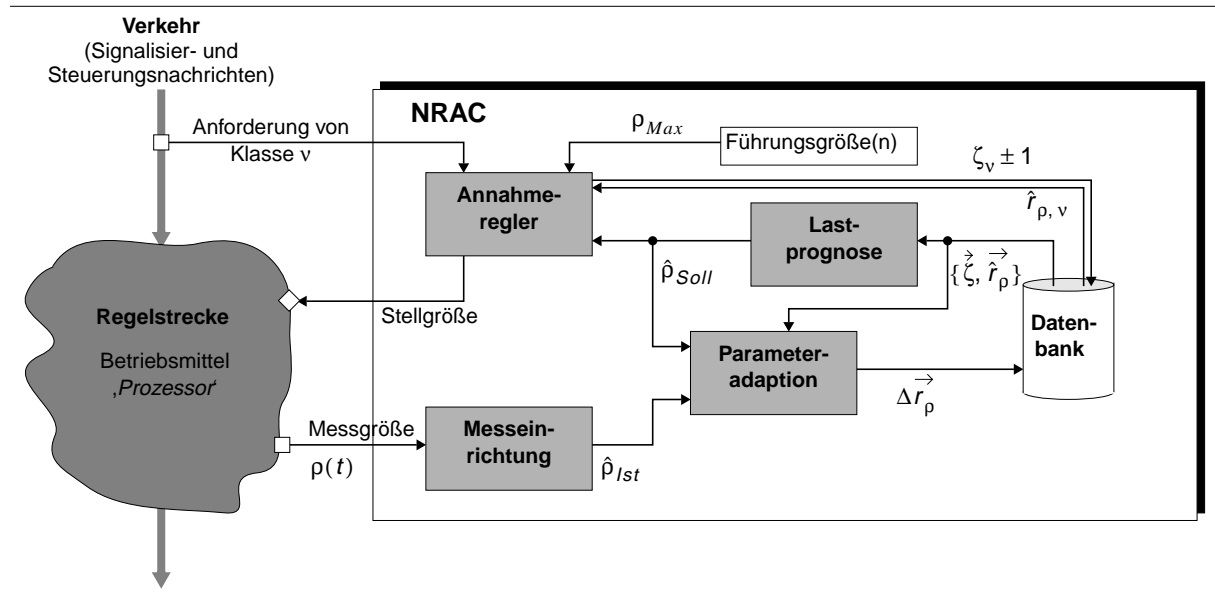
<sup>27</sup> Die Systembelastung und damit auch Verkehrsparameter lassen sich indirekt schätzen. SHARMA [484] schlägt hierzu folgendes Verfahren vor: Ein lokaler Testverkehr mit bekannten Eigenschaften wird zusätzlich in das System eingespeist. Aus Delaymessungen und den bekannten Informationen können dann unbekannte Parameter geschätzt werden.

<sup>28</sup> Damit ist die Erfassung „Call Reference“-individueller Messdaten gemeint.

## 4.4.3 Algorithmus

### 4.4.3.1 Übersicht

Bild 4-4 zeigt in einem Blockdiagramm die wichtigsten Komponenten der *Annahmekontrolle für neue Dienstanforderungen (NRAC)*. In dieser Übersicht wird deutlich, dass neben dem eigentlichen Regelkreis noch ein zweiter Kreislauf für die Parameternachführung wirksam ist. Die beiden Verfahren sind Gegenstand nachfolgender Unterabschnitte.



**Bild 4-4:** Schematische Darstellung der *Annahmekontrolle für neue Dienstanforderungen (NRAC)*

### 4.4.3.2 Annahmeregung

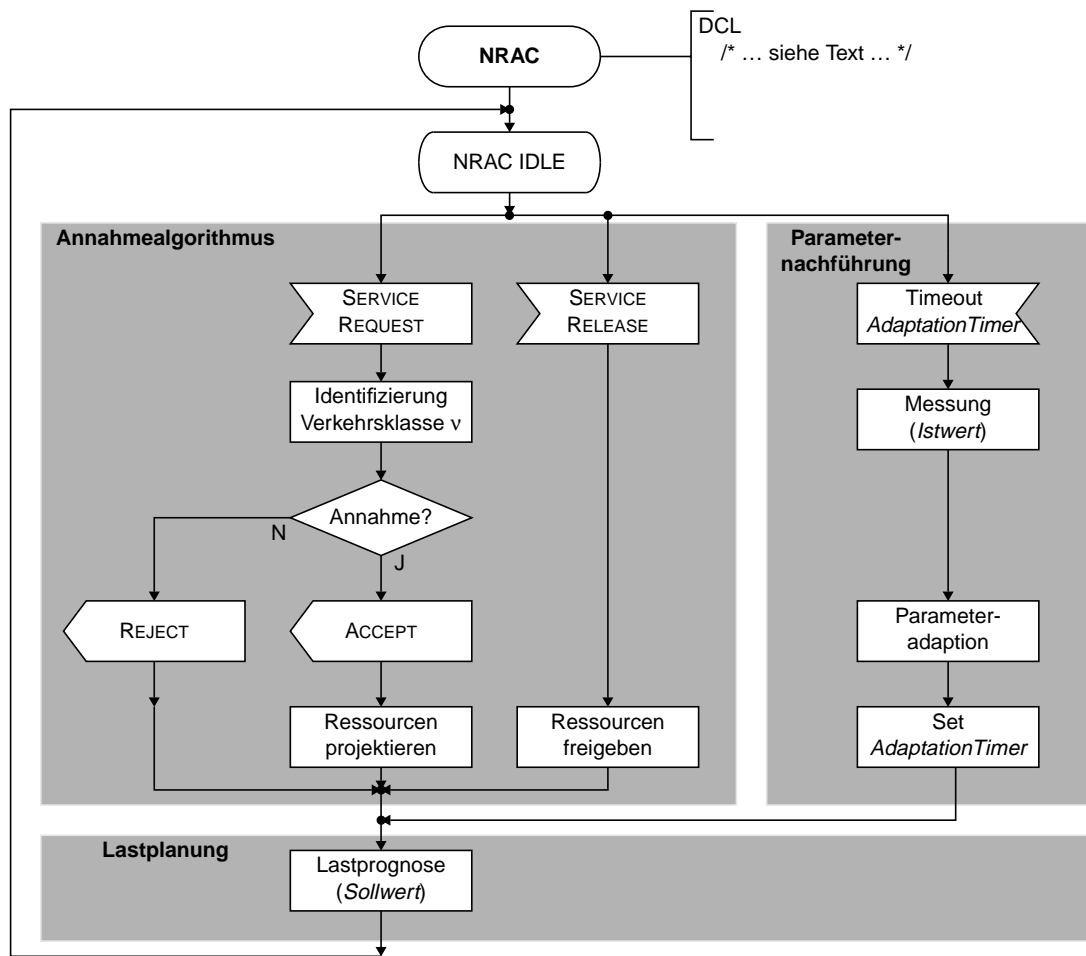
Die eigentliche Annahmeregung für neue Dienstanforderungen basiert auf nachfolgendem Verfahren. Folgende Bezeichnungen sollen eingeführt werden (siehe auch Bild 4-4):

- $\zeta_v$  ... Anzahl aktiver Kommunikationsbeziehungen der Verkehrsklasse v.
- $\hat{r}_{\rho, v}$  ... Geschätzter Ressourcenbedarf an relativer CPU-Zeit einer Kommunikationsbeziehung der Verkehrsklasse v (effektive Prozessorauslastung).

Der Ablauf auf oberster Ebene ist in Bild 4-5 dargestellt. Eine neue Anforderung wird zuerst klassifiziert. Hierzu werden bestimmte Nachrichtenfelder analysiert und anschließend eine geeignete Verkehrsklasse v identifiziert. Aus der Datenbank wird nun der geschätzte Ressourcenbedarf  $\hat{r}_{\rho, v}$  für diesen Dienst ausgelesen und auf Basis folgender Ungleichung

$$(\hat{\rho}_{Soll} + \hat{r}_{\rho, v}) \leq \rho_{Max} \quad (4-8)$$

eine Regelentscheidung getroffen. Abschließend erfolgt die Einstellung der Stellgröße sowie, bei einer positiven Annahmeentscheidung, die Inkrementierung der Zählvariablen  $\zeta_v$ . Damit ist dieser Ressourcenanteil verplant bzw. projiziert.



**Bild 4-5:** Allgemeine Ablauforganisation der NRAC in SDL-Notation

Für die Berechnung des Sollwerts  $\hat{p}_{Soll}$  wird zuerst vom *Lastprognose*-Modul der Ressourcenbedarf berechnet:

$$1. \text{ Verplante Ressourcen pro Verkehrsklasse: } \hat{R}_{\rho, v} = \zeta_v \cdot \hat{r}_{\rho, v}. \quad (4-9)$$

$$2. \text{ Verplante Ressourcen insgesamt: } \hat{R}_{\rho, ges} = \sum_{v=1}^K \hat{R}_{\rho, v}. \quad (4-10)$$

Da die Messeinrichtung aus Aufwandsgründen nur die *gesamte* Prozessorauslastung erfassen kann, muss bei der Sollwertbestimmung der Anteil der nichtvermittlungstechnischen Last  $\hat{R}_{\rho, BL}$  („Hintergrundlast“) abgeschätzt werden. Dieser Wert sei hier idealerweise als konstant angenommen. Damit ergibt sich folgender projektierter Lastsollwert:

$$\hat{p}_{Soll} = \hat{R}_{\rho, ges} + \hat{R}_{\rho, BL} = \vec{\zeta} \cdot \vec{\hat{r}}_{\rho} + \hat{R}_{\rho, BL}. \quad (4-11)$$

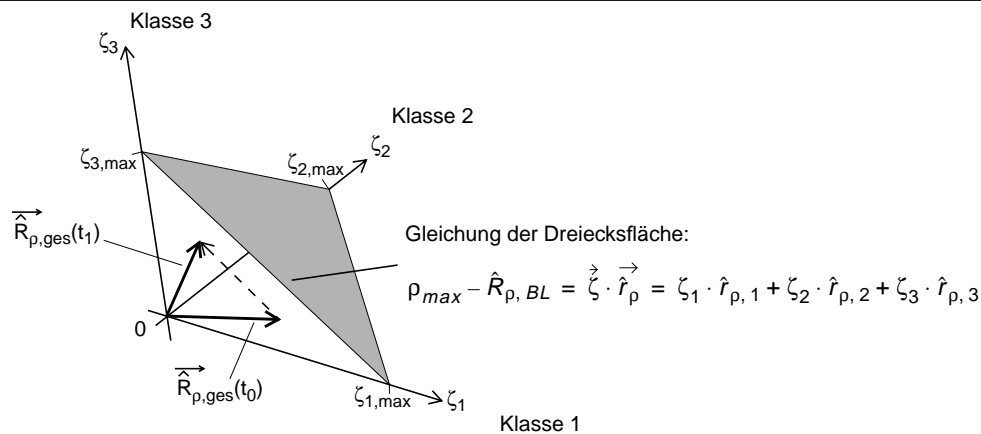
Die beiden Vektoren repräsentieren den Belegungszustand und den geschätzten Ressourcenbedarf über alle  $K$  Verkehrsklassen:

$$\vec{\zeta} = \begin{bmatrix} \zeta_1 \\ \dots \\ \zeta_K \end{bmatrix}^T \quad \text{und} \quad \vec{\hat{r}}_\rho = \begin{bmatrix} \hat{r}_{\rho, 1} \\ \dots \\ \hat{r}_{\rho, K} \end{bmatrix}. \quad (4-12)$$

Abschließend soll das Prinzip der Annahmeregulation an einer Graphik verdeutlicht werden (Bild 4-6). Es seien hier *drei Verkehrsklassen* angenommen. Als Grenzfläche für die maximal verplanbaren Ressourcen ergibt sich dann eine auf den ersten „Quadranten“ begrenzte Ebene mit der Gleichung  $\rho_{max} - \hat{R}_{\rho, BL} = \zeta_1 \cdot \hat{r}_{\rho, 1} + \zeta_2 \cdot \hat{r}_{\rho, 2} + \zeta_3 \cdot \hat{r}_{\rho, 3}$ . Der aktuelle Systemzustand kann durch den Vektor

$$\vec{\hat{R}}_{\rho, ges}(t) = \begin{bmatrix} \zeta_1(t) \cdot \hat{r}_{\rho, 1}(t) \\ \zeta_2(t) \cdot \hat{r}_{\rho, 2}(t) \\ \zeta_3(t) \cdot \hat{r}_{\rho, 3}(t) \end{bmatrix} \quad (4-13)$$

repräsentiert werden. Durch eine neue Anforderung wird in Zeile  $v$  der Betrag  $\hat{r}_{\rho, v}$  addiert. Der Zeiger wandert in Richtung dieser Verkehrsklasse und der Annahmealgorithmus sollte dafür sorgen, dass dieser innerhalb der Grenzfläche (*admission space*) bleibt, oder zumindest diese – aufgrund des statistischen Multiplexens – „nicht allzulange Zeit“ verlässt.



**Bild 4-6:** Grenzfläche der maximalen Auslastung bei *drei Verkehrsklassen*

In Abschnitt 6.3 werden Messungen und Parameterstudien für konkrete Annahmealgorithmen, d.h. wie die Regelentscheidung von Gl. (4-8) in eine bestimmte Stellgrößeneinstellung umgesetzt wird, gezeigt.

### 4.4.3.3 Nachführung von Parametern

Die Ressourcenbedarfsparameter sind zwangsläufig (zunächst) nur grobe Näherungswerte. Ein denkbarer Ausgangspunkt für eine erste Einordnung können A-Priori-Kenntnisse, die beispielsweise aus Labormessungen an Prototypsystemen gewonnen werden, sein. Diese Ausgangsschätzwerte werden im folgenden als Originalparameter  $\hat{r}_{\rho, v}^{orig}$  bezeichnet. Aufgrund dieser Unschärfe müssen die Parameter dynamisch nachgeführt werden. Die Überprüfung und Anpassung kann dabei aus Kostengründen nur von Zeit zu Zeit erfolgen.

Im rechten Zweig des SDL-Diagrammes (Bild 4-5) ist dieser Aspekt durch die Zeitgebervariable *Adaptation Timer* angedeutet. Dieser Timer sorgt für den regelmäßigen Aufruf der Parameteradaption.<sup>29</sup>

In jedem Adaptionintervall  $k$ , d.h. nach  $T_{AI}$  Zeiteinheiten, läuft folgender Prozess ab. Aus dem letzten Mess- (*Istwert*) und Prognosewert (*Sollwert*) wird die *Abweichung*  $\Delta R_\rho$  berechnet:

$$\Delta R_\rho(k) = \hat{\rho}_{Ist}(k) - \hat{\rho}_{Soll}(k). \quad (4-14)$$

Die *Gewichtungsfaktoren*  $g_v$  repräsentieren die Anteile der aktuellen Klassenbelegung:

$$g_v(k) = \frac{\zeta_v(k)}{\sum_v \zeta_v(k)} = \frac{\zeta_v(k)}{\zeta_{ges}(k)} \text{ oder } \dot{g}(k) = \frac{1}{\sum_v \zeta_v(k)} \cdot \dot{\zeta}(k). \quad (4-15)$$

Daraus werden die *Korrekturwerte*  $\Delta r_{\rho, v}$  wie folgt bestimmt:

$$\Delta r_{\rho, v}(k) = \alpha \cdot g_v(k) \cdot \Delta R_\rho(k) \text{ oder } \overrightarrow{\Delta r}_\rho(k) = \alpha \cdot \Delta R_\rho(k) \cdot \dot{g}(k). \quad (4-16)$$

Die Berechnung der Korrekturwerte nach Gl. (4-16) basiert auf der Annahme, dass proportional zur Abweichung  $\Delta R_\rho$  korrigiert werden sollte. Parameter  $\alpha$  stellt eine *Dämpfungskonstante*, die zu starke Parameterschwankungen verhindern soll, dar. Die Wahl eines praktikablen Wertes für  $\alpha$  wird später diskutiert. Das *Ziel der Parameteradaption* muss eine Minimierung der Differenz  $\Delta R_\rho$  sein:

$$\Delta R_\rho \rightarrow 0. \quad (4-17)$$

Aus den Gleichungen (4-11, 4-14 und 4-17) sowie dem Wertebereich von  $\hat{\rho}_{Ist}$  und  $\hat{\rho}_{Soll}$  ergibt sich folgender Wertebereich für obige Differenz:

$$-(1 + \hat{R}_{\rho, BL}) \leq \Delta R_\rho \leq 1. \quad (4-18)$$

Es soll hier additiv korrigiert werden.<sup>30</sup> Das Vorzeichen bestimmt die Richtung der Korrektur für alle Verkehrsklassen. Damit ergibt sich als *vorläufiger Schätzwert*  $\hat{r}'_{\rho, v}(k+1)$  für das nächste Intervall  $k+1$ :

$$\hat{r}'_{\rho, v}(k+1) = \hat{r}_{\rho, v}(k) + \Delta r_{\rho, v}(k) \text{ oder } \overrightarrow{\hat{r}}'_\rho(k+1) = \overrightarrow{\hat{r}}_\rho(k) + \overrightarrow{\Delta r}_\rho(k). \quad (4-19)$$

Da  $\hat{r}'_{\rho, v}$  nicht kleiner oder gleich Null sein kann, ist eine zusätzliche Begrenzung notwendig. Es findet deshalb noch folgende *Bereichsüberprüfung* statt:

$$\delta_L \cdot \hat{r}_{\rho, v}^{org} \leq \hat{r}'_{\rho, v}(k+1) \leq \delta_H \cdot \hat{r}_{\rho, v}^{org}. \quad (4-20)$$

<sup>29</sup> Neben den hier resultierenden äquidistanten Adaptionintervallen wäre prinzipiell auch ein ereignisgesteuerter Ansatz denkbar. Beispielsweise wird nach jeder *N-ten* REQUEST-Nachricht die Parameteradaptionfunktion aufgerufen. Realisiert wird dies mit Hilfe einer Zählvariablen  $C_{ALC}$  (*Adaption Loop Counter*). Dieses Verfahren wurde auch simuliert. Beide Varianten haben ihre Vor- und Nachteile.

<sup>30</sup> Andere bewährte Ansätze für Parameternachführungen sind beispielsweise additive Erhöhungen und multiplikative Verkleinerungen oder nichtlineare Verfahren [129, 269, 538].

Somit ergeben sich für die einzelnen Verkehrsklassen folgende *Schätzwerte*  $\hat{r}_{\rho, v}(k+1)$ :

$$\hat{r}_{\rho, v}(k+1) = \begin{cases} \delta_H \cdot \hat{r}_{\rho, v}^{org} & \text{für } \frac{\hat{r}'_{\rho, v}(k+1)}{\hat{r}_{\rho, v}^{org}} > \delta_H \\ \hat{r}'_{\rho, v}(k+1) & \text{für } \delta_L \leq \frac{\hat{r}'_{\rho, v}(k+1)}{\hat{r}_{\rho, v}^{org}} \leq \delta_H \\ \delta_L \cdot \hat{r}_{\rho, v}^{org} & \text{für } \frac{\hat{r}'_{\rho, v}(k+1)}{\hat{r}_{\rho, v}^{org}} < \delta_L \end{cases} \quad (4-21)$$

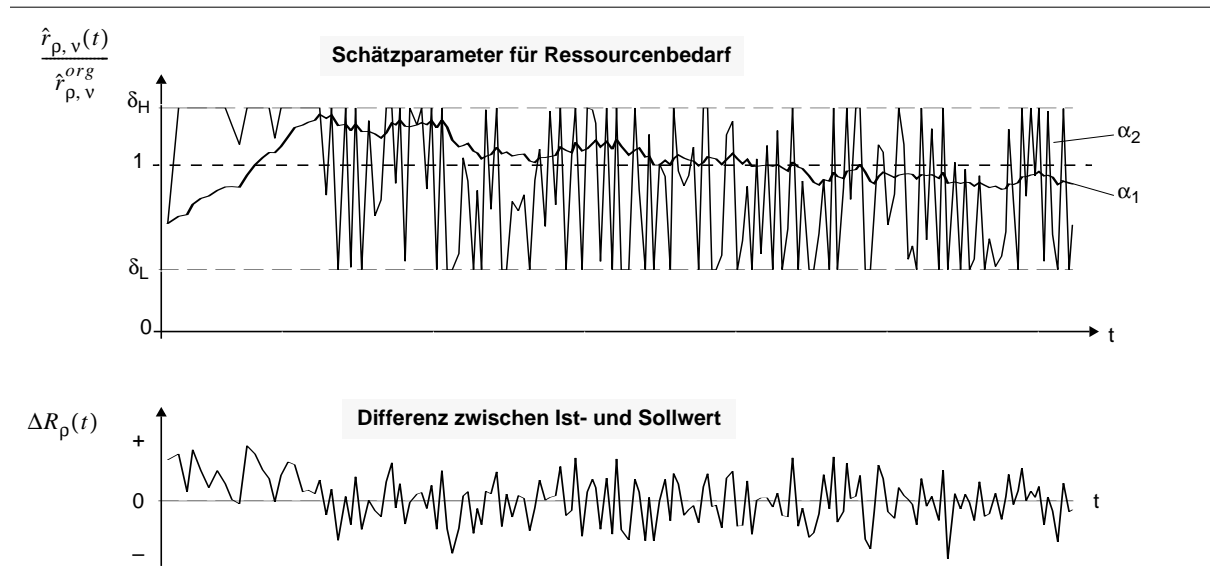
Abschließend werden die aktualisierten Parameter in der Datenbank neu eingestellt. Der gesamte Algorithmus basiert auf einfachen arithmetischen Rechnungen sowie zwei Vergleichsoperationen. In Abschnitt 6.2 wird die Anwendung dieses Verfahrens gezeigt. Zunächst sollen hier aber noch einige Überlegungen zur Parameterwahl von  $\alpha$ ,  $\delta_L$ ,  $\delta_H$  und  $\hat{r}_{\rho, v}^{org}$  diskutiert werden.

Entsprechend ihrer Definition werden die Werte der Schätzparameter  $\hat{r}_{\rho, v}^{org}$  in der Praxis sehr klein sein.<sup>31</sup> Wie in Gl. (4-18) gezeigt wurde, können deshalb u. U. die Korrekturwerte

$$|\Delta r_{\rho, v}| \gg \hat{r}_{\rho, v}^{org} \quad (4-22)$$

sein. Diese Tatsache motiviert die Einführung des Glättungsfaktors  $\alpha$ , mit einem prinzipiellen Wertebereich von  $\alpha \in (0, 1]$ . Wegen Gl. (4-22) sollte  $\alpha$  wesentlich kleiner als eins gesetzt werden, ansonsten wird eventuell zu stark korrigiert („Schmitt-Trigger-Charakteristik“). Für die Größenordnung von  $\alpha$  hat sich folgende Faustformel bewährt:

$$\alpha \approx \frac{1}{K} \sum_{v=1}^K \hat{r}_{\rho, v}^{org} \quad (4-23)$$



**Bild 4-7:** Exemplarischer Verlauf von  $\hat{r}_{\rho, v}(t)$  für zwei verschiedene Dämpfungsfaktoren  $\alpha_1$  und  $\alpha_2$  (Größenordnung:  $\alpha_1$  nach Gl. (4-23) und  $\alpha_2 \approx 20\alpha_1$ ; Begrenzungskonstanten bei dieser Messung:  $\delta_L = 0.4$ ,  $\delta_H = 1.3$ ) sowie die zugehörige Regelabweichung  $\Delta R_{\rho}(t)$

<sup>31</sup> In dem Beispiel von Abschnitt 4.4.2.1 (siehe Fußnote 19) wäre der Ausgangsschätzwert  $\hat{r}_{\rho, v}^{org} = 4 \cdot 10^{-5}$ .

Bild 4-7 zeigt aus einer exemplarischen Messung den qualitativen Verlauf von  $\hat{r}_{\rho, \nu}(k)$  innerhalb des durch  $\delta_L$  und  $\delta_H$  begrenzten Bereiches für *zwei unterschiedliche* Dämpfungskonstanten.

#### 4.4.3.4 Weitere Diskussion

Die zeitlichen Verläufe der Schätzwerte  $\hat{r}_{\rho, \nu}$  sollten protokolliert werden, um in größeren Zeitabständen gegebenenfalls die Originalwerte  $\hat{r}_{\rho, \nu}^{orig}$  korrigieren zu können. Diese Korrektur wäre eine Aufgabe aus dem Bereich des Netzmanagements oder könnte durch den Eingriff des Netzbetreibers erfolgen. Durch dieses Vorgehen könnte man langfristig zu einer stabilen Datenbasis und einer relativ genauen Abschätzung des mittleren Ressourcenbedarfs der einzelnen Verkehrsklassen kommen. Sollten die Schwankungen innerhalb einer Verkehrsklasse auch langfristig nicht nachlassen, dann wäre zu überlegen, ob eine weitere Klasse eingeführt oder eine andere Dienstaufteilung vorgenommen werden sollte.

Vergleicht man das hier vorgestellte Konzept mit klassischen Überlastabwehrverfahren, dann könnte der Ressourcenbedarfsparameter  $\hat{r}_{\rho, \nu}$  als *BHCA*-Wert interpretiert werden, wobei aber folgende Unterschiede bestehen:

- Es wird nicht mit einem *BHCA*-, sondern mit mehreren *BHCA<sub>v</sub>*-Werten gearbeitet, d.h. es werden hier mehrere Verkehrsklassen berücksichtigt.
- Eine Klasse umfasst nicht nur den üblichen Rufmix (*Call Mix*, siehe Abschnitt 5.2.3.1), sondern mehrere Dienste in ihren unterschiedlichsten Ausprägungen.

Die hier vorgestellte Berechnungsmethode schließt weiterhin nicht aus, dass, aufgrund der aktuellen Dominanz anderer Verkehrsklassen, einzelne  $\hat{r}_{\rho, \nu}$  in einem Adaptionsintervall in die falsche Richtung korrigiert werden. Dieser Nachteil muss aber hier in Kauf genommen werden, da aus Aufwandsgründen auf klassenindividuelle Messungen der  $\rho_{Ist, \nu}$  bewusst verzichtet wurde. Durch diese Beschränkung wird aber die Güte des Verfahrens nicht nachhaltig beeinträchtigt.

Für genauere Lastprognosen und auch Parameteradaptionen könnten die Ergebnisse zurückliegender Adaptionsintervalle ( $k-1, k-2, \dots$ ), entsprechend den in Abschnitt 3.4.6 vorgestellten Filterverfahren, in der Berechnung mitberücksichtigt werden. Für einen ersten Ansatz soll aber darauf verzichtet werden.

Vorläufig unklar bleibt, wie sich eine lang andauernde Überlast auf die Parameternachführung auswirkt, oder welche Konsequenzen eine fokussierte Überlast, d.h. eine Verkehrsklasse besitzt einen überproportionalen Lastanteil, auf die anderen Parameter hat.

Weiterhin kann bei diesem Ansatz keine zeitliche Lastplanung erfolgen, da über die Dauer einer Kommunikationsbeziehung keine Aussage gemacht wird. In einem erweiterten Modell müsste demzufolge der Zeitparameter  $T_S$  mitberücksichtigt werden. Auf diese Hypothese soll aber zunächst verzichtet werden.

#### 4.4.4 Abgrenzung zu bekannten Verfahren

Aus der Literatur sind *keine* Annahmeverfahren, die versuchen den *Ressourcenbedarf* eines Dienstes für die *Steuerungsebene* abzuschätzen, bekannt.<sup>32</sup> Gewisse Elemente des vorgestell-

---

<sup>32</sup> Nicht berücksichtigt sind hier SCP-Regelkreise des IN.



ten Regelkreises sind auch in anderen Verfahren zu finden, die aber andere Objekte zum Ziel haben oder beispielsweise in einer höheren Zeitebene (Netzmanagement) arbeiten. Ganz vage lassen sich Berührungspunkte bezüglich folgender Aspekte feststellen:

- *Adaption von ressourcenbeschreibenden Parametern*

In [220,327,458] werden adaptive CAC für die Nutzer-Ebene vorgestellt. Diese werden auch als dynamische CAC bezeichnet und verwenden als Lastindikator typischerweise die Zellankunftsrate [454], die Zellpufferbelegung [487] oder die Zellverlustwahrscheinlichkeit [250]. SAITO [456] verwendet den Begriff des *Self-Sizing Network*<sup>33</sup> für die adaptive Ressourcenverwaltung von VP-Bandbreiten. Diese Funktion ist dort in der Management-Ebene angesiedelt und die Adaptionintervalldauer beträgt 10 min.

KAWAHARA [296,297] beschreibt ein Annahmeverfahren für einen SCP, dessen Algorithmus nicht auf einer fix eingestellten Bedienrate beruht. Es werden keine Annahmen über die aktuelle Verkehrszusammensetzung gemacht. Die Stellgröße ‚Akzeptierbare Rufanforderungen pro Regelintervall‘ wird adaptiv, auf Basis der gemessenen Leistungskenngröße ‚Komplettierungsrate‘ und einer vorgegebenen maximalen Bearbeitungszeit, nachgeführt.<sup>34</sup> Bis auf den Aspekt der Parameternachführung unterscheidet sich dieser Vorschlag in einer Vielzahl von Punkten (z.B. nur aktiv bei Überlast) und scheint auch wegen des Vorgabeparameters ‚maximale Antwortzeit‘ (wirksam für alle Dienste!) für einen B-ISDN-Vermittlungsrechner weniger geeignet.

- *Methodische Ansätze und Werkzeuge für dynamische Parameternachführungen*

Der Wirkungskreislauf der Parameteradaption basiert vornehmlich auf der Aufbereitung und Bewertung von Messdaten stochastischer Prozesse, mit dem Idealziel einer möglichst genauen Lastprognose. Diese spezielle Problematik lässt sich auf die anwendungsunabhängige Methode der Trendanalyse einer diskreten Zeitreihe (*time series analysis*) zurückführen, die hierfür eine Fülle von Verfahren, z.B. Regressionsverfahren, Fuzzy-Logik, Neuronale Netze, Bayesian Netze usw., zur Verfügung stellt. Diese sind meines Erachtens aber für die Lastregelung in einem Vermittlungsrechner, aufgrund des Kosten-Nutzen-Verhältnisses, weniger geeignet.

- *Lastplanung*

GELENBE [201] analysiert das Rufaufbauzeitverhalten in großen ATM-Netzen. Für seine Überlastuntersuchungen unterscheidet er drei Verkehrsklassen mit unterschiedlicher Ankunftsrate, Rufdauer und Bandbreitenbedarf. Bei erfolgreichem Routing wird die entsprechende Bandbreite in allen involvierten ATM-Netzknoten für die Dauer eines Rufes reserviert. Da von bekannten Rufdauern ausgegangen wird, wäre hier eine einfache Lastplanung in der Nutzer-Ebene realisierbar, was aber dort nicht weiter betrachtet wird.

Die aufgeführten Beispiele haben deshalb keine bzw. zumindest keine größeren Gemeinsamkeiten zu dem vorgestellten NRAC-Verfahren.

---

<sup>33</sup> Folgender Ansatz liegt zugrunde: Durch regelmäßige Verkehrsmessungen werden Vergleichsdaten erfasst, anhand derer ein einfacher Algorithmus eine Ressourcenvergabe vornimmt. Als Vorteile werden Einfachheit („keine komplexen Rechnungen notwendig“) und Robustheit („eine Unschärfe (Diskrepanz) zwischen Realität und geschätztem Zustand ist (im Gegensatz zu statischen Verfahren) kein Problem“) aufgeführt.

<sup>34</sup> Ein Schätzparameter für den Ressourcenbedarf an CPU-Zeit für eine Verkehrsklasse muss dort nicht vorgegeben werden. Damit ist dieser Ansatz völlig unabhängig von der zugrunde liegenden Hardwareplattform. In [297] werden Simulationsergebnisse für ein Dienstespektrum, bestehend aus drei Verkehrsklassen, vorgestellt.

## 4.5 Behandlung neuer Dienstanforderungen

Nach dem im vorherigen Unterkapitel vorgestellten, mehr präventiv agierenden Verfahren soll nun ein weiterer neuer, nun aber reaktiv wirkender Algorithmus vorgestellt werden, dessen Schwerpunkt die Berücksichtigung der Dienstvielfalt ist.

### 4.5.1 Motivation – Prinzip der Leistungsregelung

Wie bereits in Abschnitt 4.3.4 ausgeführt wurde, sollte in einem B-ISDN-Vermittlungsrechner gerade der Diensteaspekt auch in eine Überlastabwehrstrategie einfließen. Durch Verwendung des Leistungsmaßes *Power-Produkt* als zu optimierende Größe wird diese Anforderung hier berücksichtigt. Der neu konzipierte und nachfolgend vorgestellte Regelkreis wird aus diesem Grund auch als *Power-Product Control (PPC)* bezeichnet.

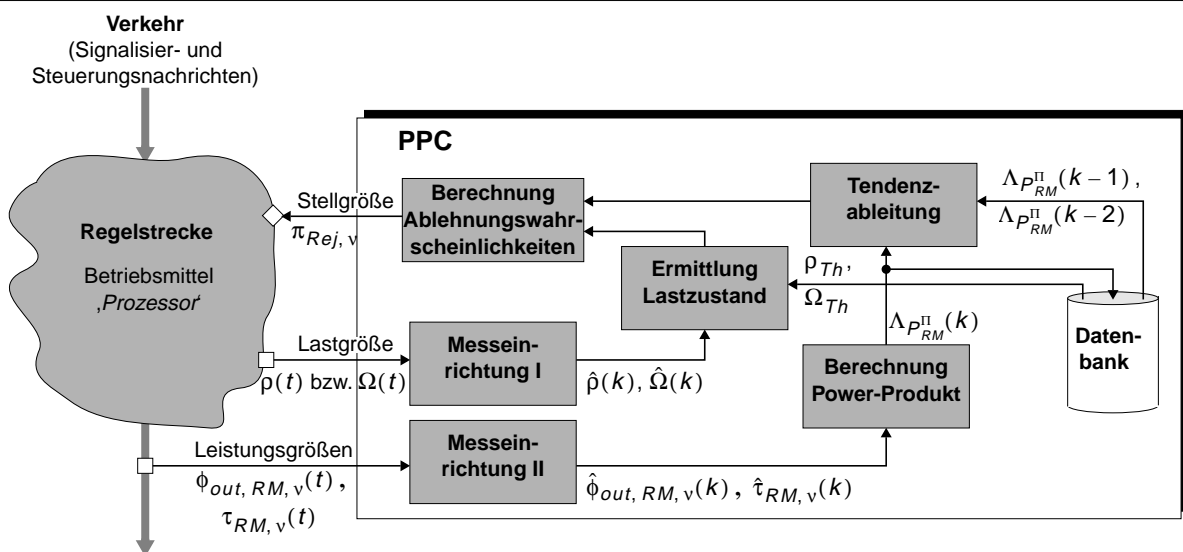
Neben der Berücksichtigung des aktuellen Dienstmixes wird auch eine schnelle Reaktion auf starke Lastimpulse gefordert. Während die *NRAC* präventiv arbeitet, reagiert die *PPC* sofort durch Zurückweisung neuer Dienstanforderungen unter Überlast.

### 4.5.2 Neuer Algorithmus

#### 4.5.2.1 Übersicht

Bild 4-8 zeigt in einem Blockdiagramm die wichtigsten Komponenten dieser *Leistungsregelung* basierend auf dem *Power-Produkt (PPC)*.

Die Regelung wird zyklisch alle  $T_{CI,PPC}$  Zeiteinheiten (ZE) aktiviert. Ziel ist die Berechnung der klassenspezifischen *Zurückweisungs-* oder *Ablehnungswahrscheinlichkeiten*  $\pi_{Rej,v}(k+1)$  (*Rejection Probability*) für das nächste Regelintervall  $k+1$  (siehe auch Bild 4-9). Die in die Berechnung eingehenden Informationen sind nachfolgend beschrieben.



**Bild 4-8:** Schematische Darstellung der lokalen *Leistungsregelung* basierend auf dem *Power-Produkt (PPC)*

In der ersten Phase erfolgt eine **Bereichsunterteilung des Lastraumes**. Eine Ablehnung neuer Dienstanforderungen bei „Niedriglast“ macht keinen Sinn, da ausreichend Betriebsmittel verfügbar sein sollten. Aus diesem Grund soll der *PPC-Regelkreis* erst ab einer gewissen Last aktiviert werden. Die Unterteilung des Lastraumes in zwei Bereiche soll mit Hilfe einer zusätzlichen Lastkenngröße (siehe oben) realisiert werden. Gute Erfahrungen wurden mit der Prozessorauslastung  $\hat{\rho}$  und der Warteschlangenbelegung neuer Dienstanforderungen  $\hat{\Omega}_{RM}$  erzielt.<sup>35</sup> Im folgenden wird das Verfahren anhand der Auslastungsgröße beschrieben. Für einen vorgegebenen Schwellwert  $\rho_{Th}$  seien die beiden *Lastzustände*  $X$  wie folgt definiert:

- $PN_B$  im *Normallastzustand* für  $\hat{\rho}(k) < \rho_{Th}$  ( $X_B = ,NL'$ ) bzw.
- $PN_B$  im *Überlastzustand* für  $\hat{\rho}(k) \geq \rho_{Th}$  ( $X_B = ,ÜL'$ ).

Für eine weitergehende Diskussion des Lastraumes und des „optimalen“ Arbeitsbereiches wird auf Abschnitt 7.2 verwiesen. Durch die Betrachtung eines Mehrklassensystems und die zusätzliche Unterscheidung verschiedener Signalisiernachrichtentypen geht hier allerdings viel an Anschaulichkeit verloren.<sup>36</sup>

In der zweiten Phase erfolgt dann die **Erfassung und Berechnung der Eingangsvariablen**. Dabei werden folgende *Leistungskenngrößen* gemessen:

- Die *Bearbeitungsrate*  $\hat{\phi}_{out, RM, v}(k)$  der REQUEST-Nachrichten der Verkehrsklasse  $v$  und
- die *Verzögerungszeit*  $\hat{\tau}_{RM, v}(k)$  der REQUEST-Nachrichten der Verkehrsklasse  $v$ .

Anschließend erfolgt die Berechnung der klassenspezifischen Leistungsmaße *Power*  $\Lambda_{P, RM, v}(k)$ , basierend auf den REQUEST-Nachrichten nach Gl. (3-10):

$$\Lambda_{P, RM, v}(k) = \frac{\hat{\phi}_{out, RM, v}(k)}{\hat{\tau}_{RM, v}(k)}. \quad (4-24)$$

*Anmerkung:* Wird in einem Intervall *keine* REQUEST-Nachricht einer bestimmten Verkehrsklasse fertig bearbeitet (z.B. zu kleine Last oder sehr viele Nachrichten anderer Klassen werden zuerst bedient), dann ist  $\hat{\phi}_{out, RM, v}(k) = 0$  und  $\hat{\tau}_{RM, v}(k)$  undefiniert. In diesem Fall sei auch der zugehörige Quotient *Power*  $\Lambda_{P, RM, v}(k)$  zu null definiert.

Aus diesen Einzel-Power-Größen erfolgt die Berechnung des übergeordneten Leistungsmaßes *Power-Produkt*  $\Lambda_{P_{RM}^{\Pi}}$  entsprechend Gl. (3-11):

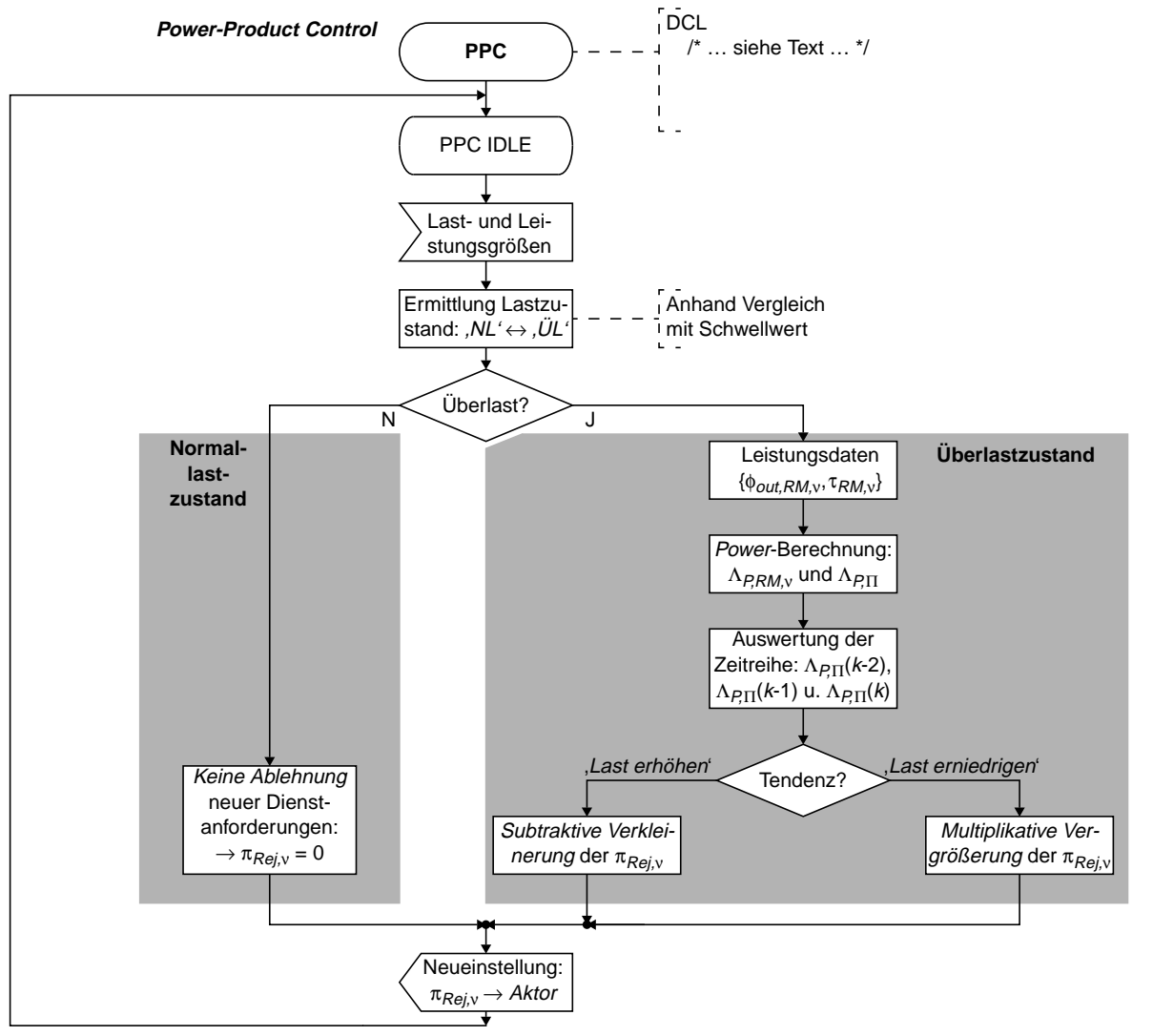
$$\Lambda_{P_{RM}^{\Pi}}(k) = \prod_{v | \Lambda_{P, RM, v}(k) > 0} \Lambda_{P, RM, v}(k). \quad (4-25)$$

Für den Sonderfall  $\Lambda_{P, RM, v}(k) = 0$  wird diese Größe im gegenwärtigen Intervall „nicht berücksichtigt“.<sup>37</sup>

<sup>35</sup> Die Verwendung eines zusätzlichen Lastindikators ist nicht unbedingt erforderlich. Rückschlüsse auf die Belastung des Prozessorknotens lassen sich auch indirekt aus den Leistungsdaten ableiten. Beim *DLB-Regelkreis* (siehe Abschnitt 4.6) beispielsweise wird diese Information aus den gemessenen Antwortzeiten  $\tau_{RM}$  gewonnen (siehe Unterabschnitte 4.6.4.3 bzw. 4.6.4.4). Beide Lösungen haben ihre Vor- und Nachteile, die später diskutiert werden sollen.

<sup>36</sup> Ein Beispiel soll dies verdeutlichen. Betrachtet werde ein System mit *drei Verkehrsklassen* mit dem zugehörigen dreidimensionalen Lastraum (siehe auch Bild 7-2, S. 156). Die Variable  $\rho$  zur Unterteilung des Lastraumes ist in erster Näherung eine Funktion von den drei Verkehrsangeboten  $\rho = f(A_1, A_2, A_3)$ . Die *Fläche*  $\rho = \rho_{Th}$ , die die *Grenze zwischen Normal- und Überlastgebiet* darstellt, entspricht dann dem Oberflächenausschnitt einer Ebene. Durch die Berücksichtigung verschiedener Nachrichtentypen setzt sich das Angebot einer Verkehrsklasse aus verschiedenen Teilangeboten zusammen. Weiterhin wird ein PN durch nichtvermittlungstechnischen Verkehr belastet. In der Realität ist diese Grenzfläche ist deshalb eine weit kompliziertere Funktion als oben beschrieben.

<sup>37</sup> Die Auswirkungen für die zugehörige *Ablehnungswahrscheinlichkeit*  $\pi_{Rej, v}$  sind durch Gl. (4-33, 4-35, 4-36) beschrieben.



**Bild 4-9:** Ablauforganisation der *PPC* in *SDL*-Notation. Konkret wurde das sogenannte *Tripel-Verfahren* (siehe Abschnitt 4.5.2.2.4) dargestellt. Die Funktion wird zyklisch alle  $T_{CI,PPC}$  Zeiteinheiten aufgerufen.

Nach Identifizierung des Lastraumes müssen auf Basis der gewonnen Leistungsinformation die Ablehnungswahrscheinlichkeiten neu berechnet werden. Hierzu wird nachfolgend ein erfolgversprechender Ansatz vorgestellt.

#### 4.5.2.2 Berechnung der Zurückweisungswahrscheinlichkeiten

Die Stellgrößen sind die *Zurückweisungswahrscheinlichkeiten*  $\pi_{Rej,v}(k+1)$  für die einzelnen Verkehrsklassen  $v$ . Diese werden jeweils für das nächste Regelintervall  $k+1$  berechnet. Der Einstellbefehl an den Aktor (siehe Bild 4-8) entspricht bei  $K$  Verkehrsklassen dem *Vektor*

$$\vec{\pi}_{Rej}(k+1) = \begin{bmatrix} \pi_{Rej,1}(k+1) \\ \dots \\ \pi_{Rej,K}(k+1) \end{bmatrix}. \quad (4-26)$$

Aufgrund des Wertebereichs des Zurückweisungsvektors  $\overrightarrow{\pi_{Rej}}$  beim Dreiklassensystem bewegt sich der Vektor in einem Würfel mit Kantenlänge eins. Sind alle Spaltenelemente von  $\overrightarrow{\pi_{Rej}}$  gleich null, dann werden alle Dienstanforderungen aller Klassen angenommen. Besitzt der Vektor den Wert  $\overrightarrow{\pi_{Rej}} = \{1, 1, 1\}$ , dann werden alle neuen Anforderungen aller Verkehrsklassen zurückgewiesen.

#### 4.5.2.2.1 Motivation

Allgemeines Ziel ist es, die Last so zu beeinflussen, dass eine möglichst optimale Leistung erzielt wird. Als leistungscharakterisierende Größe wird das *Power-Produkt*  $\Lambda_{P^{PI}, RM}$  verwendet. Die Last ergibt sich aus dem aktuellen Verkehrsangebot und dem gegenwärtigen Systemzustand. Diese Last wird im wesentlichen von außen aufgeprägt. Einer Beeinflussung sind deshalb enge Grenzen gesetzt.<sup>38</sup> Als lokaler Mechanismus werden hier gegebenenfalls neue Dienstanforderungen zurückgewiesen. Die Leistung als Funktion der Angebote der verschiedenen Verkehrsklassen besitzt ein globales Maximum, dessen Lage und Wert im allgemeinen unbekannt ist.

Die Anforderungen an Lastregelverfahren in Vermittlungsrechnern (z.B. die verfügbare CPU-Zeit ist sehr stark begrenzt; siehe auch Abschnitt 3.4.8) erlauben nicht, sehr genaue Messungen von Last- und Leistungsdaten sowie aufwendige Algorithmen einzusetzen. Aus den zyklisch ermittelnden Leistungskenngrößen kann deshalb hier nur der Wert des *Power-Produkts*, nicht aber der Ort im Lastraum, bestimmt werden. Die Richtung, in der das vermutete Optimum liegt, bleibt deshalb im Verborgenen.

Aus den genannten Gründen kann für diesen Regelkreis nur ein relativ einfaches Verfahren eingesetzt werden. Die Ausgangsbasis bildet die ermittelte *Leistungsfunktion*  $\Lambda_{P^{PI}, RM}(kT_{CI})$  (siehe Bild 4-10). Als Verfahren soll die *direkte Optimum-Suche* eingesetzt werden. Durch kleine Veränderungen der Stellgröße und direktem Vergleich der Messwerte wird versucht, sich schrittweise dem Optimum zu nähern [243]. Diese Algorithmen werden deshalb auch als *Hill Climbing* bezeichnet [53]. Da kein mathematisches Modell des funktionalen Zusammenhangs zwischen Leistung und Last vorausgesetzt wird, stellt dieses Verfahren einen heuristischen Ansatz dar [53]. Direkte Optimum-Suchverfahren basieren auf der Existenz eines globalen Maximums der Leistungsfunktion. Sattelpunkte sind ausgeschlossen [243].

Der im folgenden vorgestellte Algorithmus basiert auf der Tatsache, dass die Leistungskurve eines Vermittlungsrechners typischerweise folgenden Verlauf aufweist (siehe auch Abschnitte 3.1.2 bzw. B.4): Beim Ansteigen der Last verbessert sich üblicherweise das Leistungsverhalten bis ein Maximum erreicht wird. Eine weitere Lasterhöhung führt zu einer Verschlechterung der Leistung. Dieses Wissen wird hier für die Leistungsregelung genutzt.

Je nachdem *wieviel Information* vom gemessenen Leistungsverlauf verwendet wird, können verschiedene Verfahren definiert werden.

#### 4.5.2.2.2 Variante I – Schwellwertverfahren

Ein Algorithmus, der nur auf einem, dem aktuellen Leistungswert  $\Lambda_{P^{PI}, RM}(k)$ , beruht, ist nicht brauchbar. Aufgrund der Eigenschaft der oben angenommenen Leistungsfunktion wird hier zusätzliche Lastinformation benötigt, um den Leistungswert der zugehörigen Lastregion zuzuordnen zu können. Auch wenn der Wert des Optimums  $\Lambda_{P^{PI}, RM, max}$  bekannt sein sollte, wäre

---

<sup>38</sup> Eine direkte Rückwirkung und Beeinflussung der Verkehrsquellen schließt sich bei dieser Anwendung – „lokale Leistungsregelung in einem Vermittlungsknoten“ – aus. Die Last kann nur in begrenztem Maße reduziert, aber nicht direkt „stimuliert“ werden.

damit noch kein praktikables Verfahren realisierbar. Die notwendige zusätzliche Lastinformation könnte beispielsweise aus der gemessenen mittleren Antwortzeit  $\overline{\tau}_{RM}(k)$  abgeleitet werden. Dieser Zeitwert wird dann mit einem Schwellwert verglichen. Je nach identifiziertem Lastbereich werden dann die Zurückweisungswahrscheinlichkeiten gesetzt.

#### 4.5.2.2.3 Variante II – Differenzenverfahren

Werden die beiden letzten Leistungswerte  $\Lambda_{P^\Pi, RM}(k)$  und  $\Lambda_{P^\Pi, RM}(k-1)$  berücksichtigt, dann kann aus dem *Differenzwert*<sup>39</sup>

$$\Delta\Lambda_{P^\Pi, RM}(k) = \Lambda_{P^\Pi, RM}(k) - \Lambda_{P^\Pi, RM}(k-1) \quad (4-27)$$

eine *Tendenz* abgeleitet werden. Ein *Tendenzwert*  $\delta_{Ten}(k)$  könnte beispielsweise folgendermaßen berechnet werden:

$$\delta_{Ten}(k) = \begin{cases} 1 & \text{für } \Delta\Lambda_{P^\Pi, RM}(k) > 0 \\ 0 & \text{für } \Delta\Lambda_{P^\Pi, RM}(k) \approx 0 \\ -1 & \text{für } \Delta\Lambda_{P^\Pi, RM}(k) < 0 \end{cases} . \quad (4-28)$$

Diese Steigungsinformation  $\Delta\Lambda_{P^\Pi, RM}(k)$  alleine bietet noch keine ausreichende Informationsbasis für die Einstellung der Zurückweisungswahrscheinlichkeiten  $\overrightarrow{\pi}_{Rej}(k+1)$ . Auf die Richtung, in der man sich im letzten Regelintervall im Lastraum bewegt hatte, kann grob aus den vorgenommenen Einstellbefehlen (siehe Gl. (4-26)) rückgeschlossen werden. Die Richtung des *Differenzvektors*  $\overrightarrow{\Delta\pi}_{Rej}$

$$\overrightarrow{\Delta\pi}_{Rej}(k) = \overrightarrow{\pi}_{Rej}(k) - \overrightarrow{\pi}_{Rej}(k-1) \quad (4-29)$$

kann hierüber Aufschluss geben. Zeigt  $\overrightarrow{\Delta\pi}_{Rej}$  in Richtung Koordinatenursprung, dann wurde im letzten Regelintervall versucht, die *Last zu stimulieren*, wenn nicht, dann wurde versucht, *lastreduzierend* einzuwirken.

Dieser Algorithmus, wie auch das Schwellwertverfahren, hat den Nachteil, dass zusätzliche Informationen herangezogen werden müssen, um die Ausgangsvariablen  $\overrightarrow{\pi}_{Rej}(k+1)$  berechnen zu können. Dieser Nachteil wird in nachfolgender Variante vermieden. Aus diesem Grund werden auch hier für die ersten beiden Varianten keine genauen Berechnungsvorschriften für die Zurückweisungswahrscheinlichkeiten vorgestellt.

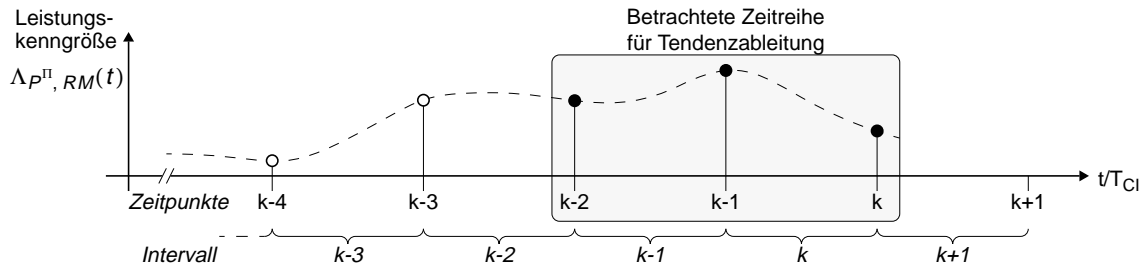
#### 4.5.2.2.4 Variante III – Tripelverfahren

Bei dieser Variante gehen die letzten drei Leistungswerte  $\Lambda_{P^\Pi, RM}(k)$ ,  $\Lambda_{P^\Pi, RM}(k-1)$  und  $\Lambda_{P^\Pi, RM}(k-2)$  in die Rechnung ein (siehe Bild 4-10). Es wird deshalb auch als *Tripel-Verfahren* [78] bezeichnet. Die Art der Lastinterpretation in nachfolgendem Algorithmus ist angelehnt an das Tripel-Verfahren von BADEL ET AL. [78]<sup>40</sup>.

<sup>39</sup> Wie bereits erwähnt wurde, liegen die Koordinaten der ermittelnden Leistungswerte im Lastraum hier nicht vor. Die *Richtungsinformation*, die aus dem *Gradienten*

$$\nabla\Lambda_{P^\Pi}(A_1, \dots, A_K) = \left\{ \frac{\partial\Lambda_{P^\Pi}}{\partial A_1}, \dots, \frac{\partial\Lambda_{P^\Pi}}{\partial A_K} \right\}$$

entnommen werden könnte, kann hier nicht bestimmt werden.



**Bild 4-10:** Verlauf der *Power-Produkt-Funktion*  $\Lambda_{P^II, RM}(t=kT_{CI})$ . Die Tendenz wird aus den letzten drei Power-Werten abgeleitet.

Bei diesem Ansatz wird davon ausgegangen, dass man die Lage des Maximums durch Vergleich dreier zeitlich aufeinanderfolgender Messwerte lokalisieren kann. Ist die Zeitreihe monoton wachsend, so befindet man sich ‚links‘ vom Extremum, ist sie monoton fallend, befindet man sich ‚rechts‘ davon. Ist der mittlere Wert größer als die beiden anderen oder sind alle annähernd gleich, so befindet man sich ‚nahe‘ oder ‚im‘ Maximum. Ist der mittlere Wert kleiner als die beiden anderen, so bewegt man sich wieder zurück Richtung Maximum.

Dieses Wissen muss nun in Regeln für die Berechnung der Zurückweisungswahrscheinlichkeiten umgesetzt werden. Bei einer Zeitreihe mit drei Elementen sind verschiedene Vergleichsmuster denkbar. Tabelle 4-1 zeigt alle typischen Muster und die Gleichungen (4-30), (4-31) und (4-32) die Vergleichsoperationen der entsprechenden Zurückweisungsstrategien.

Leistungsverlauf	Typische Muster	Strategie
„ansteigend“		Laststimulierung
„fallend“		Lastreduzierung
„ungefähr gleichbleibend“, „wieder ansteigend“ bzw. „wieder abfallend“ <sup>a</sup>		Lastbeibehaltung

a. Bei diesem Muster könnte auch mit der Strategie ‚Lastreduzierend‘ reagiert werden. Hier muss aber davon ausgegangen werden, dass die Lastrichtung in der dieses lokale Maximum durchlaufen wurde, unbekannt ist.

**Tabelle 4-1:** Charakteristische Muster für Leistungsverläufe und resultierende Strategien

Im *Überlastzustand* werden die Zurückweisungswahrscheinlichkeiten zyklisch neu berechnet. Dabei können drei Teilstrategien unterschieden werden:

- *Lastreduzierung* durch *Vergrößerung* der  $\pi_{Rej, v}$  falls folgende Bedingung erfüllt ist:<sup>41</sup>

$$\{[(k-2) > (k-1)] \wedge [(k-1) \geq (k)]\} \vee \{[(k-2) = (k-1)] \wedge [(k-1) > (k)]\}. \quad (4-30)$$

<sup>40</sup> Wesentliche Unterschiede bei BADEL: keine verschiedenen Lastbereiche werden unterschieden, neben den Mittelwerten werden auch die Varianzen in die Schätzung mit einbezogen, ereignisgesteuerte Regelung.

Anwendung: Regulierung des Multiprogramming-Grades durch das Betriebssystem.

<sup>41</sup> Aus Platzgründen wurde bei den Vergleichsbedingungen das *Power-Produkt-Symbol*  $\Lambda_{P^II, RM}$  weggelassen.

- *Laststimulierung durch Verkleinerung* der  $\pi_{Rej,v}$  bei folgender Bedingung:

$$\{[(k-2) < (k-1)] \wedge [(k-1) \leq (k)]\} \vee \{[(k-2) = (k-1)] \wedge [(k-1) < (k)]\}. \quad (4-31)$$

- *Lastbeibehaltung durch Nichtveränderung* der  $\pi_{Rej,v}$  wenn

$$\begin{aligned} & \{[(k-2) < (k-1)] \wedge [(k-1) > (k)]\} \vee \\ & \{[(k-2) > (k-1)] \wedge [(k-1) < (k)]\} \vee \\ & \{(k-2) = (k-1) = (k)\} \end{aligned} \quad (4-32)$$

zutritt.

Die Abprüfung auf Gleichheit in obigen Bedingungen sollte nicht zu hart interpretiert werden. Vielmehr sei eine gewisse Unschärfe zugelassen, d.h. auf *ungefähr gleich* wird verglichen. Die Art und Weise wie dies realisiert wird, hängt vom Implementierungsaufwand ab. Ein einfaches Rundungsverfahren durch Stellenabschneidung kann hierfür bereits ausreichend sein.

Die Vergrößerung bzw. Verkleinerung der  $\pi_{Rej,v}(k+1)$  soll nach der Strategie „je größer der Power-Anteil einer Verkehrsklasse, umso stärker wird ihre Zurückweisungswahrscheinlichkeit variiert“ erfolgen. Diese Strategie entspricht der Richtungssuche im Lastraum nach dem Ort des vermutenden Optimums. Eine Diskussion und Bewertung dieser Vorgehensweise erfolgt in Kapitel 7. Diese Strategie soll in folgende Berechnungsvorschriften umgesetzt werden (siehe auch Bild 4-9):

- *Subtraktive Verkleinerung* der  $\pi_{Rej,v}(k+1)$  zur Laststimulierung:

$$\pi_{Rej,v}(k+1) = \text{Max}\{0, \pi_{Rej,v}(k) - \delta_d \cdot g_v(k)\} \text{ mit z.B. } \delta_d = 0.25 \quad (4-33)$$

und den Gewichtungsfaktoren  $g_v$ :

$$g_v(k) = \frac{\Lambda_{P, RM, v}(k)}{\Lambda_{P_{RM}^\Sigma}(k)} = \frac{\Lambda_{P, RM, v}(k)}{\sum_v \Lambda_{P, RM, v}(k)} \text{ und } \sum_v g_v(k) = 1. \quad (4-34)$$

- *Multiplikative Vergrößerung* der  $\pi_{Rej,v}(k+1)$  zur Lastreduzierung:

$$\pi_{Rej,v}(k+1) = \text{Min}\{1, \pi_{Rej,v}(k) \cdot [1 + \delta_u \cdot g_v(k)]\} \text{ mit z.B. } \delta_u = 1.5. \quad (4-35)$$

Für den Sonderfall, dass  $\pi_{Rej,v}(k) = 0$  ist, soll folgende Einstellung vorgenommen werden:

$$\pi_{Rej,v}(k+1) = \text{Min}\{1, \delta_u \cdot g_v(k)\}. \quad (4-36)$$

Die Parameter  $\delta_d$  bzw.  $\delta_u$  stellen Dämpfungsfaktoren dar. Abschließend soll noch einmal darauf hingewiesen werden, dass im *Normallastzustand* alle Anforderungen angenommen werden, d.h.,  $\pi_{Rej,v}(k+1) = 0$  für alle  $v$ .

Ergebnisse und weitere Ausführungen zu der Leistungsregelung basierend auf der *Power-Kenngröße* werden in Kapitel 7 vorgestellt.



### 4.5.3 Diskussion der Power-Produkt-Funktion

Wie bereits erwähnt wurde, basiert das im Rahmen dieser Arbeit verwendete *Power*-Maß auf den beobachteten REQUEST-Nachrichten. Somit geht in die Leistungsfunktion die Annahme- und nicht die Komplettierungsrate ein. Begründet wird dies damit, dass dieser Regelkreis auf *neuen Verkehr* sofort reagieren soll. Gemäß der Berechnungsvorschrift ist die *Power-Produkt-Funktion* direkt abhängig von den gemessenen Leistungskenngrößen

$$\Lambda_{P_{RM}^{\Pi}} = f(\overrightarrow{\Lambda_{P, RM}}) = f(\overrightarrow{\hat{\phi}_{out, RM}}, \overrightarrow{\hat{\tau}_{RM}}), \quad (4-37)$$

wobei die Leistungskenngrößen direkt von den Lastkenngrößen *nichtlinear* abhängen:

$$\overrightarrow{\hat{\phi}_{out, RM}} = f_1(\overrightarrow{\hat{\phi}_{in, RM}}) \text{ und } \overrightarrow{\hat{\tau}_{RM}} = f_2(\overrightarrow{\hat{\phi}_{in, RM}}). \quad (4-38)$$

Durch die Rückwirkung auf den Verkehr mittels der Zurückweisungswahrscheinlichkeiten entsteht ein Regelkreis mit einer nichtlinearen Kopplung zwischen Leistung und Last.

Weiterhin umfasst der Wertebereich des *Power-Produkts* mehrere Größenordnungen<sup>42</sup>. Dieser Umstand muss ggf. im Algorithmus berücksichtigt werden. Der für die Leistungsoptimierung grundlegende Vorteil eines globalen Maximums der *Power-Produkt-Funktion* hat eine zentrale Bedeutung. Ein Regelkreis auf Basis dieser Kenngröße ist weniger als Überlastregelung für kurzfristige Lastspitzen geeignet, sondern sollte vielmehr für eine mittel- bis langfristige Leistungsoptimierung im Lastbereich „in der Nähe“ der Nennkapazität verwendet werden. Weitere Eigenschaften des *Power-Produkts* werden in Anhang B.4 diskutiert.

Wie bereits erwähnt wurde, wäre eine Verbesserung des Verfahrens möglich, wenn neben der Messung von Leistungsdaten zusätzlich die zugehörige Lastinformation erfasst werden würde. Dies würde die „Orientierung“ im Lastraum wesentlich vereinfachen. Das *Power-Produkt* kann auch als *Potentialfeld* interpretiert werden. Durch permanente Aufzeichnung des Leistungswertes mit den zugehörigen Lastkoordinaten könnte ein Vermittlungsrechner oder Prozessorknoten die ihm eigene Leistungsfunktion („Leistungsfeld“) *lernen*. Durch diesen Lernalgorithmus wird im Laufe der Zeit eine Datenbasis (sogenanntes „Expertenwissen“) verfügbar, die eine Leistungsregelung vereinfachen und vor allem effizienter machen würde. Untersuchungen zur Wirtschaftlichkeit und Leistungsfähigkeit dieses aufwendigen Verfahrens sollen aber nicht Gegenstand dieser Arbeit sein.

Ein weiterer Aspekt betrifft den Einfluss der Antwortzeiten. Von der internen Systemorganisation eines Prozessorknotens hängt es ab, ob die Ermittlung einer gemeinsamen mittleren Antwortzeit ausreichend ist oder ob klassenindividuelle Messungen sinnvoll sind. Diese Entscheidung wird beispielsweise durch Pufferabfertigungsdisziplinen, Prioritäten, Prioritätsmechanismen und Strategien zur Überlastabwehr beeinflusst. Entfallen diese Eigenschaften oder sind die Kosten für Messungen und Datenaufbereitung zu hoch, kann mit einer gemeinsamen Antwortzeitgröße gearbeitet werden.

---

<sup>42</sup> Die *Antwortzeiten*  $\tau_{RM}$  der REQUEST-Meldungen umfassen typischerweise maximal drei Größenordnungen, da die mittlere Bedienzeit im Millisekunden- und die Zeitüberwachung durch Protokolltimer im Sekundenbereich liegt ( $h \leq \tau \leq T_{303}$  bei Q.2931). Der Durchsatz  $\phi_{RM}$  liegt im Bereich von ( $0 \leq \phi \leq 1/h$ ) und umfasst damit „sehr viele“ Größenordnungen. Weiterhin gehen Divisionen und Multiplikationen als Rechenoperationen in das Power-Produkt ein.

## 4.6 Algorithmen zur Lastzuteilung von A- an B-Prozessor-knoten

In dem betrachteten generischen B-ISDN-Vermittlungsrechner arbeiten auf verschiedenen Funktionsebenen jeweils mehrere Prozessorknoten in Lastteilung. Um eine möglichst gleichmäßige Belastung des gesamten Vermittlungsrechners anzustreben, sind auf oberster Raumebene globale Lastregelkreise notwendig. Für diese Anforderung wird ebenfalls ein neues Verfahren vorgestellt.

### 4.6.1 Einführung

Ein Vermittlungsrechner mit einer Mehrprozessorarchitektur lässt prinzipiell die parallele und verteilte Bearbeitung mehrerer Dienstanforderungen zu. Für eine gute Dienstgüte und ein optimales Leistungsverhalten des Gesamtsystems ist eine ausgewogene und möglichst gleichmäßige Belastung der einzelnen Verarbeitungseinheiten unabdingbar. Dies ist die Aufgabe eines Lastverteilungsverfahrens. Je nach Steuerungsarchitektur, Funktionsverteilung und Regelungsziel wird diese Art von Mechanismus auch als *Scheduling* (z.B. Semidynamic Scheduling [527]), *Load Forwarding*, *Load Distributing* [489], *Load Sharing* [165,547], *Load Balancing* [110,501], *Load Routing* u.a. bezeichnet. Eine weitere Unterteilung findet in statische und dynamische Verfahren statt [33].

Diese Algorithmen sind seit vielen Jahren ein intensiv bearbeitetes Forschungsgebiet aus dem Bereich der Rechnerarchitekturen und Rechnervernetzung (z.B. [148,548,499,583,585,576]). Die Liste von Veröffentlichungen ist dementsprechend groß; eine gute Übersicht, Klassifizierungskriterien sowie eine Vielzahl weiterer Referenzen wird in [488] gegeben.

Die großen Unterschiede im Vergleich zur hier betrachteten Anwendung sind die Verkehrscharakterisierung und insbesondere die Echtzeitfähigkeitsanforderungen [336]. Diese speziellen Kriterien werden in Abschnitt 4.6.3 ausführlich erläutert. Zuvor werden kurz bekannte Ansätze für Vermittlungsrechner, die als parallele und verteilte Multiprozessorsysteme konzipiert sind, vorgestellt.

### 4.6.2 Bekannte Verfahren für Vermittlungsrechner

Klassische SPC-, aber auch ISDN-Netzknoten weisen oft eine hierarchische verteilte Steuerungsstruktur auf. Mit der *Lastverteilung* für Steuerungsarchitekturen dieser Systeme beschäftigt sich beispielsweise [64,113,162,334,407,577,578,579,588,591,592,595,601,602]. Weiterhin wurden spezielle Lastverteilalgorithmen für Netzelemente des IN [587,590] oder auch Paketknoten [593] entwickelt.

Ist es einer Verarbeitungseinheit gestattet, einen Auftrag an eine andere Instanz derselben Funktionsebene weiterzuleiten, dann bezeichnet man diesen Mechanismus oft als *Load Balancing*. Dabei können prinzipiell zwei Teilstrategien unterschieden werden: Die *Transfer Policy* muss entscheiden, ob eine neue Anforderung lokal bearbeitet oder weitergeleitet wird; im Falle der Weiterleitung muss die *Location Policy* entscheiden, welchem Prozessorknoten diese zugeteilt wird.<sup>43</sup> In [603,604] wird ein entsprechendes Verfahren für eine ATM-Vermittlungssteue-

rung vorgeschlagen (siehe auch Anhang A.3.1). Nachteilig beim *Load Balancing* sind, dass die Lastweiterleitung eine Blindlast darstellt, dass sich bestimmte Weiterleitungseffekte<sup>44</sup> einstellen können oder dass jeder Knoten regelmäßig über den Lastzustand aller anderen informiert werden muss.

Für die Lastverteilung in einer lose gekoppelten Vermittlungssteuerung wird in [594] ein spezielles Verfahren vorgestellt. Der Algorithmus besteht wiederum aus den zwei Teilstrategien und ist völlig verteilt realisiert. In [581] dagegen wird eine eng gekoppelte Multiprozessorsteuerung betrachtet. Ein gemeinsam genutzter Speicher dient u. a. zum Austausch von Ressourceninformationen. Die gleichmäßige CPU-Auslastung aller Prozessorknoten hat [582] als spezielles Optimierungsziel.

Besitzt eine Vermittlungssteuerung sehr viele Prozessorknoten, dann kann der Kommunikationsaufwand zum Austausch von Lastinformationen erheblich werden. Um dem entgegenzuwirken wird in [486] vorgeschlagen, dass ein Knoten nur eine bestimmte Teilmenge (auch *buddy set* genannt) informiert.

In all diesen Algorithmen wird das Kriterium der Echtzeitfähigkeit berücksichtigt, aber es wird keine dedizierte Dienstunterscheidung vorgenommen.<sup>45</sup> In einer B-ISDN-Vermittlungssteuerung sollten aber für eine ausgewogene Lastverteilung unterschiedliche Verkehrsklassen in Betracht gezogen werden.

### 4.6.3 Besonderheiten der neu entwickelten Algorithmen

Die Leistung des Gesamtsystems hängt hauptsächlich von der aktuellen Belastung der einzelnen Verarbeitungseinheiten (d.h. der momentanen Ressourcenbelegung) und von der Intensität und Zusammensetzung des neu ankommenden Verkehrs ab. Diese Informationen müssen laufend durch Sensoren erfasst, aufbereitet und an die entsprechenden Steuerungsinstanzen verteilt werden. Hier sind verschiedenste Ansätze vorstellbar, die zum einen von der Systemarchitektur<sup>46</sup> und der Funktionsverteilung abhängen und zum anderen einen Kompromiss zwischen Aufwand<sup>47</sup> und Leistungsverhalten finden müssen.

Von der Architekturseite her ist der einfachste Ansatz der, einen zentralen Prozessor für die Verteilung des Eingangsverkehrs auf nachfolgende Einheiten vorzusehen. Hierbei durchläuft der *gesamte* ankommende und abgehende *Verkehr* diese zentrale Instanz. Dieser Typ von Steuerungsstruktur wurde beispielsweise in [63] untersucht. Solche Systeme lassen relativ einfache Lastverteilungsalgorithmen zu und eignen sich prinzipiell für kleinere bis mittlere Netzknotengrößen. Auf der anderen Seite wird der Skalierbarkeitsbereich durch die zentrale Instanz sehr stark beschränkt, und es müssen erhebliche Vorkehrungen getroffen werden, um die Anforderungen an Ausfallsicherheit bzw. Zuverlässigkeit zu erfüllen. Aus diesem Grund werden für größere Vermittlungssysteme mehrere Prozessoreinheiten auf den einzelnen Funktionsebenen vorgesehen (siehe auch Anhang A.2.2 und A.3).

<sup>43</sup> Eine weitere Strategie ist die *Information Policy*, die entweder zeit- oder ereignisgesteuert ausgeführt wird. Weiterhin wird unterschieden, ob überlastete Prozessoren nach „Abnehmern“ suchen (*Sender-initiated*), unterlastete Einheiten sich aktiv anbieten (*Receiver-initiated*) oder beide Strategien realisiert sind (*Symmetrically-initiated*).

<sup>44</sup> Beispielsweise Gefahr des sogenannten *Processor Thrashings* [165]. Hierbei sind alle Prozessorknoten damit beschäftigt, ihre Aufträge auszulagern, d.h. nahezu die gesamte Rechenzeit wird für die Lastweiterleitung benötigt. Damit geht der Nutzdurchsatz auf null zurück.

<sup>45</sup> Algorithmen die mehrere Auftragsklassen (*multi class systems*) berücksichtigen, aber weniger Wert auf Echtzeitfähigkeit legen, sind dagegen aus der Computerwelt bekannt (z.B. [399]).

<sup>46</sup> Z.B. Steuerungsstruktur, Anzahl Prozessoren auf den einzelnen Funktionsebenen oder Kommunikationspfade.

<sup>47</sup> Beispielsweise Typ, Anzahl und Ort der Sensoren, Mess- und Regelintervalldauer, relativ genaue oder approximative Verfahren usw.

Hier soll die in Abschnitt 4.2 diskutierte Steuerungsstruktur betrachtet werden. Dies ist eine Architektur mit mehreren Funktionsebenen, wobei die ersten beiden Ebenen mit parallelen Verarbeitungseinheiten ausgestattet sind und nach dem Lastteilungsprinzip arbeiten. Die Vorteile solcher Vermittlungsrechner wurden bereits in Abschnitt 2.6.5 diskutiert. Auf der anderen Seite werden an die Lastverteilungsalgorithmen hier besondere Anforderungen gestellt. Jeder Prozessorknoten der A-Ebene ( $PN_{A,i}$ ) muss entscheiden, an welchen Prozessor der B-Ebene ( $PN_{B,j}$ ) eine neue Dienstanforderung weitergeleitet wird (siehe auch Bild 4-1). Die Entscheidungen für die Lastverteilung müssen hier parallel auf räumlich verteilten Prozessoren ausgeführt und ggf. auch getroffen werden. Weitere Probleme stellen z.B. die Informationsverteilung – jeder  $PN_{A,i}$  sollte wissen, wie die einzelnen  $PN_{B,j}$  momentan belastet sind – bzw. die Auswirkungen verschiedener Regelentscheidungen<sup>48</sup> dar.

Ein Ansatz, diesem Problemen auszuweichen, stellt wiederum eine zentrale Instanz („Master“) dar, an die sämtliche Lastinformationen gesendet bzw. von ihr an die entsprechenden Stellen regelmäßig verteilt werden. Diese Lösung wird beispielsweise in [410] eingesetzt.

## Neuer Algorithmus

Ein *neues Konzept* für einen *verteilten Lastverteilungsalgorithmus* wird im Rahmen dieser Arbeit vorgeschlagen. Die wichtigsten Merkmale sind:

- Von der Funktionsverteilung her sei ein B-Prozessor die am stärksten belastete Komponente im Laufe der Dienstkomplettierung, d.h. A- und C-Steuerungseinheiten seien so dimensioniert, so dass bei „voller Last“ diese noch ausreichend Reserven besitzen.
- Ein A-Prozessor muss für jede neue Dienstanforderung entscheiden, an welchen B-Prozessor diese weitergeleitet wird. Alle Folgemeldungen dieses Dienstes werden an denselben Prozessorknoten geroutet.
- Ein B-Prozessor kann eine neue Dienstanforderung nur annehmen, ablehnen oder verwerfen, aber *nicht* an einen anderen B-Prozessor *weiterleiten*. Ein sogenanntes *Load Balancing*, wie in [604] beschrieben, findet somit hier nicht statt.
- Auf eine Erfassung von Lastzustandsinformation direkt in den B-Prozessoren wird verzichtet. Dies hat den Vorteil, dass keine zusätzlichen Betriebsmittel für diese Aufgabe – im ohnehin schwächsten Glied – reserviert werden müssen und die Nennkapazität zusätzlich reduziert wird.
- Vielmehr soll jeder A-Prozessor indirekt notwendige Informationen über den Belastungszustand der B-Ebene aus dem zurückfließenden Steuerungsverkehr gewinnen. Beispielsweise kann dies durch die Messung von *Antwortzeiten*<sup>49</sup> oder *Durchsatzraten* realisiert werden.
- Dem Dienstaspekt soll durch Unterscheidung von verschiedenen Verkehrsklassen Rechnung getragen werden.

---

<sup>48</sup> Z.B. könnten mehrere A-Prozessoren gleichzeitig Anforderungen an einen bestimmten (unterlasteten) B-Prozessor senden („gewisse Effekte verstärken sich“).

<sup>49</sup> In [484] wird untersucht, wie der Zustand eines Systems indirekt aus Verkehrsinformationen geschätzt werden kann. Hierzu wird ein zusätzlicher Testverkehr, mit einer wesentlich kleineren Intensität als der eigentliche Nutzverkehr, in das System eingespeist und anschließend ausgewertet.

In dieselbe Richtung gehen auch Vorschläge, um den Belastungszustand eines paket- oder zellvermittelnden Netzes abzuschätzen. Hierzu werden von der Quelle regelmäßig Testpakete bzw. -zellen (auch *Tourist Cells*, *Sample Cells* bzw. *Scout Packets* genannt; siehe z.B. [124, 268, 594]) an die Senke gesendet, die diese wiederum zurückschickt. Da die Durchlaufzeit eine monoton steigende Funktion in Abhängigkeit der Netzbelastung ist, kann dieses *Round-Trip-Delay* als einfacher Lastindikator verwendet werden.

Sogenannte *Choke Packets* (auch *Source Quench Messages* genannt [271]) dagegen werden in verbindungslosen Netzen von überlasteten Knoten eingesetzt, um die Quellen über einen kritischen Belastungszustand zu informieren. Diese Rückwirkung findet aber erst unter Überlast statt [205].

Der vorletzte Punkt, dass jede  $A$ -Verarbeitungseinheit regelmäßig Messungen durchführen muss, um Rückschlüsse auf die Ressourcenbelegung der  $B$ -Prozessoren zu erhalten, stellt den einzigen wesentlichen Nachteil dar, der sich allerdings durch folgende Implementierungsaspekte entkräften lassen kann. Durch die vorgenommene Funktionsverteilung besitzt ein  $A$ -Prozessor ausreichend Reserven für Lastmessungen. Außerdem werden hier die Lastkenngrößen ‚Antwortzeit‘ und ‚Nutzdurchsatz‘ verwendet,<sup>50</sup> die leichter zu messen sind (z. B. im Vergleich zur Prozessorauslastung).

Neben den Vorteilen, die oben bereits angedeutet wurden, ist dieser verteilte Algorithmus prinzipiell sehr robust hinsichtlich Ausfallsicherheit, Skalier- und Erweiterbarkeit.<sup>51</sup> Aufgrund der Eigenschaften dieses Verfahrens soll dieser Last- und Leistungsregelmechanismus als *Distributed Load Balancing (DLB)* bezeichnet werden. Im folgenden Abschnitt sollen die Einzelheiten des Verfahrens näher ausgeführt werden.

## 4.6.4 *Distributed Load Balancing* – Algorithmen

Die übergeordnete Strategie muss sein, neuen Verkehr vorrangig den weniger belasteten Prozessorknoten zuzuteilen. Als Lastindikatoren werden klassenindividuelle Antwortzeiten und Nutzdurchsätze sowie daraus abgeleitete *Power*-Maße verwendet. Es werden drei Algorithmen, die jeweils einen dieser Kenngrößentypen als Optimierungsziel besitzen, vorgestellt.

### 4.6.4.1 Erfassung der Lastzustandsinformationen

Auf den Systemzustand eines  $PN_{B,j}$  wird anhand gemessener Antwortzeiten  $\tau$  und Durchsatzraten  $\phi$  rückgeschlossen. Aus der Vielzahl von Meldungstypen, die im Signalisierverkehrsstrom auftreten, werden aus Kostengründen hier von den Messeinrichtungen nur die REQUEST-Nachrichten (siehe Abschnitt 3.1.3.2) erfasst. Damit wird auf eine detaillierte Aufzeichnung des Verkehrsgeschehens bewusst verzichtet (es werden beispielsweise nicht die Ablehnungs- oder Komplettierungsrate ermittelt), aber diese reduzierte Datenbasis bietet dennoch ausreichend Information für Aussagen hinsichtlich der Belastung einzelner  $B$ -Prozessoren.

---

<sup>50</sup> Eine Meldungsrate wird üblicherweise mittels einer Zählvariablen und Antwortzeiten anhand von Zeitstempeln gemessen. Diese Aufgaben könnten hier in die Softwaremodule der Meldungsanalyse integriert werden. Bei jeder ankommenden Signalisier- und Steuerungsnachricht müssen bestimmte Datenfelder in den einzelnen Protokollschichten ausgewertet werden. Die Schicht-2-Signalisierung sei hier Aufgabe der  $A$ -Ebene. Für die Weiterleitung von Signalisiernachrichten an die  $B$ -Ebene muss jeder  $A$ -Prozessor mindestens eine Teilmenge (z. B. Protokollvariante, Meldungstyp, *Call Reference*) der Schicht-3-Signalisierung analysieren. In einer lokalen Tabelle wird von einem  $PN_{A,i}$  festgehalten, an welchen  $PN_{B,j}$  eine neue Dienstanforderung (mit *Call Reference* = CR) weitergeleitet wurde. In Spalte 1 wird die *Call Reference*, in Spalte 2 Prozessor  $B_j$  eingetragen. Damit ist sichergestellt, dass alle Folgemeldungen zum selben  $PN_{B,j}$  geroutet werden. (Dabei können natürlich die verschiedenen Verbindungsabschnitte einer Kommunikationsbeziehung auf verschiedenen  $A$ -Prozessoren enden.)

In einer weiteren Spalte kann der Zeitpunkt des Beginns („Bearbeitung einer REQUEST-Meldung“) einer Kommunikationsbeziehung eingetragen werden, anhand derer sich relativ einfach spezifische Antwortzeiten ermitteln lassen. Die Erweiterung einer Nachricht um ein zusätzliches Feld für Zeitstempelinformation kann somit hier entfallen. Anhand dieser Tabelle kann weiterhin ein  $PN_{A,i}$  feststellen, wieviele Verbindungen (Kommunikationsbeziehungen) aus seiner Sicht momentan auf jedem  $PN_{B,j}$  aktiv sind. Diese Information stellt eine weitere Möglichkeit dar, einen Rückschluss auf den Belastungszustand der  $B$ -Ebene zu bekommen.

Anmerkung: Für eine hardwaremäßige Realisierung dieser Tabelle lassen sich vorteilhaft inhaltsadressierbare Speicherbausteine (CAM) einsetzen. Solche Chips sind mittlerweile mit einer relativ großen Speicherkapazität verfügbar (z. B. [706, 707, 710]).

<sup>51</sup> Ab einer „gewissen Anzahl“ von Prozessorknoten auf der  $B$ -Ebene kann dieses Verfahren bzgl. der zu erfassenden und verarbeitenden Informationsmenge relativ aufwendig werden. Zumal hier der Verkehr ja auf mehrere Klassen abstrahiert wird und demzufolge auch klassenindividuelle Messungen durchgeführt werden müssen. Dem kann beispielsweise dadurch begegnet werden, indem jeder  $PN_{A,i}$  nur eine Teilmenge der  $B$ -Prozessoren bedient (d. h. neue Dienstanforderungen zustellt).

Damit werden von jedem  $PN_{A,i}$  in jedem Mess- und Regelintervall  $k$  folgende direkte *Leistungskenngrößen* gemessen (siehe Bild 4-11):

- Annahmerate  $\hat{\phi}_{out, RM, j, v}(k)$  einer REQUEST-Nachricht von  $PN_{B,j}$  für Verkehrsklasse  $v$ ,
- Antwortzeit  $\hat{\tau}_{RM, j, v}(k)$  einer REQUEST-Nachricht von  $PN_{B,j}$  für Verkehrsklasse  $v$ .

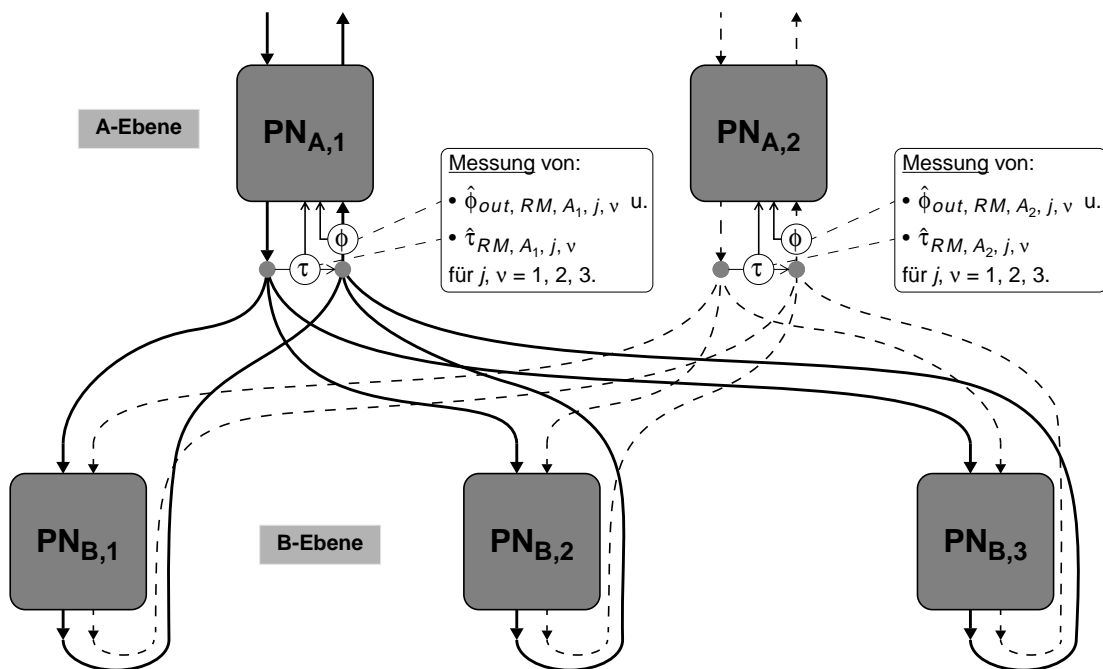
Weiterhin wird das kombinierte Leistungsmaß *Power*  $\Lambda_{P, RM, j, v}(k)$  (Gl. (3-10)) berechnet:

$$\Lambda_{P, RM, j, v}(k) = \frac{\hat{\phi}_{out, RM, j, v}(k)}{\hat{\tau}_{RM, j, v}(k)}. \quad (4-39)$$

Aus diesen einzelnen Powergrößen wird für jeden  $PN_{B,j}$  das *Power-Produkt*  $\Lambda_{P^\pi, RM, B_j}(k)$  nach Gl. (3-12) berechnet. Der Vektor

$$\vec{\Lambda}_{P^\pi, RM, A_i}(k) = \begin{bmatrix} \Lambda_{P^\pi, RM, B_1}(k) \\ \dots \\ \Lambda_{P^\pi, RM, B_M}(k) \end{bmatrix} \quad (4-40)$$

repräsentiert somit ein *Maß für die aktuelle Lastverteilung auf der B-Ebene aus Sicht eines A-Prozessors*.



**Bild 4-11:** Lastsensoren der A-Ebene (vereinfachte Darstellung ohne  $PN_C$  sowie nur drei  $PN_{B,j}$  und zwei  $PN_{A,i}$ ).

Der *Power-Vektor*  $\vec{\Lambda}_{P^\pi}$  bildet die Grundlage für die Lastverteilung und damit einer Leistungsregelung im Vermittlungsrechner. Zunächst erhebt sich für einen  $PN_{A,i}$  die Frage, wie die Größe des *Power-Produkts*  $\Lambda_{P^\pi, RM, B_j}$  eines  $PN_{B,j}$  zu interpretieren ist, d.h. in welcher Lastregion befindet sich ein *B-Prozessor*?<sup>52</sup> Um dieses Problem zu vermeiden, sollen vorbereitend Regelalgorithmen, die direkt auf den gemessenen Leistungskenngrößen beruhen, entwickelt werden.

#### 4.6.4.2 Verfahren auf Basis der Antwortzeit

Die *Antwortzeit*  $\tau(A)$  in Abhängigkeit der Last stellt üblicherweise eine *streng monoton steigende Funktion* dar. Ein Regelmechanismus, der bei Überschreiten eines bestimmten Schwellwertes  $\tau_{Th}$  aktiviert wird, stellt ein relativ einfaches und häufig eingesetztes Verfahren für Monoprozessorsysteme dar. Für das betrachtete Mehrprozessorsystem soll folgender Lastverteilungsalgorithmus untersucht werden.

Wie im vorherigen Abschnitt erläutert wurde, ermittelt jeder  $PN_{A,i}$  für jeden  $PN_{B,j}$  pro Regelintervall einen *Antwortzeitvektor*<sup>53</sup>:

$$\overrightarrow{\hat{\tau}_{RM, A_i, B_j}(k)} = \begin{bmatrix} \hat{\tau}_{RM, B_j, 1}(k) \\ \dots \\ \hat{\tau}_{RM, B_j, K}(k) \end{bmatrix}. \quad (4-41)$$

Zuätzlich wird noch für jeden  $PN_{B,j}$  die klassenunabhängige *mittlere Antwortzeit*  $\bar{\tau}$  berechnet:

$$\overline{\hat{\tau}_{RM, B_j}(k)} = \frac{1}{K} \sum_{v=1}^K \hat{\tau}_{RM, B_j, v}(k). \quad (4-42)$$

Die *Zuteilungswahrscheinlichkeit*  $\pi_{F, B_j}(k)$  (auch *Forwarding* oder *Routing Probability*) bezeichnet den Prozentsatz neuer Dienstanforderungen, die im Regelintervall  $k$  an  $PN_{B,j}$  weitergeleitet werden. Folgende *Strategie* für eine *Lastverteilung* soll zunächst angewendet werden:

$$\pi_{F, B_j}(k) \sim \frac{1}{\overline{\hat{\tau}_{RM, B_j}(k)}}, \quad (4-43)$$

d.h. je größer die Antwortzeit, umso weniger wird an neuem Verkehr zugeteilt.<sup>54</sup> Damit ergibt sich folgender normierter *Lastzuteilungsvektor*:

$$\overrightarrow{\pi_{F, A_i}(k)} = \frac{1}{\sum_{j=1}^M \frac{1}{\overline{\hat{\tau}_{RM, B_j}(k)}}} \begin{bmatrix} [\overline{\hat{\tau}_{RM, B_1}(k)}]^{-1} \\ \dots \\ [\overline{\hat{\tau}_{RM, B_M}(k)}]^{-1} \end{bmatrix} = \begin{bmatrix} \pi_{F, B_1}(k) \\ \dots \\ \pi_{F, B_M}(k) \end{bmatrix}. \quad (4-44)$$

Dieser einfache Algorithmus verteilt den Verkehr auf die  $B$ -Prozessoren, ohne die einzelnen Dienste zu unterscheiden. Messungen zu diesem Verfahren werden in Abschnitt 8.2 präsentiert. Eine Verfeinerung dieser Strategie hinsichtlich der Berücksichtigung einzelner Verkehrsklassen wäre hier, da die Messwerte bereits vorliegen, ohne Schwierigkeit möglich.

<sup>52</sup> Ist das Angebot  $A$  für einen  $PN_{B,j}$  gleich Null, dann ist auch  $\Lambda_{P,II} = 0$ . Wird die Last  $A$  erhöht, dann steigt der Wert von  $\Lambda_{P,II}$  an. Ab einem gewissen Angebot fällt  $\Lambda_{P,II}$  wieder. Es wird also prinzipiell ein Maximum durchlaufen, dessen Koordinaten idealerweise den Arbeitspunkt des Vermittlungsrechners darstellen. Bei Berücksichtigung mehrerer Verkehrsklassen ist  $\Lambda_{P,II}$  eine mehrdimensionale Lastfunktion. Somit muss die Leistungsregelung versuchen, den Lastvektor „in das richtige Zielgebiet“ („Arbeitsraum“) zu führen (siehe auch Anhang B.4, S. 245ff).

<sup>53</sup> Bei  $M$   $B$ -Prozessoren und  $K$  Verkehrsklassen können prinzipiell alle Messwerte in einer  $M \times K$ -Matrix zusammengefasst werden.

<sup>54</sup> Aus dem Gebiet der Rechnertechnik gibt es ein ähnliches Verfahren, die sogenannte *Minimum Response Time Policy* [130]. Hierbei wird versucht, die mittlere Antwortzeit für alle Prozessorknoten zu minimieren.

### 4.6.4.3 Verfahren auf Basis des Durchsatzes

Der zweite Indikator typ, die Durchsatzkenngröße, bietet sich ebenfalls für einen Lastverteilalgorithmus an. Die *Durchsatzfunktion*  $\phi(A)$  ist allerdings nur für ideale Systeme monoton steigend. In der Realität trifft diese Eigenschaft nicht mehr zu. Typischerweise weist die stationäre Durchsatzfunktion ein Maximum  $\phi_{max}(A_{opt})$  auf (siehe z.B. Abschnitt B.3.2), das als Arbeitspunkt prädestiniert ist. In welchem Lastbereich sich ein  $PN_{B,j}$  befindet, ist somit nicht direkt aus dem gemessenen Durchsatz feststellbar. Eine Lösung dieses Problems bieten beispielsweise folgende Ansätze:

- Verwendung eines zusätzlichen Lastsensortyps – z.B. die Prozessorauslastungen  $\rho_{B,j}$ .<sup>55</sup>
- Messung der Ankunftsrate  $\lambda(t)$  und Protokollierung des zeitlichen Verlaufs von  $\phi(t)$ . Damit wird ein direkter Zusammenhang zwischen Last und Leistung gewonnen.

Da bewusst auf direkte Lastmessungen in den  $PN_{B,j}$ , und auch auf die Messung des ankommenden Verkehrs verzichtet werden soll, bieten sich hier die gemessenen Antwortzeiten als zusätzlicher Informationslieferant an. Im Detail soll folgender Lastverteilalgorithmus untersucht werden.

Jeder  $PN_{A,i}$  ermittelt für jeden  $PN_{B,j}$  pro Regelintervall einen *Durchsatzvektor*:

$$\overrightarrow{\hat{\phi}_{out, RM, A_i, B_j}(k)} = \begin{bmatrix} \hat{\phi}_{out, RM, B_j, 1}(k) \\ \dots \\ \hat{\phi}_{out, RM, B_j, K}(k) \end{bmatrix}. \quad (4-45)$$

Zuätzlich wird noch für jeden  $PN_{B,j}$  der klassenunabhängige *mittlere Durchsatz*  $\bar{\phi}$  berechnet:

$$\overline{\hat{\phi}_{out, RM, B_j}(k)} = \frac{1}{K} \sum_{v=1}^K \hat{\phi}_{out, RM, B_j, v}(k). \quad (4-46)$$

Der Durchsatzvektor bewegt sich in einem  $K$ -dimensionalen Lastraum. Dieser Lastraum soll auf *zwei Lastzustände* abstrahiert werden. Diese Unterteilung soll anhand der berechneten mittleren Antwortzeit  $\overline{\hat{\tau}_{RM, B_j}}$  (siehe Gl. (4-42)) vorgenommen werden. Für einen bestimmten Schwellwert  $\tau_{Th}$  seien die beiden Zustände  $X$  wie folgt definiert:

- $PN_{B,j}$  im *Normallastzustand* für  $\overline{\hat{\tau}_{RM, B_j}(k)} < \tau_{Th}$  ( $X_{B_j} = ,NL'$ ) bzw.
- $PN_{B,j}$  im *Überlastzustand* für  $\overline{\hat{\tau}_{RM, B_j}(k)} \geq \tau_{Th}$  ( $X_{B_j} = ,ÜL'$ ).

Der *Belastungszustand der B-Ebene* aus Sicht von  $PN_{A,i}$  kann durch folgenden Vektor wiedergegeben werden:

$$\overrightarrow{X_{A_i}(k)} = \begin{bmatrix} X_{B_1}(k) \\ \dots \\ X_{B_M}(k) \end{bmatrix}. \quad (4-47)$$

Hinsichtlich der Lastverteilstrategien können nun prinzipiell die in Tabelle 4-2 gegenübergestellten drei Fälle unterschieden werden.

Ergebnisse zu diesem einfachen Lastzuteilungsverfahren werden in Abschnitt 8.3 vorgestellt. Wiederum werden hier vorhandene klassenspezifische Lastdaten nicht genutzt. Diese Informationen werden aber bei dem im nächsten Abschnitt vorgestellten Verfahren berücksichtigt.

<sup>55</sup> Mögliche Interpretation:  $\phi$  und  $\rho$  beide „klein“  $\rightarrow$  „Normallastzustand“,  $\phi$  „klein“ und  $\rho$  „groß“  $\rightarrow$  „Überlastzustand“.



Belastungs- zustand	Strategie	Berechnung der Zuteilungswahrscheinlichkeiten $\pi_{F,B,j}$
① Kein $PN_{B,j}$ in Überlast	„Wer weniger hatte, bekommt jetzt mehr“: $\pi_{F, B_j}(k) \sim \frac{1}{\hat{\phi}_{out, RM, B_j}(k)}$	$\vec{\pi}_{F, A_i}(k) = \frac{1}{\sum_{j=1}^M \frac{1}{\hat{\phi}_{out, RM, B_j}(k)}} \begin{bmatrix} \frac{1}{[\hat{\phi}_{out, RM, B_1}(k)]^{-1}} \\ \dots \\ \frac{1}{[\hat{\phi}_{out, RM, B_M}(k)]^{-1}} \end{bmatrix}$
② Alle $PN_{B,j}$ in Überlast	„Wer mehr hatte, bekommt jetzt weniger“: $\pi_{F, B_j}(k) \sim \frac{1}{\hat{\phi}_{out, RM, B_j}(k)}$	$\vec{\pi}_{F, A_i}(k) = \frac{1}{\sum_{j=1}^M \frac{1}{\hat{\phi}_{out, RM, B_j}(k)}} \begin{bmatrix} \frac{1}{\hat{\phi}_{out, RM, B_1}(k)} \\ \dots \\ \frac{1}{\hat{\phi}_{out, RM, B_M}(k)} \end{bmatrix}$
③ Mindestens ein $PN_{B,j}$ in Normal- und mindestens ein $PN_{B,j}$ in Überlast	1. „Wer überlastet ist bekom- mt keinen neuen Verkehr“: $\pi_{F, B_j}(k) = 0$ wenn $X_{B,j} = \text{ÜL}'$ .  2. „Neue Dienstanforderun- gen werden entsprechend Strategie ① auf die nichtüber- lasteten $PN_{B,j}$ aufgeteilt“: $\pi_{F, B_j}(k) \sim \frac{1}{\hat{\phi}_{out, RM, B_j}(k)}$ wenn $X_{B,j} = \text{NL}'$ .	$\vec{\pi}_{F, A_i}(k) = \frac{1}{\sum_{\text{alle } X_{B_j} = \text{NL}} \frac{1}{\hat{\phi}_{out, RM, B_j}(k)}} \cdot \begin{bmatrix} \dots \\ \frac{1}{[\hat{\phi}_{out, RM, B_j}(k)]^{-1}} \quad \forall X_{B_j} = \text{NL} \\ \dots \\ 0 \quad \forall X_{B_j} = \text{ÜL} \\ \dots \end{bmatrix}$

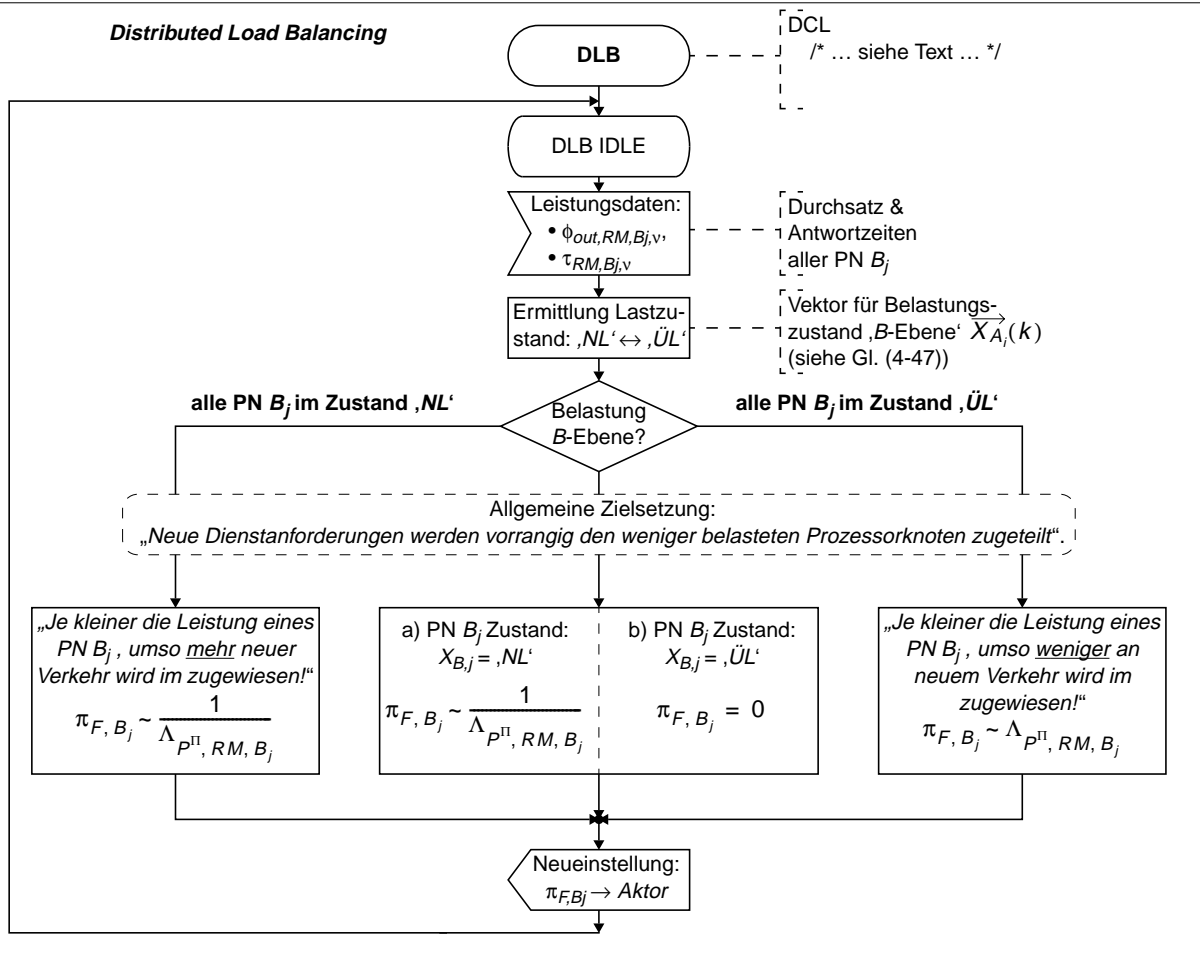
**Tabelle 4-2:** Lastverteilstrategien bei Verwendung der Durchsatzkenn-  
größe als Lastindikator

#### 4.6.4.4 Verfahren auf Basis von Power-Kenngrößen

Aus den gemessenen Antwortzeiten und Durchsatzwerten wird für jeden  $PN_{B,j}$  das kombinierte Leistungsmaß *Power-Produkt*  $\Lambda_{P^{\Pi}, RM, B_j}$  nach Gl. (4-39) berechnet. Der *Power-Vektor* (siehe Gl. (4-40)) gibt Auskunft über die aktuelle Belastung der  $B$ -Ebene. Die *Power-Produkt-Funktion*  $\Lambda_P(A)$  besitzt den qualitativ gleichen Verlauf wie die Durchsatzfunktion (siehe z.B. Anhang B.4). Für die Leistungsregelung stellt sich somit das schon diskutierte Problem, wie bei der Aufgabe festzustellen, in welchem Gebiet des Lastraumes sich die Regelung gerade befindet und in welcher Richtung der optimale Arbeitspunkt liegt. Wiederum sollen auf explizite Lastmessungen in den  $B$ -Prozessoren verzichtet und die bereits verfügbare mittlere Antwortzeit  $\overline{\hat{\tau}_{RM, B_j}}$  (siehe Gl. (4-42)) für eine *Lastbereichsunterscheidung* verwendet werden.

Bild 4-12 zeigt die allgemeine Ablauforganisation des *DLB*-Verfahrens. Als Ausgangspunkt eines ersten Verfahrens sollen die Lastzustandsdefinitionen und der entsprechende Lastvektor  $\vec{X}_{A_i}$  (siehe Gl. (4-47)) des vorherigen Abschnittes verwendet werden. Wiederum sind die drei

Fallunterscheidungen, die zu unterschiedlichen Lastzuteilstrategien führen, sinnvoll (siehe Tabelle 4-3, S. 123).



**Bild 4-12:** Allgemeine Ablauforganisation des *DLB* in SDL-Notation. Funktion wird zyklisch alle  $T_{CI,DLB}$  Zeiteinheiten aufgerufen.

In das *Power-Produkt* gehen die Anteile der unterschiedlichen Verkehrsklassen ein. Es findet somit eine Berücksichtigung des Dienst- und Verkehrsmixes statt. Ergebnisse und weitere Verfeinerungen zu dieser Lastzuteilungsstrategie werden in Abschnitt 8.4 präsentiert.

### 4.6.5 Weitere Diskussion

Die vorgestellten Algorithmen für eine globale Leistungsregelung basieren auf einer möglichst ausgewogenen Lastverteilung in der *B*-Ebene. Neue Dienstanforderungen werden selbst bei kompletter Überlastung aller *B*-Prozessoren nicht durch die  $PN_{A,i}$  zurückgewiesen. Dieser Überlastabwehrmechanismus ist hier Aufgabe der lokalen Lastregelung eines  $PN_{B,j}$  (siehe auch Abschnitte 4.4 und 4.5) und damit kein Bestandteil des globalen Verfahrens. Durch eine Erweiterung des Funktionsumfangs eines *A*-Prozessors könnte bereits eine Zurückweisung oder auch Verwerfung neuer Dienstanforderungen in der *A*-Ebene vorgenommen werden. Hierdurch würden keine Blindlasten in der *B*-Ebene anfallen und deshalb eine erhebliche Leistungssteigerung des gesamten Vermittlungsrechners erzielt werden. Diese Realisierungsvariante ist aus Sicht einer globalen Leistungsregelung für ein Mehrprozessorsystem weniger interessant – die

Belastungs- zustand	Strategie	Berechnung der Zuteilungswahrscheinlichkeiten $\pi_{F,B,j}$
① Kein $PN_{B,j}$ in Überlast	„Wer weniger hatte, bekommt jetzt mehr“: $\pi_{F, B_j}(k) \sim \frac{1}{\Lambda_{P^\Pi, RM, B_j}(k)}$	$\longrightarrow \pi_{F, A_i}(k) = \frac{1}{\sum_{j=1}^M \frac{1}{\Lambda_{P^\Pi, RM, B_j}(k)}} \begin{bmatrix} [\Lambda_{P^\Pi, RM, B_1}(k)]^{-1} \\ \dots \\ [\Lambda_{P^\Pi, RM, B_M}(k)]^{-1} \end{bmatrix}$
② Alle $PN_{B,j}$ in Überlast	„Wer mehr hatte, bekommt jetzt weniger“: $\pi_{F, B_j}(k) \sim \Lambda_{P^\Pi, RM, B_j}(k)$	$\longrightarrow \pi_{F, A_i}(k) = \frac{1}{\sum_{j=1}^M \Lambda_{P^\Pi, RM, B_j}(k)} \begin{bmatrix} \Lambda_{P^\Pi, RM, B_1}(k) \\ \dots \\ \Lambda_{P^\Pi, RM, B_M}(k) \end{bmatrix}$
③ Mindestens ein $PN_{B,j}$ in Normal- und mindestens ein $PN_{B,j}$ in Überlast	1. „Wer überlastet ist bekom- mt keinen neuen Verkehr“: $\pi_{F, B_j}(k) = 0$ wenn $X_{B,j} = \text{ÜL}'$ .  2. „Neue Dienstanforderun- gen werden entsprechend Strategie ① auf die nichtüber- lasteten $PN_{B,j}$ aufgeteilt“: $\pi_{F, B_j}(k) \sim \frac{1}{\Lambda_{P^\Pi, RM, B_j}(k)}$ wenn $X_{B,j} = \text{NL}'$ .	$\longrightarrow \pi_{F, A_i}(k) = \frac{1}{\sum_{\text{alle } X_{B_j} = \text{NL}} \frac{1}{\Lambda_{P^\Pi, RM, B_j}(k)}} \cdot \begin{bmatrix} \dots \\ [\Lambda_{P^\Pi, RM, B_j}(k)]^{-1} \quad \forall X_{B_j} = \text{NL} \\ \dots \\ 0 \quad \forall X_{B_j} = \text{ÜL} \\ \dots \end{bmatrix}$

**Tabelle 4-3:** Lastverteilstrategien bei Verwendung des *Power-Produkts* als Lastindikator

Problematik entspricht der einer Monoprozessorstruktur – und soll deshalb hier nicht weiter verfolgt werden. Diese angedeutete Lösung berührt auch in erheblichem Maße die Frage, in welchen Verarbeitungseinheiten eines verteilten Systems die Protokolle der Schicht-3-Signalisierung implementiert werden sollten.<sup>56</sup>

<sup>56</sup> Problem: Schicht-3-Signalisierung sei hier die eigentliche Aufgabe der *B*-Prozessoren. Für eine Rufzurückweisung bereits auf der *A*-Ebene muss man sich allerdings der entsprechenden Mechanismen dieser Signalisierprotokolle bedienen. Dies geht weit über die in Fußnote 50 (S. 117) diskutierte und im Rahmen dieser Arbeit notwendige Nachrichtenanalyse hinaus.

# Kapitel 5

## Modellierung und angewandte Analyseverfahren

Im Rahmen der Leistungsuntersuchung wurden *Verkehrsmodelle* entwickelt und bestimmte Methoden als *Analyseverfahren* angewandt. Die Modellierung kann in die zwei Bereiche Verkehrsquellen und Vermittlungsrechner getrennt werden. Die Verkehrsmodellierung beinhaltet hier die Aspekte Dienste, Signalisierszenarien und Rufwiederholungen. Durch die Einführung neuer Dienste und Netztechnologien zeichnet sich eine fundamentale Änderung der Verkehrsmuster ab [101].

Die ausreichend genaue Nachbildung des realen Ablaufgeschehens in einer Vermittlungssteuerung führt oft zu komplexen und schwer handhabbaren Modellen. Für diese Modellstrukturen sowie für die Leistungsuntersuchung dynamischer Regelvorgänge hat sich als Methode die *in-stationäre Simulationstechnik* bewährt. Angewandte Methoden und Werkzeuge werden in diesem Kapitel erläutert.

### 5.1 Prinzipien der verkehrstheoretischen Modellierung und Analyse

Untersuchungen an realen Systemen sind, sofern ein System überhaupt bereits realisiert wurde, oft aus technischen, wirtschaftlichen oder Komplexitätsgründen nicht möglich. Die Analyse eines Vorganges erfordert deshalb zuerst eine Abstrahierung und Abbildung des Untersuchungsobjekts aus der physikalischen Welt in einen Modellbereich. Für das Verständnis vieler Überlegungen sollte man sich daher immer vor Augen halten, dass das entwickelte Modell niemals der Wirklichkeit völlig entspricht. Es spiegelt vielmehr nur die Eigenschaften wieder, die entsprechend dem Anwendungszweck nachgebildet werden sollen. Bezüglich der Modellkomplexität muss ein Kompromiss zwischen einfacher Handhabbarkeit, z.B. geschlossen lösbares Gleichungssystem oder schnell ablaufendes Simulationsprogramm, und der Forderung nach einem sehr detaillierten Modell gefunden werden.

Die Systemabläufe der in dieser Arbeit betrachteten Einrichtung sind teilweise zufälliger Natur (z.B. Bearbeitungszeit einer Signalisiernachricht). Um quantitative Aussagen über die Verkehrsleistung erzielen zu können, ist deshalb eine verkehrstheoretische Modellierung, unter Verwendung stochastischer Prozesse, notwendig [42]. Der Abstraktionsgrad bestimmt sich dabei aus den zu ermittelnden Kenngrößen, der gewünschten Aussagegenauigkeit und dem vorhandenen Wissen über das tatsächliche System.

## 5.2 Modellierung von Anwendungen und Diensten

Ein Kommunikationsnetz mit seinen Netzelementen stellt seinen Benutzern je nach Netztechnologie, -typ, Einsatzzweck usw. gewisse Dienste zur Verfügung. Für die Abwicklung eines Dienstes werden in einem ressourcenreservierenden, d.h. einem verbindungsorientierten Netz, Steuerungsinformationen in Form von Signalisierungsmeldungen zwischen Teilnehmern und Netzknoten (Dienstnutzern und Diensterbringern) ausgetauscht.<sup>1</sup> Die Folge von Meldungen, die von Beginn bis zum erfolgreichen oder nicht erfolgreichen Ende einer Dienstbearbeitung ausgetauscht werden, wird als Meldungsszenario, oder kurz Szenario bezeichnet. Je nach Umfang und Definition eines Signalisierprotokolles können allein bei der Bearbeitung eines einzigen Dienstes bereits eine Vielzahl von Szenarien unterschieden werden.

### 5.2.1 Prinzipien der Quellmodellierung

Modelle von Verkehrsquellen sollen das Verhalten bzw. die Eigenschaften von Teilnehmern, Anwendungen, Diensten, Endgeräten, Codierern usw., je nach gewünschter Betrachtungsebene und Abstraktionsgrad, beschreiben. Ein Quellmodell soll Verkehr mit einer bestimmten Charakteristik für das zu untersuchende Systemmodell erzeugen.

#### 5.2.1.1 Modellierungshierarchie in ATM-Systemen

In Untersuchungen von Abläufen, Algorithmen und Komponenten der Nutzer-Ebene (User Plane) ist die Unterteilung in Zell-, Burst-, Dialog-, Verbindungs- bzw. Rufaktivitätsebenen bekannt [328,515]. Darüber ist die Anwendungsebene, die oft auch als Dienst- bzw. Sitzungsebene bezeichnet wird. Jede dieser Modellierungsebenen repräsentiert prinzipiell eine eigene Zeitebene. Eine Analyse konzentriert sich deshalb i.d.R. immer nur auf eine dieser Ebenen. Für die Untersuchung von Effekten in der Steuerungs-Ebene (Control Plane) macht eine Modellierung auf Zellebene wenig Sinn. Als elementare Informationseinheit wird vielmehr eine einzelne Signalisiernachricht oder ein kompletter Ruf betrachtet.

Neue Anwendungen sind durch komplexeres Quellverhalten gekennzeichnet. Typisch dabei sind etwa ein büschelförmiger Charakter oder Langzeit-Abhängigkeiten. In beiden Fällen können klassische, auf Markoff-Modellen [383] beruhende Analyseverfahren entweder gar nicht oder mit Modifikationen (z.B. Markoff-Modulierter Poisson-Prozess, MMPP) angewandt werden. Weitere Merkmale können nichtstationäres oder netzzustandsabhängiges Quellverhalten sein, welche ebenfalls verfeinerte Analysemethoden bedingen (z.B. unterbrochener Poisson-Prozess [526]). Dies trifft insbesondere für Betrachtungen des Steuerdaten-Verkehres zu (Überlasteffekte, fensterbasierte Regelungsmechanismen u. ä. m.).

#### 5.2.1.2 Quellmodelle für die Steuerungs-Ebene

Prinzipiell lassen sich alle bekannten Verkehrsquellen der Nutzer-Ebene auch zur Modellierung des Steuerungsverkehrs verwenden. Anstelle eines IP-Paketes oder einer ATM-Zelle wird dann ein Ankunftsereignis als Signalisiernachricht, Ruf- oder Verbindungsanforderung inter-

---

<sup>1</sup> Bei verbindungslosen Netzen werden in einem Netzknoten die Steuerungsnachrichten intern erzeugt. Bei jedem Eintreffen eines Paketes wird i.d.R. eine gewisse Steuerungslast anfällig.

pretiert. Steht aber ein bestimmtes Teilnehmerverhalten, Rufmodell (*Call Model*) oder Signalisierprotokoll im Vordergrund, dann kommen dedizierte Quellmodelle zur Anwendung [347].

Die Funktionsweise eines Quellmodelles lässt sich durch einen endlichen Automaten bzw. eine darauf beruhende SDL-Spezifikation beschreiben. Zustandsübergänge werden entweder zufällig nach einer bestimmten Verteilung oder durch spezielle Ereignisse (z.B. Systemrückwirkungen) ausgelöst. Beispiele sind in [105, 146, 214, 351, 352, 549, 556] zu finden.

## 5.2.2 Modellierung klassischer Fernsprechrufe

Jahrzehntelange Erfahrungen mit dem Fernsprechverkehr sind in ein sehr genaues Quellmodell dieses Dienstes eingeflossen [10]. Ein klassischer Fernsprechruf ist durch eine mittlere Belegungsdauer von ungefähr drei Minuten charakterisiert [75, 552]. Die Gesprächsdauer kann durch eine negativ-exponentielle Verteilung angenähert werden, und die Rufankünfte lassen sich durch einen Poisson-Prozess beschreiben [104]. Weiterhin sind Rufe, die zu verschiedenen Kommunikationsbeziehungen gehören, unabhängig, Teilnehmer können gewissen Teilnehmerklassen (*residential*, *small business*, usw.) zugeordnet werden, und typische Merkmale einer Teilnehmerklasse sind für alle zugehörigen Teilnehmer repräsentativ [105].

Ein Quellmodell für den klassischen Telefondienst (*POTS*) lässt sich deshalb unter vorgenannten Voraussetzungen komplett wie folgt beschreiben: mittlere Belegungsdauer (*holding time*) mit zugehörigem mittlerem Ankunftsabstand (*interarrival time*).

Diese hier beschriebene, relativ einfache Möglichkeit der Verkehrscharakterisierung lässt sich aber nun nicht ohne weiteres auf neuere Dienste übertragen [626].

## 5.2.3 Bekannte Ansätze zur Modellierung neuer Dienste

Mit der Entwicklung des Schmalband-ISDN wurden auch verstärkt Teilnehmermodelle für Nicht-POTS-Dienste vorgeschlagen (z.B. [179, 251, 381]). Für die Leistungsuntersuchung von Rechnernetzen wurden ebenfalls Modelle von Anwendungen und Diensten entwickelt (z.B. [331]). Durch das breite Dienstespektrum, das durch das B-ISDN abgedeckt werden soll, wurden dann für diese Technologie detaillierte Verkehrsmodelle und Dienstcharakteristiken notwendig (z.B. [90]). Auf dem Gebiet des Internet entstanden durch das rapide Wachstum der WWW-Nutzung entsprechende Modelle (z.B. [264]). Dedizierte Modelle für die Dienstabwicklung in Signalisiersystemen sind beispielsweise in [351, 555] beschrieben.

Die genannten Beispiele repräsentieren nur eine kleine Auswahl von Entwicklungen zur Dienstmodellierung in jüngerer Zeit. Für die Charakterisierung von Diensten – speziell in der *Steuerungsebene* – besitzen nachfolgend diskutierte Punkte einen wesentlichen Einfluss.

### 5.2.3.1 Verkehrsklassen

Um die Leistungsfähigkeit verschiedener Systeme objektiv miteinander vergleichen zu können, ist die Definition von Referenzlasten notwendig. Für PSTN-Systeme wurden diesbezüglich sogenannte Verkehrsgemische, bestehend aus Anteilen verschiedener Verkehrsklassen, in ITU Q.543 [648] spezifiziert. Die Unterteilung erfolgt hier in *Ruftypen* wie beispielsweise lokaler oder ankommender Verkehr.

Weitere bekannte Klassifizierungen basieren auf der Unterscheidung von *Teilnehmerarten* ([90]: private Teilnehmer, kleinere, mittlere bzw. große Firmen; [351]: privat, geschäftlich oder mobil), *Belegungsauern* [201] und *Diensten*. In dieser Arbeit werden verschiedene Dienste mit ähnlichen Merkmalen auf wenige Verkehrsklassen abgebildet (siehe Abschnitt 5.2.6.5).

### 5.2.3.2 Szenariodetaillierung

Für die Komplettierung eines Dienstes werden temporäre Betriebsmittelbelegungen während der Dauer einer Kommunikationsbeziehung notwendig. Aus Sicht eines Netzknotens könnte der erforderliche Ressourcenbedarf durch genau *eine Anforderungsmeldung* modelliert werden. Das andere Extrem wäre die Nachbildung des vollständigen Signalisierablaufes mit dem entsprechenden Betriebsmittelbedarf für jede einzelne Signalisiernachricht. Aufgrund der Komplexität wird der zweite Quellmodelltyp nicht verwendet.

Für PSTN-Untersuchungen wurde fast immer ein Ruf durch eine Steuerungs-Nachricht modelliert. Beispiele für detaillierte Modelle von Signalisierabläufen sind in [521,523] („*Subcall-Oriented Modelling*“; neben der Rufankunft werden weitere zugehörige Folgeereignisse berücksichtigt), [105] (ISDN), [623] (SS Nr. 7), [322] (B-ISDN) oder in [136,262,291,472] (Breitband-IN) zu finden.

Da ein wesentliches Ziel dieser Arbeit die Berücksichtigung eines Dienstespektrums ist, müssen einer bestimmten Signalisiernachricht unterschiedliche Betriebsmittelanforderungen zugeordnet werden können. Weiterhin soll die Dauer einer Kommunikationsbeziehung Berücksichtigung finden. Dies erfordert zumindest eine REQUEST- und eine RELEASE-Meldung in einem Szenario. Die verwendeten Signalisierabläufe sind in Abschnitt 5.2.6.5 und Anhang C.3.3 dokumentiert.

### 5.2.3.3 Verteilung der Belegungszeiten

Die CPU ist die wichtigste Ressource in einem Vermittlungsrechner. Die Inanspruchnahme dieses Betriebsmittels durch einen Dienst ist somit ein wichtiger Modellierungsaspekt. Charakteristische Merkmale diesbezüglich können unter anderem aus dem Bearbeitungsverhalten von Signalisiernachrichten abgeleitet werden. In dieser Arbeit soll die Dauer einer Kommunikationsbeziehung als weiterer Parameter in die Modellierung miteinfließen. Solche dienstspezifischen Werte können beispielsweise aus Belegungszeiten von Verbindungskapazitäten für durchschaltvermittelnde Netze gewonnen werden.

Beim klassischen Fernsprechruf wird die Verteilung der Belegungsauern unabhängig von der Verteilung der Zwischenankunftsabstände von Rufen angenommen und hat einen „morphologischen“ Charakter [104] für diesen Dienst. Eine Approximation durch die negativ-exponentielle Verteilung ist deshalb gerechtfertigt. Diese Charakterisierung trifft allerdings für Internet-Dienste weniger zu. Hier wurden wesentlich längere Belegungszeiten beobachtet [75,216,388]. Um verschiedene Dienstypen durch ein Teilnehmermodell nachbilden zu können, wurde beispielsweise eine Approximation durch Überlagerung mehrerer Normalverteilungen<sup>2</sup> vorgeschlagen [100,291].

---

<sup>2</sup> Motivation: Die einer Leistungsuntersuchung zugrundeliegende Verteilungsfunktion der Belegungszeiten hat signifikanten Einfluss auf die erzielten Ergebnisse. Die traditionelle POTS-Annahme liefert in IN- bzw. ISDN-Untersuchungen zu optimistische Resultate [100]. Gerade der Anteil von sehr kurzen Rufen in einem Call Mix darf nicht vernachlässigt werden. BOLOTIN [100] schlägt deshalb (für Untersuchungen von Signalisiernetzen) eine Überlagerung mehrerer Normalverteilungen vor. In einer aus Messdaten gewonnenen Parametrisierung werden zwei Normalverteilungen mit Mittelwerten von 25s bzw. 4min vorgeschlagen.

Zusammengefasst kann festgehalten werden, dass sich die *mittlere Dauer* einer Kommunikationsbeziehung für verschiedene Dienste sehr stark unterscheiden kann. Dabei sind Unterschiede für Extremfälle mit einem Faktor 100 bis 1000 nicht unwahrscheinlich.<sup>3,4</sup> Dieses dienstspezifische Merkmal wird in der hier vorgenommenen Modellierung berücksichtigt.

## 5.2.4 Modellierung des Signaliserverkehrs unter Berücksichtigung neuer Dienste

Speziell seit Beginn der Untersuchungen zum Signalisiersystem Nr. 7 ist eine Vielzahl von Modellen zu Signalisierabläufen neuer Dienste entstanden. Eine Vielzahl bekannter Lastmodelle ist hierzu in [575] dargestellt. Diese Modelle können einen guten Ausgangspunkt für die Untersuchung von Lastregelkreisen von Steuerungssystemen in Breitbandnetzen bilden. Gegebenenfalls sollten aber noch komplexere Breitbanddienste berücksichtigt werden.

In den ITU-Empfehlungen E.711-713 [621,622,623] sind Anregungen zur Signalisiermodellierung im Fernsprechnetz (POTS) und ISDN zu finden. Aspekte der Rufmodellierung im B-ISDN werden in E.716 [624] erläutert, und es wird auch dort Schicht drei als Modellierungsebene für Signaliserverkehr vorgeschlagen.

## 5.2.5 Anforderungen an die Dienstmodellierung in Breitbandnetzen

Folgende Entwicklungen und Tendenzen haben einen signifikanten Einfluss und sollten zukünftig in der Modellierung von Verkehrsquellen und Diensten für die Steuerungs-Ebene berücksichtigt werden:

- Allgemein wird ein *Lastanstieg* aufgrund folgender Aspekte erwartet:
  - Zunehmend verbesserte Unterstützung von Diensten durch neue und erweiterte Signalisierstandards.
  - Stetig neu hinzukommende Dienste, d.h. die Anzahl verfügbarer Dienste steigt [365].
  - Der Kreis von Diensteanbietern wird nach Wegfall von Monopolen immer größer.
  - Es wird erwartet, dass Dienste häufiger genutzt werden [367].<sup>5</sup>
  - Die Heterogenität der Nutzerzusammensetzung nimmt zu.
- Die *Anforderungen* erhöhen sich zunehmend aufgrund folgender Punkte:
  - Bereitstellung verschiedener Dienstklassen in der Nutzer-Ebene. Diese Unterscheidung kann auch in einem begrenzten Bereich für die Steuerungs-Ebene sinnvoll sein. Mögliche Konsequenzen wären beispielsweise die Einführung von Dienstprioritäten.
  - Unterschiedliche Charakteristiken und Lastprofile eines Dienstes in verschiedenen Netztypen (z.B. klassischer Basisruf in einem Festnetz oder im Mobilfunknetz).
  - Ausprägungen eines bestimmten Dienstes in Form unterschiedlichster Signalisierzenarien (z.B. Anzahl von Meldungen oder Dauer einer Kommunikationsbeziehung).

---

<sup>3</sup> Z.B. schneller Datentransfer mit wenigen Sekunden auf der einen Seite und Internetsitzungen oder Konferenzschaltungen im Minuten-/Stundenbereich auf der anderen Seite.

<sup>4</sup> In [343] werden folgende Faktoren für die erzeugte Signalisierlast von Rufen unterschiedlicher Anwendungen angegeben: Ein *PCS-Call* (*Personal Communication Service*) umfasst 3-4 mal soviel Verkehr wie ein typischer Mobilfunkruf, der selbst wiederum 4-11 mal soviel Last als ein durchschnittlicher *ISDN-Call* erzeugt.

<sup>5</sup> Beispielsweise ist laut [90] die Anzahl von BHCA's pro B-ISDN-Teilnehmer bis zu zehnmals größer als im N-ISDN.



Diese Aspekte müssen je nach Untersuchungsobjekt unterschiedlich stark in die Dienstmodellierung miteinfließen. Wie bereits in Abschnitt 5.2.3.2 erwähnt wurde, bilden Signalisierabläufe den Schwerpunkt der Dienstmodellierung in dieser Arbeit.

## **5.2.6 Modellierung des Steuerungsverkehrs in dieser Arbeit**

Zunächst werden die für diese Arbeit relevanten Aspekte erläutert und bekannte Ergebnisse aus der Literatur kurz dargestellt. In Abschnitt 5.2.6.5 wird dann die Modellierung des Signalisierverkehrs im Rahmen dieser Arbeit vorgestellt.

### **5.2.6.1 Problematik**

In der gegenwärtigen Erprobungs- und Aufbauphase des B-ISDN ergeben sich folgende Schwierigkeiten bei der Suche nach einem geeigneten Quellmodell für die Steuerungs-Ebene. Zum jetzigen Zeitpunkt fehlen realistische Daten aus Feldmessungen, derzeitige Komponenten haben fast durchweg prototypischen Charakter (z.B. Signalisierungsmöglichkeiten oder Unterstützung von Wählverbindungen), sehr wenige Teilnehmer und Endgeräte sind bisher angeschlossen und viele Dienste sind noch nicht realisiert.

Der in dieser Arbeit verwendete Ansatz einer Verkehrsmodellierung wird deshalb zukünftig revidiert und dem jeweiligen Umfeld angepasst werden müssen. Das derzeitige Vorgehen in der Literatur basiert auf Messungen ähnlicher Dienste aus anderen Netzen oder auf der Schätzung und Hochrechnung von Parametern.

Gerade bei der Untersuchung von Überlastabwehrverfahren und Lastregelkreisen muss der Dienstaspekt kritisch betrachtet werden. Macht es aus Sicht der Überlastabwehr überhaupt Sinn, die einzelnen Dienste zu unterscheiden und auch unterschiedlich zu behandeln? Zum einen ist fraglich, ob verwertbare und sinnvolle Informationen über Dienste der Lastregelfunktion überhaupt zugänglich sein werden, zum anderen, ob dieser nicht unerhebliche Anteil an Steuerungskosten für die Informationsgewinnung auch eine hoffentlich zu erwartende Leistungssteigerung rechtfertigt. Es stellt sich also für jede zu konzipierende B-ISDN-Vermittlungssteuerung die Frage, ob nicht ein einfacher, konservativer Ansatz für eine Lastregelung bereits ausreichend ist und befriedigende Ergebnisse liefert.

Da ein essentieller Vorteil dieser Netztechnologie aber die Unterstützung verschiedenster Dienstgütern in der Nutzer-Ebene ist, sollte ein differenzierteres Konzept gewählt werden. Die unterschiedliche Behandlung von Nutzdatenverkehr mit garantierten Dienstgütern darf nicht durch die Steuerungs-Ebene verschlechtert, sondern muss vielmehr auch in dieser Ebene berücksichtigt und unterstützt werden.

### **5.2.6.2 Signalisierszenarien ausgewählter Dienste**

Um einen Eindruck in die Vielfalt möglicher Ausprägungen von Signalisierabläufen zu geben, sind in Anhang C.3.2 zwei Szenarien verschieden komplexer Dienste dokumentiert.

### 5.2.6.3 Repräsentative Auswahl bestimmter Dienste

Aufgrund der mittlerweile enormen Vielfalt an Diensten ist für eine Leistungsuntersuchung eine Beschränkung auf wenige Dienste unerlässlich. Aus der Literatur sind hierzu eine Vielzahl repräsentativer Mengen bekannt, wobei sich die Selektion am jeweiligen Untersuchungsobjekt orientiert. Beispiele für Dienstauswahlen sind in PASCH [421]<sup>6</sup>, ONVURAL [50]<sup>7</sup> oder ULMER [532]<sup>8</sup> zu finden.

Am vielversprechendsten für Untersuchungen in der Steuerungs-Ebene ist der sehr detaillierte Ansatz von LIU [360]<sup>9</sup>, da hier Signalisierszenarien auf Ebene von Schicht-3-Nachrichten betrachtet werden.<sup>10</sup> Als Verkehrsmaß werden für den Dienstemix *Hourly Call Attempts* (HCA) definiert: eine *Peak Per-Customer* „Ruftrate“ sowie eine *Per-Minute Channel Switching Rate*.

### 5.2.6.4 Lastmodellierung

Durch die Festlegung einer repräsentativen Auswahl von Diensten steht auch die Charakteristik des Verkehrs fest. Über die Lastzusammensetzung ist aber damit noch nichts gesagt. Alle Aspekte und Anforderungen der Lastmodellierung lassen sich allgemein im sogenannten Verkehrsbezugsmodell zusammenfassen [90]. Darunter versteht man die angemessene Berücksichtigung folgender Faktoren:

- Vielzahl von Diensten,
- Verbreitungsgrad von Diensten,
- Benutzungshäufigkeit eines Dienstes,
- Berücksichtigung verschiedener Varianten (z. B. Transitruf, lokaler Ruf),
- Einsatz eines Vermittlungssystems im Zugangs- oder Kernnetz (d. h. der Netzebene mit unterschiedlicher Zusammensetzung der Verkehrsströme),
- Verbindungsdauer oder die Berücksichtigung von Verkehr aus Nicht-B-ISDN-Netzen (z. B. IP-basierte Netze oder UMTS-Mobilfunknetz).

Diese Kriterien sind grundsätzlich bei der Erstellung und Auswahl eines Teilnehmermodells einzubeziehen.

---

<sup>6</sup> Betrachtet wird ein zukünftiger Teilnehmer eingebunden in eine Freizeit- und Geschäftsumgebung, wobei angenommen wird, dass sich der Großteil des erzeugten Verkehrs auf vier Klassen abbilden lässt, der mehrheitlich folgenden Anwendungen entstammt: *Voice Telephony*, *Video Telephony*, *File Access* und *Character Traffic* (z. B. E-Mail, ftp).

<sup>7</sup> Klassen: *Voice* und *Video Telephony*, *Document* und *Video Retrieval*, *Fax*, *File Transfer* und *Data on Demand*. Weiterhin werden acht Teilnehmerklassen (*user classes*) unterschieden.

<sup>8</sup> Für die Quellmodellierung auf Verbindungsebene werden elf ausgewählte Dienste des B-ISDN berücksichtigt. Für die in der Arbeit durchgeführten Leistungsuntersuchungen werden dann zwei repräsentative Dienstkombinationen mit jeweils vier Diensten (1.: Sprache, Datenübertragung, Videokonferenz, Fernsehen; 2.: Bildtelefonie, LAN-Kopplung, Videokonferenz, HQ-Video) betrachtet. Ankunftsabstände und Belegungszeiten werden für alle Dienste als negativ-exponentiell verteilt angenommen. Die Bearbeitungszeit einer Signalisiermeldung wird auf 10 µs geschätzt.

<sup>9</sup> Simulation der *Control Plane* eines generischen, hybriden B-ISDN-Systems (Knoten ist für die Vermittlung von POTS-, N-ISDN- und B-ISDN-Verkehr konzipiert). Folgende Dienste werden berücksichtigt: POTS (*Voice Telephony*), N-ISDN (*Voice Telephony*, *Data*, *File Transfer* und *Video*); jeweils mit Unterscheidung in D- und B-Kanal-Dienst), B-ISDN (alle N-ISDN-Dienste sowie *Video*). Es werden weiterhin acht Netzbereiche mit unterschiedlicher Dienstzusammensetzung unterschieden. In jedem Netzbereich gibt es insgesamt 1536 Teilnehmer die jeweils einem der drei Netztypen zugeordnet sind. Sprachtelefonie wird weiterhin unterteilt in lokale Rufe, Fern- und IN-Rufe. Zusätzlich werden Rufwiederholungen für blockierte Rufe modelliert.

<sup>10</sup> Die Bedienzeiten werden aufgrund Programmcodeumfang, Prozessortyp und Taktfrequenz geschätzt. Somit benötigt jeder Dienst unterschiedlich viel an Prozessorressourcen. Prozessorhintergrundlast findet in Form eines festen und eines zur Anzahl von Rufen in Bearbeitung proportionalen Anteils Berücksichtigung.

### 5.2.6.5 Gewählte Dienstmodellierung und Verkehrsklassen

Die im Rahmen dieser Arbeit gewählte *Modellierungsebene* ist die Ebene der Schicht-3-Signalschichtnachrichten. Diese Festlegung wurde bereits ausführlich in Abschnitt 4.4 motiviert. Dagegen wird nicht jede Signalmeldung innerhalb eines Szenarios unterschieden, sondern es erfolgt hier eine Abstraktion auf drei generische Nachrichtentypen: REQUEST (Anforderung eines Dienstes), FOLLOWMESSAGE (Meldungen während der „Lebensdauer“ eines Dienstes) und RELEASE (Beendigung eines Dienstes).

Betrachtet man verschiedene Dienste, dann ist eine hohe Korrelation aufeinanderfolgender Meldungen in einzelnen Rufphasen feststellbar (siehe auch Bild 4-3).<sup>11</sup> Zwischen den Aktivitätsphasen bestehen relativ lange Ruhepausen.<sup>12</sup> Diese Eigenschaften sollen für die Untersuchung in einem generischen Quellmodell wie folgt berücksichtigt werden.

Ein Generator erzeugt Ankunftsereignisse nach einer vorgegebenen Verteilungsfunktion. Jedes Ereignis repräsentiert eine neue Dienstanforderung (REQUEST) einer bestimmten Verkehrsklasse. Alle Antworten der Vermittlungssteuerung werden einer gesonderten Komponente zugeführt, die diese in Folgemeldungen (FOLLOWMESSAGE bzw. RELEASE) umsetzt. Die Anzahl von Folgemeldungen kann deterministisch oder zufallsabhängig konfiguriert werden. Die Verteilung der Zwischenankunftsabstände kann vorgegeben werden.<sup>13</sup> Alle Meldungen eines individuellen Szenarios besitzen dieselbe Referenznummer (vergleichbar der *Call Reference*) und sind somit eindeutig zuordenbar. Diese Eigenschaft ist wichtig für das Routing von Meldungen innerhalb eines Multiprozessor-Steuerungssystems.

Das Konzept dieses Quellmodells wurde bereits in [470] vorgestellt. Ein weiteres Merkmal ist die Berücksichtigung der Möglichkeit von wiederholten Rufversuchen, falls eine Anforderung zurückgewiesen wurde (siehe Abschnitt 5.3.2).

Für die Untersuchung der in dieser Arbeit vorgestellten Algorithmen zur Last- und Leistungsregelung wurden drei Muster an Signalisierabläufen (siehe auch Anhang C.3), die die folgenden drei Verkehrsklassen repräsentieren, definiert (Tabelle 5-1).

Klasse	Beschreibung	Beispiele	Priorität
1	Komplexere Dienste (Multimedia-, Multiconnection- oder Multiparty-Call)	Konferenzdienste, IN-Dienste	2
2	Basisdienste (z. B. Singleconnection Calls, wichtige Dienste)	Sprachtelefonie, Maildienste	2
3	Verkehr aus dem Netzzinneren	alle Dienste	1

**Tabelle 5-1:** Ausgewählte Verkehrsklassen

Der Überlastabwehr kommt in den Vermittlungsknoten am Netzrand eine besondere Bedeutung zu, da sie für die Regulierung des Zuflusses an Verkehr in ein Netz mitverantwortlich ist. Aus diesem Grund wird hier der sichtbare Ursprung des Verkehrs, d. h. aus dem Netz oder von Endgeräten bzw. von einem Zugangsnetz, in der Klassifizierung berücksichtigt. Dieses Merk-

<sup>11</sup> Das Hinzufügen einer weiteren Verbindung in einem *Multiconnection Call* löst beispielsweise einen Burst an Signalmeldungen aus.

<sup>12</sup> Ein bekanntes Quellmodell, das diese Eigenschaften erfüllt, ist die *On-Off*- oder *Talkspurt-Silence*-Quelle. Sie hat allerdings hier die Nachteile, dass während der Aktivitätsphase die Ankunftsereignisse deterministisch verteilt sind, und aufgrund der Überlastabwehr ein Ruf ein abgeschlossenes Signalisierszenario darstellen soll. Hierzu müsste dieses Zweizustandsmodell entsprechend erweitert werden.

<sup>13</sup> Diese Zeiten repräsentieren die „Gesprächsphasen“, also den Austausch von Nutzerdaten während der Dauer einer Kommunikationsbeziehung (z. B. 3 min bei Telefongespräch oder 2 h bei einem Filmabruf).

mal soll sich auch in einer globalen Priorität für die Bearbeitung in einer Vermittlungssteuerung niederschlagen. Aufgrund der bereits erfolgten Betriebsmittelbelegung sollte der Verkehr vom Netzzinneren eine höhere Priorität besitzen. Unter Hochlast ist es dagegen sinnvoll, die Prioritätsverhältnisse gegebenenfalls temporär zu modifizieren.

Jede Verkehrsklasse wird durch ein repräsentatives Signalisierszenario beschrieben. Diese unterscheiden sich in Meldungsanzahl, Zwischenankunftsabständen und im Ressourcenbedarf der einzelnen Meldungen. In Anhang C.3.3 werden diese näher erläutert.

## 5.3 Teilnehmerverhalten, Endgeräteaspekte und Protokollmechanismen

Ein realistisches Teilnehmermodell ist eine notwendige Grundlage für Überlastabwehruntersuchungen. Im Zuge der „alles-integrierenden Netze“ erhebt sich allerdings dabei die grundsätzliche Frage, ob überhaupt noch ein generisches Benutzermodell existiert.

### 5.3.1 Problematik

Bereits in Abschnitt 3.1.2.2 wurden Überlasteffekte erläutert, die durch die Wechselwirkung zwischen Verkehrsquellen und Netzelementen entstehen können. Am bekanntesten sind die sogenannten Rufwiederholungen (*Repeated Call Attempts*) als Reaktion des menschlichen Teilnehmers auf Rufblockierung. Da dieser Mitkopplungseffekt die Verkehrscharakteristik und damit das zeitliche Lastprofil signifikant beeinflussen kann, darf in Untersuchungen zu Lastregelkreisen dieses Phänomen prinzipiell nicht außer acht gelassen werden.

Mit der Einführung diensteintegrierender Netze wurde dieses Problem verschärft, da nun für viele Dienste ein Rechner als Endgerät fungiert. Dies hat zur Konsequenz, dass dann nicht mehr der menschliche Benutzer direkt Signalisierereignisse triggert, sondern diese programmgesteuert ausgelöst werden.<sup>14</sup> Das Endgeräteverhalten hängt nun entscheidend davon ab, welche Vorgaben von einem Netzbetreiber für einen bestimmten Dienst vorgegeben werden und wie die jeweilige Software realisiert wird.<sup>15</sup> Auf jeden Fall unterscheidet sich die Charakteristik eines Wiederholversuchs vom Typ einer Verkehrsquelle. Bekannte Modelle für Rufwiederholungen für menschliche Teilnehmer können deshalb nicht ohne weiteres auf rechnergestützte Endgeräte übertragen werden.

Eine weitere Quelle für Wiederholversuche steckt auch in den Signalisierprotokollen selbst. Beispielsweise überwacht Q.2931 [670] die Antwortzeiten der meisten Signalisiernachrichten. Bei Ablauf des Timers für die SETUP-Nachricht wird beispielsweise dann genau *ein Wiederholversuch* gestartet, bevor die Anwendung benachrichtigt wird.

---

<sup>14</sup> Bekannt ist dieses Problem vom klassischen Faxdienst. Bei nicht erfolgreichem Senderuf muss eine gewisser Zeitabstand zum nächsten Rufwiederholversuch eingehalten werden.

<sup>15</sup> Liegen beispielsweise für einen Dienst keine Vorgaben vor, dann könnte man das Endgerät dazu bringen, permanent wiederholte Anforderungen bis zu einem erfolgreichen Aufbau abzusetzen.

## 5.3.2 Modellierung von Rückkopplungseffekten

Durch bestimmte Ereignisse getriggert führt Verkehr zu stoßartigen Lasten. Aufgrund der im vorherigen Abschnitt diskutierten Ausführungen ergeben sich entsprechend Konsequenzen für ein Quellmodell, das schwerpunktmäßig diesen Effekt berücksichtigen soll. Dabei erscheint eine Unterteilung in verschiedene Ebenen zweckmäßig. Folgende Strukturierung bietet sich an: Signalisier-, Anwendungs- und Teilnehmerebene.<sup>16</sup> Dadurch entsteht ein hierarchisches Lastmodell, bei welchem je nach gewünschtem Dienstespektrum oder Endgerätetyp die ebenspezifischen Effekte berücksichtigt werden sollten.

Im Rahmen dieser Arbeit wurden Untersuchungen zu Rufwiederholungen durchgeführt. Dabei kann das realisierte Quellmodell Wiederholungen bei den Ereignissen „Timerablauf“ und „Zurückweisung einer Anforderung“ absetzen. Die Anzahl von Wiederholversuchen sowie die Verteilung ist flexibel konfigurierbar. Ergebnisse zu diesen Experimenten sind hauptsächlich in [418] enthalten.

In den in Kapitel 6 bis 8 enthaltenen Messungen wurde dagegen dieser Effekt bewusst vernachlässigt, da in der Regel das angelegte Verkehrsprofil bereits aus einem idealisierten Lastsprung besteht.<sup>17</sup>

## 5.4 Modellierung von Vermittlungsrechnern

Neben den Verkehrsquellen bildet die Modellierung der Steuerung eines B-ISDN-Knotens den zweiten Schwerpunkt der Modellierungsarbeit. Hierbei reicht das Spektrum bezüglich der Modellierungstiefe von relativ einfachen, analytisch handhabbaren Modellen bis zu sehr aufwendigen Simulationsmodellen.

### 5.4.1 Einführung

Vermittlungsrechner repräsentieren oft eine generische Steuerungsarchitektur und sind deshalb für verschiedene Netztechnologien verwendbar. Verstärkt wird diese Tendenz durch den zunehmenden Einsatz kommerzieller Rechner- und Echtzeit-Betriebssysteme. Bereitgestellte Dienste oder Funktionszuordnungen eines Prozessorknotens sind weitestgehend in der Software realisiert. Die Hardwarestruktur bildet oft den Ausgangspunkt der Modellierung eines Vermittlungsrechners. Ein Mikroprozessor wird dann durch eine Bedieneinheit modelliert.<sup>18</sup>

Durch die vielen Gemeinsamkeiten zwischen einem Vermittlungsrechner und einem allgemein einsetzbaren Computersystem können viele Ansätze oder Werkzeuge aus dem Bereich der Lei-

---

<sup>16</sup> Findet eine Modellierung unterhalb Schicht drei statt, dann sind beispielsweise auch Wiederholversuche der Schicht-2-Signalisierung (speziell SSCOP) oder der Bitübertragungsschicht zu berücksichtigen.

<sup>17</sup> Die aufgeführten Teilnehmerrückwirkungen stellen Mitkopplungen dar. Deshalb kann es unter Umständen zu lawinenartigem Verkehrszuwachs kommen (siehe z. B. [73]). Die *Zeitpunkte*, an denen solche Ereignisse ausgelöst werden, sind bei einem stochastischen System *nicht vorhersagbar*. Deshalb ist die Methode der instationären Simulation für Untersuchungen solcher Effekte weniger geeignet. Durch einen Simulationstrace kann zwar der Prozessverlauf für dieses Ereignis gemessen werden, statistisch gesicherte Aussagen sind aber dann nicht möglich.

<sup>18</sup> Um Untersuchungen zu vereinfachen, werden auch Prozesse oder Protokolle oft durch eine eigenständige Bedieneinheit nachgebildet (z. B. [322]). Dabei entsteht ein Warteschlangennetz mit paralleler Bearbeitung. Da aber Prozesse auf einem Monoprocessorsystem nur seriell ausgeführt werden können, sind die erzielten Ergebnisse kritisch zu betrachten.

stungsuntersuchung von Rechnersystemen eine gute Basis für die Analyse von Vermittlungssteuerungen bilden.

Die Randbedingungen für eine Modellierung richten sich aber auch hier nach den Untersuchungszielen, vorhandenen Werkzeugen und wirtschaftlichen Gesichtspunkten. Eine gute Einführung und Übersicht zur Modellbildung von Vermittlungssteuerungen ist in [35,149,186,329,330] gegeben. Die Arbeit von DUTKIEWICZ [163] beschäftigt sich speziell mit Aspekten der Modellierung und Analyse von Steuerungsstrukturen für ISDN-Netzknotten.

## 5.4.2 Einfache Modelle für analytische Untersuchungen

Für eine erste Approximation eines Vermittlungsrechners werden Bediensysteme mit einer Bedieneinheit (*Single Server*), z.B. das *M/M/1*-, *M/D/1*- oder das *M/G/1*-System, verwendet. Diese Modelle liefern gute Ergebnisse wenn beispielsweise nur das *Black-Box*-Verhalten<sup>19</sup> eines Vermittlungsrechners bekannt ist und *stationäre Leistungsdaten* gesucht sind. Dieser Modelltyp wurde hauptsächlich in den ersten SPC-Untersuchungen (z.B. [495,549]) eingesetzt, ist aber auch heute noch sehr häufig anzutreffen.

Um die inneren Abläufe nachzubilden, sind Erweiterungen notwendig. Beispielsweise durch Rückkopplungen [318,452,480,544] (mit oder ohne Verzögerungsglied) oder die Erweiterung einer Bedieneinheit auf mehrere Bedienphasen zu einem *Prozessormodell* (sogenannter *N-Server*) [330].<sup>20</sup> Beide Maßnahmen können auch kombiniert werden. Weiterhin werden ein- und mehrstufige Modelle unterschieden [332].

Aus Gründen der analytischen Handhabbarkeit werden bei diesen Modellen die Bedienzeiten häufig als negativ-exponentiell verteilt oder deterministisch angenommen.

Da in dieser Arbeit ein Mehrklassensystem (*multi-class network*) betrachtet wird, wurden auch Warteschlangennetze mit sogenannten *BCMP*-Eigenschaften [81] näher betrachtet. Diese Netzwerke erlauben eine exakte Lösung (Produkt-Form) für allgemeine Bedienzeitverteilungen und bestimmte Abfertigungsdisziplinen. Dieses Modell wird auch für ATM-Vermittlungsrechner verwendet [509,559]. Für diese Arbeit scheidet dieser Ansatz jedoch aus, da u.a. Mechanismen der Lastregelung zu berücksichtigen sind.

## 5.4.3 Komplexere Modelle

Ein wesentliches Kriterium bei der Auslegung eines Lastregelkreises ist sein dynamisches Verhalten. Hierzu müssen *zeitabhängige* Leistungsdaten ermittelt werden. Weiter sind insbesondere interne Betriebsmechanismen und Steuerungsverfahren zu berücksichtigen. Diese Eigenschaften lassen sich nicht mehr in einem einfachen Modell realisieren. Zwangsläufig nimmt dann die Modellkomplexität zu.

Beispiele für Modelle, die Überlastabwehrstrategien berücksichtigen und noch einer analytischen bzw. numerischen Lösung zugänglich sind, werden in [73,523] analysiert. Die Hauptschwierigkeit stellt hier der „explodierende“ Zustandsraum dar.

---

<sup>19</sup> Z.B. wenn nur mittlere Gesamtverzögerungszeiten oder Durchsatzwerte für bestimmte Lastpunkte vorliegen.

<sup>20</sup> Eine an einem Vermittlungsrechner ankommende Signalisiernachricht löst intern die Bearbeitung einer Prozess- oder Meldungskette aus [136,185,472]. Diese Prozesskette wird dann auf eine mehrphasige Bedieneinheit als Prozessormodell abgebildet. In [258] wird beispielsweise ein *N-Server* zur Modellierung eines Paketknotens verwendet.

Modelle für hierarchische Steuerungsstrukturen und Mehrprozessorsystemen sind u. a. in [160, 162] zu finden. Detailliertere Modelle von ATM-Vermittlungssteuerungen sind in [469, 509, 559] beschrieben, wobei diese nur für stationäre Analysen konzipiert sind. Bei der Vermittlung von IP-Verkehr in ATM-Netzen werden durch den Auf- und Abbau von ATM-Wählverbindungen spezielle Lastprofile an die Steuerung erzeugt. In [401] wird hierzu ein entsprechendes Systemmodell (*MAP/G/I/K*) für den ATM-Vermittlungsrechner vorgeschlagen.

## 5.4.4 Entworfen und verwendete Simulationsmodelle

### 5.4.4.1 Anforderungen

Einfache Modelle erleichtern wesentlich die Analyse, liefern aber unter Umständen zu grobe Ergebnisse, da sie die internen Systemabläufe nicht in der notwendigen Genauigkeit wiedergeben können. Die Leistungsanalysen im Rahmen dieser Arbeit sind weitestgehend Untersuchungen im instationären Bereich. Für zeitabhängige Funktionsverläufe ist die Simulationmethode prädestiniert und oft auch die einzig brauchbare Methode. Folgende Anforderungen sollen berücksichtigt werden:

- Vermittlungsrechner besitzen Multitasking-Betriebssysteme mit Prioritätensteuerung.
- Unterschiedlichen Diensten und auch sonstigen Funktionen können unterschiedliche Prioritäten zugewiesen werden.
- Die Bearbeitung einer neuen Anforderung zerfällt in eine Kette von unterschiedlichen Aufgaben, die mit verschiedenen Prioritäten bearbeitet werden können.
- Die Interprozess- wie auch die Interprozessorkommunikation<sup>21</sup> erfolgt in der Regel nachrichtenbasiert.
- Neben dem Nutzverkehr soll auch eine nichtvermittlungstechnische Lastkomponente berücksichtigt werden.
- Mechanismen der Last- und Leistungsregelung.

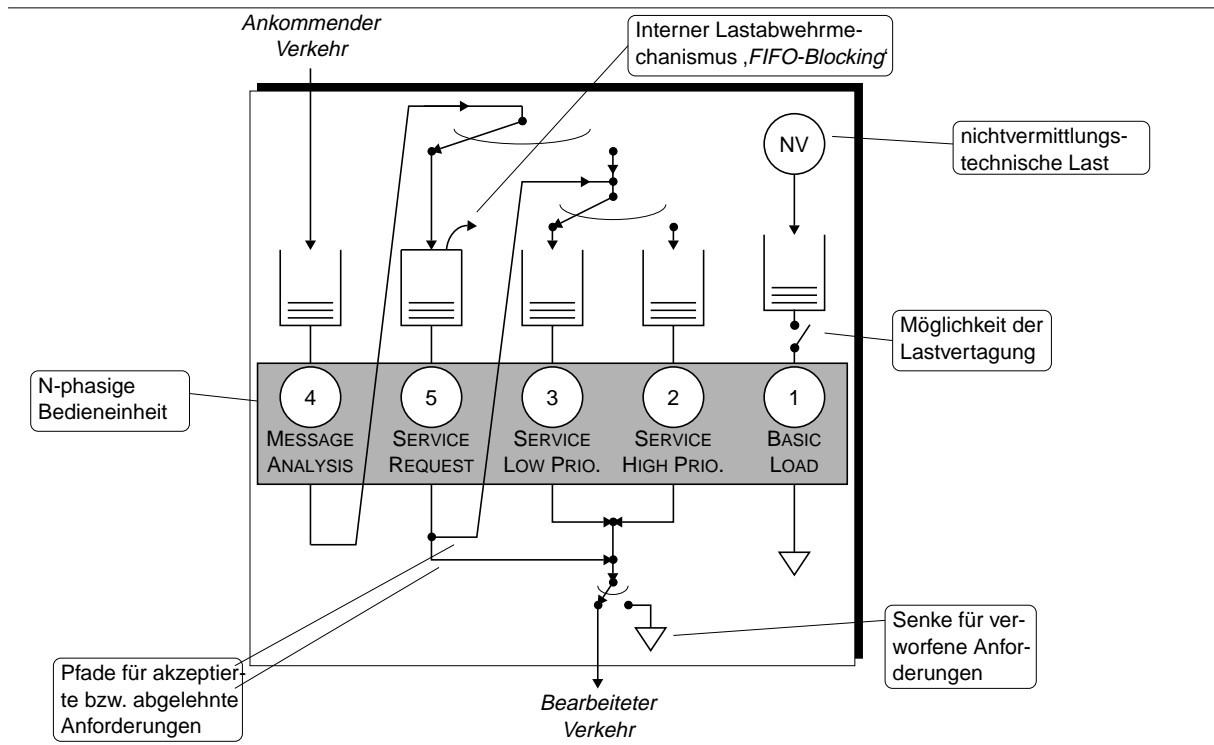
### 5.4.4.2 Der *N-Server* als Elementarmodell eines Prozessorknotens

Aus den genannten Gründen soll ein mehrphasiger Bedienserver (auch *N-Server* genannt) mit unterschiedlichen Prioritäten als zentrale Modellkomponente eingesetzt werden. Nicht berücksichtigt werden hier die interne Synchronisation zusammengehöriger Nachrichten oder Fluss-Steuerungsverfahren zwischen verschiedenen Prozessorknoten. Solche Mechanismen wurden beispielsweise im Modell von FAIß [174] berücksichtigt.

Das verwendete Prozessormodell ist in Bild 5-1 dargestellt. Der Vorschlag basiert im wesentlichen auf den Erfahrungen von [469, 472].

---

<sup>21</sup> Sowohl die Kommunikation zwischen Prozessorknoten innerhalb der zentralen Vermittlungssteuerung wie auch zu peripheren Steuereinheiten der Nutzer-Ebene.



**Bild 5-1:** N-Server als Warteschlangenmodell eines Prozessorknotens

Der ankommende Verkehr wird aus Signalisiernachrichten gebildet. Ein bestimmter Teil der Protokollinformation muss von jeder Nachricht extrahiert und analysiert werden. Danach ist bekannt, ob es sich um eine neue Dienstanforderung oder um eine Nachricht für einen bereits in Bearbeitung befindlichen Dienst handelt. Diese wird anschließend der entsprechenden Folgephase zugestellt.

Phase	Priorität	Beschreibung
MESSAGE ANALYSIS	4 mittel	Analyse von Protokollinformationen
SERVICE REQUEST	5 niedrig	Aufruf von Funktionen für neue Dienstanforderungen
SERVICE LOW PRIORITY	3 mittel	Bearbeitung von Diensten niedriger Priorität
SERVICE HIGH PRIORITY	2 mittel	Bearbeitung von Diensten hoher Priorität
BASIC LOAD	1 hoch	Nichtvermittlungstechnische Funktionen

**Tabelle 5-2:** Beschreibung der einzelnen Bedienphasen

Für eine neue Dienstanforderung werden zum einen eine Reihe typischer ATM-Funktionen wie z.B. die CAC aufgerufen, zum anderen muss die lokale Lastregelung entscheiden, ob die Anforderung angenommen werden kann (NRAC). Weiterhin werden Ressourcen belegt wie beispielsweise Speicherbereiche für Ruf- und Verbindungsdaten. Eine neue Anforderung wird entweder *angenommen*, *zurückgewiesen* oder *verworfen*. Jede dieser Entscheidungen führt zu einem unterschiedlichen Bedarf an Prozessorzeit, bis der Dienst beendet wird.

Akzeptierte Anforderungen, wie auch Nachrichten für bestehende Kommunikationsbeziehungen, werden entsprechend ihrer Dienstpriorität an die jeweilige Phase weitergeleitet. Dort fin-



det die eigentliche Bearbeitung des Dienstes, d.h. hier Dienstvermittlung und -steuerung, statt.<sup>22</sup>

Tabelle 5-2 zeigt die vorgenommene Priorisierung der Bedienphasen. Die höchste Priorität besitzt hier die nichtvermittlungstechnische Last, da diese auch die sicherungstechnischen Funktionen enthält, die den zuverlässigen Betrieb des Vermittlungssystems garantieren müssen. Gerade unter Überlast ist es wirtschaftlich sinnvoll, der Komplettierung bestehender Dienste eine höhere Priorität zu geben als der Bearbeitung neuer Anforderungen.<sup>23</sup> Entsprechend sind die Prioritäten gewählt. Weiterhin kann Phase 4 die Bearbeitung eines Auftrages in Phase 3 unterbrechen. Ansonsten sind alle Prioritäten nichtunterbrechend (*non-preemptive*).

Die Abfertigungsdisziplin ist FIFO für alle Warteschlangen. Eine bekannte Überlastabwehrmaßnahme ist die Umschaltung auf LIFO-Strategie für den Puffer der ‚SERVICE REQUEST‘-Phase. Dieser Mechanismus wurde bereits anderweitig ausführlich untersucht (siehe auch Abschnitt 3.4.9.2) und ist nicht Gegenstand dieser Arbeit.

Ergänzt wird dieses Modell noch um einen Kontrollpfad mit Komponenten für Regelkreise zur Last- und Leistungsregelung. Mit Hilfe dieses Elementarmodelles können relativ einfach komplexere Modelle für Mehrprozessorsysteme aufgebaut werden. Liegt bei diesen Steuerungsstrukturen Funktionsteilung vor, dann ändern sich gegebenenfalls Anzahl und Bedeutung von Bedienphasen sowie die interne Struktur. In Anhang C sind zwei Beispielmmodelle dargestellt, die die Grundlage für die durchgeführten Messungen bilden.

## 5.4.5 Bedienzeitverteilung

Unter Bedienzeit wird die Belegungsdauer des Betriebsmittels Prozessor verstanden. Bedienzeiten werden entweder direkt aus Labormessungen gewonnen (z.B. [180,470]), oder sie müssen geschätzt werden. Die Bearbeitungszeiten für Aufträge in Vermittlungsrechnern werden üblicherweise als deterministisch angenommen. Dies wird dadurch begründet, dass der Umfang an Programmcode<sup>24</sup> bzw. die Befehlsanzahl für einen bestimmten Dienstbearbeitungsvorgang nahezu immer gleich ist.

In der Realität sind diese Zeiten jedoch fast nie konstant, da der aktuelle Systemzustand oder bestimmte Algorithmen (z.B. Suchvorgänge) einen Einfluss auf die Ausführungszeiten besitzen. In vielen Leistungsuntersuchungen werden deshalb die CPU-Belegungszeiten durch eine negativ-exponentielle Verteilung modelliert. Dieser Ansatz wird auch dann oft verwendet, wenn keine meldungstypspezifischen Bedienzeiten unterschieden werden, sondern alle Signalmeldungen derselben Bedienzeitverteilung gehorchen sollen [523].

Andererseits hat sich bei bestimmten Systemen in der Praxis gezeigt, dass die Verteilung der Bedienzeiten eine kleinere Varianz besitzt, als es die negativ-exponentielle Verteilung impliziert [269]. Für die Modellierung von Bedienzeiten eines Vermittlungsrechners wäre deshalb eine Verteilung, deren Eigenschaften zwischen denen der deterministischen und der negativ-exponentiellen liegen, der Realität am nächsten (z.B. Erlang-k-Verteilung oder Reihenschaltung von D- und M-Phase).

---

<sup>22</sup> Bei Konzeption des Netzknotens als SSP sind die Dienststeuerungsfunktionen in zentrale Rechner ausgelagert. Für die Modellierung hätte dies folgende Konsequenzen: zusätzlicher Server als SCP, Erweiterung des Signalisierenszenarios sowie prinzipiell kleinere Bearbeitungszeiten in den Dienstbearbeitungsphasen des SSP (siehe auch [472]).

<sup>23</sup> Durch die Bevorzugung von bereits akzeptierten Diensten wird eine *Blindlastminimierung* erreicht.

<sup>24</sup> In [360] wird eine Bedienzeitverteilung aus der Verteilung der Größe des Objektcodes verschiedener Meldungstypen abgeleitet.

## 5.5 Methoden zur Leistungsuntersuchung

In [163,265] wird eine Übersicht gängiger Methoden zur stationären Analyse von Steuerungsstrukturen, allerdings ohne Mechanismen zur Last- und Leistungsregelung, gegeben. Für die Untersuchung von Überlastabwehrverfahren werden hauptsächlich die nachfolgend beschriebenen Analysemethoden eingesetzt.

### 5.5.1 Analytische Verfahren

Eine analytisch geschlossene Lösung ist prinzipiell das Ziel jeder Systemcharakterisierung. Diese Beschreibungsform scheint aber für die meisten Systeme nicht zu existieren, oder sie ist aufgrund der Komplexität oder mangels mathematischer Methoden nicht zugänglich. Falls für ein Warteschlangensystem eine Zustandsbeschreibung noch möglich ist, dann lassen sich die Gleichungssysteme in der Regel durch numerische Verfahren lösen. In [20,73,261,522] werden instationäre Ergebnisse von Überlastabwehrstrategien mittels dieser Methode berechnet.

Ist man nur am ersten Moment von Kenngrößen interessiert, dann kann eine Signalflussanalyse oder eine Fluid-Flow-Approximation bereits zu guten Ergebnissen führen. Bei der Fluid-Flow-Approximation wird der Verkehr durch einen kontinuierlichen Flüssigkeitsstrom mit zufallsabhängiger Intensität nachgebildet [616]. Überlastuntersuchungen, die auf diesen Ansätzen beruhen, sind beispielsweise in [146,172,173,189,406] zu finden.

### 5.5.2 Simulationstechniken

Die Simulation ist die geeignete Methode für aufwendige Modelle. Da durch dieses Verfahren auch *zeitabhängige Eigenschaften* von Prozessverläufen gemessen werden können, ist die Simulationsmethode die Grundlage der Leistungsuntersuchungen im Rahmen dieser Arbeit. Da für diese spezielle Simulationstechnik zunächst ein Konzept festgelegt und anschließend eine Erweiterung der IND-Simulationsplattform [315] implementiert werden musste, bildete die Bereitstellung dieses Werkzeuges einen nicht unerheblichen zeitlichen Anteil an dieser Arbeit.

#### 5.5.2.1 Ereignisgesteuerte Simulation

Für verkehrstheoretische Modelle hat sich die ereignisgesteuerte, zeitreue Simulation bewährt. Da in diesen Systemen nur zu diskreten Zeitpunkten – „Eintreten eines Ereignisses“ – Zustandsänderungen stattfinden, ist es ausreichend, nur diese zu betrachten. Die bei einer Simulation durchgeführten Messungen liefern Stichproben von Zufallsgrößen. Für die Schätzung von Vertrauensintervallen [335,461] sollten diese das Ergebnis unabhängiger Beobachtungen sein, damit die Annahme, „*alle Messwerte sind statistisch unabhängig und folgen derselben Verteilung*“, begründet ist.

Eine Einführung und gute Übersicht zu dieser Simulationstechnik ist beispielsweise in [21,43,46,242] zu finden.

### 5.5.2.2 Simulation stationärer Vorgänge

Ein stochastischer Prozess besitzt die Eigenschaft der Stationarität, wenn seine statistischen Merkmale invariant gegenüber Zeitverschiebungen sind. Nahezu bei allen Simulationen verkehrstheoretischer Modelle werden stationäre Prozessverläufe vorausgesetzt und stationäre Kennwerte ermittelt. Begründet wird dies dadurch, dass bei einer angelegten stationären Last auch die Systemkenngrößen stationäres Verhalten aufweisen.<sup>25</sup>

### 5.5.2.3 Simulation transienter Vorgänge

Reale Systeme zeigen dagegen häufig kein stationäres Verhalten. Beispielsweise werden durch die in dieser Arbeit betrachteten Regelmechanismen instationäre Vorgänge ausgelöst. Neben der Instationarität als inneres Merkmal eines Systems bildet die externe Lasterzeugung durch instationäre Prozesse die zweite Ursache dafür, dass nicht stationär gemessen werden darf. Gerade bei Regelkreisen sind Aussagen über das zeitliche Verhalten von starkem Interesse. Viele Algorithmen besitzen zwar dieselben stationären Durchsatzkurven, unterscheiden sich aber erheblich bezüglich ihres transienten Verhaltens [318].

Während bei einer stationären Simulation Zeitmittelwerte bestimmt werden, interessieren bei einer instationären Simulation die Schar- oder Ensemblemittelwerte von stochastischen Prozessen [15]. Diese Anforderung hat entscheidenden Einfluss auf die Organisation der Zufallsexperimente innerhalb einer Simulation und der Messmethodik.

## 5.6 Technik der instationären Simulation

Die instationäre Simulation bildet die zentrale Methode zur Untersuchung von Überlastvorgängen. Bevor das eingesetzte Werkzeug vorgestellt wird, sollen eine kurze Einführung und Erfahrungen zu diesem Verfahren gegeben werden.

### 5.6.1 Übersicht

Typische Anwendungsfelder instationärer Simulationen sind verkehrstheoretische Untersuchungen von *Regelkreisen* (z.B. Überlastabwehr als systeminterner Regelkreis, Rufwiederholungen oder Flusskontrolle als externer Regelkreis), *Steuerungsstrategien* (z.B. Scheduling, Lastverteilung, adaptives Routing), *Reaktionen auf Strukturänderungen* (z.B. Ausfälle von Systemkomponenten, Topologieänderungen) oder *Systemreaktionen auf Laständerungen* (z.B. Einschaltvorgänge, Lastimpulse).

Die Anforderung nach instationären Messungen besteht seit Anwendung der Simulationsmethode. Während zur Technik der stationären Simulation bereits eine Vielzahl an Veröffentlichungen erschienen ist, gibt es auf dem Gebiet der instationären Simulation relativ wenige Publikationen (z.B. [59,284,356,524]). Die umfassendste systematische Aufarbeitung zu dieser Methodik stammt von LOVEGROVE [363,364].

---

<sup>25</sup> Durch eine sogenannte Warmlaufphase wird das System zuerst in den eingeschwingenen Zustand gebracht.

Für das realisierte Werkzeug zur Simulation instationärer Vorgänge wurde zunächst ein Konzept ausgearbeitet, das hauptsächlich die Gebiete Lasterzeugung, Organisation einer Simulation und Messmethodik umfasst. Dieses wurde bereits in [473,474] ausführlich dokumentiert und soll deshalb hier nicht näher erläutert werden.<sup>26</sup>

## 5.6.2 Erfahrungen auf dem Gebiet der Überlastabwehr

Für die Analyse von Überlastabwehrmechanismen, also der Leistungsuntersuchung von Lastregelkreisen, ist die Methode der instationären Simulation geradezu prädestiniert, wenn nicht sogar unverzichtbar. Betrachtet man aber Veröffentlichungen auf diesem Gebiet, dann ist auffällig, wie wenig diese Methode offenbar angewandt wird. Folgende Gründe scheint es dafür zu geben:

- *Mechanismen* zur lokalen Lastregelung unterliegen in der Regel *nicht der Standardisierung*, sondern sind vielmehr herstellereigenspezifische Lösungen. Diese Funktion ist vordergründig nicht die Vermittlungsfunktion, die zum Gewinn beiträgt, und wird deshalb zunächst eher mit niedriger Priorität behandelt. Oft wird auch ein konservatives Vorgehen gepflegt, d.h. ein Verfahren muss zuerst sicher funktionieren, eine Optimierung wäre wünschenswert, steht aber nicht an dringlichster Stelle. Auf der anderen Seite scheint auch die Zeit für notwendige und relativ lange dauernde Parameterstudien zu kostbar zu sein.
- *Methoden zur Leistungsuntersuchung* von Systemen werden oft nicht beherrscht [317], sind unbekannt oder es fehlt an Erfahrung auf den Gebieten Modellierung von Systemen, Simulationstechnik oder statistische Grundlagen für aussagesichere Messdaten.
- Problem der *Verfügbarkeit* eines geeigneten Werkzeugs. *Software-Tools* zur instationären Simulation sind in der Regel Speziallösungen.
- *Aufwandsgründe*: Für Modellierung, Modellimplementierung, Rechenläufe und Parameterstudien ist viel Zeit erforderlich.

Es lässt sich auch beobachten, dass veröffentlichte Untersuchungen zu dynamischem Zeitverhalten fast immer unzureichend (z.B. *Traces* als repräsentative Messresultate, fast nie Abschätzung von Konfidenzbereichen, oft nur stationäre Durchsatzkurven) sind.

Wie bereits angedeutet wurde, kann die Simulationsmethode entfallen, wenn analytische bzw. numerische Lösungsverfahren anwendbar oder ausreichend sind. Dies war bei vielen rechnergesteuerten Vermittlungssystemen der ersten Generation der Fall, da dort die Randbedingungen (z.B. Verkehrsmixe, Systemstrukturen, Einsatzorte) und Algorithmen einfacher waren.<sup>27</sup>

Auf der anderen Seite werden an die Lastregelung von Breitbandsystemen erhöhte Anforderungen, z.B. durch die notwendige Unterstützung verschiedenster Dienstgüten für verschiedene Verkehrsklassen, gestellt. Die hierfür notwendigen Regelstrategien und Algorithmen erfordern die instationäre Simulation als Methode zum Leistungsnachweis der Verfahren.

---

<sup>26</sup> Die Methode der instationären Simulation wurde am IND bereits in früheren Arbeiten eingesetzt [122, 182, 340, 438, 516, 520, 551].

<sup>27</sup> Trotzdem wurde auch damals teilweise schon instationär simuliert (z.B. [246]).

### 5.6.3 Flexibles Werkzeug für die instationäre Untersuchung von Vermittlungsrechnern

Aus wirtschaftlichen Gründen sollte jedes Werkzeug so flexibel konzipiert sein, dass es für ein breites Anwendungsspektrum einsetzbar ist. Im Rahmen dieser Arbeit sind diese Anforderungen durch die unterschiedlichen Steuerungs- und Regelkreisstrukturen gegeben. Für die C++-Simulationsbibliothek des IND [315] wurden folgende Erweiterungen entwickelt und implementiert:

- *Bibliothek mit Komponenten zur Technik der instationären Simulation* [418,473,474]  
Beinhaltet Klassen von Generatoren zur Verkehrserzeugung, Klassen von Strukturkomponenten (z.B. Bedieneinheit, Warteschlangen, Schalter), Messgeräten, Statistiken und Klassen zur Simulationssteuerung.
- *Bibliothek mit Komponenten für Lastregelkreise* [63,199,418]  
Beinhaltet viele Klassen zur Sensorik (z.B. unterschiedliche Typen von Lastindikatoren), Messwertaufbereitung (z.B. Filter- und Schätzverfahren), Aktorik (z.B. temporäre Lastvertagung, Prioritätensteuerung) und verschiedenen Reglertypen. Zusätzlich sind auch mehrstufige, hierarchische Regelungen möglich.<sup>28</sup> Weiterhin lassen sich durch bestimmte Zähler- und Filtertypen *Repeated Call Attempts* simulieren.

Zusätzlich kann der Aufbau einer bestimmten Modellstruktur in einem begrenzten Bereich dynamisch realisiert werden, da alle Komponenten parametrisierbar sind (z.B. Anzahl und Typen von Sensoren, Reglerkennlinien, Warteschlangeneigenschaften).

Damit wurde ein flexibles Werkzeug realisiert, das eine Basis für die simulative Leistungsuntersuchung beliebiger Steuerungsstrukturen von Vermittlungsrechnern mit beliebigen Regelkreisstrukturen darstellt.

---

<sup>28</sup> Über definierte Schnittstellen erfolgt der Austausch von Simulationsnachrichten zwischen verschiedenen Modellkomponenten des Datenpfades (sogenanntes *Portkonzept*). Dieses Konzept wird auch für Kontrollpfadelemente verwendet, was wiederum die Flexibilität und damit die Realisierung beliebiger Regelkreisstrukturen unterstützt.

# Kapitel 6

## Untersuchung der Annahmekontrolle für neue Kommunikationsbeziehungen

Für die Last- und Leistungsoptimierung eines generischen B-ISDN-Vermittlungsrechners wurde in Kapitel 4 ein spezielles Konzept einer mehrstufigen Regelkreisstruktur vorgeschlagen. Dabei wurde u. a. ein neuartiges Verfahren zur *Annahmekontrolle neuer Dienstanforderungen* (NRAC) präsentiert (siehe Abschnitt 4.4). Dieser präventiv wirkende Regelkreis soll in diesem Kapitel näher untersucht und bewertet werden.

### 6.1 Fallstudien im Überblick

#### 6.1.1 Zielsetzung

Wie bereits in Abschnitt 4.4.3 ausgeführt wurde, werden bei der NRAC die klassenspezifischen Ressourcenbedarfsparameter regelmäßig nachgeführt. Damit ist die NRAC als adaptives Verfahren tolerant gegenüber Parameterunschärfen, die auch aufgrund von Informationsdefiziten zwangsläufig hier nicht vermeidbar sind, und entlastet somit den Netzbetreiber. Erwartet wird deshalb vor allem ein robustes Verhalten.

Durch Messungen soll der funktionelle Nachweis für diesen Lastregelkreis erbracht werden. Entsprechend dem zweigeteilten Algorithmus werden aus Gründen der Übersicht und Parametervielfalt die Simulationen in folgende zwei Hauptgruppen unterteilt:

- Automatische Parameternachführung und Lastprognose
- Annahmealgorithmus für neue Dienstanforderungen

Für diese Fallstudien ist es ausreichend, ein Teilmodell des generischen Vermittlungsrechners zu verwenden. Da dieser Mechanismus als lokaler Lastregelkreis für einen *B*-Prozessorknoten  $PN_B$  konzipiert ist, besteht der Modellkern aus dieser Prozessoreinheit. Zusätzlich wird noch ein *A*-Prozessorknoten  $PN_A$  vorgeschaltet. Das Simulationsmodell ist in Anhang C.1.1 abgebildet und näher dokumentiert.

Alle Ergebnisse entstammen instationären Simulationen. Es wurden jeweils 95%ige Konfidenzbereiche geschätzt.<sup>1</sup> Nachfolgende Schaubilder repräsentieren nur einen kleinen, aber repräsentativen Ausschnitt der durchgeführten Messreihen.

---

<sup>1</sup> Falls die Vertrauensintervalle relativ klein waren, dann wurden diese aus Gründen der Übersichtlichkeit, zumal auch noch relativ viele Messzeitpunkte genommen wurden, weggelassen. Dies gilt für alle Messungen im Rahmen dieser Arbeit.

## 6.1.2 Fixe Modell- und Verkehrsparameter

Die Grenze für eine sinnvolle zeitliche Auflösung wird durch die „kleinsten“ Elementarereignisse vorgegeben. In diesem Fall sind dies die Meldungsbearbeitungszeiten in den Prozessorknoten. Es sei angenommen, dass der Mittelwert im Bereich weniger Millisekunden liege. Diese Annahme ist auf jeden Fall für die aktuelle Mikroprozessorgeneration gültig und kann auch durch zahlreiche Veröffentlichungen belegt werden (z.B. [201,262,565]). Für die instationären Statistiken wurden in der Regel Messintervalle von  $T_{MI} = 0.5ZE$  eingestellt.<sup>2</sup>

Der simulierte Verkehr setzt sich aus den in Abschnitt 5.2.6.5 beschriebenen drei Klassen zusammen. Die Verkehrsklassen besitzen eine unterschiedliche Charakteristik und einen unterschiedlichen Betriebsmittelbedarf (siehe Anhang C.3.3.1). In Prozessorknoten  $PN_B$  wird zusätzlich eine Hintergrundlast berücksichtigt. Alle weiteren Parametereinstellungen sind bei den jeweiligen Messungen angemerkt.

## 6.2 Automatische Parameternachführung

Das Leistungsverhalten der dynamischen Anpassung der Regelparameter wird direkt von folgenden Aspekten beeinflusst:

- Lastprognosealgorithmus,
- Adaptionsgesetz,
- Adaptionsintervalllängen und
- Mess- bzw. Schätzverfahren.

Während die ersten beiden Punkte wesentlich von den verfügbaren Informationen abhängen, wird die Regelgüte bei den anderen Punkten mehr vom Aufwand, und damit von den Kosten, die man in die Berechnung investiert, beeinflusst.

### 6.2.1 Lernen der sogenannten Originalparameter

Den Ausgangspunkt für die Schätzung der effektiven Prozessorbelastung einer Verkehrsklasse ist Gleichung (4-6). Deren Wert stellt den ersten Ansatz für den sogenannten Originalschätzparameter  $\hat{r}_{p,v}^{orig}$  dar. Während sich ein ungefährender Anhaltspunkt für den CPU-Bedarf  $T_{H,v}$  eines Dienstes theoretisch noch grob aus der Systemrealisierung (Hardwareplattform, Umfang des Programmcodes, Betriebssystemeigenschaften) abschätzen ließe, muss für die mittlere Sitzungsdauer  $T_{S,v}$  ein hypothetischer Wert veranschlagt werden. Dieses „Nichtwissen“ wurde durch die zugrundeliegende Verkehrsmodellierung zumindest teilweise berücksichtigt. In Tabelle 6-1 sind die entsprechenden Modellparameter zusammengestellt.

Da die Zeitebenen der Bedien- und Sitzungsdauern normalerweise um mehrere Größenordnungen differieren, wurden für überschaubare Simulationszeiten die Werte  $T_{S,v}$  kleiner als in der Realität zu erwarten gewählt.<sup>3</sup>

---

<sup>2</sup> Zeitangaben: In dieser Arbeit sind alle Zeitwerte mit der neutralen Einheit *Zeiteinheit* (ZE) bezeichnet. Es werden keine absoluten Angaben gemacht, um keinen Bezug zur aktuellen Prozessortechnologie und Halbleitergeneration herzustellen. Gewisse Zeitparameter von Diensten (z.B. mittlere Rufdauer) sind davon natürlich weniger betroffen.

Verkehrsklasse	Mittlere Gesamtbearbeitungszeit	Mittlere Gesamtdauer	Verhältnis $T_S:T_H$	Effektive Prozessorbelastung
1	$\bar{T}_{H,1} = 0.050\text{ZE}$	$\bar{T}_{S,1} = 70\text{ZE}$	1400	$\hat{r}_{\rho,1}^{org} = 0.07142\%$
2	$\bar{T}_{H,2} = 0.020\text{ZE}$	$\bar{T}_{S,2} = 80\text{ZE}$	4000	$\hat{r}_{\rho,2}^{org} = 0.02500\%$
3	$\bar{T}_{H,3} = 0.025\text{ZE}$	$\bar{T}_{S,3} = 40\text{ZE}$	1600	$\hat{r}_{\rho,3}^{org} = 0.06250\%$
BL	–	–	–	$\hat{r}_{\rho,BL}^{org} = 5.00\%$

**Tabelle 6-1:** Parameter für Ressourcenabschätzung ( $PN_B$ )

Wie in Abschnitt 4.4.3.3 ausführlich erläutert wurde, ist ein wichtiger Aspekt dieses Verfahrens die dynamische Nachführung der Schätzwerte  $\hat{r}_{\rho,v}$  für die einzelnen Verkehrsklassen  $v$ . Durch diese Adaptivitätseigenschaft des Algorithmus wird vor allem den Anforderungen *Güte* und *Robustheit* Rechnung getragen. Auf Grund stochastischer Verkehrsschwankungen und auch aus Kostengründen erfolgt eine Parameteranpassung nur von Zeit zu Zeit. Bei nachfolgenden Ergebnissen wurden die Adaptionintervalle mittels einer Zählvariablen  $C_{ALC}$  realisiert. Dies hat den Vorteil, dass die Intervalldauer lastabhängig ist.<sup>4</sup>

### 6.2.1.1 Stationäre Normallast

Das Adaptionsgesetz für die (neuen) Schätzwerte  $\hat{r}_{\rho,v}(k+1)$  ist in Gl. (4-21) beschrieben. Der Wertebereich wird durch die Begrenzungskonstanten  $\delta_L$  und  $\delta_H$  beschränkt. Prinzipiell wird in Abhängigkeit der Größe der Abweichung zwischen Ist- und Sollwert korrigiert (siehe Gl. (4-14)). Da aus Gründen des Aufwands hier nur auf die Gesamtabweichung reagiert wird, sind die klassenindividuellen Differenzwerte unbekannt. Bei der Festlegung der Originalschätzwerte kann nun die Situation eintreten, dass bestimmte Originalwerte zu groß und die anderen zu klein festgelegt wurden. Da aber für alle Klassen immer in dieselbe Richtung korrigiert wird, kann es theoretisch sehr lange dauern, bis einigermaßen zuverlässige Werte vorliegen.<sup>5</sup> Diesem Nachteil kann aber relativ einfach begegnet werden, in dem zu Beginn alle Werte wesentlich zu groß oder zu klein (pessimistische bzw. optimistische Schätzung) eingestellt werden. Bild 6-1 zeigt beispielhaft in einem *Simulationstrace*<sup>6</sup> den Einschwingvorgang für zu optimistisch angenommene Parameter.

Bei dieser Messung wurden die Ausgangswerte halb so groß wie der tatsächlich zu erwartende Ressourcenbedarf eingestellt. Selbst bei dieser äußerst groben Abschätzung wird die „Nähe“ der tatsächlichen Bedarfswerte nach relativ kurzer Zeit erreicht.<sup>7</sup> In der Messung von Bild 6-1

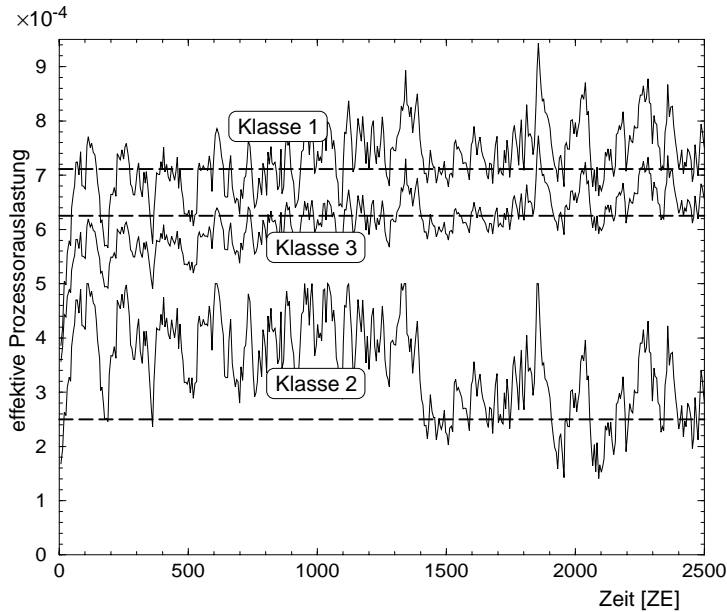
<sup>3</sup> Dieser „Skalierungsfaktor“ stellt aber für diese Untersuchungen keine wesentliche Einschränkung dar.

<sup>4</sup> Bei jeder REQUEST-Nachricht wird die Zählvariable  $C_{ALC}$  (*Adaption Loop Counter*) inkrementiert. Erreicht der Zähler die voreingestellte Konstante  $N_{ALC}$ , dann wird die Parameteradaptionfunktion aufgerufen und  $C_{ALC}$  wieder zurückgesetzt. Wenn die mittlere Ankunftsrate  $\bar{\lambda}_{RM} = (1/K) \sum_v \lambda_{RM,v}$  von REQUEST-Nachrichten aller  $K$  Verkehrsklassen ist, dann wird die Zählvariable  $C_{ALC}$  durchschnittlich alle  $1/\bar{\lambda}_{RM}$  Zeiteinheiten hochgezählt. Damit ist das Adaptionintervall last- und zeitabhängig:  $T_{AI}(\lambda_{RM,v}(t)) = N_{ALC}/\bar{\lambda}_{RM}$ . Je größer die Last, umso häufiger erfolgt eine Anpassung der Parameter.

<sup>5</sup> Die „Einschwingdauer“ auf stabile Werte hängt zum einen von der Größe des Gesamtangebotes ab und zum anderen vom aktuellen Verkehrsmix. Je nach Dominanz einer bestimmten Verkehrsklasse wird der entsprechende Wert für die effektive Prozessorauslastung „am besten“ adaptiert. D.h., erste gute Näherungswerte lassen sich gewinnen in dem das System nacheinander mit Verkehr der einzelnen Klassen belastet wird.

<sup>6</sup> Unter einem *Trace* (dt. Übersetzung: Spur, Nachzeichnung) wird in der Simulationstechnik die einmalige Simulation des Prozessverlaufs für einen bestimmten Zeitabschnitt verstanden. Dabei wird der (detaillierte) Zeitverlauf ausgewählter Messgrößen ausgegeben. Es findet somit keine Mittelwertbildung statt. Angaben über statistische Aussage-sicherheiten können deshalb auch nicht gemacht werden.





**Bild 6-1:** Beispiel für die Parameternachführung.

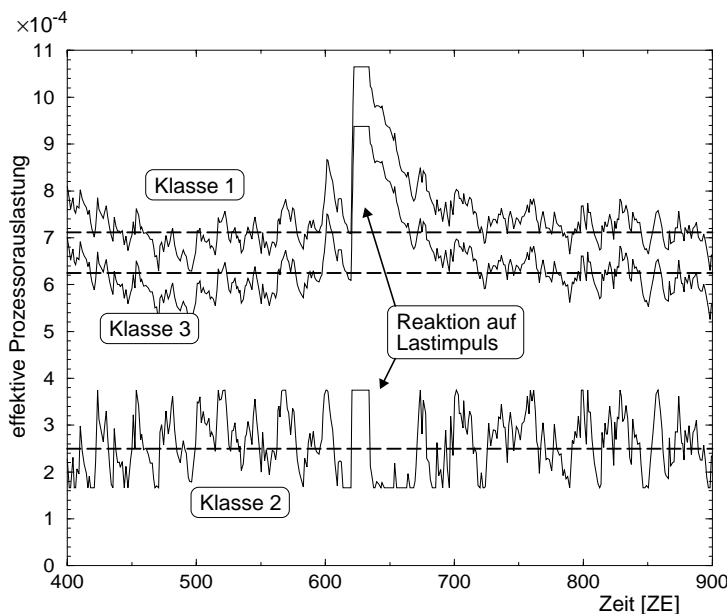
Darstellung der berechneten Schätzwerte  $\hat{r}_{\rho, v}$  der drei Verkehrsklassen und ihrer Sollwerte (gestrichelte Linien). Die Ausgangsschätzwerte (Originalparameter  $\hat{r}_{\rho, v}^{org}$ ) wurden auf 50% des tatsächlichen Ressourcenbedarfs eingestellt. Weitere Parametereinstellungen:

Adaptionszykluslänge  $N_{ALC} = 85$ , Dämpfung  $\alpha = 5 \cdot 10^{-4}$ , Begrenzungswerte  $\delta_L = 0.25$  und  $\delta_H = 4.0$ . Die Ankunftsraten  $\lambda_{RM, v}$  sind hier für alle Verkehrsklassen  $v$  gleich groß.

wurden die Begrenzungswerte relativ pessimistisch gewählt. Eine zusätzliche Verbesserung ließe sich durch einen engeren Begrenzungsbereich erzielen.

### 6.2.1.2 Verhalten unter starker Überlast

Im vorigen Abschnitt wurde das Einschwingverhalten für stationäre Verkehrswerte untersucht. Da aber in der Realität Lastschwankungen auftreten, soll hier die Reagibilität der Parameteradaption auf Lastimpulse überprüft werden. Bild 6-2 zeigt ein Beispiel mit Überlastimpulsen, die durch Verkehrsklasse zwei verursacht werden. Hier wurde nun der Begrenzungsbereich etwas enger gewählt. Nach einer kurzzeitigen starken Abweichung pendeln sich die Schätzwerte wieder auf ihre Sollwerte ein.



**Bild 6-2:** Parameternachführung bei Lastimpulsen.

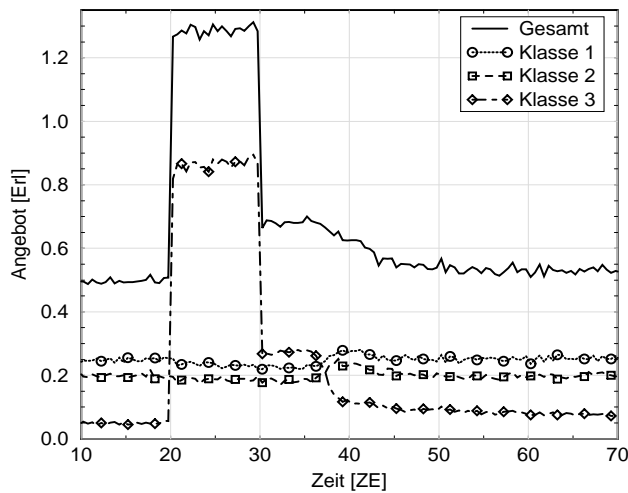
Darstellung der berechneten Schätzwerte  $\hat{r}_{\rho, v}$  der drei Verkehrsklassen und ihrer Sollwerte (gestrichelte Linien). Ein Lastimpuls entsteht hier durch Anstieg der Ankunftsrate für Klasse 2 für zehn Zeiteinheiten auf zehnfache Last (Rate  $\lambda_{RM, 2, NL} = 5$  bzw.  $\lambda_{RM, 2, \ddot{U}L} = 50$ ; Klasse 1 und 3:  $\lambda_{RM, 1} = 3.3$ ,  $\lambda_{RM, 3} = 5$ ).

Weitere Parametereinstellungen: Dämpfung  $\alpha = 5 \cdot 10^{-4}$ , Adaptionszykluslänge  $N_{ALC} = 17$ , Begrenzungswerte  $\delta_L = 0.67$  und  $\delta_H = 1.5$ .

<sup>7</sup> Die Qualität der Adaption hängt vornehmlich von den Parametern  $N_{ALC}$ ,  $\alpha$ ,  $\delta_L$  und  $\delta_H$  ab. Weiterhin spielen natürlich die Ankunftsraten  $\lambda_{RM, v}$  und die Ausgangsschätzwerte eine Rolle.

## 6.2.2 Einfluss der Adaptionsintervalldauer und anderer Parameter

Neben der Wirkung der Begrenzungswerte wurde der Einfluss der Parameter Adaptionsintervalldauer  $N_{ALC}$  und Dämpfungsfaktor  $\alpha$  analysiert. Bei diesen instationären Untersuchungen wurde jeweils ein Lastimpuls durch Klasse-3-Nachrichten angelegt. Dieser wird durch Anstieg der Ankunftsrate für zehn Zeiteinheiten auf 25fache Last gebildet. Bild 6-3 zeigt das resultierende mittlere Gesamtangebot sowie die zugehörigen Teilangebote. Die Annahmekontrolle wurde für die hier durchgeführten Parameteruntersuchungen deaktiviert, d.h. alle Anforderungen werden akzeptiert.

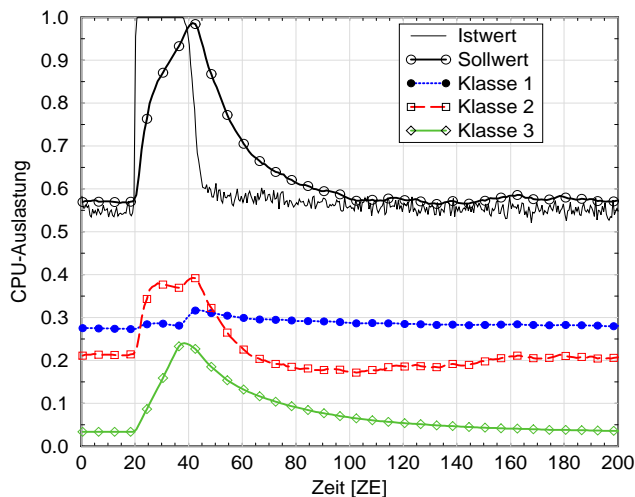


**Bild 6-3:** Lastprognose und Parameternachführung.

Darstellung des gemessenen Angebotes. Verkehrsklasse drei verursacht eine kurzfristige Überlast.

Die Variable  $N_{ALC}$  für die Adaptionsintervalldauer wurde innerhalb eines großen Bereiches variiert. Wie zu erwarten, ist die Abweichung zwischen Ist- und prognostiziertem Sollwert umso geringer, je kleiner die Adaptionszyklen sind. Aufgrund der stochastischen Verkehrsschwankungen muss allerdings ein Kompromiss bezüglich der Wahl von  $N_{ALC}$  gefunden werden.

Bild 6-4 zeigt eine exemplarische Messung mit  $N_{ALC} = 100$ . Selbst für diese starke Lastspitze folgt der berechnete Sollwert relativ gut der tatsächlichen mittleren Prozessorauslastung. Obwohl die mittleren Angebote von Klasse eins und zwei nahezu konstant sind (siehe Bild 6-3), sind es die prognostizierten Teillasten nach den Lastwechseln nicht. Dies rührt daher, dass die

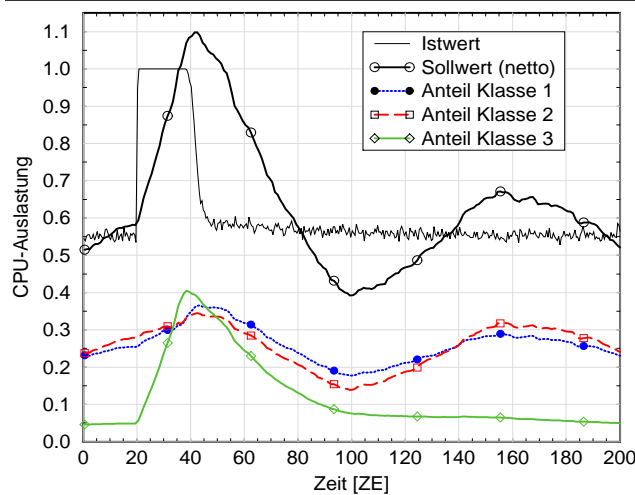


**Bild 6-4:** Lastprognose und Parameternachführung.

Darstellung des gemessenen Ressourcenverbrauchs  $\hat{p}_{Ist}$  (Istwert) und der verplanten Ressourcen  $\hat{R}_{p, ges}$  (Sollwert) sowie die Anteile der einzelnen Verkehrsklassen  $\hat{R}_{p, v}$ . Verkehrsklasse drei verursacht eine kurzfristige Überlast ( $\lambda_{RM,1} = 5$ ,  $\lambda_{RM,2} = 10$ ,  $\lambda_{RM,3,NL} = 2$  bzw.  $\lambda_{RM,3,ÜL} = 50$ ). Weitere Parametereinstellungen: Adaptionszykluslänge  $N_{ALC} = 100$ , Dämpfung  $\alpha = 5 \cdot 10^{-4}$ , Begrenzungswerte  $\delta_L = 0.6$  und  $\delta_H = 2.0$ .

Parameternachführung auf der berechneten Gesamtabweichung (siehe auch Abschnitt 6.2.1.1) basiert. Dadurch wird die zu erwartende Belastung durch diese Verkehrsklassen leicht überschätzt. Dieser Effekt wirkt sich nach einem Lastsprung allerdings nicht nachteilig aus, vielmehr verbessert sich das Einschwingverhalten. Nach der Hochlastphase pendeln sich die Ressourcenbedarfsschätzwerte wieder ein.

Neben der Adaptionzykluslänge ist der Dämpfungsfaktor  $\alpha$  für ein stabiles Regelkreisverhalten verantwortlich. Wiederum wurden verschiedene Fallstudien durchgeführt. In Bild 6-5 wurde bewusst eine Messung mit zu geringer Dämpfung dargestellt.



**Bild 6-5:** Lastprognose und Parameter-nachführung.

Darstellung des gemessenen Ressourcenverbrauchs  $\hat{\rho}_{Ist}$  (*Istwert*) und der verplanten Ressourcen  $\hat{R}_{\rho, ges}$  (*Sollwert*) sowie die Anteile der einzelnen Verkehrsklassen  $\hat{R}_{\rho, v}$ . Klasse drei verursacht hier eine Überlast ( $\lambda_{RM,1} = 5$ ,  $\lambda_{RM,2} = 10$ ,  $\lambda_{RM,3,NL} = 2$  bzw.  $\lambda_{RM,3,ÜL} = 50$ ). Weitere Parametereinstellungen: Adaptionzykluslänge  $N_{ALC} = 1000$ , Dämpfung  $\alpha = 10^{-2}$ , Begrenzungswerte  $\delta_L = 0.6$  und  $\delta_H = 2.0$ .

Wie dem Schaubild entnommen werden kann, treten hier langsam abklingende Schwingungen mit einer relativ niedrigen Frequenz auf. Dieses instabile Verhalten ist in einem Lastregelkreis nicht erwünscht. Durch eine stärkere Dämpfung, d.h. kleineres  $\alpha$ , wird diesen Oszillationseffekten entgegengewirkt.<sup>8</sup>

### 6.2.3 Mess- und Schätzverfahren

Die Güte eines Regelkreises hängt von vielen Parametern ab. Wichtige Kenngrößen wurden bereits in den vorangegangenen Abschnitten diskutiert. Entsprechende Aufmerksamkeit sollte aber auch der Sensorik eines stochastischen Regelsystem gewidmet werden. Diese Aussage gilt für alle Algorithmen dieser Arbeit. Stellvertretend wird dieser Aspekt hier im Zusammenhang des NRAC-Verfahrens erläutert.

Aufgrund des zufallsabhängigen Verlaufs vieler Größen und wegen Stabilitätsforderungen ist eine Filterung und Schätzung gemessener Lastindikatoren unbedingt notwendig (siehe auch Abschnitt 3.4). Dabei ist für die Länge der Mess- und Schätzintervalle ein Kompromiss zwischen Reagibilität und Vorhersagegenauigkeit (Prädiktion) zu finden.

Im Rahmen dieser Arbeit wurden verschiedene bekannte Verfahren implementiert und deren Güte untersucht [199,471]. Die fensterbasierte Mittelwertbildung („gleitendes Fenster“) ist für eine gute Approximation des *aktuellen Wertes* einer Systemvariablen geeignet. Stehen *Stabilitätskriterien* im Vordergrund, dann bietet sich eine exponentielle Glättung (*exponential smoothing*) mit ihrer Tiefpasscharakteristik an. Diese rekursive Berechnungsmethode wurde hier auch für alle Sensoren eingesetzt. Die Gewichtung erfolgte dabei mit 70% des aktuellen und

<sup>8</sup> In allen nachfolgenden Simulationen wurde  $\alpha = 5 \cdot 10^{-4}$  gewählt.

mit 30% des letzten Wertes. Als Schätzintervalllänge wurde je nach Sensortyp das 100 bis 1000fache der kürzesten mittleren Bedienzeit einer Meldung in einer bestimmten Phase eingestellt (sogenanntes Elementarereignis).

Das gewählte Filterverfahren und die Parameterwahl haben sich in den durchgeführten Untersuchungen bewährt.

## 6.2.4 Lastprognosealgorithmus

Der Prognosealgorithmus soll den erwarteten mittleren Bedarf an Betriebsmitteln für einen Dienst einer bestimmten Verkehrsklasse abschätzen. Dabei steht die wichtige Ressource *CPU* im Mittelpunkt des NRAC-Verfahrens. Dieser Algorithmus lässt sich aber prinzipiell noch für zusätzliche Parameter erweitern. Diese Thematik soll im zweiten Unterabschnitt weiter ausgeführt werden.

### 6.2.4.1 Effektive Prozessorbelastung

Der Algorithmus für die Lastvorhersage nach Gl. (4-11) soll anhand leichter Lastschwankungen demonstriert werden. Der Sollwert  $\hat{\rho}_{Soll}$  hängt dabei unmittelbar vom Verkehrsmix und den Ressourcenparametern ab.

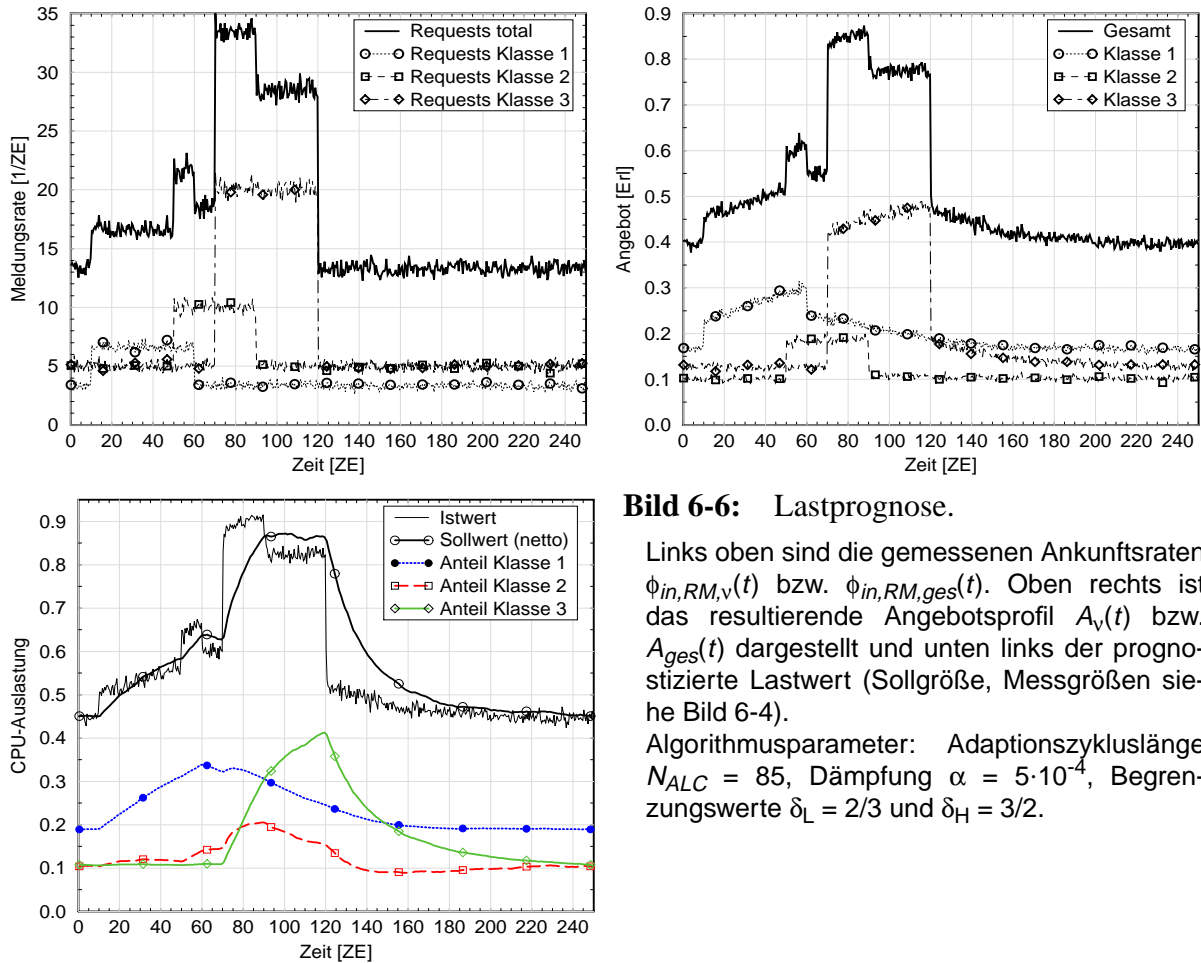
In den einzelnen Fallstudien wurde der Verkehrsmix zeitlich variiert. Dabei besitzt jede Verkehrsklasse eine Hochlastphase. Die verschiedenen Hochlastphasen überlappen sich teilweise (Lastprofil siehe Bild 6-6). Die von der NRAC für den Prozessor prognostizierte Last ist im Bild links unten dargestellt. Die Abweichung zwischen berechnetem Sollwert  $\hat{\rho}_{Soll}$  und gemessener CPU-Auslastung  $\hat{\rho}_{Ist}$  ist relativ gering. Diese Abweichung hängt hauptsächlich von der Verkehrsdynamik ab. Für die Verwendung als Eingangsvariable einer Lastregelung ist der Sollwert somit brauchbar. Untersuchungen zu dem Annahmealgorithmus werden in Abschnitt 6.3 präsentiert.

### 6.2.4.2 Ausblick – Weitere Planungsparameter

Gemäß seiner Definition erlaubt der vorgestellte Planungsparameter der effektiven Prozessorauslastung zum einen nur eine Aussage über die Ressource *Prozessor*, und zum anderen ist der Planungszeitraum auf die Gegenwart begrenzt. Je weiter der Blick in die Zukunft geht, umso ungenauer wären Aussagen über die zu erwartende Belastung.

Eine Ursache sind die unbekanntes Charakteristiken zukünftiger Dienste. Eine Beschreibung ist deshalb gegenwärtig nur durch Hypothesen möglich. Je mehr Wissen über das zu erwartende mittlere zeitliche Verkehrsprofil eines Dienstes bekannt sein wird, umso genauer können ein zeitliches Belastungsprofil berechnet und Lastprognosen für die nähere Zukunft gemacht werden. Damit müsste eine Verbesserung des NRAC-Verfahrens möglich sein, was einer progressiven Lastregelung zu Gute kommen sollte.

Der zweite Aspekt bezieht sich auf die Berücksichtigung weiterer Betriebsmitteltypen. Denkbar wäre die Miteinbeziehung der Ressourcen Meldungspuffer, verfügbare Protokolltimer oder Speicher für Verbindungsdaten [471]. Zu erwartende Engpässe bei diesen speziellen Betriebsmitteln würden dann auch in die Entscheidung eingebunden werden, ob eine neue Anforderung angenommen werden kann oder nicht.



**Bild 6-6:** Lastprognose.

Links oben sind die gemessenen Ankunftsrate  $\phi_{in,RM,v}(t)$  bzw.  $\phi_{in,RM,ges}(t)$ . Oben rechts ist das resultierende Angebotsprofil  $A_v(t)$  bzw.  $A_{ges}(t)$  dargestellt und unten links der prognostizierte Lastwert (Sollgröße, Messgrößen siehe Bild 6-4).

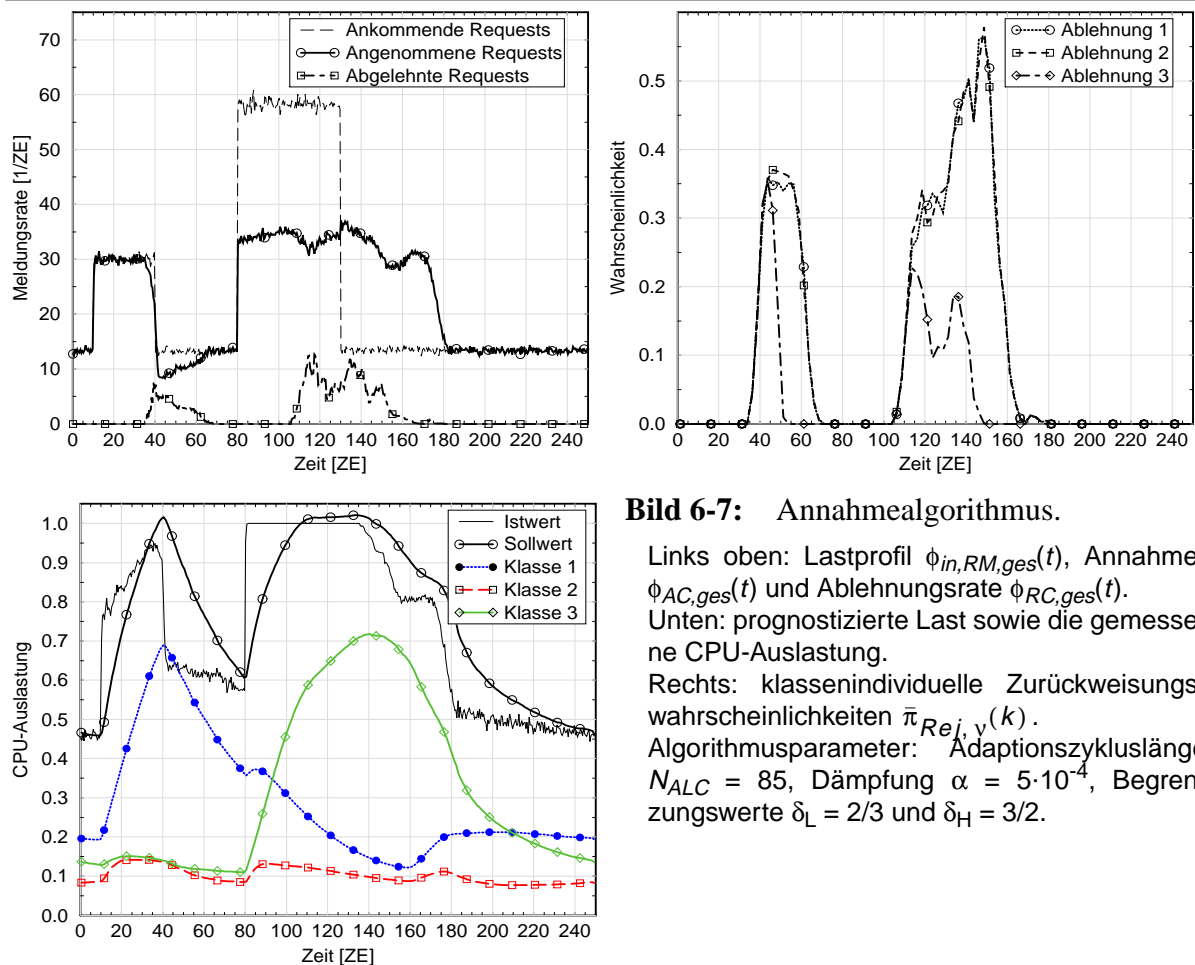
Algorithmusparameter: Adaptionzykluslänge  $N_{ALC} = 85$ , Dämpfung  $\alpha = 5 \cdot 10^{-4}$ , Begrenzungswerte  $\delta_L = 2/3$  und  $\delta_H = 3/2$ .

## 6.3 Annahmealgorithmus

Die Annahmeregulung für neue Dienstanforderungen basiert auf Gl. (4-8). Übersteigt der prognostizierte Ressourcenbedarf einen vorgegebenen Schwellwert, dann werden neue Anforderungen zurückgewiesen oder verworfen. Die Aktivierung dieses Regelkreises muss nicht exakt an den genauen Schwellwert gekoppelt sein, da eine gewisse Überschreitung dieser Grenze aufgrund des stochastischen Verkehrs sinnvoll ist. Dieser Tatsache wird durch Filterung, und damit einer Glättung der gemessenen Istgröße  $\hat{\rho}_{Ist}$ , begegnet. Weiterhin kann durch unterschiedliche Schwellwerte für verschiedene Verkehrsklassen eine Priorisierung vorgenommen werden.

### 6.3.1 Zurückweisung neuer Dienstanforderungen

Basierend auf der Lastprognose wird für jede neue Dienstanforderung entschieden, ob eine Reservierung des Betriebsmittels CPU vorgenommen wird, oder ob die Anforderung nicht angenommen werden kann. Die Stellgrößen sind die *Zurückweisungswahrscheinlichkeiten*  $\pi_{Rej,v}$  für die einzelnen Verkehrsklassen  $v$ . Für die Überprüfung dieses Verfahrens wurden verschie-



**Bild 6-7:** Annahmealgorithmus.

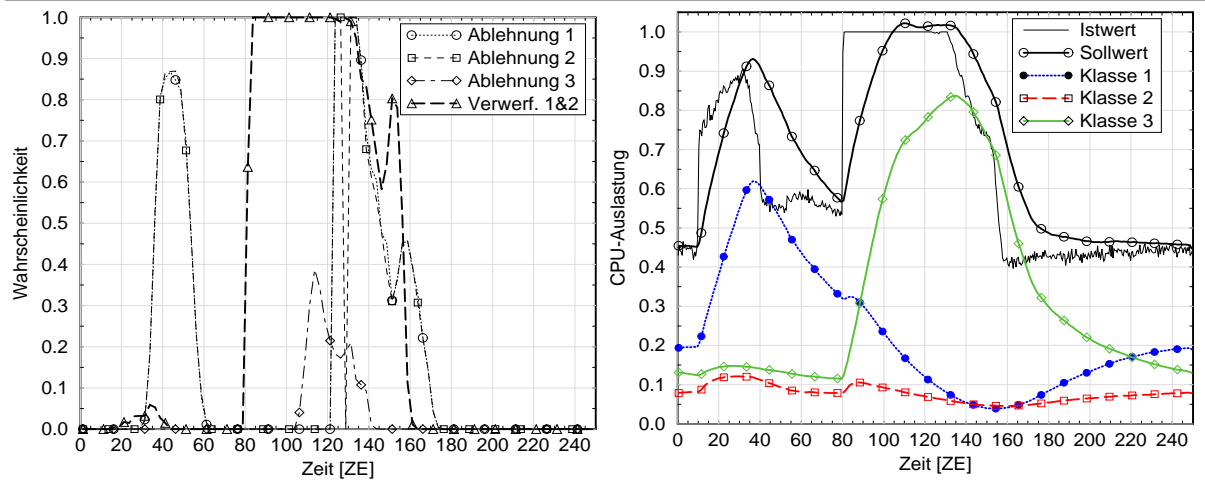
Links oben: Lastprofil  $\phi_{in, RM, ges}(t)$ , Annahmewahrscheinlichkeit  $\phi_{AC, ges}(t)$  und Ablehnungswahrscheinlichkeit  $\phi_{RC, ges}(t)$ .  
 Unten: prognostizierte Last sowie die gemessene CPU-Auslastung.  
 Rechts: klassenindividuelle Zurückweisungswahrscheinlichkeiten  $\bar{\pi}_{Rej, v}(k)$ .  
 Algorithmusparameter: Adaptionszykluslänge  $N_{ALC} = 85$ , Dämpfung  $\alpha = 5 \cdot 10^{-4}$ , Begrenzungswerte  $\delta_L = 2/3$  und  $\delta_H = 3/2$ .

dene Verkehrsmuster untersucht. Bild 6-7 zeigt repräsentative Messungen für ein Lastprofil bestehend aus zwei Hochlastphasen. Der erste Lastimpuls wird durch Anforderungen von Klasse eins, der zweite von Klasse-3-Nachrichten verursacht. Die resultierende prognostizierte Sollgröße weist während der Einschwingvorgänge einen konservativen Wert auf. Ab folgenden Schwellwerten  $\rho_{Soll, Th, v}$  werden neue Anforderungen zurückgewiesen: Für Klasse eins und zwei ab einem Wert von 70%, für Klasse 3 erst ab 90%. Damit wird auch eine weitere Priorisierung für diese Verkehrsklasse erreicht.<sup>9</sup> In Bild 6-7 sind die sich bei dieser Messung ergebenden mittleren Zurückweisungswahrscheinlichkeiten  $\bar{\pi}_{Rej, v}(k)$  dargestellt.

### 6.3.2 Zusätzliche Verwerfung

Bei großen Verkehrsmengen ist es sinnvoll, neben der Zurückweisung von Anforderungen auch die Verwerfung in Betracht zu ziehen. Damit werden weitere Ressourcen für den Nutzverkehr verfügbar und die Warteschlangenlängen bzw. Antwortzeiten werden nicht zu groß. Der im vorherigen Abschnitt verwendete Algorithmus soll deshalb noch um eine Verwerfungsstrategie ergänzt werden. Als weiterer Lastindikator wird hierzu der Pufferfüllgrad von Warteschlange zwei überwacht. Ab einem bestimmten Schwellwert  $\Omega_{Th, 2}$  werden alle neuen Anforderungen von Klasse eins und zwei verworfen. Klasse-3-Verkehr bleibt hiervon unberührt, wobei der Zurückweisungsalgorithmus nach wie vor aktiviert ist.

<sup>9</sup> Zur Erinnerung: Die Bearbeitung von Klasse-3-Nachrichten in bestimmten Bedienphasen besitzt eine höhere Priorität als die der anderen Klassen (siehe Abschnitt 5.4.4.2).

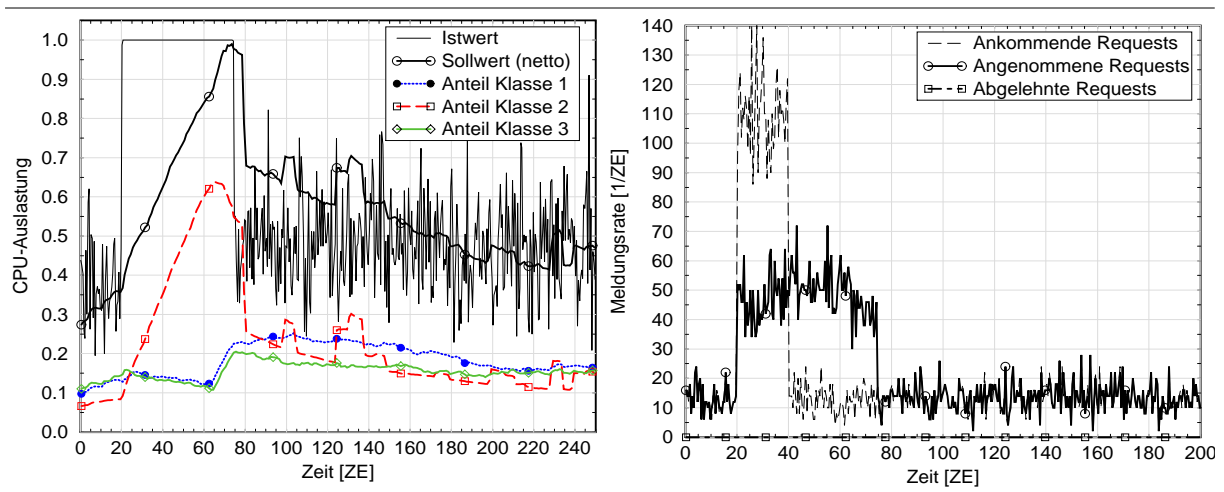


**Bild 6-8:** Annahmealgorithmus – Zusätzliche Verwerfung von Anforderungen als Ergänzung zur Zurückweisungsstrategie.

In Bild 6-8 sind die mittleren Zurückweisungs-  $\bar{\pi}_{Rej,v}(k)$  und Verwerfungswahrscheinlichkeiten  $\bar{\pi}_{Dis,v}(k)$  dargestellt. Während für den ersten, kleineren Lastimpuls die Zurückweisung dominiert, werden beim zweiten Impuls der Großteil aller neuen Anforderungen von Klasse eins und zwei verworfen. Damit wird das Antwortzeitverhalten deutlich besser. Auch die Abweichung des prognostizierten Lastwertes vom Istwert ist relativ gering.

### 6.3.3 Verhalten unter „extremer Überlast“

Die bisher verwendeten Verkehrsmuster sind bewusst idealisierte Lastprofile. Solche extremen Schwankungen sind in der Realität äußerst selten. Trotzdem muss ein Lastregelkreis robust gegenüber solchen Verkehrsverhältnissen sein. Aus diesem Grund wurden weitere Messungen mit sehr kurzen, aber starken Lastspitzen durchgeführt.<sup>10</sup>



**Bild 6-9:** Reaktion auf eine kurze, heftige Überlastspitze

<sup>10</sup> Überlastfaktor  $\zeta_{\dot{U}L}$  sehr groß, Dauer  $\Delta_{\dot{U}L}$  relativ klein (Definition siehe Abschnitt 4.3.2).

Aufgrund des im Verhältnis zur Überlastdauer relativ großen Mess- und Regelintervalles reagiert die Annahmekontrolle auf kurze Lastspitzen sehr träge. Wie Bild 6-9 entnommen werden kann, ist das Einschwingen des Sollwertes stark gedämpft. Dadurch spricht die Lastregelung in diesem speziellen Fall nicht an. In dieser Parameterstudie werden deshalb auch keine Anforderungen abgelehnt.

## 6.4 Bewertung und Ausblick

Die Annahmekontrolle für neue Dienstanforderungen basiert auf einer relativ einfachen Dienstcharakterisierung in Form einer spezifischen Ressourcenbedarfsabschätzung für verschiedene Verkehrsklassen. Anhand der Anzahl aktiver Kommunikationsbeziehungen lässt sich damit die Belastung eines Vermittlungsprozessors abschätzen. Dieser Sollwert bildet die Entscheidungsgrundlage über die Annahme oder Ablehnung einer neuen Anforderung. Die Funktionsweise der Annahmekontrolle wurde für verschiedene Verkehrsverhältnisse, Parametereinstellungen und Reglerkennlinien getestet. Weiterhin wurde die dynamische Parameterachführung überprüft. Dieser Lastregelkreis weist dabei ein robustes und befriedigendes Leistungsverhalten auf.

Das vorgestellte *NRAC*-Verfahren bildet die Grundlage einer *CAC* für steuerungstechnische Betriebsmittel in einem B-ISDN-Vermittlungsrechner. Durch die Abstraktion auf mehrere Verkehrsklassen kommt sie der Dienstproblematik entgegen. Werden weitere dienstcharakteristische Größen in die Annahmekontrolle aufgenommen, dann wird durch diese Verfeinerung eine Erhöhung der Vorhersagegenauigkeit erzielt. Wie bereits erwähnt wurde, könnte durch den Parameter mittlere Dienstdauer eine zeitliche Planung des Ressourcenbedarfs vorgenommen werden.

Prinzipiell ist die *NRAC* ein präventives Verfahren für den Niedriglastbereich und für relativ schwache Lastschwankungen konzipiert. In dieser Last- und Zeitebene sollte für die Softwarebearbeitung komplexerer Algorithmen auch noch ausreichend *CPU-Zeit* in einem realen System zur Verfügung stehen.



# Kapitel 7

## Untersuchung der Algorithmen für die lokale Leistungsregelung eines Dienstbearbeitungsprozessors

Für jede neue Dienstanforderung muss entschieden werden, ob diese akzeptiert werden kann oder nicht. Diese Entscheidung wird lokal von den entsprechenden Prozessorknoten getroffen. Während die im vorherigen Kapitel untersuchte NRCA mehr präventiv agiert, werden nun in diesem Kapitel Ergebnisse zu dem in Abschnitt 4.5 vorgestellten Verfahren diskutiert, das eine Maximierung des Leistungsmaßes *Power-Produkt* zum Ziel besitzt. Dieser neue Algorithmus ist deshalb gerade für Mehrklassensysteme prädestiniert.

### 7.1 Übersicht

Als Verkehrsmodell wird, wie in Kapitel 6, die Dualprozessorstruktur (siehe Anhang C.1.1) betrachtet. Der Schwerpunkt liegt dabei auf Knoten  $PN_B$  mit seiner lokalen Leistungsregelung, der *Power-Product Control (PPC)*. Zunächst werden Aspekte des Arbeitsbereiches und der funktionalen Verlaufs des *Power-Produkts* für das betrachtete Dreiklassensystem diskutiert. Die Leistungsoptimierung, oder vielmehr die Optimumsuche, ist für den vorliegenden Systemtyp mit seinen speziellen Randbedingungen durchaus nicht einfach. Diese Problematik wird in Abschnitt 7.3 näher beleuchtet. Anschließend werden die dynamischen Eigenschaften für relativ kurze Hochlastimpulse vorgestellt.

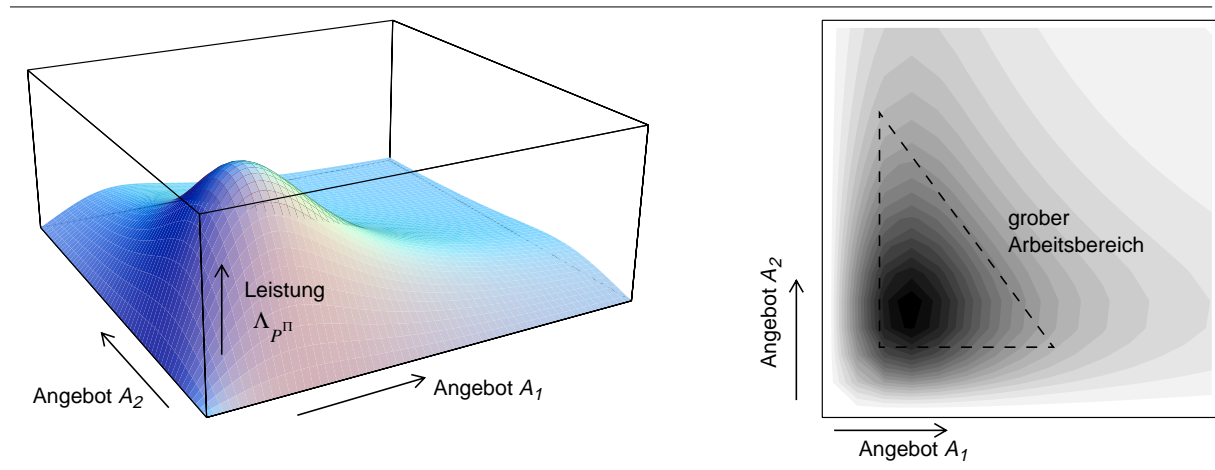
Die *PPC* lässt sich auch als *reine Überlastabwehr* verwenden. Für eine, nur für diesen Arbeitsbereich ausgelegten Leistungsregelung, lässt sich der *PPC*-Algorithmus entscheidend vereinfachen. Dieser Ansatz wird in Abschnitt 7.5 vorgestellt und diskutiert. Abschließend wird eine Einordnung und Bewertung der *PPC* vorgenommen.

### 7.2 Führungsgröße ‚*Power-Produkt*‘

#### 7.2.1 Arbeitsbereich

Aufgrund des Funktionsverlaufes von  $\Lambda_{P^1, RM}$  in Abhängigkeit der Last, liegt der optimale Arbeitsbereich dieser Leistungsregelung im Normallastbereich unterhalb der Nennkapazität. Die Problematik dieses Regelungszieles wird in Anhang B.4 für ein  $M_1M_2/M_1M_2/1/K$ -Modell, d.h. ein System mit zwei Verkehrsklassen, diskutiert. Es wird gezeigt, dass prinzipiell ein glo-

bales Maximum existiert und dieses bei entsprechender Systemdimensionierung im Normallastbereich liegt. Bild 7-1 zeigt den qualitativen Verlauf dieses Leistungsmaßes und skizziert den Arbeitsbereich dieses Regelkreises. Aufgrund der Sensitivität dieses Leistungsmaßes ist dieser Regelkreis mehr der „mittelfristigen“ Zeitebene zuzuordnen und weniger für eine kurzfristige Reaktion auf Laststöße geeignet.



**Bild 7-1:** Exemplarischer Verlauf des *Power-Produktes* bei zwei Verkehrsklassen  $\Lambda_{P^n} = f(A_1, A_2)$  (linkes Schaubild). Im Konturdiagramm ist der prinzipielle Arbeitsbereich des entsprechenden Leistungsregelkreises eingezeichnet (rechtes Schaubild).

Bild 7-1 zeigt die stationäre „Leistungsfläche“ eines Idealsystems. In der Realität sind die Anforderungen an diese Art von Leistungsregelung deutlich komplexer. Besonders folgende Randbedingungen kommen in einem B-ISDN-Vermittlungsrechner erschwerend zum tragen:

- *Mehrere Verkehrsklassen ( $K > 2$ )*  
Die Regelung bewegt sich nicht mehr innerhalb einer (begrenzten) Lastfläche, sondern in einem Lastraum (bzw. Hyperraum). Dadurch geht prinzipiell Anschaulichkeit verloren.
- *Dynamik*  
Alle Parameter sind zeitabhängig; auch der optimale Arbeitspunkt. Selbst bei stationärem Signalisierverkehr ist die Leistung aufgrund der Hintergrundlast zeitabhängig.<sup>1</sup>
- *Berücksichtigte Verkehrsparameter*  
Die Dienstmodellierung unterscheidet verschiedene Nachrichtentypen innerhalb eines Dienstszenarios (siehe Anhang C.3.3.1). In die *Power*-Maße gehen hier aber nur die REQUEST-Nachrichten, und beispielsweise nicht die Komplettierungsraten, ein.
- *Lage des Optimums*  
Die Lage des optimalen Arbeitspunktes ist normalerweise unbekannt. Eine adaptive Leistungsregelung kann nur versuchen, sich im Wirkbetrieb an diesen „heranzutasten“.<sup>2</sup>

Die Lage des Arbeitsbereiches bzw. Lastraumes für diesen Leistungsregelkreis soll im folgenden näher spezifiziert werden. Die Funktion  $\Lambda_{P^n, RM} = f(\vec{A}_{RM}) = f(A_{RM,1}, \dots, A_{RM,K})$  habe ein *globales Maximum*. Dieses liege bei  $\vec{A}_{RM, opt}$  – der entsprechende Lastwert beträgt

<sup>1</sup> Um dies an einem Beispiel zu veranschaulichen: Im Zweiklassensystem verursacht die Hintergrundlast zeitliche Niveauänderungen der Leistungsfläche.

<sup>2</sup> Alternative: Durch zahlreiche Lastmessungen an einem Laborsystem mit verschiedenen Verkehrsgemischen wird versucht, den optimalen Arbeitsbereich zu lokalisieren.

$$A_{RM, opt} = \sum_{v=1}^K A_{RM, v} \quad (7-1)$$

– und teile den Lastraum in die folgenden beiden Bereiche:

- Bereich I („links“ vom Maximum) für  $0 \leq A_{RM} < A_{RM, opt}$  und
- Bereich II („rechts“ vom Maximum) für  $A_{RM, opt} \leq A_{RM} < \infty$ .

Für ein Dreiklassensystem entspricht dem Optimum der dreidimensionale Vektor

$$\vec{A}_{RM, opt} = \begin{bmatrix} A_{RM, opt, 1} \\ A_{RM, opt, 2} \\ A_{RM, opt, 3} \end{bmatrix}, \quad (7-2)$$

wobei hier die Zeitabhängigkeit vernachlässigt sei. Es ist sinnvoll, die Leistungsregelung erst bei Überschreitung eines gewissen Verkehrsangebotes zu aktivieren. Dieser Vorgang wird üblicherweise durch einen bestimmten Lastindikator getriggert, der auf irgendeine Weise mit diesem Minimalangebot  $A_{RM, min}$  korreliert. Dabei sei  $A_{RM, min} < A_{RM, opt}$ <sup>3</sup>

## 7.2.2 Schnittbilder des Leistungsprofils und stationäre Werte

In Tabelle 7-1 sind die mittleren Gesamtbearbeitungszeiten  $\bar{T}_{H, v}$  und die minimalen mittleren Antwortzeiten  $\bar{\tau}_{RM, v}$  einer REQUEST-Nachricht im  $PN_B$  für Dienste einer bestimmten Verkehrsklasse angegeben.

Verkehrs- klasse	CPU- priorität	Mittlere Gesamtbear- beitungszeit [ZE]	Mittlerer Ankunftsab- stand (Normallast) [ZE]	Angebot [Erl]	Minimales RE- QUEST-Delay [ZE]
1	3	$\bar{T}_{H,1} = 0.050$	$\bar{T}_{A, RM, 1} = 0.333$	$A_1 = 0.15$	$\bar{\tau}_{RM, 1} = 0.013$
2	3	$\bar{T}_{H,2} = 0.020$	$\bar{T}_{A, RM, 2} = 0.133$	$A_2 = 0.15$	$\bar{\tau}_{RM, 2} = 0.013$
3	2	$\bar{T}_{H,3} = 0.025$	$\bar{T}_{A, RM, 3} = 0.166$	$A_3 = 0.15$	$\bar{\tau}_{RM, 3} = 0.013$
BL	1	$\bar{T}_{H, BL} = 0.00125$	$\bar{T}_{A, BL} = 0.025$	$A_{BL} = 0.05$	–

**Tabelle 7-1:** Last- und andere Parameterwerte für den gemeinsamen Koordinatenpunkt aller drei Schnittdiagramme

Für das vorliegende System soll nun das *globale Maximum* lokalisiert werden, d.h. gesucht ist der Lastvektor  $\vec{A}_{RM, opt}$  und das Extremum

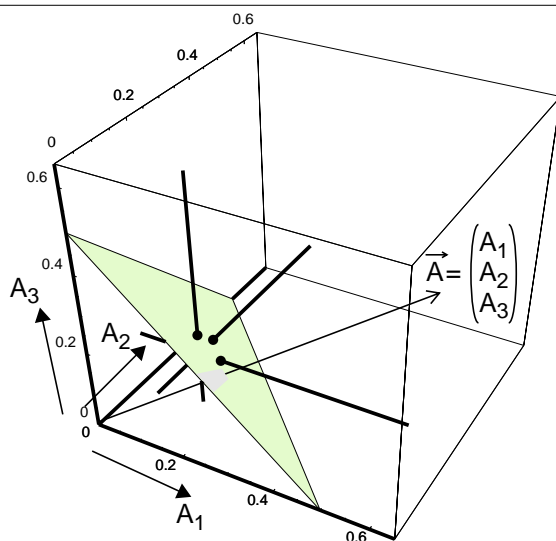
$$\Lambda_{P^n, RM, opt} = f(\vec{A}_{RM, opt}) = f(A_{RM, 1}, A_{RM, 2}, A_{RM, 3}). \quad (7-3)$$

Dies ist hier, aufgrund des aufwendigen Modells, nur durch Messungen möglich. Ordnet man im dreidimensionalen Raum (kartesische Koordinaten) den drei Achsen die Lastangebote der drei Verkehrsklassen zu (z.B.  $x$ -Achse  $\rightarrow A_1$ , usw.), dann entspricht der *Power*-Kenngröße ein Skalar „an der Spitze des Lastvektors“.<sup>4</sup> Gemessen wird unter stationären Bedingungen, d.h. die mittleren Verkehrsangebote sind jeweils konstant. Anschaulich bedeutet dies, dass eine

<sup>3</sup> Arbeitsbereich I entspricht somit beim Dreiklassensystem (siehe Bild 7-2) dem durch folgende fünf Ebenen begrenzten Raum:  $A_{ges} = A_{RM, min}$ ,  $A_{ges} = A_{RM, opt}$ ,  $A_1 = 0$ ,  $A_2 = 0$  und  $A_3 = 0$ .

<sup>4</sup> Der Wert der *Power*-Kenngröße ist hier die eigentlich vierte Dimension. Anschaulich könnte der *Power*-Wert für das Dreiklassensystem z.B. als Temperatur- oder Druckverteilung in einem Zimmer interpretiert werden.

Messung nur die Verhältnisse für exakt einen Punkt im Raum (Koordinate  $\{A_1 = \text{konst.}, A_2 = \text{konst.}, A_3 = \text{konst.}\}$ ) liefern kann. Da jede Simulation relativ viel Rechenzeit beansprucht (mehrere Stunden), kann nur ein kleiner Teil des Raumes vermessen werden. Aus diesem Grund wurden deshalb nur drei Schnittdiagramme ermittelt, was aber für eine grobe Lokalisierung völlig ausreichend ist. Wie bereits diskutiert wurde, wird das Maximum im Normallastbereich, und zwar in der Nähe von  $A_{ges} = 0.5$ , vermutet. Alle Schnittbilder sollen deshalb dieses „Raumgebiet“ des vermuteten Optimums durchschneiden (Bild 7-2). Aus diesem Grund wurde die Last wie folgt variiert: jeweils zwei Verkehrsklassen lieferten einen Anteil von 15% und das Angebot der dritten Klasse wurde im Bereich von 1% ... 60% verändert (Tabelle 7-1 gibt die Parameterwerte für den Schnittpunkt aller drei Ebenen). Durch den Hintergrundlastanteil von  $A_{BL} = 5\%$  ergab sich somit jeweils ein Gesamtangebot von  $A_{ges} \in [0.36, 0.95]$ .



**Bild 7-2:** Lastraum eines Dreiklassensystems.

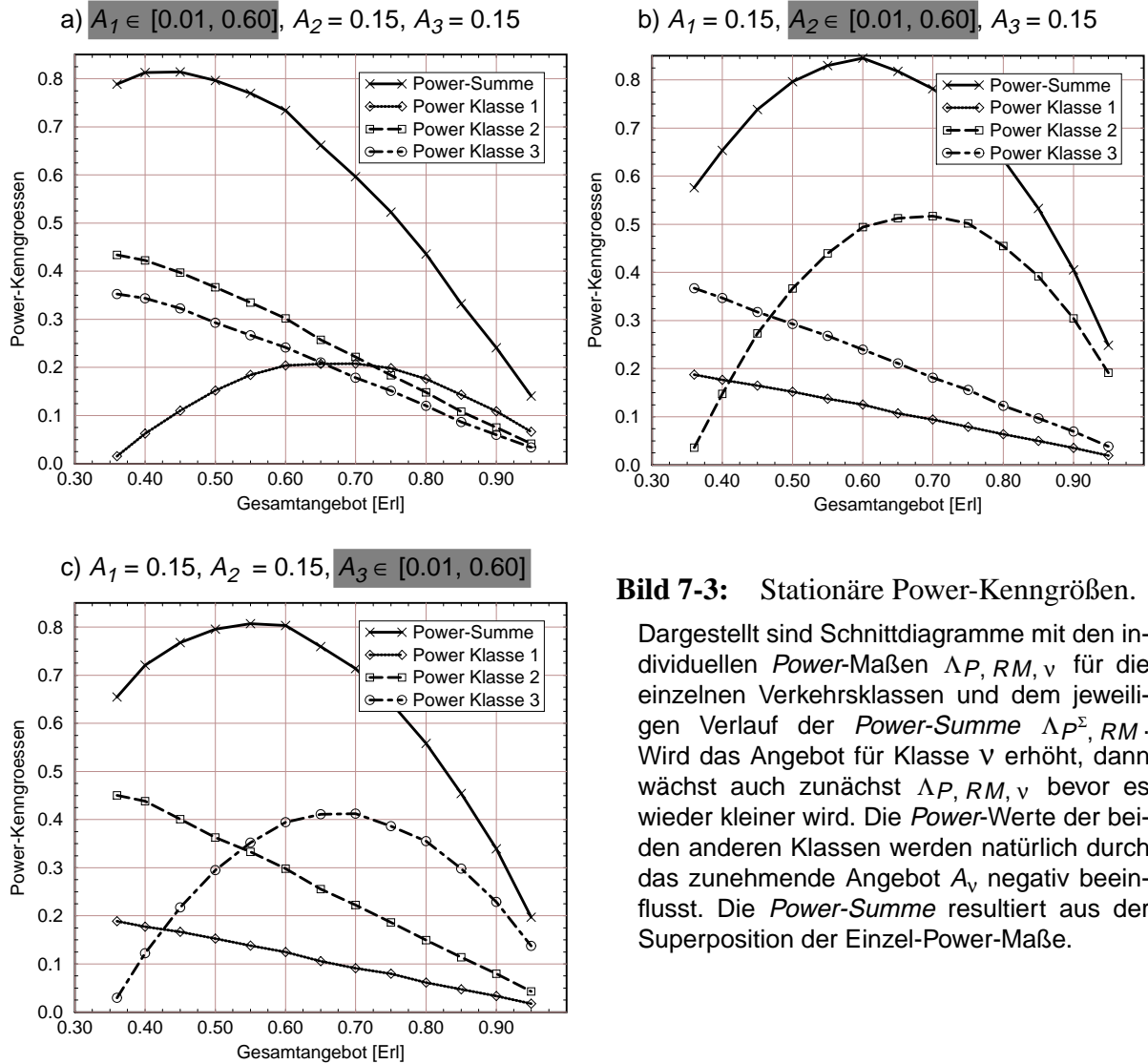
Eingezeichnet sind die drei „Lastachsen“ die den Schnittdiagrammen von Bild 7-3 zugrunde liegen.<sup>1</sup> Für eine Messreihe wird der Lastvektor entlang einer Achse bewegt. Das jeweilige Maximum des *Power-Produkts*  $\Lambda_{P^{\text{II}}, RM}$  (siehe Bild 7-4) liegt dabei ungefähr bei  $A_{ges} \approx 0.55$  (eingezeichnete Punkte).<sup>2</sup> Das globale Maximum ist in der Nähe dieser Punkte.

- 1 Die Hintergrundlast von  $A_{BL} = 0.05$  wurde hier nicht berücksichtigt.
- 2 Das eingezeichnete Polygon entspricht der Lastebene  $A_{ges} = A_1 + A_2 + A_3 - A_{BL} = 0.5$ . Die Schnittpunkte mit den Achsen sind durch Punkte markiert.

Weiterhin wurde die Annahmeregulierung der *PPC* deaktiviert, d.h. es wurden keine Dienstanforderungen zurückgewiesen. Bild 7-3 zeigt zunächst die drei Schnittdiagramme<sup>5</sup> mit den individuellen *Power-Maßen*  $\Lambda_{P, RM, v}$  für die einzelnen Verkehrsklassen und der *Power-Summe*  $\Lambda_{P^{\Sigma}, RM}$ .<sup>6</sup>

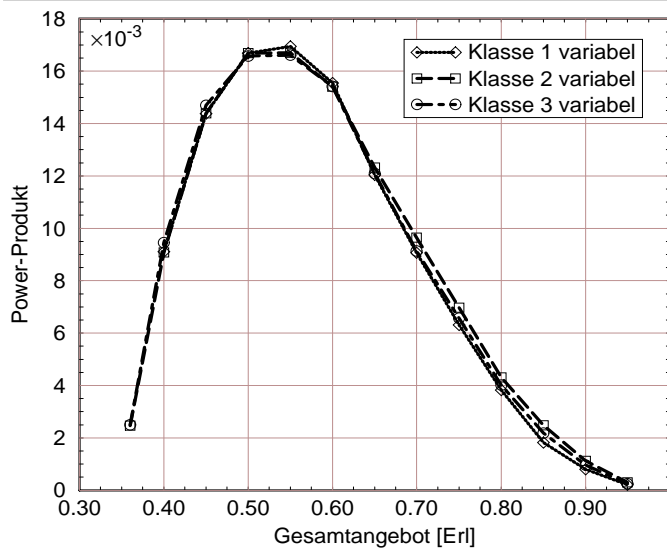
Bild 7-4 zeigt die drei *gemessenen* Verläufe der Kenngröße *Power-Produkt*  $\Lambda_{P^{\text{II}}, RM}$ . Diese entsprechen den Produkten der klassenindividuellen *Power-Maße* aus den drei Schnittdiagrammen (Bild 7-3). Überraschend ist der nahezu identische Verlauf in allen drei Ebenen. Das Maximum liegt bei ungefähr  $A_{ges} \approx 0.55$  und beträgt ungefähr  $\Lambda_{P^{\text{II}}, RM} = 16.8 \cdot 10^{-3}$ . Ob diese Extrema auch mit dem globalen Maximum identisch sind ist hiermit aber noch nicht bewiesen.<sup>7</sup> Dies müsste durch weitere Messungen im Gebiet um diesen Punkt überprüft werden (siehe auch Bild 7-2).

- 5 Alle *Power-Kenngrößen* beziehen sich auf die REQUEST-Nachrichten. In den Diagrammen werden diese aber jeweils über dem Gesamtangebot (d.h. alle Signalisier- und Steuernachrichten) aufgetragen und nicht über dem Teilangebot aus den REQUEST-Nachrichten resultierend. Dies ist aber hier ohne Bedeutung (entspricht nur einer linearen Verzerrung der Kurven).
- 6 Die *Power-Summe* ist hier nur der Vollständigkeit halber aufgetragen. Für die Leistungsregelung hat sie in einem Mehrklassensystem keine praktische Bedeutung, da ihre Maxima (und damit auch das globale Maximum) allesamt auf dem Rand des „Lastwürfels“ liegen (siehe auch Bild B-18, S. 251). In diesen Gebieten sind aber immer die Angebote zweier Klassen gleich Null. Der Regelkreis würde somit versuchen, den Verkehr zweier Klassen „abzuwürgen“ und man würde letztlich ein Einklassensystem betrachten. Dies widerspricht aber der hier formulierten Zielsetzung für einen B-ISDN-Vermittlungsrechner.
- 7 Veranschaulichung am Beispiel eines *Zweiklassensystems*: In der Umgebung des globalen Maximums kann hier der Verlauf des *Power-Produkts* durch einen *Paraboloiden* approximiert werden. Hier sind ebenfalls eine Vielzahl von Schnittebenen denkbar, die zwar denselben Funktionsverlauf besitzen, aber nicht durch das Extremum gehen.



**Bild 7-3:** Stationäre Power-Kenngrößen.

Dargestellt sind Schnittdiagramme mit den individuellen  $Power$ -Maßen  $\Delta P, RM, v$  für die einzelnen Verkehrsklassen und dem jeweiligen Verlauf der  $Power$ -Summe  $\Delta P^{\Sigma}, RM$ . Wird das Angebot für Klasse  $v$  erhöht, dann wächst auch zunächst  $\Delta P, RM, v$  bevor es wieder kleiner wird. Die  $Power$ -Werte der beiden anderen Klassen werden natürlich durch das zunehmende Angebot  $A_v$  negativ beeinflusst. Die  $Power$ -Summe resultiert aus der Superposition der Einzel- $Power$ -Maße.



**Bild 7-4:**  $Power$ -Produkt  $\Delta P^{\Pi}, RM$  für drei Verkehrsklassen.

Bei den einzelnen Leistungskurven sind jeweils die Angebote zweier Klassen fix und das dritte Angebot wird variiert. Jeder der drei eingezeichneten Funktionsverläufe lässt sich durch Multiplikation der Einzel- $Power$ -Maße  $\Delta P, RM, v$  aus den Teildigrammen von Bild 7-3 berechnen.

Da die Leistungsregelung nur versuchen kann, das System in die Nähe des globalen Maximums zu „dirigieren“, ist die *exakte Lage* hier nicht von entscheidender Bedeutung. Das Vermittlungssystem kann hier nur auf die externen Bedingungen, d.h. die momentan ankommende Last, *reagieren*. Nur wenn auch direkt die Verkehrsquellen gesteuert werden könnten, wäre man in der Lage, das System in den optimalen Arbeitspunkt zu bringen. Dies macht hier aber natürlich nur bedingt Sinn, da „einem Teilnehmer keine ‚Lastproduktion‘ vom Netz vorge-schrieben werden kann“.<sup>8</sup> Festzuhalten ist, dass die durchgeführten Messungen die Existenz eines derartigen Extremums klar andeuten und dieses im besagten Arbeitsbereich liegt.

## 7.3 Optimumsuche

Ideal wäre es, wenn die Leistungsregelung es schaffen würde, den Arbeitspunkt in das Optimum zu bringen oder zumindest in der Nähe des Optimums halten könnte. Ein Maß für die Güte des Regelkreises wäre dann beispielsweise wie schnell und wie gut der optimale Arbeitspunkt angesteuert wird. Diese wünschenswerte Vorgabe kann hier allerdings aus vielerlei Gründen nicht vollständig erfüllt werden:

- Eine Optimumsuche bedeutet eine permanente Veränderung der Lastkoordinaten, d.h. je nach aktuellem Arbeitspunkt müssen Teillasten erhöht oder verkleinert werden, gegebenenfalls nach Betrag und Richtung unterschiedlich für die einzelnen Verkehrsklassen. Es versteht sich, dass aufgrund der Charakteristik der hier vorliegenden Lastquellen (Teilnehmer, Endgeräte, Anwendungen) diese Anforderung nicht erfüllt werden kann.
- Die stochastischen Eigenschaften des Verkehrs erschweren eine schnelle Regelung, da Rauschfilterung und Mittelwertbildungen notwendig sind.
- Eingriffe in die Lastwerte sind hier nur durch Zurückweisung und Verwerfung neuer Dienstanforderungen möglich. Diese Maßnahmen haben eine andere Rückwirkung auf das Leistungsverhalten, wie wenn direkt die Verkehrsparameter der Quellen gesteuert werden würden.
- Da die Leistungsregelung hier eine integrierte Funktion im Vermittlungsprozessor ist, wurde der vorgestellte Algorithmus bewusst einfach gehalten. Jeder weitere Programmbefehl kostet zusätzlich Ressource *CPU*, die zu Lasten der verfügbaren Vermittlungsleistung geht.

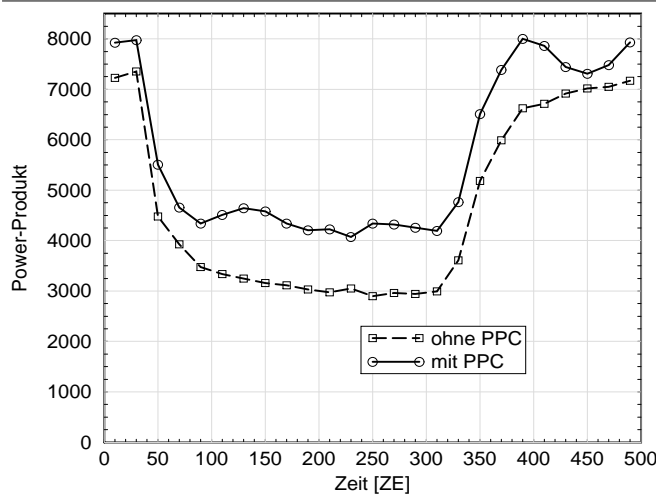
Trotz dieser Einschränkungen wird durch das in Abschnitt 4.5.2.2.4 vorgestellte *Tripel-Verfahren* eine Optimumsuche unterstützt. Nachweisen lässt sich dies durch einen Simulationstrace, wobei der zeitliche Verlauf der klassenindividuellen Zurückweisungswahrscheinlichkeiten und des *Power-Produkts* als Reaktion auf einen Lastwechsel (Anstieg oder Abfall) gemessen werden. Im Lastraum bedeutet dies den Wechsel von einem stationären Arbeitspunkt  $\vec{A}_{RM,I}$ , mit den stationären Zurückweisungswahrscheinlichkeiten  $\vec{\pi}_{Rej,I}$ , zu einem anderen stationären Arbeitspunkt  $\vec{A}_{RM,II}$ . Da die mittleren Ankunfts-raten von REQUEST-Nachrichten von außen aufgeprägt sind und hier nicht vom Vermittlungsrechner geändert werden können, lassen sich nur die Nutz-zu-Blindlast-Anteile der einzelnen Verkehrsklassen variieren. Dies geschieht durch die Ablehnung neuer Anforderungen und lässt sich am Übergangsverhalten der Zurück-

---

<sup>8</sup> Diese Aussage gilt natürlich nicht unbedingt für ein entsprechendes Regelverfahren auf einer anderen *Raumbene*. Z.B. bei einer *Verkehrslenkung* und damit einer *Lastverteilung im Netz* könnte die Leistung einzelner Knoten hinsichtlich des Power-Produkt-Kriteriums optimiert werden, da fast immer ausreichend Verkehr zur Verfügung steht.

weisungswahrscheinlichkeiten, die nun auf einen neuen stationären Punkt  $\vec{\pi}_{Rej, II}$  einschwingen, darstellen.

Damit wird die Gesamtleistung des Vermittlungsrechners für die neuen Lastverhältnisse angepasst. Da in das vorgestellte *Power-Produkt*  $\Lambda_{P^II, RM}$  aber nur die REQUEST-Nachrichten eingehen, lässt sich mit diesem speziellen Leistungskennwert die Optimumsuche nicht nachweisen. Verdeutlicht werden soll dies an einer Messung mit aktiviertem und deaktiviertem Zurückweisungsmechanismus (Bild 7-5).



**Bild 7-5:** Übergangsverhalten des *Power-Produkts*  $\Lambda_{P^II, RM}(t)$  für einen leichten Lastwechsel.

Gemessen wurde dieser Leistungskennwert bei aktiviertem und deaktiviertem Zurückweisungsmechanismus.

Eingestelltes Verkehrsprofil: Normallast gleich 70%, Hochlast von 80% im Zeitraum  $30 < t < 320$ , d.h. der Arbeitspunkt wird während der Hochlastphase vom Optimum weg bewegt („die Leistung bricht deshalb kurzzeitig ein“).

Mit aktiviertem Zurückweisungsverfahren ist das Leistungsverhalten deutlich besser. Durch die Optimumsuche ist dem *Power-Produkt*, begründet auch durch die hier relativ kurzen Mess- und Regelintervallauern, eine leichte Schwingung überlagert.

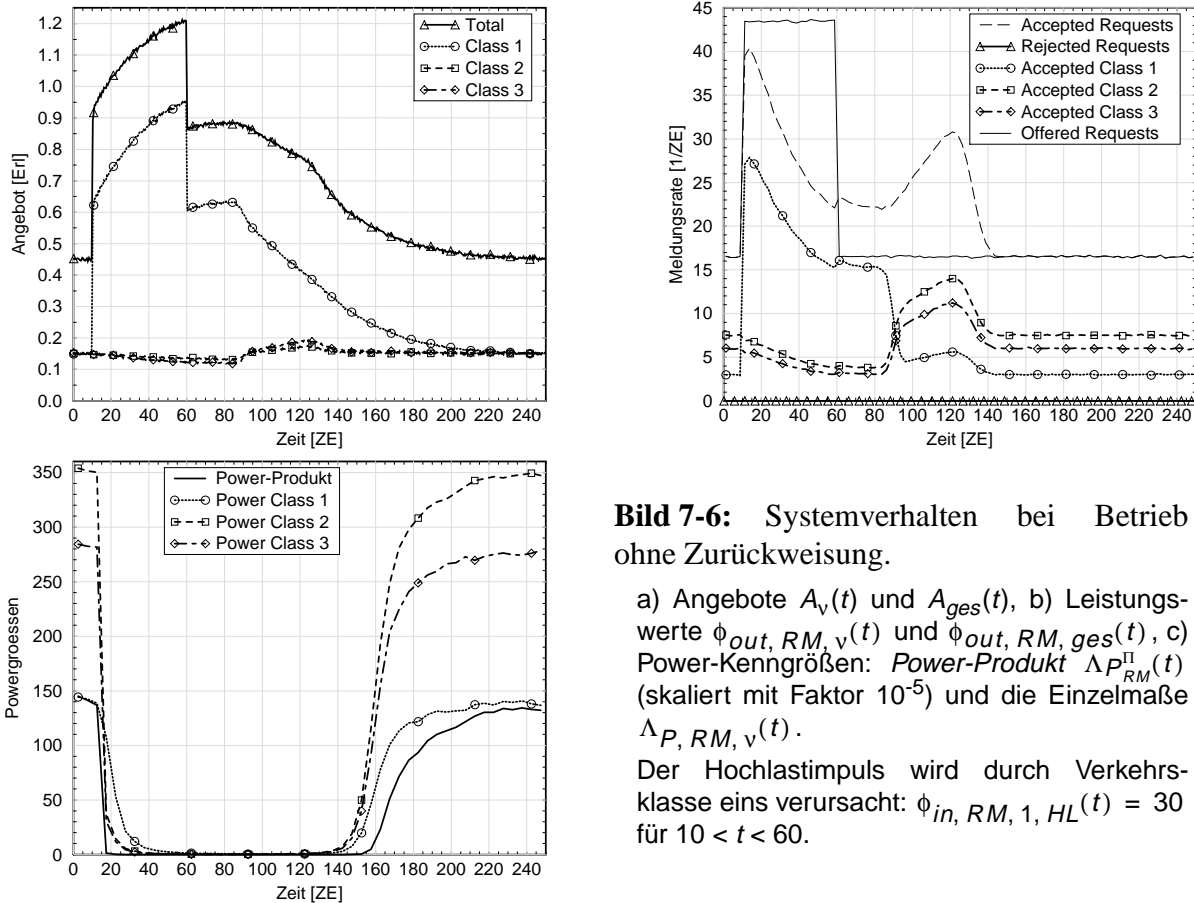
## 7.4 Kurze Hochlastimpulse

Der *PPC*-Algorithmus soll nun für stärkere Lastschwankungen untersucht werden. Im Vordergrund steht dabei weniger eine Leistungsoptimierung, vielmehr wird für diese Verkehrsprofile ein robustes Verhalten erwartet.

### 7.4.1 Annahme aller Anforderungen

Um die *PPC* bewerten zu können, wurden auch Messungen mit deaktivierter Aktorik durchgeführt, d.h. alle Anforderungen werden von der Vermittlungssteuerung akzeptiert. In Bild 7-6 sind Last- und Leistungswerte für einen kurzen Lastimpuls durch Klasse eins dargestellt.

Im Grundlastbereich ist die CPU annähernd mit 50% ausgelastet. Nach dem Lastsprung wird das Angebot  $A_{ges}(t)$  größer eins und das System wird für die kurze Hochlastphase instabil. Dadurch wachsen die Warteschlangen in den Puffern an und die Antwortzeiten werden unzulässig groß. Wie Bild 7-6 entnommen werden kann, geht die Leistung in diesem Bereich stark zurück, da die *Power*-Werte fast auf null absinken.



**Bild 7-6:** Systemverhalten bei Betrieb ohne Zurückweisung.

a) Angebote  $A_v(t)$  und  $A_{ges}(t)$ , b) Leistungswerte  $\phi_{out, RM, v}(t)$  und  $\phi_{out, RM, ges}(t)$ , c) Power-Kenngrößen: Power-Produkt  $\Delta P_{RM}^{\Pi}(t)$  (skaliert mit Faktor  $10^{-5}$ ) und die Einzelmaße  $\Delta P_{P, RM, v}(t)$ .

Der Hochlastimpuls wird durch Verkehrsklasse eins verursacht:  $\phi_{in, RM, 1, HL}(t) = 30$  für  $10 < t < 60$ .

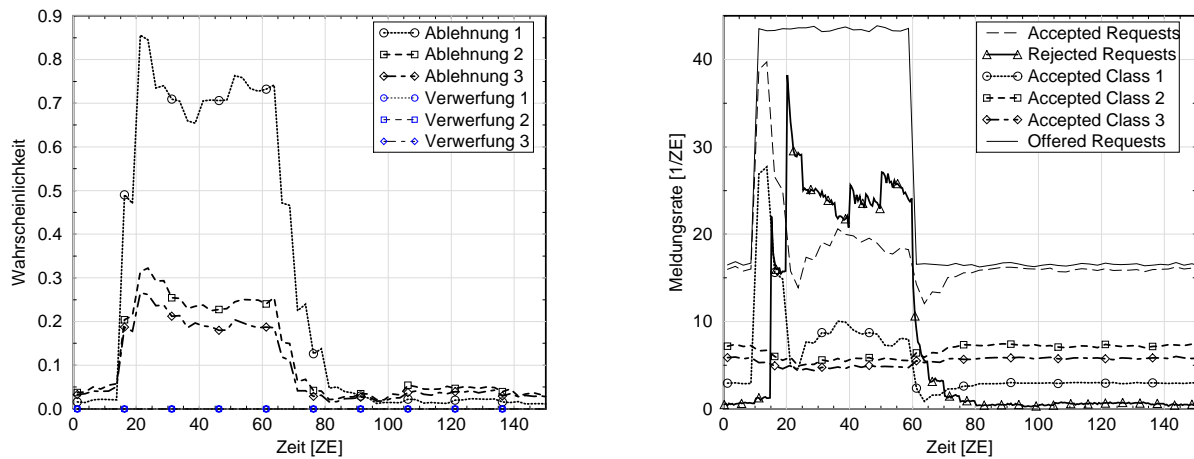
## 7.4.2 Zurückweisung neuer Anforderungen

Wie den Ergebnissen des vorherigen Abschnittes entnommen werden kann, ist für eine Leistungssteigerung eine Begrenzung der Last in Hochlastphasen unerlässlich.

Wiederum wird ein Lastimpuls durch Klasse-1-Anforderungen erzeugt.<sup>9</sup> Die gemessene Prozessorauslastung  $\hat{\rho}(k)$  stellt die zusätzliche Lastkenngröße dar. Die Trennung in Normal- und Hochlastbereich erfolgt mit dem Schwellwert  $\rho_{Th}$  von 0.6. In Bild 7-7 sind die mittleren Zurückweisungswahrscheinlichkeiten  $\pi_{Rej, v}(k)$  und die mittleren Meldungsraten dargestellt. Während der Hochlastphase werden ungefähr 70% aller Klasse-1-Anforderungen zurückgewiesen. Verworfen werden hier keine Nachrichten.

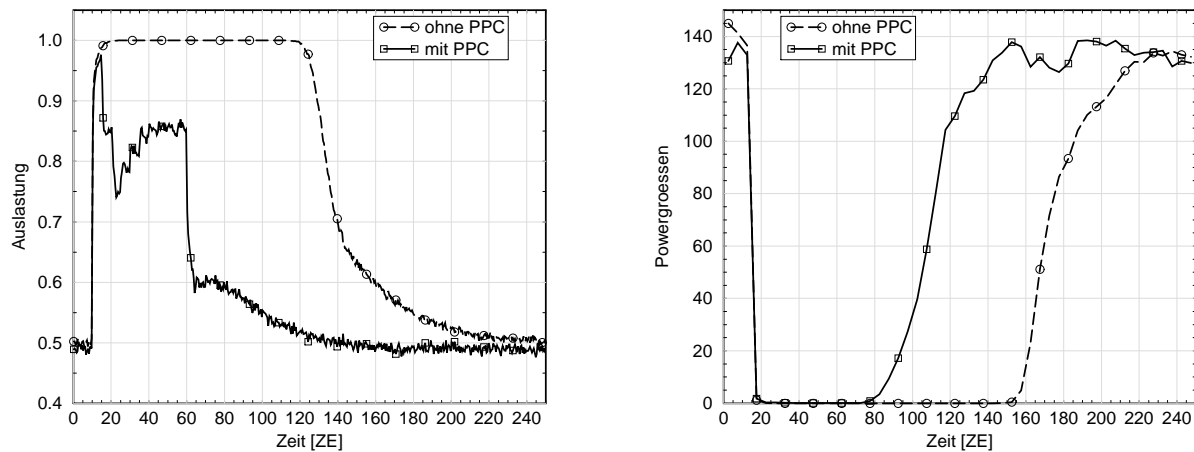
<sup>9</sup> Verkehrsparameter des Lastimpulses:  $\phi_{in, RM, 1, HL}(t) = 30$  für  $t \in [10, 60]$ .





**Bild 7-7:** Systemverhalten bei Betrieb mit *PPC*. a) Zurückweisungswahrscheinlichkeiten  $\pi_{Rej, v}(k)$  und b) Leistungswerte  $\phi_{out, RM, v}(t)$  und  $\phi_{out, RM, ges}(t)$ .

Die Zurückweisung unter Hochlast wirkt sich positiv auf die Prozessorbelastung und die erzielte Leistung aus. In Bild 7-8 sind diese Kenngrößen für den  $PN_B$  mit *PPC* und ohne Leistungsregelung gegenübergestellt. Mit aktivierter Regelung bleibt das System auch unter Hochlast stabil. Die CPU besitzt noch ausreichend Reserven. Betrachtet man die Leistung  $\Lambda_{P_{RM}}^{\Pi}(t)$ , dann bricht diese aufgrund der großen Verkehrsmenge durch den Lastimpuls ebenfalls sehr stark ein. Nach der Hochlastphase steigt die Leistung wesentlich früher an als ohne Zurückweisungsmechanismus.



**Bild 7-8:** Ergebnisse mit bzw. ohne *PPC*. a) CPU-Gesamtauslastung  $\rho_{ges}(t)$  und b) *Power-Produkt*  $\Lambda_{P_{RM}}^{\Pi}(t)$ .

Die hier betrachteten Lastmuster repräsentieren sogenannte „*Worst-Case*“-Szenarien, da diese Form von Lastsprüngen in der Realität eher die Ausnahme ist.<sup>10</sup> Bei diesem Lastmuster steht deshalb weniger die Leistungsoptimierung im Vordergrund. Vielmehr sollen Stabilitätskriteri-

<sup>10</sup> Solche Lastanstiege treten beispielsweise bei Mitkopplungseffekten zwischen Teilnehmer und Netz auf („lawinarartiges“ Anwachsen der Last durch Rufwiederholungen).

en und Dienstgüteparameter eingehalten werden. Eine zielgerichtete Optimumsuche ist auch aufgrund der kurzen Hochlastphase nicht möglich.

## 7.5 PPC als reine Überlastabwehr

Für den dedizierten Einsatz der *PPC* im Überlastbereich, d.h. der Arbeitsbereich des Regelkreises wird hier nun weiter beschränkt, lässt sich das Verfahren vereinfachen.

### 7.5.1 Vereinfachte Berechnung der Zurückweisungswahrscheinlichkeiten

Soll die *PPC* in einem Regelkreis eingesetzt werden, der hauptsächlich auf Überlasten reagieren soll, dann lässt sich der vorgestellte Algorithmus entscheidend vereinfachen. Unter Überlast versteht man Angebote, die die Nennkapazität eines Vermittlungsrechners übersteigen. Wie den Leistungsmessungen des vorliegenden Steuerungssystems entnommen werden kann (siehe Abschnitt 7.2.2), liegt das Optimum für das *Power-Produkt*  $\Lambda_{P, RM}$  mit großer Wahrscheinlichkeit in einem Lastbereich mit  $A_{ges} < 60\%$ . Nennkapazitäten liegen typischerweise oberhalb dieses Wertes. D.h. unter Überlast befindet sich das System in einem Lastbereich, in dem das *Power-Produkt* eine monoton fallende Funktion ist (vgl. auch Bild 7-4).

Diese Eigenschaft lässt den Algorithmus entscheidend vereinfachen, da nun die Richtung in welcher das Maximum liegt, bekannt ist. Es ist daher naheliegend, die Zurückweisungswahrscheinlichkeiten  $\pi_{Rej, v}$  einer Verkehrsklasse zu vergrößern, wenn ihr *Power-Maß*  $\Lambda_{P, RM, v}$  abfällt. Folgender einfache Algorithmus soll deshalb für diesen Einsatzzweck angewendet werden:

$$\pi_{Rej, v}(k+1) = f(\Lambda_{P, RM, v}(k)) = 1 - \frac{\Lambda_{P, RM, v}(k)}{\hat{\Lambda}_{P, RM, max, v}}. \quad (7-4)$$

Diese Geradengleichung erfüllt obige Forderung. Mit  $\hat{\Lambda}_{P, RM, max, v}$  ist der geschätzte maximale *Power-Wert* einer Verkehrsklasse gemeint. Diese Größe ist normalerweise unbekannt und muss deshalb approximiert werden. Messungen entsprechend Bild 7-3 können hier erste Anhaltspunkte für Maximumwerte liefern. Wenn der ungefähre Ressourcenbedarf an CPU-Zeit und die ungefähre minimale Durchlaufzeit bekannt sind, lässt sich folgende Abschätzung vornehmen:

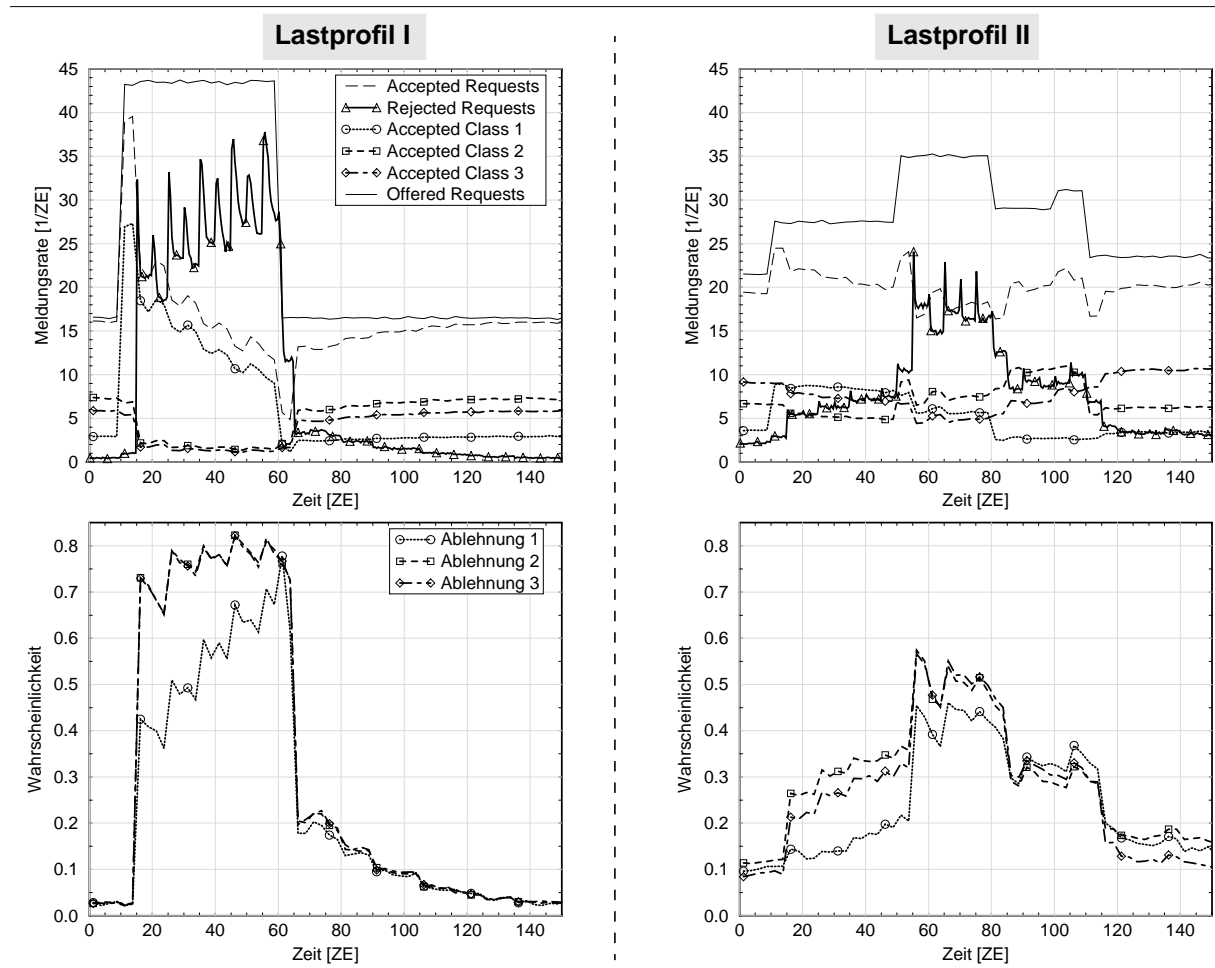
$$\hat{\Lambda}_{P, RM, max, v} \approx \frac{1}{4} \cdot \frac{\hat{\Phi}_{out, RM, max, v}}{\tilde{\tau}_{RM, min, v}} \quad \text{mit} \quad \hat{\Phi}_{out, RM, max, v} = \frac{1}{\hat{T}_{A, RM, v}}. \quad (7-5)$$

Diese Schätzung beruht auf der Tatsache, dass bei Einklassensystemen mit großem Warteraum die optimale Leistung bei ungefähr 50%iger Auslastung erzielt wird (vgl. Abschnitte B.2.1 und B.3.1). In diesem Arbeitsbereich entspricht der Durchsatz der halben maximalen Bedienrate und die Antwortzeit dem doppelten Minimalwert. Dieser Ansatz wurde hier verwendet.

Sollte sich erweisen, dass ein Schätzwert  $\hat{\Lambda}_{P, RM, max, v}$  zu klein gewählt wurde, dann könnte dieser im Wirkbetrieb, durch Vergleich mit den Messwerten  $\Lambda_{P, RM, v}(k)$ , korrigiert und damit adaptiv nachgeführt werden.

## 7.5.2 Messungen und Bewertung

Die PPC soll durch die zusätzliche Lastkenngröße  $\hat{\rho}(k)$  mit Schwellwert  $\rho_{Th} = 0.8$  aktiviert werden. In Bild 7-9 ist die Überlastreaktion auf zwei verschiedene Verkehrsprofile<sup>11</sup> dargestellt.



**Bild 7-9:** Überlastabwehrverhalten für zwei verschiedene Verkehrsprofile. Abgebildet sind Meldungsraten  $\phi_{out, RM, v}(t)$ ,  $\phi_{out, RM, ges}(t)$  und Zurückweisungswahrscheinlichkeiten  $\pi_{Rej, v}(k)$ .

Die Hochlastphasen besitzen eine unterschiedliche Dauer sowie verschiedene Lastintensitäten. Die Mess- und Regelintervalldauer betragen hier fünf Zeiteinheiten. Dadurch wird unmittelbar auf Lastsprünge reagiert, andererseits weist die Ablehnungsrate  $\phi_{Rej, ges}(t)$  dadurch leichte Schwingungen auf. Die Prozessorauslastung wird für beide Lastprofile in einen sicheren Arbeitsbereich unter 90% geregelt (Anmerkung: Ergebnisse für  $\rho(t)$  hier nicht dargestellt). Die PPC mit dem vereinfachten Algorithmus ist somit auch für die Abwehr von Überlasten in einem Mehrklassensystem geeignet.

<sup>11</sup> Lastprofil I: Hochlastphase durch Klasse 1. Lastprofil II: Hochlastimpulse durch alle drei Klassen (zeitlich nacheinander sowie teilweise überlappend).

## 7.6 Bewertung und Ausblick

Nahezu alle Verfahren der Überlastabwehr und Leistungsregelung für Vermittlungsrechner haben die vorrangige Zielsetzung, *systemspezifische* Leistungskenngrößen zu verbessern. Die Akzeptanz und der Erfolg einer Netztechnologie ist aber wesentlich davon abhängig, wie sich das Netz aus Teilnehmersicht in seinem Leistungsverhalten und Dienstgüteprofil darstellt. Gerade deshalb sollte im B-ISDN mit seinem Dienstklassenkonzept dieser Aspekt von einer Last- und Leistungsregelung nicht unberücksichtigt bleiben.

Durch die Definition des *Power-Produkts* und seiner Verwendung als zu optimierende Leistungskenngröße, wird diese Forderung bezüglich des Dienstaspekts im vorgestellten *PPC*-Algorithmus berücksichtigt. Dieses Verfahren wurde simulativ in zahlreichen Parameterstudien überprüft und seine Funktion nachgewiesen. Auch für große Lastsprünge weist dieser Mechanismus ein robustes Verhalten auf und wird nicht instabil.

Wie bereits in Kapitel 4 motiviert wurde, hat die *PPC* zum Ziel, einen Kompromiss zwischen den Dienstgüteanforderungen der Teilnehmer und der Gewinnmaximierung (hoher Durchsatz) der Vermittlungsstelle (bzw. des Netzes) zu finden. Vergleicht man die *PPC* deshalb mit Verfahren, die eine Maximierung des Durchsatzes oder eine Minimierung der Antwortzeiten zum Ziel haben, dann ist die *PPC* bei Berücksichtigung des jeweiligen (d.h. einzigen) Leistungskennwertes schlechter. Werden jedoch Nutzdurchsatz und Antwortzeitverhalten als kombiniertes Leistungsmaß verwendet, dann ist die *PPC* den einseitig optimierenden Algorithmen überlegen.

Aufgrund der speziellen Eigenschaften der *Power-Produkt-Funktion* ist die *PPC* mehr für den Lastbereich in der Gegend der Nennkapazität und einen größeren Zeitbereich prädestiniert.

# Kapitel 8

## Untersuchung der Verfahren zur globalen Lastverteilung

Der Vermittlungsrechner für die Leistungsuntersuchung der *DLB*-Algorithmen entspricht einer hierarchischen Steuerstruktur mit kombinierter Last- und Funktionsteilung. In diesem Kapitel sollen Ergebnisse der in Abschnitt 4.6 diskutierten Lastzuteilungsalgorithmen vorgestellt, die einzelnen Verfahren gegebenenfalls verfeinert und abschließend bewertet werden.

### 8.1 Übersicht

Das den Leistungsuntersuchungen dieses Kapitels zugrundeliegende Simulationsmodell wird in Anhang C.1.2 näher erläutert. Konkret wird ein Systemmodell bestehend aus zwei  $PN_{A,i}$ , drei  $PN_{B,j}$  und einem  $PN_C$  betrachtet. Gemessen wird das dynamische Verhalten auf sich zeitlich ändernde Lastprofile. Als Ursache für die temporär variierende Lastzusammensetzung wird entweder ein inhomogener interner „Störverkehr“, d.h. die mittlere Last durch nichtvermittlungstechnische Prozesse ist für die einzelnen  $PN_{B,j}$  verschieden, oder ein inhomogener externer Verkehrsmix eingepreßt. Zusätzlich wirkt die lokale Überlastabwehr in der *B*-Ebene als Störgröße für die globale Leistungsregelung.

#### 8.1.1 Feste Parametereinstellungen

Die mittleren Bearbeitungszeiten eines Dienstes einer bestimmten Verkehrsklasse für einen  $PN_B$  und auch die Prozessorprioritäten sind in Tabelle 7-1 (siehe S. 155) aufgeführt.

		Mittlere Gesamtbearbeitungszeit in den einzelnen PN [ZE]			Mittlere Ankunftsrate [ZE <sup>-1</sup> ]	Mittleres Gesamtangebot pro PN			Minimales REQUEST-Delay [ZE]
		$\bar{T}_{H, A_i, v}$	$\bar{T}_{H, B_j, v}$	$\bar{T}_{H, C, v}$	$\bar{\Phi}_{in, RM, A_i, v}$	$A_{A_i}$	$A_{B_j}$	$A_C$	$\bar{\tau}_{min, RM, B_j, v}$
Verkehrsklasse	1	0.008	0.050	0.003	3.3	2.6%	11.1%	2.0%	0.016
	2	0.003	0.020	0.003	3.3	1.0%	4.4%	2.0%	0.016
	3	0.003	0.025	0.003	3.3	1.0%	5.5%	2.0%	0.016
	BL	–	0.00125	–	–	–	5.0%	–	–
Summe pro PN:						4.6%	26.1%	6.0%	

**Tabelle 8-1:** Mittlere Bearbeitungszeiten und Verkehrsangebot für die Normalstarbeitspunkte der einzelnen Prozessorknoten (PN)

In Tabelle 8-1 sind weiterhin die mittleren Ankunftsraten für einen  $PN_{A,i}$  im Normallastbereich angegeben. Diese *homogene Verkehrsverteilung* soll die stationäre Grundlast repräsentieren. Wie dieser Tabelle entnommen werden kann, erfahren die  $PN_{B,j}$  bei gleichmäßiger Lastverteilung eine Auslastung von unter 30%. Die  $PN_{A,i}$  als auch der  $PN_C$  stellen hier keinen Systemengpass dar. Die Mess- und Regelintervalldauer beträgt in allen Fallstudien fünf Zeiteinheiten.

## 8.1.2 Bewertungskriterien für die Gesamtleistung

Die Gesamtleistung eines Vermittlungsrechners ist ein Maß für die Güte eines Verfahrens. Die gleichmäßige CPU-Auslastung aller  $B$ -Prozessoren wäre beispielsweise ein denkbares, vor allem in der Rechnertechnik oft verwendetes Optimierungskriterium. Da dieser Lastkenngrößen-typ aber bewusst hier nicht als Eingangsgröße in die Leistungsregelung eingeht, erhebt sich die Frage, welche sonstigen Kenngrößen für eine Bewertung geeignet wären.

Als gängiges Maß für die Bewertung der Steuerleistung aus Sicht eines Vermittlungssystems bieten sich die *Komplettierungsraten* eines  $PN_{B,j}$

$$\overrightarrow{\phi_{out, B_j}(t)} = \begin{bmatrix} \phi_{out, B_j, 1}(t) \\ \dots \\ \phi_{out, B_j, K}(t) \end{bmatrix} \quad (8-1)$$

bzw. die *Annahmeraten*

$$\overrightarrow{\phi_{out, RM, B_j}(t)} = \begin{bmatrix} \phi_{out, RM, B_j, 1}(t) \\ \dots \\ \phi_{out, RM, B_j, K}(t) \end{bmatrix} \quad (8-2)$$

an. Die Annahmeraten sollen hier das Leistungsmaß für den Durchsatz darstellen. Als Gesamtdurchsatz für die einzelnen Verkehrsklassen ergibt sich dann

$$\overrightarrow{\phi_{out, RM}(t)} = \sum_{j=1}^M \overrightarrow{\phi_{out, RM, B_j}(t)} = \sum_{j=1}^M \begin{bmatrix} \phi_{out, RM, B_j, 1}(t) \\ \dots \\ \phi_{out, RM, B_j, K}(t) \end{bmatrix} \quad (8-3)$$

und als *totale Nutzdurchsatzgröße*  $\phi_{out, RM, ges}(t)$  die Summe über aller Klassen

$$\phi_{out, RM, ges}(t) = \sum_{j=1}^M \sum_{v=1}^K \phi_{out, RM, B_j, v}(t). \quad (8-4)$$

Aus Teilnehmersicht sind die *Antwortzeiten* eines  $PN_{B,j}$  für erfolgreiche Dienstanforderungen

$$\overrightarrow{\tau_{RM, B_j}(t)} = \begin{bmatrix} \tau_{RM, B_j, 1}(t) \\ \dots \\ \tau_{RM, B_j, K}(t) \end{bmatrix} \quad (8-5)$$

wichtiger. Über alle  $B$ -Prozessorknoten ergeben sich dann für die einzelnen Verkehrsklassen die mittleren Antwortzeiten

$$\overrightarrow{\bar{\tau}_{RM}(t)} = \frac{1}{M} \cdot \sum_{j=1}^M \overrightarrow{\tau_{RM, B_j}(t)} = \frac{1}{M} \cdot \sum_{j=1}^M \begin{bmatrix} \tau_{RM, B_j, 1}(t) \\ \dots \\ \tau_{RM, B_j, K}(t) \end{bmatrix}, \quad (8-6)$$

bzw. als Gesamtzeitmittelwert  $\overline{\tau_{RM, ges}(t)}$

$$\overline{\tau_{RM, ges}(t)} = \frac{1}{M \cdot K} \cdot \left( \sum_{j=1}^M \sum_{v=1}^K \tau_{RM, B_j, v}(t) \right). \quad (8-7)$$

Aus dieser mittleren Antwortzeit  $\overline{\tau_{RM, ges}(t)}$  über alle  $PN_{B,j}$  und der totalen Nutzdurchsatzgröße  $\phi_{out, RM, ges}(t)$  nach Gl. (8-4) lässt sich ein *Gesamt-Power-Maß* für die  $B$ -Ebene definieren:

$$\Lambda_{P, RM, ges}(t) = \frac{\phi_{out, RM, ges}(t)}{\overline{\tau_{RM, ges}(t)}}. \quad (8-8)$$

Sollen die *Power-Produkte* der einzelnen  $PN_{B,j}$  stärker betont werden, dann können aus diesen

$$\overrightarrow{\Lambda_{P^\Pi, RM}(t)} = \begin{bmatrix} \Lambda_{P^\Pi, RM, B_1}(t) \\ \dots \\ \Lambda_{P^\Pi, RM, B_M}(t) \end{bmatrix} \quad (8-9)$$

weitere kombinierte Leistungsmaße, z.B. für eine mehr knotenindividuelle Bewertung, abgeleitet werden. Als skalare Größe für die *Gesamtleistung* lassen sich aus dem *Power-Produkt-Vektor*  $\overrightarrow{\Lambda_{P^\Pi, RM}}$  verschiedene Kennwerte definieren:

- Betrag  $\Lambda_{P^\Pi, \parallel, RM}$  des Vektors:

$$\Lambda_{P^\Pi, \parallel, RM}(t) = \left| \overrightarrow{\Lambda_{P^\Pi, RM}(t)} \right| = \sqrt{\sum_{j=1}^M [\Lambda_{P^\Pi, RM, B_j}(t)]^2} \quad (8-10)$$

- Summe  $\Lambda_{P^\Pi, \Sigma, RM}$  aller Spaltenelemente:

$$\Lambda_{P^\Pi, \Sigma, RM}(t) = \sum_{j=1}^M \Lambda_{P^\Pi, RM, B_j}(t) \quad (8-11)$$

- Produkt  $\Lambda_{P^\Pi, \Pi, RM}$  aller Spaltenelemente:

$$\Lambda_{P^\Pi, \Pi, RM}(t) = \prod_{j=1}^M \Lambda_{P^\Pi, RM, B_j}(t) \quad (8-12)$$

In dieser Arbeit wird in der Regel das *Power-Produkt-Produkt*  $\Lambda_{P^\Pi, \Pi, RM}$  als Maßzahl für die Gesamtleistung eines Vermittlungsrechners verwendet. Motiviert wird dies dadurch, dass der Verlauf der Summenfunktion  $\Lambda_{P^\Pi, \Sigma, RM}$  nicht ganz „unproblematisch“ ist (siehe später), und dass der Betrag  $\Lambda_{P^\Pi, \parallel, RM}$  relativ aufwendig zu berechnen ist.

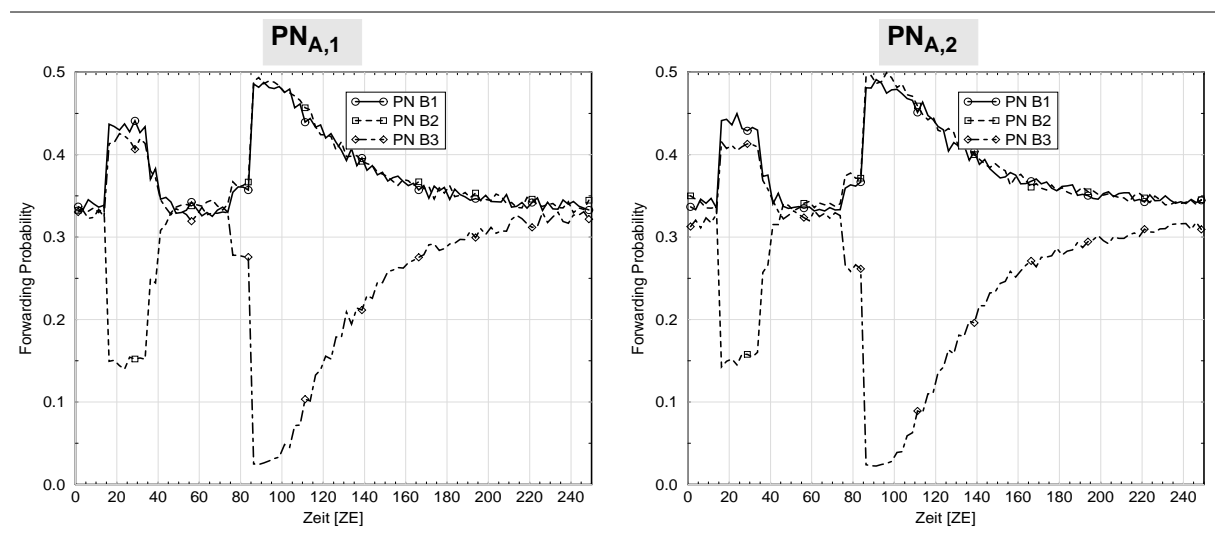
Alle in diesem Abschnitt definierten Maße für die Bewertung der Gesamtleistung der  $B$ -Ebene besitzen ihre Berechtigung. Ihre Anwendung orientiert sich an der Regelkreisstruktur und vor allem am Optimierungsziel der Regelung. Weitere Maßdefinitionen mit z.B. mehr klassenspezifischer Betonung wären ebenfalls denkbar. Die vorgestellten Leistungsmaße sind allesamt Zeitfunktionen. Wird die lokale Überlastabwehr in den  $PN_{B,j}$  deaktiviert, dann erfolgt die globale Lastverteilung verlustfrei. In diesem Fall sollte beispielsweise die stationäre Gesamtleistung in Form der *Power-Produkt-Summe*, unabhängig vom Verfahren, für ein ausreichend langes Messintervall  $T$ , ungefähr gleich groß sein:

$$\overline{\Lambda_{P^\Pi, \Sigma, RM}} = \frac{1}{T} \cdot \int_{t=t_0}^{t_0+T} \Lambda_{P^\Pi, \Sigma, RM}(t) dt. \quad (8-13)$$

Bei einer dynamischen Betrachtung sind natürlich Unterschiede zu erwarten.

## 8.2 Verfahren auf Basis der Antwortzeit

Ein einfaches Lastzuteilungsverfahren, basierend auf dem Indikator *Antwortzeit*, wurde in Abschnitt 4.6.4.2 vorgestellt. In einem ersten Ansatz werden keine Verkehrsklassen unterschieden. Aufgrund interner Lastimpulse durch nichtvermittlungstechnische Prozesse auf verschiedenen  $PN_{B,j}$  wird die Regelung bei nachfolgender Messung aus ihrem Gleichgewichtszustand gebracht. Die lokalen Lastregelkreise der *B*-Prozessoren wurden für diese Fallstudie deaktiviert.



**Bild 8-1:** Lastzuteilungsvektoren  $\overrightarrow{\pi_{F,A_i}}(k)$  der beiden  $PN_{A,i}$ . Interne Überlast in  $PN_{B,2}$  für  $t \in [10,30]$  und in  $PN_{B,3}$  für  $t \in [70,80]$ .

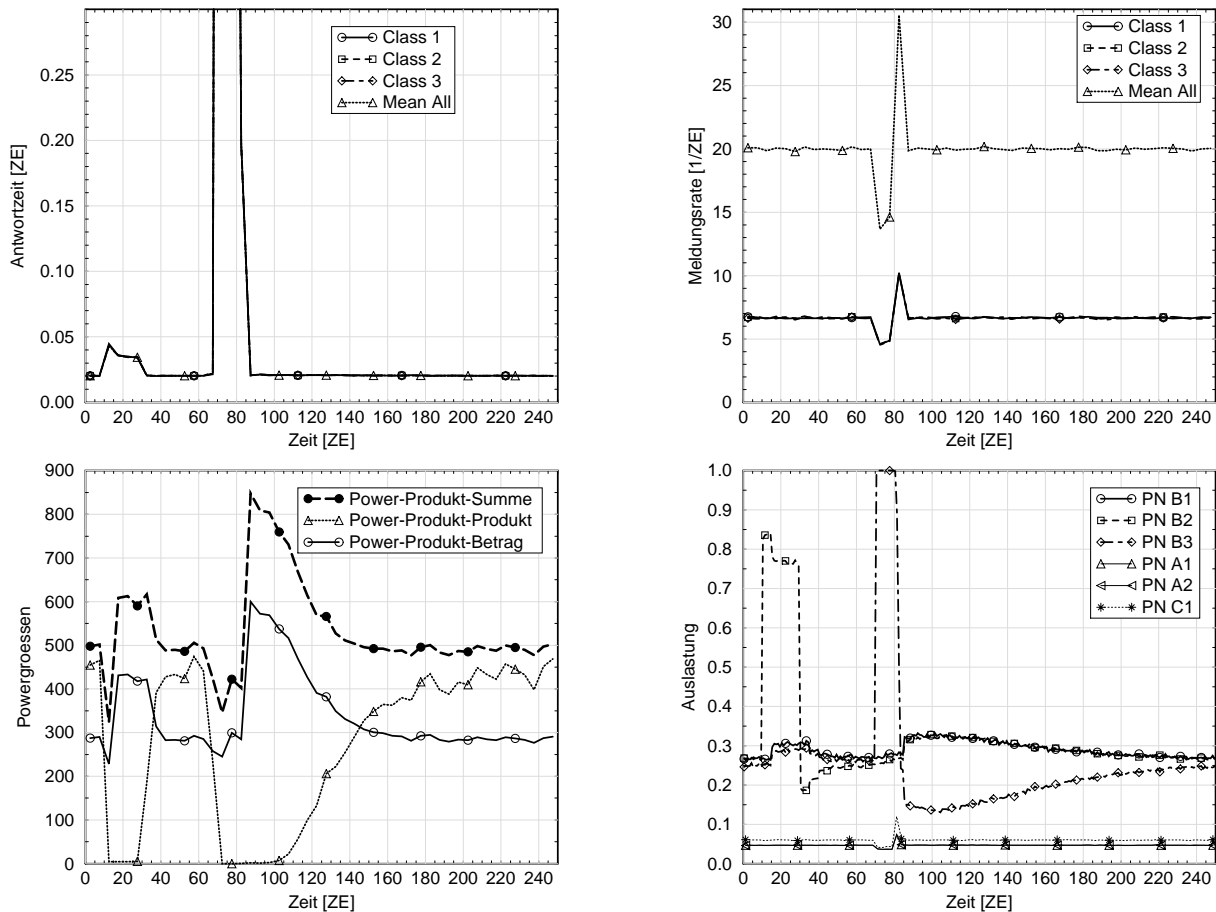
Da extern ein homogenes Verkehrsgemisch anliegt, sind die Zuteilungswahrscheinlichkeiten im Normalfall  $\pi_{F,A_i,B_j}(k) = 1/3$ . Treten interne Lastschwankungen auf, werden diese anhand der gemessenen Antwortzeiten sichtbar und es wird mit einer anderen Verkehrszuteilung reagiert (Bild 8-1). Aus den beiden Diagrammen wird auch ersichtlich – die Aktorikeinstellungen  $\overrightarrow{\pi_{F,A_i}}(k)$  sind nahezu identisch –, dass beide  $PN_{A,i}$  denselben Belastungszustand der *B*-Ebene ermitteln, obwohl sie diese Information aus disjunkten Verkehrsströmen extrahieren.

Dass die Auswahl einer geeigneten Kenngröße für die Bewertung der Gesamtleistung durchaus nicht unproblematisch ist, soll anhand dieser Parameterstudie verdeutlicht werden. In Bild 8-2 sind hierzu die in Abschnitt 8.1.2 definierten Leistungsmaße dargestellt. Zusätzlich sind die CPU-Auslastungen aller Prozessorknoten angegeben. Da keine Verluste auftreten, wachsen die Wartezeiten während der Hochlastphasen an. Die mittlere Gesamtantwortzeit  $\overline{\tau_{RM,ges}}(t)$  ist deshalb als Leistungsmaß hier weniger geeignet. Aufgrund der homogenen Verteilung des externen Verkehrs sind auch die mittleren Antwortzeiten und Durchsatzraten für alle Verkehrsklassen gleich groß.

Die Annahmeraten  $\overrightarrow{\phi_{out,RM}}(t)$  bzw.  $\phi_{out,RM,ges}(t)$  sollten während einer internen Hochlastphase einbrechen und anschließend kurzzeitig über dem stationären Mittelwert liegen.<sup>1</sup> In



Bild 8-2 (rechts oben) ist dieses Zeitverhalten für die zweite Lastspitze sichtbar. Der erste Lastimpuls ist zu schwach, um bei der vorliegenden Messintervalldauer und der resultierenden Zeitaufösung detektiert werden zu können.



**Bild 8-2:** Gesamtleistung des Vermittlungsrechners: a) mittlere Antwortzeiten  $\bar{\tau}_{RM}(t)$  und  $\bar{\tau}_{RM, ges}(t)$ ; b) mittlere Annahmeraten  $\bar{\phi}_{out, RM}(t)$  und  $\bar{\phi}_{out, RM, ges}(t)$ ; c) Powermaße mit Betrag  $\Lambda_{P^{\Pi}, \parallel, RM}(t)$ , Summe  $\Lambda_{P^{\Pi}, \Sigma, RM}(t)$  und dem Produkt  $\Lambda_{P^{\Pi}, \Pi, RM}(t)$  (skaliert mit Faktor  $10^{-4}$ ) sowie d) die Auslastungen aller Prozessorknoten  $\rho_{PN_i}(t)$ .<sup>2</sup>

Da die stationären Durchsatzwerte sich nicht ändern und die Durchlaufzeiten während der Hochlastphasen prinzipiell nicht besser werden, sollte sich dies auch in der Verschlechterung der Power-Maße bemerkbar machen. Das erwartete Zeitverhalten weist das Power-Produkt-Produkt  $\Lambda_{P^{\Pi}, \Pi, RM}(t)$  auf. Dagegen sind hier die Power-Produkt-Summe  $\Lambda_{P^{\Pi}, \Sigma, RM}(t)$  und der Power-Produkt-Betrag  $\Lambda_{P^{\Pi}, \parallel, RM}(t)$  weniger als Gesamtleistungsmaß geeignet, da diese hier einen konträren Leistungsverlauf besitzen. Die Ursache hierfür liegt bei den B-Prozessorindividuellen Power-Kenngrößen.

<sup>1</sup> Nichtvermittlungstechnische Aufträge werden mit höherer Priorität bedient als neue Dienstanforderungen.

<sup>2</sup> In allen Schaubildern dieses Kapitels sind die berechneten Vertrauensintervalle u. a. aus Gründen der Übersichtlichkeit nicht eingezeichnet (siehe auch Anhang D.2).

## 8.3 Verfahren auf Basis der Durchsatzkenngröße

Wie in Abschnitt 4.6.4.3 erläutert wurde, findet bei diesem Verfahren eine Unterteilung des Lastraumes in zwei Bereiche statt. Dies geschieht auf Basis der mittleren Antwortzeit  $\overline{\tau}_{RM, B_j}$  und einem vorgegeben Schwellwert  $\tau_{Th}$ .<sup>3</sup> Als erste *Näherung* für diesen *Schwellwert* kann folgende Einstellung, die auf der geschätzten minimal möglichen Antwortzeit  $\overline{\tau}_{min, RM}$  basiert, vorgenommen werden:<sup>4</sup>

$$\tau_{Th} \approx 2 \cdot \overline{\tau}_{min, RM}. \quad (8-14)$$

Zu diesem Lastverteilverfahren wurden auch zahlreiche Parameterstudien durchgeführt. Der völlig verteilte Algorithmus zeigt auch hier das erwartete Verhalten und sorgt für eine gleichmäßige Belastung der B-Ebene. Exemplarische Messresultate hierzu werden im Zusammenhang mit dem Vergleich aller Verfahren präsentiert (siehe Abschnitt 8.4).

## 8.4 Verfahren auf Basis von Power-Maßzahlen sowie Vergleich der drei Algorithmen

Im folgenden liegt der Schwerpunkt auf der *Power-Produkt-Methode*, wobei deren Leistungsdaten mit den beiden anderen Algorithmen verglichen werden sollen.

### 8.4.1 Inhomogener Verkehr und lokale Überlastabwehr

Neben der Grundlast (siehe Tabelle 8-1) werden folgende Lastimpulse erzeugt: a) zehnfache Last  $\phi_{in, RM, A_1, 1}(t)$  für  $t \in [10, 40]$  und b) 15fache Last  $\phi_{in, RM, A_2, 2}(t)$  für  $t \in [30, 60]$ . Das inhomogene Verkehrsgemisch resultiert somit zum einen aus einem überdurchschnittlichen Angebot aus Klasse-1- bzw. Klasse-2-Anforderungen, die wiederum an verschiedenen  $PN_{A,i}$  ankommen. In allen  $PN_{B,j}$  ist eine lokale Überlastabwehr aktiviert, die ab einer 80%igen CPU-Auslastung gleichmäßig für alle Klassen Anforderungen zurückweist.

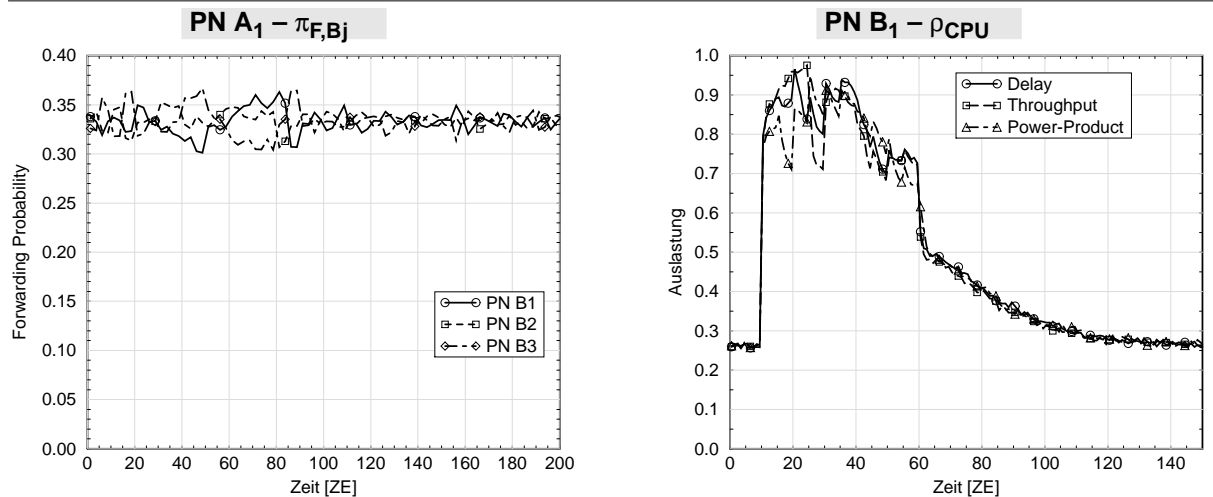
Trotz stark unterschiedlich belasteter  $PN_{A,i}$  wird, wie auch zu erwarten, der Verkehr nahezu symmetrisch auf die  $PN_{B,j}$  verteilt. Wie Bild 8-3 entnommen werden kann, liegen alle Zuteilungswahrscheinlichkeiten im Mittel bei 33%.<sup>5</sup> Im rechten Diagramm ist die Prozessorauslastung  $\rho_{B_1}(t)$  für die drei verschiedenen Verfahren dargestellt. Aufgrund der Überlastabwehr liegt die Auslastung während der Hochlastphase knapp über 80%.

Leichte Unterschiede zwischen den einzelnen Strategien werden bei den Annahme- und Ablehnungsraten sichtbar (siehe Bild 8-4). Diese Differenzen basieren auf den relativ kleinen Unterschieden bei den berechneten Zuteilungswahrscheinlichkeiten. Diese Schwankung ist bei der *Power-Produkt-Methode* am größten.

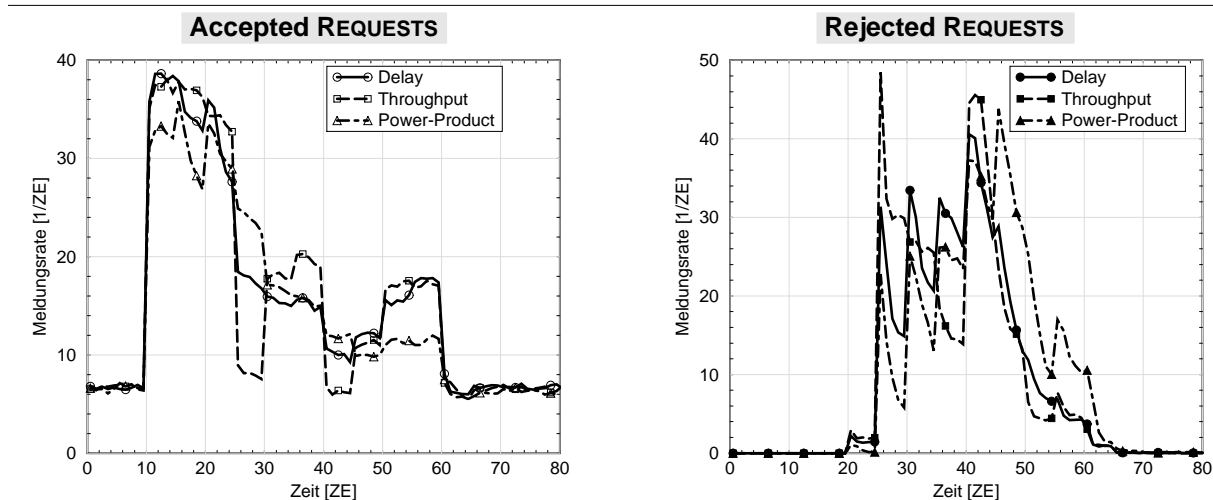
<sup>3</sup> Anhaltspunkte für mittlere Antwortzeiten werden i. d. R. auf Messungen basieren.

<sup>4</sup> Begründung: Bei bestimmten Systemen liegt der optimale Arbeitspunkt (für maximales Power) bei *doppelter minimal möglicher Antwortzeit*  $\tau_{opt} = 2 \cdot \tau_{min}$  (siehe auch Anhang B.2.1, Fußnote 2 auf S. 235).

<sup>5</sup> Aus Symmetriegründen sind hier jeweils nur Diagramme mit den Ergebnissen eines PN dargestellt.



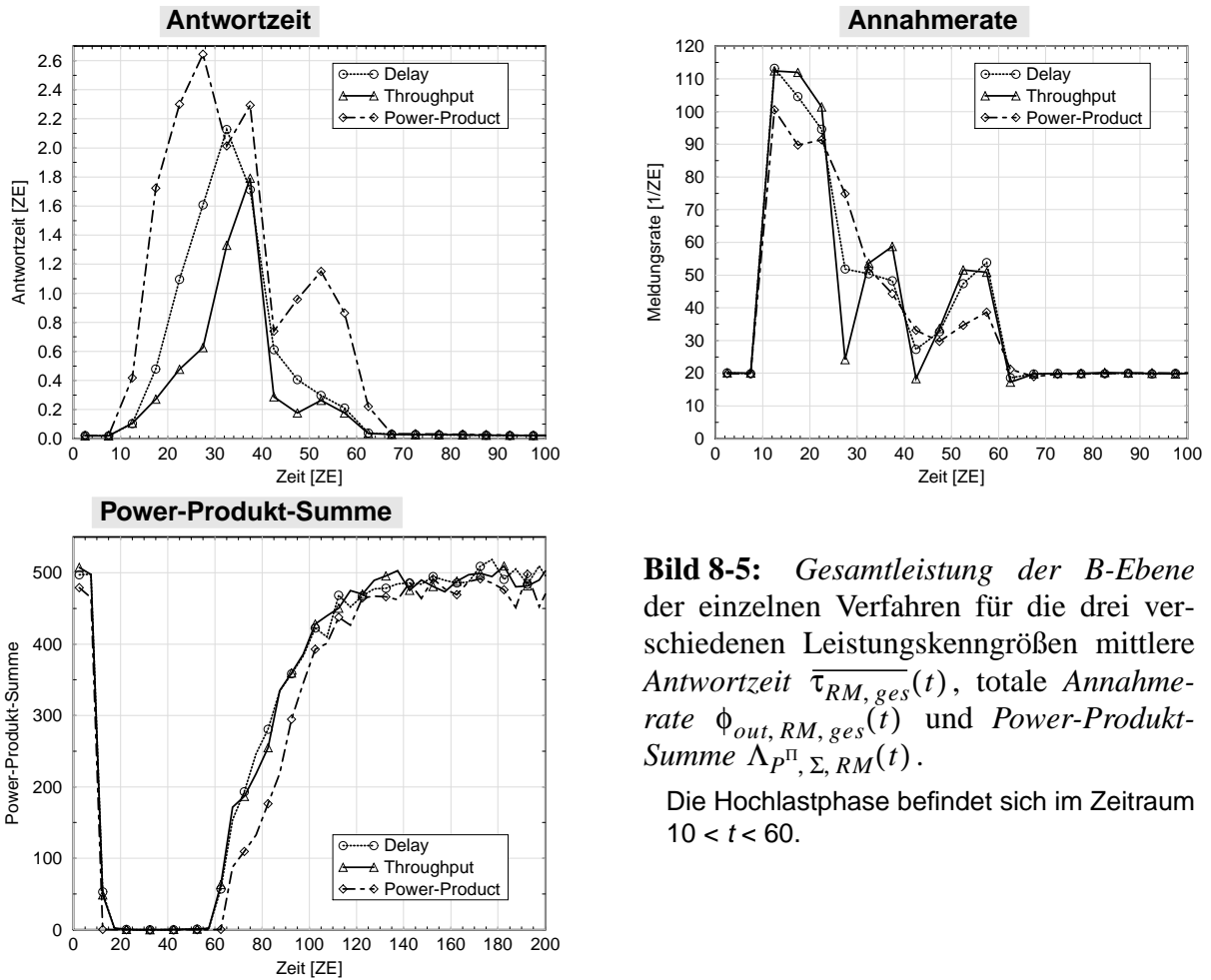
**Bild 8-3:** Lastzuteilungsvektor  $\overrightarrow{\pi_{F, A_1}(k)}$  von  $PN_{A,1}$  nach der Power-Produkt-Strategie und CPU-Auslastung  $\rho_{B,1}(t)$  von  $PN_{B,1}$  für alle drei verschiedenen Verfahren.



**Bild 8-4:** Angenommene  $\phi_{out, RM_{AC}, B_3, ges}(t)$  und abgelehnte Anforderungen  $\phi_{out, RM_{RC}, B_3, ges}(t)$  für  $PN_{B,3}$ .

In der Gesamtleistung der  $B$ -Ebene des Vermittlungsrechners gibt es deshalb auch kleine Variationen zwischen den einzelnen Verfahren (siehe Bild 8-5). Bei der Gesamtleistungskenngröße *Power-Produkt-Summe*  $\Lambda_{P^{\Pi}, \Sigma, RM}$  weist das *Power-Produkt-Verfahren* etwas schlechtere Werte während der Ausschwingphase auf. Unter stationärer Hoch- und natürlich auch Grundlast sind allerdings kaum Unterschiede feststellbar. Das *Durchsatzverfahren* weist für die Gesamtleistungskenngröße *mittlere Antwortzeit*  $\overline{\tau_{RM, ges}}(t)$  die kleinsten Werte auf. Bei der totalen *Annahmerate*  $\phi_{out, RM, ges}(t)$  sind allerdings die größten Schwankungen sichtbar.

Durch die symmetrische Lastzuteilung sind die Unterschiede in der Gesamtleistung zwischen den einzelnen Verfahren relativ klein. Betrachtet man Aufwands- und Kostenargumente dann weist das *Durchsatzverfahren* allerdings hier klare Vorteile auf.



**Bild 8-5:** Gesamtleistung der B-Ebene der einzelnen Verfahren für die drei verschiedenen Leistungskenngrößen mittlere Antwortzeit  $\bar{\tau}_{RM, ges}(t)$ , totale Annahmerate  $\phi_{out, RM, ges}(t)$  und Power-Produkt-Summe  $\Lambda_{P^{\Pi}, \Sigma, RM}(t)$ .

Die Hochlastphase befindet sich im Zeitraum  $10 < t < 60$ .

## 8.4.2 Inhomogener Verkehr und deaktivierte Überlastabwehr

Ohne lokale Überlastabwehr in den  $PN_{B,j}$  werden alle neuen Dienstanforderungen akzeptiert und somit entsteht auch keinerlei Blindlast. Da auch der extern ankommende Verkehr nahezu symmetrisch aufgeteilt wird, ist die mittlere Belastung für alle  $PN_{B,j}$  gleich. Einzige Unterschiede resultieren aus den unterschiedlichen Lastzuteilungsverfahren. Die Einstellung der Zuteilungswahrscheinlichkeiten basiert zum einen auf der verwendeten Leistungskenngröße und zum anderen auf der Berechnungsvorschrift.

Messungen haben die symmetrische Verkehrsaufteilung und die resultierende gleichmäßige Belastung der B-Ebene bestätigt. Zwischen den verschiedenen Verfahren sind hier keine signifikanten Unterschiede feststellbar.

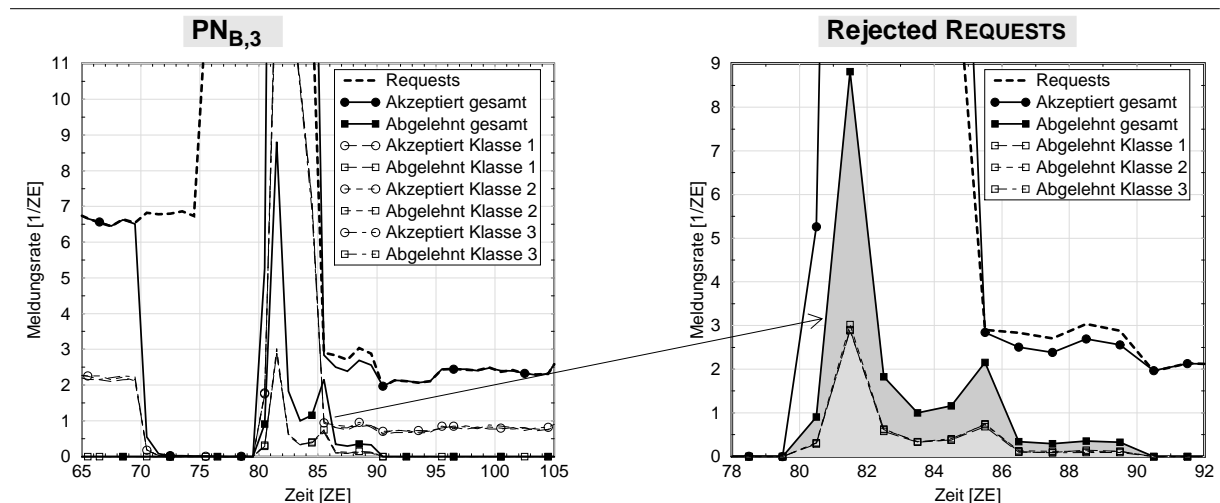
## 8.4.3 Interne Lastschwankungen

Das externe Verkehrsangebot sei homogen verteilt und stationär (Parametereinstellungen siehe Tabelle 8-1). Hochlastphasen werden durch Erhöhung der nichtvermittlungstechnischen Last in den  $PN_{B,j}$  erzeugt (Grundlast beträgt  $A_{BL, B_j} = 5\%$ ). Eine lokale Überlastabwehr ist aktiviert. Folgende zwei interne Lastimpulse liegen den Messungen zugrunde:

- $PN_{B,2}$ :  $A_{BL, B_2}(t) = 62.5\%$  für  $t \in [10, 30]$  und
- $PN_{B,3}$ :  $A_{BL, B_3}(t) = 100\%$  für  $t \in [70, 80]$ .

In Anhang D sind für diese Fallstudie die Messdaten von Sensoren, Aktoren sowie verschiedener Last- und Leistungsgrößen zusammengestellt. Beispielsweise sind in Bild D-6 (S. 269) die resultierenden Zuteilungswahrscheinlichkeiten  $\overrightarrow{\pi_{F, A_1}}(k)$  von  $PN_{A,1}$  für die verschiedenen Verfahren dargestellt. Da das externe Verkehrsgemisch homogen ist, ist die Belastung für die beiden A-Prozessoren (siehe auch rechte Spalte in Bild D-6), und damit auch die berechneten Zuteilungswahrscheinlichkeiten, annähernd gleich.

In Bild 8-6 sind exemplarisch die Meldungsraten des Knotens  $PN_{B,3}$  für die zweite Hochlastphase aufgeführt (*Power-Produkt-Methode*). Aufgrund der lokalen Überlastabwehr wird ein gewisser Anteil von Anforderungen zurückgewiesen. Wie das rechte Diagramm auch verdeutlicht, sind – aufgrund der homogenen Verkehrsverteilung – die instationären Ablehnungsraten  $\phi_{out, RM_{RC, B_j, v}}(t)$  für alle Verkehrsklassen gleich groß.



**Bild 8-6:** Meldungsraten ( $\phi_{in, RM, B_j, ges}(t)$ ,  $\phi_{out, RM_{AC}, B_j, ges}(t)$  und  $\phi_{out, RM_{RC}, B_j, ges}(t)$ ; sowie individuell  $\phi_{out, RM_{AC}, B_j, v}(t)$  und  $\phi_{out, RM_{RC}, B_j, v}(t)$ ) von  $PN_{B,3}$  (links, siehe auch Bild D-7). Um die Ablehnungsraten hervorzuheben ist rechts eine Ausschnittsvergrößerung dargestellt (graue Schattierungen).

Für die Bewertung der *Gesamtleistung der B-Ebene* sind in Bild D-9 (S. 272) verschiedene Maßzahlen zusammengestellt. Dabei sei besonders auf folgenden Unterschied zwischen *Power-Produkt-Summe*  $\Lambda_{P^\Pi, \Sigma, RM}$  und *Power-Produkt-Produkt*  $\Lambda_{P^\Pi, \Pi, RM}$  hingewiesen. Während der Hochlastphasen ist prinzipiell eine Leistungsver schlechterung zu erwarten. Dieses Verhalten zeigt deutlich das *Power-Produkt-Produkt* während die *Power-Produkt-Summe* nach einem kurzen Einbruch einen Leistungsanstieg verzeichnet.

## 8.5 Bewertung und Ausblick

Herausragende Merkmale der *DLB-Algorithm*en sind die völlig verteilte Regelstruktur ohne die Notwendigkeit einer zentralen Instanz sowie die indirekte Ermittlung des Belastungszu-

stands eines Prozessorknotens außerhalb dieser Verarbeitungseinheit in Form von *Black-Box*-Messungen. Dadurch entsteht bei dieser globalen Leistungsregelung keinerlei Blindlast was wiederum das Leistungsverhalten positiv beeinflusst. Oberste Zielsetzung dieses Regelkonzeptes ist die gleichmäßige Belastung aller *B*-Prozessoren durch Optimierung des knotenindividuellen Leistungsverhaltens. Dieser Aspekt stand bei den hier durchgeführten Parameterstudien im Vordergrund. Da für die Regelung aber auch klassenindividuelle Leistungskenngrößen erfasst werden, wäre auch eine Optimierung des Leistungsverhaltens für die verschiedenen Verkehrsklassen denkbar. Hierzu müssten die Algorithmen entsprechend geändert werden. Da das klassenindividuelle Last- und Leistungsverhalten in dieser generischen Steuerungsstruktur bereits in das *NRAC*- und *PPC*-Verfahren einging, wurde bewusst bei den *DLB*-Algorithmen dieser Aspekt ausgespart.

Die durchgeführten Simulationsstudien haben den Nachweis für die erfolgreiche Funktion dieses total verteilten Verfahrens erbracht. Als Kriterien standen hierbei die dynamischen Eigenschaften, was beispielsweise die Attribute Übergangverhalten, Oszillationen und Stabilität betrifft, im Vordergrund. Zusätzlich wurden die drei vorgestellten Algorithmen aber auch bezüglich ihrer stationären Leistungsdaten verglichen.

Tabelle 8-2 gibt eine Übersicht zur qualitativen Bewertung der verschiedenen Algorithmen. Wird ein vergleichsweise schnell reagierender Regelkreis gewünscht, dann weisen das *Power-Produkt*- und das Durchsatzverfahren leichte Vorteile auf. Werden aber keine besonderen Anforderungen an die Reagibilität gestellt, dann bietet sich das Antwortzeitverfahren an, da auch hier die Implementierung vergleichsweise geringen Aufwand erfordert.

Algorithmus (Optimierungsziel)	Merkmale				
	Messaufwand	Komplexität (Rechenzeit)	Übergangverhalten	Stabilität	Gesamtleistung <sup>a</sup>
<i>Antwortzeit</i> $\tau_{RM}$	0	++	-	+	++
<i>Durchsatz</i> $\phi_{RM}$	+	+	+	0	++
<i>Power-Produkt</i> $\Lambda P^{\Pi}, RM$	-	-	+	0	++

a. Beachte: Regelstrategie hat als Zielsetzung für eine gleichmäßige Lastverteilung in der *B*-Ebene zu sorgen.

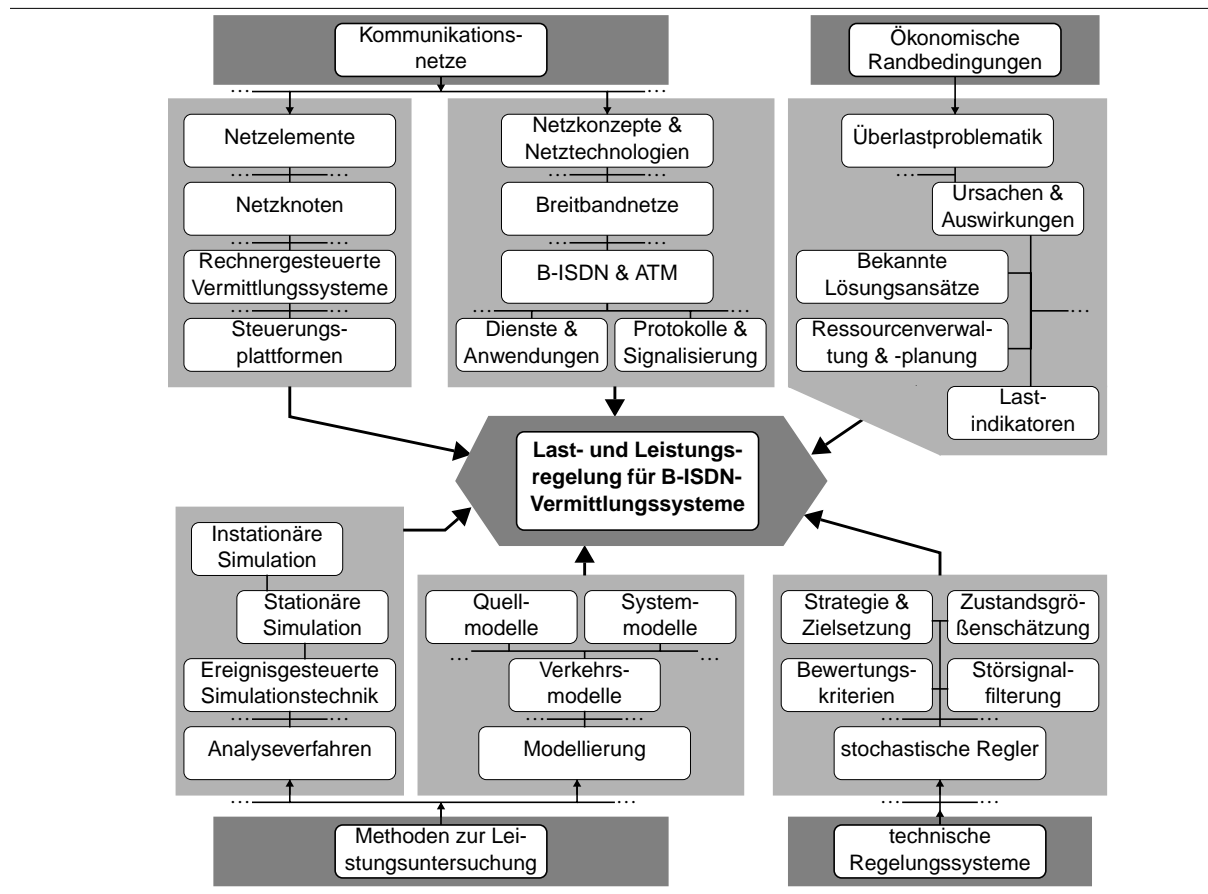
**Tabelle 8-2:** Qualitative Bewertung und Vergleich der drei *DLB*-Algorithmen anhand ausgewählter Merkmale

Neben diesen Merkmalen wird aber in der Praxis auch das Anwendungsumfeld – d.h. Raum-, Zeit- und Lastebene der Regelkreisstruktur sowie gegebenenfalls weitere Regelungen – ein maßgebliches Kriterium für die Auswahl eines bestimmten Verfahrens sein.

# Kapitel 9

## Zusammenfassung und Ausblick

In der vorliegenden Arbeit wurde die Problematik der Last- und Leistungsregelung in dienstintegrierenden Breitband-Vermittlungssystemen untersucht. Dabei lag der Schwerpunkt auf der universellen Netztechnologie B-ISDN, die wiederum ATM als Übermittlungsverfahren verwendet. Die Überlastabwehr ist eine Steuerungsfunktion in einem Netzknoten, deren Konzeption eine äußerst vielschichtige Aufgabe darstellt. In Bild 9-1 sind hierzu die wichtigsten Randbedingungen nach strukturellen Gesichtspunkten zusammengestellt.



**Bild 9-1:** Wichtige Randbedingungen für die Konzeption einer Überlastabwehr für B-ISDN-Netzknoten

Seit dem Einsatz rechnergesteuerter Vermittlungssysteme spielt der Entwurf leistungsfähiger Überlastabwehralgorithmen eine wichtige Rolle in Forschung und Entwicklung. Betrachtet man deren Entwicklung im Lauf der Jahre, dann kann festgestellt werden, dass für neue Systemarchitekturen und Netzgenerationen weniger grundsätzlich neue Lastregelverfahren entworfen wurden. Vielmehr wurden oft bestimmte Elemente, Teilstrategien oder Mechanismen

bekannter Algorithmen wiederverwendet. Eine Analyse der Überlastabwehrproblematik für B-ISDN/ATM-Netzknoten muss deshalb Erfahrungen und Algorithmen vergangener System- und Netzgenerationen berücksichtigen. Mit aus diesem Grund wurde für diese Arbeit eine umfangreiche und ausführliche Literaturrecherche durchgeführt.

Prinzipielle Aspekte, die die Systemplattform betreffen, wurden in Kapitel 2 erläutert. Dies betrifft u. a. die Einordnung in die Kategorie speichervermittelnder Netzknoten, die Darstellung der vielfältigen Aufgaben einer Vermittlungssteuerung sowie allgemeiner Steuerungsstrukturmerkmale. Grundsätzlich *neue Anforderungen* an die Überlastabwehr resultieren aus dem breiten Spektrum von Anwendungen und Diensten, die durch das B-ISDN unterstützt werden sollen. Die Ablaufsteuerung eines Dienstes manifestiert sich aus Sicht des Vermittlungsrechners in den hierzu notwendigen Signalisierabläufen. Aus diesem Grund wurden in Kapitel 2 auch Zeichengabeprotokolle des B-ISDN sowie ihre neuen Funktionalitäten dargestellt.

Wie in Bild 9-1 angedeutet wird, bildet der zugrundeliegende Systemaufbau eine wichtige Anforderung im Kontext einer Überlastabwehr. Aus diesem Grund wurden gängige ATM-Produkte und Knotenkonzepte bezüglich ihres architekturellen Aufbaus untersucht und anhand ihrer Steuerungsstruktur systematisiert.<sup>1</sup> Für diese Klassifizierung wurde hierzu ein einfaches Schema, das wesentlich auf Aspekten der Funktions- und Lastteilung basiert, definiert.

Die Überlastproblematik resultiert letztendlich aus der Tatsache, dass Ressourcen aus ökonomischen und technischen Gründen in einem abgeschlossenen System nur in endlichem Maße verfügbar sind. In einem verbindungsorientierten Netz beginnt die Abwicklung eines Dienstes durch eine vom Teilnehmer initiierte Anforderungsmeldung. Zur Komplettierung eines Dienstes sind, temporär oder für die gesamte Lebensdauer des Dienstes, vermittlungs- und steuerungstechnische Betriebsmittel in einem Netzknoten bereitzustellen. Dabei interessieren hier in dem betrachteten Teilsystem *Vermittlungsrechner* primär nur die steuerungstechnischen Betriebsmittel. Die sich aus dieser Ressourcenbedarfs- und damit der Überlastproblematik ergebenden Fragestellungen wurden ausführlich in Kapitel 3 erläutert.

Der Schwerpunkt lag hierbei auf der Darstellung der sich durch die Integration eines breiten Dienstespektrums im B-ISDN ergebenden zusätzlichen Anforderungen. So wird beispielsweise die Dienstvielfalt und auch die gebotene Funktionalität durch zusätzliche Dienstmerkmale immer größer, oder das Nutzungsprofil ändert sich zunehmend. Weiterhin ist die Verkehrscharakterisierung und die Ressourcenbedarfsabschätzung für einen bestimmten Dienst äußerst schwierig. Bei dieser Netztechnologie können auch starke Wechselwirkungen zwischen Nutzer- und Steuerungs-Ebene, beispielsweise aufgrund der unterschiedlichen Ressourcenverwaltungen, auftreten und sich negativ auf das Leistungsverhalten auswirken.

Neben diesen mehr systeminternen Ursachen hat sich auch die spezifische Umwelt gravierend verändert. So kommen zunehmend rechnerbasierte Endgeräte zum Einsatz, die ein grundlegend anderes ‚Teilnehmerverhalten‘ repräsentieren können als menschliche Benutzer. Die in Kapitel 3 durchgeführten Studien wurden durch eine Patentrecherche sowie die Zusammenstellung wichtiger bekannter Lösungsansätze ergänzt.

Die in den Kapiteln 2 und 3 präsentierten Untersuchungen bilden den Ausgangspunkt für die Aufzeigung tragfähiger Lösungsansätze zu Lastkontrollverfahren in einem B-ISDN/ATM-Vermittlungsrechner. In Kapitel 4 wurde hierzu ein *neues Konzept einer mehrstufigen Regelkreisstruktur für eine generische Steuerungsarchitektur* vorgestellt. Hierunter sind kooperierende Regelungssysteme zu verstehen, die sich hauptsächlich hinsichtlich ihres räumlich und zeitlich begrenzten Wirkungskreises unterscheiden lassen. Aus diesem Rahmenwerk wurden anschließend drei ausgewählte, vielversprechende Verfahren detaillierter vorgestellt. Diese neu entwick-

---

<sup>1</sup> Eine repräsentative Auswahl ist in Anhang A zusammengestellt.



kelten Algorithmen werden im Zusammenhang mit ihrer Leistungsbewertung weiter unten erläutert.

Zur tiefergehenden Analyse von Verfahren zur Last- und Leistungsregelung hat sich die Simulationsmethode bewährt. Diese erfordert zunächst eine *verkehrstheoretische Modellierung* des betrachteten Systems sowie des Verkehrs. Kapitel 5 beschäftigt sich deshalb zu Beginn mit der Modellierungsproblematik des Steuerungsverkehrs im Kontext des B-ISDN. Für diese Arbeit wurden Gruppen ‚ähnlicher Dienste‘ auf mehrere repräsentative Verkehrsklassen abgebildet. Der Ablauf eines Dienstes selbst erfolgt durch ein Szenario generischer Signalisier Nachrichten.

Gerade für Überlastabwehrverfahren müssen strukturelle und betriebsorganisatorische Aspekte im Systemmodell eines Vermittlungsrechners berücksichtigt werden. Hierzu wurde ein entsprechendes Prozessmodell entwickelt, das so flexibel konzipiert wurde, dass es als Elementarzelle in Modellen für Mehrprozessorarchitekturen eingesetzt werden kann.

Weiterhin wurden in Kapitel 5 die Gründe für diesen ungewöhnlich detaillierten Modellierungsansatz motiviert. Für die Untersuchung bestimmter Teilaspekte ist dabei der gewählte Modellierungsgrad sicherlich zu fein. Diesem Umstand kann aber dadurch begegnet werden, dass sich hier die Eigenschaften eines Simulationsmodells in gewissem Umfang dynamisch konfigurieren lassen.<sup>2</sup> In Anhang C sind ergänzend repräsentative Signalisierszenarien und die konkreten Simulationsmodelle dargestellt.

Zur Bewertung eines Lastregelverfahrens stehen die dynamischen Eigenschaften im Vordergrund. Dies bedingt die Messung zeitabhängiger Prozessverläufe und damit den Einsatz der Technik der *instationären Simulation*. Für die Untersuchung stationärer Vorgänge steht am Institut für Nachrichtenvermittlung und Datenverarbeitung der Universität Stuttgart eine leistungsfähige Simulationsumgebung zur Verfügung. Im Rahmen dieser Arbeit wurde diese Bibliothek um zahlreiche Komponenten zur transienten Simulation generischer Vermittlungsrechner erweitert. Somit wurde ein flexibles Werkzeug geschaffen, das auch zukünftig für die instationäre Untersuchung anderer Architekturen und Regelkreisstrukturen genutzt werden kann.

Wie bereits oben erwähnt wurde, sind im Rahmen dieser Arbeit *drei neue Verfahren* zur Last- und Leistungsregelung für einen generischen B-ISDN-Vermittlungsrechner entworfen worden. Allen gemeinsam ist, dass *mehrere Verkehrsklassen* in den jeweiligen Algorithmen berücksichtigt werden. Hierdurch wird dem Diensteaspekt besser Rechnung getragen. Diese Eigenschaft hebt sich auch von klassischen Überlastabwehrstrategien ab, in welche fast ausschließlich nur eine Verkehrsklasse (‚Rufe‘) eingehen.

In Kapitel 6 erfolgte die Vorstellung einer *Annahmeregulierung für neue Dienstanforderungen*, die auf einer Ressourcenabschätzung steuerungstechnischer Betriebsmittel basiert. Dieser präventiv wirkende Lastregelkreis entspricht einer Rufannahmesteuerung (*Call/Connection Admission Control, CAC*) für die Steuerungs-Ebene. Im Mittelpunkt steht dabei die kritischste Ressource *CPU*; weitere Betriebsmittel könnten bei Bedarf in einem erweiterten Algorithmus berücksichtigt werden. Ein Dienst einer bestimmten Verkehrsklasse wird hier durch eine sogenannte *effektive Prozessorbelastung*, die auf der Abschätzung der ungefähren mittleren Belegungszeit und Kommunikationsdauer basiert, charakterisiert. Übersteigt der prognostizierte Lastwert einen vorgegebenen Grenzwert, dann wird eine neue Anforderung zurückgewiesen. Damit erfolgt prinzipiell eine Lastreduktion aus Systemsicht.

Während dieses ‚CAC‘-Verfahren bereits im Normallastbereich aktiv ist, wird der zweite Regelkreis erst ab einer bestimmten Lastschwelle aktiviert. Die Hauptzielsetzung ist die lokale

---

<sup>2</sup> Bis auf bestimmte Strukturmerkmale sind nahezu alle Eigenschaften in einer Eingabedatei parametrisiert. Erst zur Laufzeit wird dann das konkrete Simulationsmodell dynamisch aufgebaut.

*Leistungsregelung* eines Prozessorknotens für die Dienstbearbeitung. Dabei wird – im Gegensatz zu den meisten bekannten Regelverfahren – ein *abgeleitetes Leistungsmaß*, das einen Kompromiss zwischen Teilnehmer- und Netzsicht hinsichtlich der Leistung bietet, verwendet. Zurückgegriffen wurde hierbei auf die bewährte *Power-Definition* [208], die den Nutzdurchsatz mit der Kenngröße Antwortzeit kombiniert. Werden die *Power-Maße* der einzelnen Verkehrsklassen geeignet miteinander verknüpft, dann entsteht eine Gesamtleistungskenngröße die zur Maximierung der Steuerungsleistung herangezogen werden kann. Da sich hierfür das Produkt bewährt hatte, wurde dieser Regelkreis auch als *Power-Product Control* bezeichnet.

In Kapitel 7 wurde dieser Algorithmus untersucht, bewertet und ergänzend noch ein vereinfachtes Verfahren vorgestellt, welches zur reinen Überlastabwehr besonders gut geeignet ist.

Während die Annahmekontrolle und das Leistungsoptimierungsverfahren von ihrem räumlichen Wirkungskreis auf einen lokalen Prozessorknoten begrenzt sind, ist der dritte Algorithmus als übergeordneter, *globaler Regelkreis* für den *gesamten Vermittlungsrechner* konzipiert. Von architektureller Seite wird hierbei eine *Multiprozessorsteuerungsstruktur mit mehreren Funktionsebenen und Lastteilungsprinzip* betrachtet. Die beiden unteren Funktionsebenen sind für Signalisierung (*A-Ebene*) und Dienstbearbeitung (*B-Ebene*) verantwortlich.<sup>3</sup> Da die Dienstbearbeitungsprozessoren den Leistungsengpass darstellen, ist die Hauptzielsetzung dieser Regelung für eine gleichmäßige Belastung der *B-Ebene* zu sorgen.

Hierzu wurde ein neues Verfahren mit folgenden Merkmalen entworfen. Die Regeleinrichtung wurde auf die *A-Ebene* ausgelagert und damit räumlich vollständig von der Regelstrecke (*B-Ebene*) getrennt. Damit entsteht keinerlei Blindlast durch dieses Verfahren! Da auf der *A-Ebene* ebenfalls mehrere Prozessorknoten in Lastteilung arbeiten und keine zentrale Instanz benötigt wird, liegt hierbei ein *völlig verteilter Ansatz* vor. Dies konnte durch die Wahl bestimmter Leistungskenngrößen als Lastindikatoren unterstützt werden. Dem Dienstaspekt wurde wiederum durch das spezifische indirekte Leistungsmaß *Power-Produkt* Rechnung getragen.

In Kapitel 8 wurde die Leistungsfähigkeit dieses *Distributed Load Balancing* unter Beweis gestellt und auch mit Verfahren, die auf anderen Kenngrößen beruhen, verglichen.

Damit wurde insgesamt ein Rahmenwerk einer mehrstufigen Regelkreisstruktur zur Lastkontrolle in Breitband-Vermittlungssteuerungen, unter Berücksichtigung deren spezifischer An- und Herausforderungen, entworfen und analysiert. Die Verwendung solcher Algorithmen in der Vermittlungstechnik verbietet prinzipiell aufwendige und rechenzeitintensive Verfahren; vielmehr sind echtzeitfähige, einfache und robuste Mechanismen gefragt. Auch diesen Randbedingungen wurde hier Rechnung getragen.

Weiterhin lassen sich die verschiedenen Teilstrategien auch isoliert in anderen Anwendungsumgebungen einsetzen. Dabei lässt sich durch die Berücksichtigung weiterer Parameter oder anderer Erweiterungen sicherlich noch eine Verbesserung hinsichtlich des Leistungsverhaltens erzielen.

## **Ausblick**

Neben diesen Aspekten sollen abschließend noch verschiedene Bereiche, in welchen noch starker Bedarf an Forschungs- und Entwicklungsarbeit erwartet wird, identifiziert werden.

Betrachtet man die ATM-spezifischen Aspekte in einem Netzknoten, beispielsweise die Wechselwirkungen zwischen Nutzer- und Steuerungs-Ebene, dann drängt sich der Wunsch nach einer ganzheitlichen Ressourcenverwaltung auf. In solch einer ‚integrierenden CAC‘ sollten

---

<sup>3</sup> Aufgaben mit mehr zentralem Charakter (z.B. Verwaltung bestimmter Ressourcen, Datenbanken) und hoher Zugriffsrate sind in höhere Funktionsebenen ausgelagert.

sowohl vermittlungs- als auch steuerungstechnische Betriebsmittel gleichermaßen berücksichtigt werden. Damit verbunden ist die Frage, welche Verkehrsparameter in diese Steuerungsfunktion eingehen sollten.

Unabhängig davon bleibt die Problematik der Identifizierung eines Dienstes durch die Überlastabwehr sowie die Auswahl geeigneter Dienstgruppen bzw. Verkehrsklassen. Gerade hier spielt auch das Kosten-zu-Nutzen-Verhältnis eine entscheidende Rolle, da ein Mehr an Information nur durch stark erhöhten Rechenzeitaufwand zu erzielen ist.

Ein weiterer Bereich stellt die Systemarchitektur einer Vermittlungssteuerung dar. Durch die nach wie vor anhaltenden Fortschritte in der Halbleitertechnologie und dem damit verbundenen Leistungszuwachs wird die Steuerungsstruktur auch zukünftig architekturellen Veränderungen unterworfen sein. Wie in der Arbeit mehrfach erwähnt, stellt der Plattformaspekt eine nicht unwesentliche Randbedingung für den Entwurf einer Überlastabwehr dar.

Ein wirklich optimales Leistungsverhalten lässt sich nur erzielen, wenn ein System in seiner Gesamtheit, d.h. hier alle Netzelemente, Netzbereiche und Endgeräte, in die Betrachtung mit eingeht.<sup>4</sup> Dies bedingt eine netzweite, globale Lastregelung für die Steuerungs-Ebene unterhalb der Zeitebene des Netzmanagements. Sinnvollerweise sollte sich deren räumlicher Wirkungskreis auch über Netzgrenzen hinwegsetzen.

Diese Problematik hat u.a. ihre Ursache darin, dass die Überlastabwehr eine nicht der Standardisierung unterworfenen, und damit eine herstellereigenspezifische Steuerungsfunktion ist. Dadurch treten in diesem Zusammenhang Kompatibilitätsprobleme zwischen Produkten verschiedener Hersteller auf und der Wunsch nach einer verbesserten Interoperabilität wird zunehmend deutlicher.

In eine ähnliche Richtung geht die Anforderung, dass vermehrt Interworking-Funktionen bereitgestellt werden müssen. Hierunter sind heterogene Vermittlungssysteme zu verstehen, die für eine Kopplung verschiedener Netztechnologien und -dienste, beispielsweise zwischen Mobilfunk- und Festnetzen, ausgelegt sind. Große Bedeutung erlangt dieses Thema derzeit durch den starken Wunsch des Vermittelns von Internet-Verkehr über ATM-Transportnetze. Auch hier kann keine isolierte Betrachtung der Überlastabwehr vorgenommen werden.

---

<sup>4</sup> In dieser Arbeit wurde die lokale Überlastabwehr für ein isoliertes Vermittlungssystem, beispielsweise ohne Rückwirkung auf die Verkehrsquellen, betrachtet.

# Literaturverzeichnis

## Lehrunterlagen und Fachbücher

- [1] – : *Das ATM-Handbuch – Grundlagen, Planung, Einsatz*. Hrsg. J. Claus, G. Siegmund, Hüthig Verlag, Heidelberg, 1995.
- [2] AHAMED, S. V.; LAWRENCE, V. B.: *Intelligent Broadband Multimedia Networks*. Kluwer Academic Publishers, Boston/London/Dordrecht, 1997.
- [3] AKIMARU, H.; KAWASHIMA, K.: *Teletraffic – Theory and Applications*. Springer-Verlag, Berlin/Heidelberg/New York, 1993.
- [4] ALI, S. R.: *Digital Switching Systems – System Reliability and Analysis*. McGraw-Hill, New York, 1998.
- [5] ALTEHAGE, G. (Hrsg.): *Digitale Vermittlungssysteme für Fernsprechen und ISDN*. R. v. Decker's Verlag, Heidelberg, 1991.
- [6] BERGMANN, K. (Begr.); SLABON, R. W. (Hrsg.): *Lehrbuch der Fernmeldetechnik*. Bd. 1, 5. Aufl., Fachverlag Schiele & Schön, Berlin, 1986.
- [7] BESIER, H.: *Digitale Vermittlungstechnik*. R. Oldenbourg Verlag, München/Wien, 1981.
- [8] BINDER-HOBACH, J.: *Das Vermittlungssystem Alcatel 1000 S12*. R. v. Decker's Verlag, G. Schenck, Heidelberg, 1993.
- [9] BOCKER, P.: *ISDN: digitale Netze für Sprach-, Text-, Daten-, Video- und Multimedia-kommunikation*. Springer-Verlag, 4. Auflage, Berlin/Heidelberg/New York, 1997.
- [10] BOUCHER, J. R.: *Voice Teletraffic Systems Engineering*. Artech House, 1988.
- [11] BRAMMER, K.; SIFFLING, G.: *Kalman-Bucy-Filter*. 4. Aufl., R. Oldenbourg Verlag, München/Wien, 1994.
- [12] BROWN, R. G.: *Smoothing, Forecasting and Prediction of Discrete Time Series*. 2. Ed., Prentice-Hall, Englewood Cliffs, N.J., 1964.
- [13] CHAPUIS, R. J.; JOEL, A. E.: *Electronics, Computers and Telephone Switching – A Book of Technological History; as Vol. 2: 1960-1985 of „100 Years of Telephone Switching“*. 'Studies in Telecommunication', Vol. 13, North-Holland, Amsterdam, 1990.
- [14] CHEN, T. M.; LIU, S. S.: *ATM Switching Systems*. Artech House, Boston/London 1995.
- [15] ÇINLAR, E.: *Introduction to Stochastic Processes*. Prentice Hall, Englewood Cliffs, New Jersey, 1975.
- [16] COOVER, E. R.: *ATM Switches*. Artech House, Boston/London 1997.
- [17] CREMER, M.: *Der Verkehrsfluß auf Schnellstraßen: Modelle, Überwachung, Regelung*. Springer-Verlag, Berlin/Heidelberg/New York, 1979.
- [18] FANTAUZZI, G.: *Digital Switching Control Architectures*. Artech House, Boston/London 1990.
- [19] FENEYROL, M.: *Telecommunication in the 21th Century – The Real and the Virtual*. Springer-Verlag, London/Berlin/Heidelberg/New York, 1997.

- [20] FILIPIAK, J.: *Modelling and Control of Dynamic Flows in Communication Networks*. Springer-Verlag, Berlin/Heidelberg, 1988.
- [21] FISHMAN, G. S.: *Principles of Discrete Event Digital Simulation*. John Wiley & Sons, New York, 1978.
- [22] GEORG, O.: *Telekommunikationstechnik: Eine praxisbezogene Einführung*. Springer-Verlag, Berlin/Heidelberg/New York, 1996.
- [23] GERKE, P. R.: *Rechnergesteuerte Vermittlungssysteme*. Springer-Verlag, Berlin/Heidelberg/New York, 1972.
- [24] GERKE, P. R.: *Digitale Kommunikationsnetze: Prinzipien, Einrichtungen, Systeme*. Springer-Verlag, Berlin/Heidelberg/New York, 1991.
- [25] GROSS, D.; HARRIS, C. M.: *Fundamentals of Queueing Theory*. 3rd Ed., John Wiley & Sons, New York, 1998.
- [26] HAFNER, K.; LYON, M.: *ARPA KADABRA – Die Geschichte des Internet*. dpunkt-Verlag, Heidelberg, 1997.
- [27] HÉBUTERNE, G.: *Traffic Flow in Switching Systems*. Artech House, Boston/London 1987.
- [28] HILLS, M. T.; KANO, S.: *Programming Electronic Switching Systems – Real-Time Aspects and their Language Implications*. Published by Peter Peregrinus Ltd., 1976.
- [29] ISERMANN, R.: *Digitale Regelsysteme. Band 1: Grundlagen, Deterministische Regelungen*. 2. Aufl., Springer-Verlag, Berlin/Heidelberg/New York, 1987.
- [30] ISERMANN, R.: *Digitale Regelsysteme. Band 2: Stochastische Regelungen, Mehrgrößenregelungen, Adaptive Regelungen, Anwendungen*. 2. Aufl., Springer-Verlag, Berlin/Heidelberg/New York/London/Paris/Tokyo, 1987.
- [31] JOEL, A. E. (ED.): *Electronic Switching System: Central Office Systems of the World*. IEEE Press, New York, 1976.
- [32] JOEL, A. E. (JR.); ET AL.: *A History of Engineering and Science in the Bell System – Switching Technology (1925 - 1975)*. Bell Telephone Laboratories, Inc. 1982.
- [33] KAMEDA, H.; LI, J.; KIM, C.; ZHANG, Y.: *Optimal Load Balancing in Distributed Computer Systems*. Springer-Verlag, Berlin, Heidelberg, New York, 1997.
- [34] KLEINROCK, L.: *Queueing Systems. Vol. 1: Theory*. John Wiley & Sons, New York/Chichester/Brisbane/Toronto, 1975.
- [35] KÜHN, P. J.: *Modellbildung von Vermittlungssteuerungen*. Unterlagen zum Hochschulkolleg ‚Methoden der Verkehrsanalyse‘, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1985.
- [36] KÜHN, P. J.: *Hochgeschwindigkeits-Kommunikationsnetze – Architekturen, Technologie und Leistung*. Vorlesung, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1995.
- [37] KÜHN, P. J.: *Rechnerkommunikation*. Unterlagen zum Hochschulkolleg, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1991.
- [38] KÜHN, P. J.: *Kommunikationstechnik. Teil 3: Vermittlungssysteme*. Unterlagen zum Hochschulkolleg, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1996.
- [39] KÜHN, P. J.: *Softwaretechnik für die Telekommunikation*. Unterlagen zur Lehrveranstal-

- tung, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1998.
- [40] KÜHN, P. J.: *Technische Informatik II – Rechnerarchitektur und Rechnernetze*. Unterlagen zur Lehrveranstaltung, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1997.
- [41] KÜHN, P. J.: *Technische Informatik III – Kommunikationsnetze*. Unterlagen zur Lehrveranstaltung, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1998.
- [42] KÜHN, P. J.: *Wartezeitprobleme der Daten- und Nachrichtenverkehrstheorie*. Vorlesung, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1994.
- [43] KÜHN, P. J.; RATH, T.; TRAN-GIA, P.: *Methodik der stationären Systemsimulation*. Vorlesungsumdruck zu [42], Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1994.
- [44] KUNZE, H.: *Anwendung der Elektronik in der Vermittlungstechnik. Teil II: Grundbegriffe der zentral gesteuerten Vermittlungssysteme*. Fachverlag Schiele & Schön, Berlin, 1971.
- [45] KYAS, O.: *ATM-Netzwerke: Aufbau – Funktion – Performance*. 3. Aufl., DATACOM-Buchverlag, Bergheim, 1996.
- [46] LAW, A. M.; KELTON, W. D.: *Simulation, Modeling & Analysis*. McGraw-Hill, 1991.
- [47] LEE, B. G.; KANG, M.; LEE, J.: *Broadband Telecommunications Technology*. Artech House, Boston/London 1993.
- [48] MALONE, M. S.: *The Microprocessor: A Biography*. Springer-Verlag, New York, 1995.
- [49] McDONALD, J. C. (ED.): *Fundamentals of Digital Switching*. Plenum Press, 2nd Edition, New York/London 1990.
- [50] ONVURAL, R. O.: *Asynchronous Transfer Mode Networks: Performance Issues*. Artech House, Boston/London 1994.
- [51] ONVURAL, R. O.; CHERUKURI, R.: *Signaling in ATM Networks*. Artech House, Boston/London 1997.
- [52] PAUL, H. A. (Hrsg.): *Analoge Vermittlungstechnik für Telefonverkehr einschließlich Sonderdienste und Breitbandanwendungen*. R. v. Decker's Verlag, G. Schenck, Heidelberg, 1990.
- [53] PEARL, J.: *Heuristics – Intelligent Search Strategies for Computer Problem Solving*. Addison-Wesley, 1984.
- [54] RATHGEB, E.; WALLMEIER, E.: *ATM – Infrastruktur für die Hochleistungskommunikation*. Springer-Verlag, Berlin/Heidelberg/New York, 1997.
- [55] RENOARD, H. E., VON: *Fachwörterbuch elektronische Medien und Dienste*. Springer-Verlag, Berlin/Heidelberg/New York, 1997.
- [56] SCHRICK, K.-W. (Hrsg.): *Anwendungen der Kalman-Filter-Technik – Anleitung und Beispiele*. R. Oldenbourg Verlag, München/Wien, 1977.
- [57] SIEGMUND, G.: *Grundlagen der Vermittlungstechnik*. 2. Aufl., R. v. Decker's Verlag, Heidelberg, 1992.
- [58] SIEGMUND, G.: *ATM – Die Technik des Breitband-ISDN*. R. v. Decker's Verlag, 2. Aufl.,

Heidelberg, 1994.

- [59] TRAN-GIA, P.: *Simulation instationärer Prozesse*. Hochschulkolleg ‚Methoden der Verkehrsanalyse‘, Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1984.

## Veröffentlichungen und wissenschaftliche Arbeiten

- [60] – : Special Issue on Congestion Control in Computer Networks. *IEEE Transactions on Communications*, Vol. 29, No. 4, April 1981.
- [61] – : Packet/ATM. *Bell Labs Technology – Trends & Developments*, Vol. 1, No. 3, December 1997, pp. 12.
- [62] ABE, T.; SUMIMOTO, J.-I.: Standardization of Internet Performance Metrics. *NTT REVIEW*, Vol. 10, No. 2, March 1998, pp. 50-54.
- [63] ABRAHAM, J.: *Leistungsregelung in einem Vermittlungsrechner*. Diplomarbeit Nr. 1538 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, März 1998.
- [64] AHLFORS, U.; NYBERG, C.: Modelling Optimal Overload Control in Distributed Control Systems. *12th International Conference on Computer Communication*, Seoul, Korea, August 1995, pp. 583-588.
- [65] AHLFORS, U.; NYBERG, C.: Near-Optimal Load Control of a Message-Passing Multiprocessor System: A Decentralized Markov Decision Approach. *15th International Teletraffic Congress*, Washington, June 1997, pp. 583-592.
- [66] ALCATEL: IP@ATM™ – IP Networking in the WAN with the 1100HSS®. *Alcatel White Paper*, 1998. → <http://www.alcatel.com/telecom/edd/publicat/whitpap/ipatm.pdf>
- [67] ANDRIES, R.: Main Algorithms used in Comprehensive Overload Control Strategy in ALCATEL 1000 S12. *1st St. Petersburg Regional International Teletraffic Seminar*, St. Petersburg, Russia, June 1993, pp. 177-187.
- [68] ANGELIN, L.; ARVIDSSON, Å.: A Congestion Control Mechanism for Signalling Networks based on a State Machine Controlled by Network Delays. *International Teletraffic Seminar*, Bangkok, Nov./Dec. 1995, Paper #31.
- [69] ARITAKA, N.; ITO, K.; NIHEI, M.: ATM Multimedia System for Private Network. *FUJITSU Scientific & Technical Journal*, Vol. 32, No. 1, June 1996, pp. 46-52.
- [70] ARLITT, M. F.; WILLIAMSON, C. L.: Internet Web Servers: Workload Characterization and Performance Implications. *IEEE/ACM Transactions on Networking*, Vol. 5, No. 5, October 1997, pp. 631-645.
- [71] ARVIDSSON, Å.; PETERSSON, S.; ANGELIN, L.: Congestion Control in Intelligent Networks for Real Time Performance and Profit Optimisation. *10th ITC Specialists Seminar on Control in Communications*, Lund, Sweden, September 1996, pp. 347-358.
- [72] ARVIDSSON, Å.; PETERSSON, S.; ANGELIN, L.: Profit Optimal Congestion Control in Intelligent Networks. *15th International Teletraffic Congress*, Washington, June 1997, pp. 911-920.
- [73] VAN AS, H.: *Modellierung und Analyse von Überlast-Abwehrmechanismen in Paketver-*

- mittlungsnetzen*. 38. Bericht über verkehrstheoretische Arbeiten (Dissertationsschrift), Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1984.
- [74] ASH, G. R.; CHEN, J.; FISHMAN, S. D.; ZERLING, D.: Real-Time Inter-Network Routing in the AT&T Network. *15th International Teletraffic Congress (ITC '15)*, Vol. 2, Washington, June 1997, pp. 1055-1066.
- [75] ATAI, A.; GORDON, J.: Impacts of Internet Traffic on LEC Networks and Switching Systems. *Bellcore*, White Paper 00C-1013, June 1996.
- [76] AXELL, J.; HELLSTRAND, F.: ATM Traffic Management and Resource Optimization. Ericsson Review, *The Telecommunications Technology Journal*, Vol. 75, No. 1, 1998, pp. 18-23.
- [77] AZMOODEH, M.; DAVISON, R.: Performance Management of Public ATM Networks – A Scaleable and Flexible Approach. *Proceedings of IEEE*, Vol. 85, No. 10, October 1997, pp. 1639-1645.
- [78] BADEL, M.; GELENBE, E.; LEROUDIER, J.; POTIER, D.: Adaptive Optimization of a Time-Sharing System's Performance. *Proceedings of the IEEE*, Vol. 63, No. 6, June 1975, pp. 958-965.
- [79] BAKSHI, Y.; DIAZ, A. H.; MEIER-HELLSTERN, K.; MILITO, R. A.; SKOOG, R.: Overload Control in a Distributed System. *15th International Teletraffic Congress*, Washington, DC, June 1997, pp. 571-582.
- [80] BALBONI, G. P.; MINERVA, R.; PAGLIALUNGA, A.; PELAGGI, A.; MONTESI, S.: From Early ATM Systems to Advanced Switching: Issues in Control Architecture. *15th International Switching Symposium (ISS '95)*, Vol. 1, Berlin, April 1995, pp. 474-478.
- [81] BASKETT, F.; CHANDY, K. M.; MUNTZ, R. R.; PALACIOS, F. G.: Open, Closed, and Mixed Networks of Queues with Different Classes of Customers. *Journal of the ACM*, Vol. 22, No. 2, April 1975, pp. 244-260.
- [82] BASU, K.; GORMAN, G.; AVELLANEDA, O.; MACTAVISH, N.: A Real-Time Simulator for Performance Evaluation and Overload Control Design of an SPC System. *10th International Teletraffic Congress*, Montreal, June 1983, Paper 5.2-#9.
- [83] BECKER, D.: Einsatz von Mikrorechnern in verteilten Steuerungen von Nachrichtensystemen. *Elektrisches Nachrichtenwesen*, Band 52, Nr. 3, 1977, S. 231-236.
- [84] BECKER, D.; SCHMIDT, P.: Zur Integration von Fernmeldediensten in digitalen Netzen – Teil 1: Digitale Netze und Vermittlungsstellen. *Nachrichtentechnische Zeitschrift (ntz)*, Bd. 34, Nr. 5, Mai 1981, S. 288-292.
- [85] BECKER, D.; SAVARY, H. F.; VERHILLE, H.: Architektur des Breitbandvermittlungssystems von Alcatel. *Elektrisches Nachrichtenwesen*, Band 64, Nr. 2/3, 1990, S. 147-155.
- [86] BELLCORE: Broadband Switching System Generic Requirements. GR-1110-CORE, Issue 1, July 1995.
- [87] Bellovin, S.: Stock-Market Overloads. *ACM SIGSOFT, Software Engineering Notes*, Vol. 19, Issue 44, Nov. 1997. Bzw.: <http://catless.ncl.ac.uk/Risks/19.44.html> (Forum on Risks to the Public in Computers and Related Systems).
- [88] BENNETT, R. L.; POLICELLO, G. E.: Switching Systems in the 21st Century. *IEEE Communications Magazine*, Vol. 31, No. 3, March 1993, pp. 24-28.
- [89] BERGER, A. W.: Determination of Load-Service Curves for Distributed Switching Systems: Probabilistic Analysis of Overload Control Schemes. *13th International Tele-*



- traffic Congress, Proceedings in 'North-Holland Studies in Telecommunication', Vol. 14, Copenhagen, June 1991, pp. 435-440.*
- [90] BERMEJO, L.; PARMENTIER, P.; PETIT, G. H.: Dienstcharakteristiken und Verkehrsmodelle in einem Breitband-ISDN. *Elektrisches Nachrichtenwesen*, Band 64, Nr. 2/3, 1990, S. 132-138.
- [91] BERNDT, H.; STORM, J.: Zentralsteuerwerk SSP103D im Digitalvermittlungssystem EWSD. *telcom report*, 4. Jhrg., Nr. 6/81, Dezember 1981, S. 455-460.
- [92] BERNDT, H.: SSP112D — Mikroelektronisches Zentralsteuerwerk für Digitalkommunikationssysteme. *telcom report*, 4. Jhrg., Beiheft *Digitalvermittlungssystem EWSD*, 1981, S. 33-37.
- [93] BHARATH-KUMAR, K.; JAFFE, J. M.: A New Approach to Performance-Oriented Flow Control. *IEEE Transactions on Communications*, Vol. 29, No. 4, April 1981, pp. 427-435.
- [94] BLAIOTTA, D.; ET AL.: Signalling/IN Server Architecture for a Broadband SSP. *4th IFIP International Conference on Broadband Communications (BC '98)*, Stuttgart, April 1998, pp. 171-182.
- [95] BLANKERS, P.; HU, Y.: Broadband Capabilities within Intelligent Networks. *5th IEEE Intelligent Network Workshop (IN '96)*, Melbourne, Vol. 2, April 1996, Paper #9.2.1.
- [96] BLAU, S.; ROTH, J.: AXD 301 – A New Generation ATM Switching System. *Ericsson Review, The Telecommunications Technology Journal*, Vol. 75, No. 1, 1998, pp. 10-17.
- [97] BODAMER, S.: *Mittlere Durchlaufzeiten in gesättigten M/G/1-Prioritätensystemen mit Rückkopplungen*. Diplomarbeit Nr. 1316 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, August 1994.
- [98] BODAMER, S.; RENGER, T.; RÖBLER, G.: Traffic Management in an ATM Multi-Service Switch for Workgroups. *4th IFIP International Conference on Broadband Communications (BC '98)*, Stuttgart, April 1998, pp. 341-355.
- [99] BOLLA, R.; DAVOLI, F.; MARCHESE, M.: Bandwidth Allocation and Admission Control in ATM Networks with Service Separation. *IEEE Communications Magazine*, Vol. 35, No. 5, May 1997, pp. 130-137.
- [100] BOLOTIN, V. A.: Modelling Call Holding Time Distributions for CCS Network Design and Performance Analysis. *IEEE Journal on Selected Areas in Communication*, Vol. 12, No. 3, April 1994, pp. 433-438.
- [101] BOLOTIN, V. A.: New Subscriber Traffic Variability Patterns for Network Traffic Engineering. *15th International Teletraffic Congress (ITC '15)*, Vol. 2, Washington, June 1997, pp. 867-878.
- [102] BONAMI, B.; COTTON, J. M.; DENENBERG, J. N.: SYSTEM 12 – Systemstruktur. *Elektrisches Nachrichtenwesen*, Band 56, Nr. 2/3, 1981, S. 126-134.
- [103] BONAMI, R.; CAGNAC, T.; YELLOZ, G.: Der Rechner ITT 3202. *Elektrisches Nachrichtenwesen*, Band 52, Nr. 2, 1977, S. 104-109.
- [104] BONATTI, M.; ROVERI, A.: A Traffic Model for Design Choice of ISDN System Architectures. *4th ITC Specialists Seminar on ISDN Traffic Issues*, Brussels, May 1986.
- [105] BONATTI, M.; GIACOBBO SCAVO, G.; ROVERI, A.; VERRI, L.: Terminal Exchange Access System for NB-ISDN: Key Issues for a Traffic Reference Model. *12th International Teletraffic Congress, Proceedings in 'North-Holland Studies in Telecommunication',*

- Vol. 12, Torino, June 1988, pp. 702-711.
- [106] BORCHEING, J. W.; FORYS, L. J.; FREDERICH, A. A.; HEJNY, G. J.: Coping with Overload. *Bell Laboratories Record*, Vol. 59, No. 6, July/August 1981, pp. 183-185.
- [107] BOTSCH, D.: Eigenschaften von Prozessen der Vermittlungstechnik. *telefon report*, 12. Jhrg., Nr. 4, Dezember 1976, S. 160-165.
- [108] BOTSCH, D.: Die EWSD-Software — Ein Realzeitprogrammiersystem in der höheren Programmiersprache CHILL. *telcom report*, 2. Jhrg., Nr. 3/79, 1979, S. 184-189.
- [109] BOTSCH, D.: Das System EWSD. *telcom report*, 4. Jhrg., Beiheft *Digitalvermittlungssystem EWSD*, 1981, S. 7-12.
- [110] BOZYIGIT, M.; MELHI, M.: Load Balancing Framework for Distributed Systems. *International Journal of Computer Systems, Science & Engineering*, Vol. 12, No. 5, September 1997, pp. 287-293.
- [111] BRETECHER, Y.; VILAIN, B.: The Intelligent Network in a Broadband Context. *15th International Switching Symposium (ISS'95)*, Vol. 2, Berlin, April 1995, pp. 57-61.
- [112] BRICCOLI, A.: Comparison of Regulation Methods for Traffic Overloads in SPC Systems. *9th International Teletraffic Congress*, Torremolinos, October 1979, Session 53.
- [113] BRICCOLI, A.; ET AL.: Performance Design of a Distributed Switching System. *12th International Teletraffic Congress, Proceedings in 'North-Holland Studies in Telecommunication'*, Vol. 12, Torino, June 1988, pp. 292-301.
- [114] BROUX, J. A.: Erste Erfahrungen mit dem Fernsprechvermittlungssystem 10C. *Elektrisches Nachrichtenwesen*, Band 43, Nr. 4, 1968, S. 330-336.
- [115] BUHRGARD, M.; WESTIN, T.; WICKLUND, G.: A Scaleable ATM Switching Solution. *Ericsson Review, The Telecommunications Technology Journal*, Vol. 75, No. 1, 1998, pp. 24-31.
- [116] BURKE, P. J.: Automatic Overload Controls. *5th International Teletraffic Congress*, New York, June 1967, p. 214.
- [117] BURNS, J. E.; GHOSAL, D.: Automated Detection and Control of Media Stimulated Focused Overloads. *15th International Teletraffic Congress*, Washington, June 1997, pp. 889-900.
- [118] CARBAUGH, D. H.; DREW, G. G.; GHIRON, H.; HOOVER, E. S.: No. 1 ESS Call Processing. *The Bell System Technical Journal*, Vol. XLIII, No. 5, Part 2, September 1964, pp. 2483-2531.
- [119] CARL, D.: Erweiterung existierender Fernsprechnetze zu Intelligenten Netzen. *Tutorium zur KiVS '93, Kommunikation in verteilten Systemen*, München, März 1993, S. 91-106.
- [120] CARMAGNOLA, V.; FERRETTI, M.; SCARRONE, E.: An Integrated IN/B-ISDN Reference Architecture for the Support of Multimedia Services. *5th IEEE Intelligent Network Workshop (IN'96)*, Melbourne, Vol. 2, April 1996, Paper #5.2.2.
- [121] CARRUET, V.; RIDEAU, A.: Digital-Vermittlungsstelle ITT1220 – Software-Struktur und -Methodik. *Elektrisches Nachrichtenwesen*, Band 54, Nr. 3, 1979, S. 199-207.
- [122] CESAR, B.: *Methoden zur zeitreuen Simulation instationärer Prozesse in verkehrstheoretischen Untersuchungen*. Semesterarbeit Nr. 795 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, August 1985.

- [123] CHAO, H. J.; CHOE, B.-S.; PARK, J.-S.; UZUN, N.: Design and Implementation of Abacus Switch: A Scalable Multicast ATM Switch. *IEEE Journal on Selected Areas in Communication*, Vol. 15, No. 5, June 1997, pp. 830-843.
- [124] CHATTERJEE, S.; BASSIOUNI, M. A.: Congestion Free Multimedia Communications over Broadband ATM Networks. *2nd International Conference on Broadband Islands*, Athens, June 1993, pp. 235-246.
- [125] CHEMOUIL, P.; FILIPIAK, J.: Kalman Filtering of Traffic Fluctuations for Real-Time Network Management. *Annales des Télécommunications*, Vol. 44, No. 11-12, November/December 1989, pp. 633-640.
- [126] CHEN, T. M.; LIU, S. S.: Management and Control Functions in ATM Switching Systems. *IEEE Network Magazine*, Vol. 8, No. 4, July/August 1994, pp. 27-40.
- [127] CHEN, X.; TANTIPRASUT, D. T.: Evolution of ATM Interworking. *Bell Labs Technical Journal*, Vol. 2, No. 2, Spring 1997, pp. 82-110.
- [128] CHEN, X.: Scalable Call Processing Architectures for ATM Switches. *Bell Labs Technical Journal*, Vol. 3, No. 1, January-March 1998, pp. 62-80.
- [129] CHIU, D.-M.; JAIN, R.: Analysis of the Increase and Decrease Algorithms for Congestion Avoidance in Computer Networks. *Computer Networks and ISDN Systems*, Vol. 17, No. 1, 1989, pp. 1-14.
- [130] CHOW, Y.-C.; KOHLER, W. H.: Models for Dynamic Load Balancing in a Heterogeneous Multiple Processor System. *IEEE Transactions on Computer*, Vol. 28, No. 5, May 1979, pp. 354-361.
- [131] CIDON, I.; GOPAL, I.; GUÉRIN, R.: Bandwidth Management and Congestion Control in plaNET. *IEEE Communications Magazine*, Vol. 29, No. 10, October 1991, pp. 54-64.
- [132] COAN, B.A.; HEYMAN, D.: Reliable Software and Communication III: Congestion Control and Network Reliability. *IEEE Journal on Selected Areas in Communication*, Vol. 12, No. 1, January 1994, pp. 40-45.
- [133] COHEN, J. W.: Basic Problems of Telephone Traffic Theory and the Influence of Repeated Calls. *Philips Telecommunication Review*, Vol. 18, 1957, pp. 49-100.
- [134] COUTURIER, G. W.; THOMPSON, A. W.: Distributed Processing Control Complexes for Telephone Switching Systems. *13th International Conference on Communications (ICC '77)*, Chicago, June 1977, Paper #20.5, pp. 71-75.
- [135] CSELÉNYI, I.; SZABÓ, R.; SZABÓ, I.; LATOUR-HENNER, A.; BJÖRKMAN, N.: Experimental Platform for Telecommunication Resource Management. *Computer Communications*, Vol. 21, No. 17, November 1998, pp. 1624-1640.
- [136] CZERNEK, M.: *Untersuchungen an einem Breitband-Vermittlungssystem im Intelligenzen Netz*. Diplomarbeit Nr. 1537 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, März 1998.
- [137] DAISENBERGER, G.; OEHLERICH, J.; WEGMANN, G.: STATOR – STATistical Overload Regulation – and TAIL – Time Account Input Limitation – Two Concepts for Overload Regulation in SPC Systems. *11th International Teletraffic Congress*, Kyoto, September 1985, Paper #2.1B-4.
- [138] DAISENBERGER, G.; OEHLERICH, J.; WEGMANN, G.: Two Concepts for Overload Regulation in SPC Switching Systems: STATOR and TAIL. *TELECOMMUNICATION JOURNAL*, Vol. 56, No. 5, 1989, pp. 306-313.

- [139] DARTOIS, J. P.; DAVIES, A. C.; TRELUT, J.; VAN OS, L.; VERBEECK, M.: Zehn Jahre Betriebserfahrung mit dem System METACONTA. *Elektrisches Nachrichtenwesen*, Band 52, Nr. 2, 1977, S. 110-118.
- [140] DAS, S.; STRUNK, K.; VERSTRAETE, F.: SYSTEM 12 – Hardware-Struktur. *Elektrisches Nachrichtenwesen*, Band 56, Nr. 2/3, 1981, S. 135-147.
- [141] DASCH, B.: Überlaststrategien in der Paketvermittlung am Beispiel des Systems EWSP. *Kommunikation in verteilten Systemen*, ITG/GI-Fachtagung, Stuttgart, Februar 1989, S. 370-384.
- [142] DAVIES, D. W.: The Control of Congestion in Packet-Switching Networks. *IEEE Transactions on Communication*, Vol. 20, No. 3, June 1972, pp. 546-550.
- [143] DAVIS, M. E.; PEHLERT, W. K.: Evaluation of No. 1 PSS Congestion Control Strategies. *IFIP WG7.3/TC6 2nd International Symposium on the Performance of Computer-Communication Systems*, Zurich, Switzerland, March 1984, pp. 449-461.
- [144] DELODDERE, D.: Evolution of a Narrowband Exchange towards Broadband. *15th International Switching Symposium (ISS '95)*, Vol. 1, Berlin, April 1995, pp. 196-200.
- [145] DE LOS MOZOS MARQUES, J. R.; BUCHHEISTER, A.: SYSTEM 12 – Verkehrs- und Belegungskapazität. *Elektrisches Nachrichtenwesen*, Band 56, Nr. 2/3, 1981, S. 207-217.
- [146] DENZEL, W.: *Untersuchungen zu Vermittlungssystemen mit verteilter Steuerung*. 42. Bericht über verkehrstheoretische Arbeiten (Dissertationsschrift), Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1986.
- [147] DEUSSING, P.; ECKHARDT, B.: The Distributed 1PSS Architecture: A High Reliable Switch for High-Performance Packet Switching Network. *8th International Conference on Computer Communications (ICCC '86)*, Munich, September 1986, pp. 353-358.
- [148] DEVARAKONDA, M. V.; IYER, R. K.: Predictability of Process Resource Usage: A Measurement-Based Study on UNIX. *IEEE Transactions on Software Engineering*, Vol. 15, No. 12, December 1989, pp. 1579-1586.
- [149] DIETRICH, G.: Verkehrsmodell für zentralgesteuerte Vermittlungssysteme. *Elektrisches Nachrichtenwesen*, Band 50, Nr. 1, 1975, S. 30-36. Bzw.: Telephone Traffic Model for Common Control Investigations. *7th International Teletraffic Congress (ITC-7)*, Vol. 1, Stockholm, June 1973, Paper #331.
- [150] DIETRICH, G.; SALADE, R.: Teilbelegungstreue Simulation der Steuerung von Vermittlungssystemen. *Elektrisches Nachrichtenwesen*, Band 52, Nr. 1, 1977, S. 61-68.
- [151] DOSHI, B. T.; HEFFES, H.: Analysis of Overload Control Schemes for a Class of Distributed Switching Machines. *10th International Teletraffic Congress*, Montreal, June 1983, Paper 5.2-#2.
- [152] DOSHI, B. T.; HEFFES, H.: Overload Performance of Several Processor Queueing Disciplines for the M/M/1 Queue. *IEEE Transactions on Communication*, Vol. 34, No. 6, June 1991, pp. 538-546.
- [153] DOSHI, B. T.; DRAVIDA, S.; JOHRI, P.; RAMAMURTHY, G.: Memory, Bandwidth, Processing and Fairness Considerations in Real Time Congestion Controls for Broadband Networks. *13th International Teletraffic Congress, Proceedings in 'North-Holland Studies in Telecommunication'*, Vol. 14, Copenhagen, June 1991, pp. 153-159.
- [154] DOSHI, B. T.; JOHRI, P. K.: Communication Protocols for High Speed Packet Networks. *Computer Networks and ISDN Systems*, Vol. 24, 1992, pp. 243-273.

- [155] DOUGLAS JOHNSON, G.: No. 4 ESS – Long Distance Switching for the Future. *Bell Laboratories Record*, Vol. 51, September 1973, pp. 226-232.
- [156] DRWIEGA, T.: Shared Memory Contention and its Impact on Multi-Processor Call Control Throughput. *15th International Teletraffic Congress*, Vol. 1, Washington, June 1997, pp. 593-602.
- [157] DUMORTIER, P.: Verfahren zur Verbesserung des Datendurchsatzes im Internet. *Alcatel Telecom Rundschau*, 4. Quartal, 1997, S. 300-306.
- [158] DUMORTIER, P.: Toward a New IP and ATM Routing Paradigm. *IEEE Communications Magazine*, Vol. 36, No. 1, January 1998, pp. 82-86.
- [159] DUNCAN, T.; HUEN, W. H.: Software Structure of No. 5 ESS — A Distributed Telephone Switching System. *IEEE Transactions on Communications*, Vol. 30, No. 6, June 1982, pp. 1379-1385.
- [160] DUTKIEWICZ, E.; ANIDO, G.: Performance Evaluation, Optimisation and Load Balancing in Distributed Switch Control Systems. *Proc. of the Australian Telecommunication Networks & Applications Conference*, Melbourne, December 1994, pp. 731-736.
- [161] DUTKIEWICZ, E.; ANIDO, G.: Performance Evaluation of Hierarchical Control Systems in Switching Exchanges. *Australian Telecommunications Research Journal*, Vol. 28, No. 1, 1994, pp. 15-24.
- [162] DUTKIEWICZ, E.; ANIDO, G.: Optimisation of Distributed Exchange Control Systems. *Australian Telecommunications Research Journal*, Vol. 29, No. 2, 1995, pp. 1-13.
- [163] DUTKIEWICZ, E.: *Modelling and Analysis of Hierarchical Control Systems Operating in a Multi-Service Telecommunications Environment*. Ph.D. Thesis, University of Wollongong, Australia, February 1996.
- [164] DZIONG, Z.; MONTANUY, O.; MASON, L. G.: Adaptive Traffic Admission in ATM Networks – Optimal Estimation Framework. *14th International Teletraffic Congress*, Vol. 2, Antibes Juan-les-Pins, June 1994, pp. 1065-1076.
- [165] EAGER, D. L.; LAZOWSKA, E. D.; ZAHORJAN, J.: Adaptive Load Sharing in Homogeneous Distributed Systems. *IEEE Transactions on Software Engineering*, Vol. 12, No. 5, May 1986, pp. 662-675.
- [166] EARLE VAUGHAN, H.: Development History of No. 1 ESS Call – Software. *IEE Conference on Switching Techniques for Telecommunications Networks*, London, April 1969, pp. 475-478.
- [167] EBERDING, H.: Die Software im System EWSD. *telcom report*, 4. Jhrg., Beiheft *Digitalvermittlungssystem EWSD*, 1981, S. 13-18.
- [168] ECKBERG, A. E.; WIRTH, P. E.: Switch Overload and Flow Control Strategies in an ISDN Environment. *5th ITC Seminar on Traffic Engineering for ISDN Design and Planning, Proceedings in 'North-Holland Studies in Telecommunication'*, Vol. 9, Lake Como, Italy, May 1987, pp. 425-434.
- [169] EISENBERG, M.: A Strict Priority Queueing System with Overload Control. *10th International Teletraffic Congress*, Montreal, June 1983, Paper 1.3-#2.
- [170] ENDICOTT, D.; FROST, V.: Performance Experiences with Wide Area High-Speed Networks. *IEEE Communications Magazine*, Vol. 35, No. 8, August 1997, pp. 106-108.
- [171] ENG, K. Y.; PASHAN, M. A.: Advances in Shared-Memory Designs for Gigabit ATM Switching. *Bell Labs Technical Journal*, Vol. 2, No. 2, Spring 1997, pp. 175-187.

- [172] ERRAMILI, A.; FORYS, L. J.: Oscillations and Chaos in a Flow Model of a Switching System. *IEEE Journal on Selected Areas in Communication*, Vol. 9, No. 2, February 1991, pp. 171-178.
- [173] ERRAMILI, A.; FORYS, L. J.: Traffic Synchronization Effects in Teletraffic Systems. *13th International Teletraffic Congress, Proceedings in 'North-Holland Studies in Telecommunication'*, Vol. 14, Copenhagen, June 1991, pp. 201-206.
- [174] FAIß, J.: *Modellierung und Leistungsuntersuchung des Steuerrechners einer B-ISDN-Vermittlungsstelle*. Diplomarbeit Nr. 1226 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, Juli 1997.
- [175] FALCHI, M.: IP and ATM Integration: QoS Issues with Enhanced Buffer Scheduling. *16th International Switching Symposium (ISS'97)*, Vol. 1, Toronto, September 1997, pp. 459-464.
- [176] FANG, R.; HANSON, R.: AVANZA: Punkt-zu-Punkt-Lösungen von Alcatel. *Alcatel Telecom Rundschau*, 4. Quartal, 1996, S. 257-264.
- [177] FAREL, R. A.; GAWANDE, M.: Design and Analysis of Control Strategies for Transaction Network Databases. *13th Int. Teletraffic Congress, Proceedings in 'North-Holland Studies in Telecommunication'*, Vol. 14, Copenhagen, June 1991, pp. 115-120.
- [178] FERNOW, J. P.; EL-SAYED, M. L.: Stability of Adaptive Congestion Controls in Packet Networks. *Computer Networks and ISDN Systems*, Vol. 10, 1985, pp. 7-18.
- [179] FICHE, G.; LE CORGUILLE, D.; LE PALUD, C.: Traffic Models for an ISDN Switching System. *12th International Teletraffic Congress, Proceedings in 'North-Holland Studies in Telecommunication'*, Vol. 12, Torino, June 1988, pp. 687-694.
- [180] FISCHER, W.; GOELDNER, E.-H.: Performance of the ISDN D-Channel Protocol – A Simulation Study. *8th International Conference on Computer Communications*, Munich, September 1986, pp. 612-618.
- [181] FISCHER, W.; FUNDNEIDER, O.; GOELDNER, E.-H.; LUTZ, K. A.: A Scalable ATM Switching System Architecture. *IEEE Journal on Selected Areas in Communication*, Vol. 9, No. 8, October 1991, pp. 1299-1307.
- [182] FLINNER, U.: *Erweiterung eines Simulationsprogrammes für gekoppelte Netze auf instationäre Simulation*. Semesterarbeit Nr. 996 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, Juni 1990.
- [183] FOLKESTAD, A.; EMSTAD, P. J.: A Token Based SDP Load Control Scheme. *10th ITC Specialists Seminar on Control in Communications*, Lund, Sweden, September 1996, pp. 151-162.
- [184] FOLKESTAD, A.; OSLAND, P.-O.; EMSTAD, P. J.: On Load Control for Service Data Points. *4th International Conference on Intelligence in Networks*, Bordeaux, November 1996, pp. 217-222.
- [185] FONTANA, B.; VILLÉN-ALTAMIRANO, PETIT, G. H.: Modelle und Verfahren zur Untersuchung der Verkehrsleistung von System-12-ISDN-Vermittlungsstellen. *Elektrisches Nachrichtenwesen*, Band 61, Nr. 1, 1987, S. 104-109.
- [186] FORYS, L. J.: Modelling of SPC Switching Systems. *1st ITC Seminar on Modelling of SPC Exchanges and Data Networks*, Delft, 1977, pp. 83-100.
- [187] FORYS, L. J.: Performance Analysis of a New Overload Strategy. *10th International Teletraffic Congress*, Montreal, June 1983, Paper 5.2-#4.

- [188] FORYS, L. J.: New Overload Issues in a Divested Environment. *11th International Teletraffic Congress*, Kyoto, September 1985, Paper #2.1B-5.
- [189] FORYS, L. J.; IM, C. S.; HENDERSON, W.: Analysis of Load Box Testing for Voice Switches. *12th International Teletraffic Congress, Proceedings in 'North-Holland Studies in Telecommunication'*, Vol. 12, Torino, June 1988, pp. 334-340.
- [190] FOWLER, H. J.; LELAND, W. E.: Local Area Network Traffic Characteristics, with Implications for Broadband Network Congestion Management. *IEEE Journal on Selected Areas in Communication*, Vol. 9, No. 7, September 1991, pp. 1139-1149.
- [191] FRANZ, R.; GRADISCHNIG, K. D.: ATM-based SS7 for Narrowband Networks – A Step towards Narrowband-Broadband Convergence. *16th International Switching Symposium (ISS '97)*, Vol. 2, Toronto, September 1997, pp. 3-10.
- [192] FRATTA, L.; GALLASSI, G.; MUSUMECI, L., VERRI, L.: Congestion Control Strategies in ATM Networks. *European Transactions on Telecommunications*, Vol. 3, No. 2, May/April 1992, pp. 183-193.
- [193] FRENCH, K. J.: Die Einführung des Elektronischen Fernsprech-Wählsystems EWS im Fernsprechnet der Deutschen Bundespost. *telefon report*, 13. Jhrg., Nr. 4, Dezember 1977, S. 140-145.
- [194] FUHRMANN, S.; KOGAN, Y.; MILITO, R. A.: An Adaptive Autonomous Network Congestion Controller. *10th ITC Specialists Seminar on Control in Communications*, Lund, Sweden, September 1996, pp. 337-346.
- [195] GALLASSI, G.; RIGOLIO, G., VERRI, L.: Resource Management and Dimensioning in ATM Networks. *IEEE Network Magazine*, Vol. 4, No. 3, May 1990, pp. 8-17.
- [196] GALLETI, M.; GROSSINI, F.: Performance Simulation of Congestion Control Mechanisms for Intelligent Networks. *International Zurich Seminar on Digital Communications – Intelligent Networks and their Applications*, Zurich, March 1992, pp. 391-406.
- [197] GAO, K. Z.; BASU, K.; RACZ, S.: An Overload Control Algorithm in a Wireless Switch. *15th Int. Switching Symposium (ISS '95)*, Vol. 2, Berlin, April 1995, pp. 142-146.
- [198] GAVIOUS, A.; ROSBERG, Z.: A Restricted Complete Sharing Policy for a Stochastic Knapsack Problem in B-ISDN. *IEEE Transactions on Communications*, Vol. 42, No. 7, July 1994, pp. 2375-2379.
- [199] GEISBUSCH, L.: *Überlastabwehr in einem B-ISDN-Vermittlungsrechner*. Semesterarbeit Nr. 1522 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, Dezember 1997.
- [200] GELENBE, E.; MANG, X.; ÖNVURAL, R.: Bandwidth Allocation and Call Admission Control in High Speed Networks. *IEEE Communications Magazine*, Vol. 35, No. 5, May 1997, pp. 122-129.
- [201] GELENBE, E.; KOTIA, S.; KRAUSS, D.: Call Establishment Overload in Large ATM Networks. *IEEE ATM '97 Workshop*, Lisboa, Portugal, May 1997, pp. 560-569 bzw. in *Performance Evaluation*, Vol. 31, No. 1-2, November 1997, pp. 33-49.
- [202] GENDA, K.; YAMANAKA, N.: TORUS: Terabit-per-Second ATM Switching System Architecture Based on Distributed Internal Speed-Up ATM Switch. *IEEE Journal on Selected Areas in Communication*, Vol. 15, No. 5, June 1997, pp. 817-829.
- [203] GERKE, P.: System IV — Konzeption für ein Vermittlungssystem mit gespeichertem Steuerprogramm. *Informationen Fernsprech-Vermittlungstechnik*, Siemens, 6. Jhrg.,

- Nr. 1/2, 1970, S. 11-21.
- [204] GERLA, M.; KLEINROCK, L.: Flow Control: A Comparative Survey. *IEEE Transactions on Communications*, Vol. 33, No. 10, October 1985, pp. 1058-1066.
- [205] GERLA, M.; KLEINROCK, L.: Congestion Control in Interconnected LANs. *IEEE Network Magazine*, Vol. 2, No. 1, January 1988, pp. 72-76.
- [206] GERSHT, A.; LEE, K. J.: A Congestion Control Framework for ATM Networks. *IEEE Journal on Selected Areas in Communications*, Vol. 9, No. 7, September 1991, pp. 1119-1129
- [207] GIBBENS, R. J.; KELLY, F. P.; KEY, P. B.: A Decision-Theoretic Approach to Call Admission Control in ATM Networks. *IEEE Journal on Selected Areas in Communication*, Vol. 13, No. 6, August 1995, pp. 1101-1114.
- [208] GIESSLER, A.; HÄNLE, J.; KÖNIG, A.; PADE, E.: Free Buffer Allocation – An Investigation by Simulation. *Computer Networks*, Vol. 1, No. 3, July 1978, pp. 191-208.
- [209] GINZBOORG, P.: Methods to Synchronize the IN SCP's Overload Protection Mechanism. *IFIP TC6 Conference on Intelligent Networks and New Technologies*, Lyngby, Denmark, August 1995, pp. 155-174.
- [210] GITTEN, L. J.; JANIK, J.; PRELL, E. M.; JOHNSON, J. L.: 5ESS™ System Evolution. *11th International Switching Symposium (ISS '84)*, Florence, Italy, May 1984, Paper #41A1.
- [211] GLYNN, P. W.; WHITT, W.: Indirect Estimation via  $L = \lambda W$ . *Operations Research*, Vol. 37, No. 1, January/February 1983, pp. 82-103.
- [212] GOLDA, O.: Interactive Broadband Applications: The Role of IN. *5th IEEE Intelligent Network Workshop (IN '96)*, Melbourne, Vol. 2, April 1996, Paper #9.2.2.
- [213] GONZÁLEZ SOTO, O.; GARCÍA HIGUERA, J. A.; DÍAZ BERZOSA, C.; MARTÍNEZ MÍGUEZ, L.: Flexible Modelling and Simulation for Control Processor Analysis. *9th International Teletraffic Congress*, Vol. 2, Torremolinos, October 1979. bzw.: *Electrical Communication*, Vol. 55, No. 1, 1980, pp. 26-36.
- [214] GONZÁLEZ SOTO, O.: On Traffic Modelling and Characterization of ISDN Users. *1st International IFIP TC6 Conference on Data Communications in the ISDN Era*, Tel-Aviv, Israel, March 1985, pp. 127-140.
- [215] GOODE, C. W. B.: Breitbanddienste und -anwendungen. *Elektrisches Nachrichtenwesen*, Band 64, Nr. 2/3, 1990, S. 124-131.
- [216] GORDON, J. J.; MURTI, K.; RAYES, A.: Overview of Internet Traffic Issues on the PSTN. *15th International Teletraffic Congress*, Washington, June 1997, pp. 643-652.
- [217] GORDON PEARCE, J.: Multi-Controls — A Possible Basis for the Next Generation of Switching System Organization. *7th International Switching Symposium (ISS '74)*, München, September 1974, pp. 142/1-142/6.
- [218] GRADISCHNIG, K. D.: Trends of Signalling Protocol Evolution in ATM Networks. *15th International Switching Symposium*, Berlin, Vol. 2, April 1995, pp. 310-314.
- [219] GRAVEY, A.; HÉBUTERNE, G.: Mixing Time and Loss Priorities in a Single Server Queue. *13th International Teletraffic Congress*, Copenhagen, June 1991, pp. 141-152.
- [220] GRIFFITHS, R.; KEY, P.: Adaptive Call Admission Control in ATM Networks. *14th Int. Teletraffic Congress*, Vol. 2, Antibes Juan-les-Pins, June 1994, pp. 1089-1098.
- [221] GUARENE, E.; FASANO, P.; VERCELLONE, V.: IP and ATM Integration Perspectives.



- IEEE Communications Magazine*, Vol. 36, No. 1, January 1998, pp. 74-80.
- [222] GÜNTHER, K. D.: Prevention of Deadlocks in Packet-Switched Data Transport Systems. *IEEE Transactions on Communications*, Vol. 29, No. 4, April 1981, pp. 512-524.
- [223] GUO, M.-H.; CHANG, R.-S.: Multicast ATM Switches: Survey and Performance Evaluation. *ACM SIGCOMM Computer Communication Review*, Vol. 28, No. 2, April 1998, pp. 98-131.
- [224] HAAR, J. A.; TAYLOR, F. F.; ULRICH, W.: Organisation of No. 1 ESS Central Processor. *The Bell System Technical Journal*, Vol. XLIII, No. 5, Part 1, September 1964, pp. 1845-1922.
- [225] HAAS, M.: *Mechanismen zur Leistungsregelung von Rechensystemen*. Dissertation an der Fakultät für Informatik, Universität Karlsruhe, 1990.
- [226] HAAS, Z.: Adaptive Admission Congestion Control. *ACM SIGCOMM Computer Communication Review*, Vol. 21, No. 5, October 1991, pp. 58-76.
- [227] HAENSCHKE, D. G.; KETTLER, D. A.; OBERER, E.: Network Management and Congestion in the U.S. Telecommunications Network. *IEEE Transactions on Communications*, Vol. 29, No. 4, April 1981, pp. 376-385.
- [228] HAHNE, E. L.; KALMANEK, C. R.; MORGAN, S. P.: Fairness and Congestion Control on a Large ATM Data Network with Dynamically Adjustable Windows. *13th International Teletraffic Congress, Proceedings in 'North-Holland Studies in Telecommunication'*, Vol. 14, Copenhagen, June 1991, pp. 867-872.
- [229] HÅGÅRD, G.; WOLF, M.: Multiprotocol Label Switching in ATM Networks. *Ericsson Review, The Telecommunications Technology Journal*, Vol. 75, No. 1, 1998, pp. 32-39.
- [230] HAN, M.-K.; CHO, W.-S.; YOON, Y.-I.: A Backup and Recovery Technique for Main-Memory Databases in ATM Switching Systems. *4th International Conference on Telecommunications*, Vol. 2, Melbourne, April 1997, pp. 549-553.
- [231] HANSELKA, P.; OEHLERICH, J.; WEGMANN, G.: Adaptation of the Overload Regulation Method STATOR to Multiprocessor Controls and Simulation Results. *12th International Teletraffic Congress*, Torino, June 1988, pp. 4.3A4.
- [232] HANSON, R. H.: Kompromisse zwischen verbindungslosem und verbindungsorientiertem Netzbetrieb. *Alcatel Telecom Rundschau*, 4. Quartal, 1997, S. 261-267.
- [233] HANSSON, A.: Evolution of Intelligent Network Concepts. *Computer Communications*, Vol. 18, No. 11, November 1995, pp. 793-801.
- [234] HARIU, T.; AKIBA, J.; KISHIDA, T.; KONDO, Y.: Study on Call Model of Advanced Intelligent Network for Controlling B-ISDN. *5th IEEE Intelligent Network Workshop (IN'96)*, Melbourne, Vol. 2, April 1996, Paper #5.2.3.
- [235] HARVEY, C.; GRIFFITHS, P. H.: Controlling Overloads in Multiprocessor Systems. *4th International Conference on Software Engineering for Telecommunication Switching Systems*, Coventry, United Kingdom, July 1981, pp. 167-171.
- [236] HASEGAWA, H.; YAMANAKA, N.; SHIOMOTO, K.: Architectures of ATM Nodes with Light-Weight Flow Control for High Speed, Multi-Protocol ATM-WAN. *16th International Switching Symposium (ISS'97)*, Vol. 1, Toronto, September 1997, pp. 443-449. Bzw.: ATM Nodes with Light-Weight Flow-Control for High Speed, Multi-Protocol ATM-WAN. *IEICE Transactions on Communications (Special Issue on 'ATM Switching Systems for future B-ISDN')*, Vol. E81-B, No. 2, Februar 1998, pp. 392-401.

- [237] HASHIDA, O.; NAKAJIMA, S.; OKADA, K.; SHINOHARA, M.: Congestion Mechanisms and Overload Control in Telephone Networks. *9th International Teletraffic Congress*, Torremolinos, October 1979.
- [238] HE, L.; WONG, A. K.: Connection Admission Control Design for GlobeView<sup>®</sup>-2000 ATM Switches. *Bell Labs Technical Journal*, Vol. 3, No. 1, January-March 1998, pp. 94-111.
- [239] HÉBUTERNE, G.; ROMOEU, L.; KUNG, R.: Load Regulation Schemes for the Intelligent Network. *13th International Switching Symposium (ISS '90)*, Vol. 5, Stockholm, May/June 1990, pp. 159-164.
- [240] HÉBUTERNE, G.; MONIN, W.: Service Specific Connection Oriented Protocol (SSCOP) with no Credit Limitation. *International IFIP-IEEE Conference on Broadband Communications*, Canada, April 1996, pp. 283-293.
- [241] HEFFES, H.: Analysis of Overload Performance for a Class of M/D/1 Processor Queueing Disciplines. *11th International Teletraffic Congress*, Vol. 1, Kyoto, September 1985, Paper #2.1B-1.
- [242] HEIDELBERGER, P.; LAVENBERG, S. S.: Computer Performance Evaluation Methodology. *IEEE Transactions on Communications*, Vol. 33, No. 12, December 1984, pp. 1195-1220.
- [243] HEIß, H.-U.: *Verhinderung von Überlast in Rechensystemen*. Dissertation an der Fakultät für Informatik, Universität Karlsruhe, 1987.
- [244] HENDERSON, T. R.: Design Principles and Performance Analysis of SSCOP: A New ATM Adaptation Layer Protocol. *ACM SIGCOMM Computer Communication Review*, Vol. 25, No. 2, April 1995, pp. 47-59.
- [245] HENRION, M. A.; SCHREINER, S. M.; VAN GOETHEM, J.: Anwendung von Mikroprozessoren bei großen Digital-Vermittlungsstellen. *Elektrisches Nachrichtenwesen*, Band 54, Nr. 3, 1979, S. 190-198.
- [246] HENTSCHKE, S.: Predictive Processor Overload Control Strategies for SPC Switching Systems. *ntz Archiv*, Bd. 3, H. 5, 1981, S. 121-127.
- [247] HIGGINS, W. H. C.: A Survey of Bell System Progress in Electronic Switching. *The Bell System Technical Journal*, Vol. XLIV, No. 6, July/August 1965, pp. 937-997.
- [248] HIGUERA, J. A. G.; BERZOSA, C. D.: Progressive Method of Overload Control for Processors. *9th International Teletraffic Congress*, Vol. 2, Torremolinos, October 1979.
- [249] HILLS, M. T.; LORÉTAN, R. P.: The Future Direction of S.P.C. Systems. *8th International Switching Symposium (ISS '76)*, Vol. 2, Kyoto, October 1976, Paper #412-2.
- [250] HIRAMATSU, A.: Integration of ATM Call Admission Control and Link Capacity Control by Distributed Neural Networks. *IEEE Journal on Selected Areas in Communication*, Vol. 9, No. 7, September 1991, pp. 1131-1138.
- [251] HOFSTETTER, H.; WEBER, D.: Traffic Models for Large ISDN-PABX's. *Computer Networks and ISDN Systems*, Vol. 10, 1985, pp. 195-202.
- [252] HOLMSTRÖM, S.: Optimal Overload Control of a Multiprocessor System with a Regulator Simulating Market Mechanisms. *13th Nordic Teletraffic Seminar*, Espoo, Finland, August 1996, pp. 313-325.
- [253] HOOGENDOORN, C. H.; MAHER, A. T.: Enhanced SW Architecture for an ATM Universal Communication Node. *14th International Switching Symposium (ISS '92)*, Vol. 1,

- Yokohama, October 1992, pp. 440-444.
- [254] HÖLLER, J.: Voice and Telephony Networking over ATM. *Ericsson Review, The Telecommunications Technology Journal*, Vol. 75, No. 1, 1998, pp. 40-45.
- [255] HOOVER, E. S.; ECKHART, B. J.: Automatic Overload Controls. *5th International Teletraffic Congress*, New York, June 1967, p. 235.
- [256] HOUCK, D. J.; MEIER-HELLSTERN, K. S.; SAHEBAN, F.; SKOOG, R. A.: Failure and Congestion Propagation through Signaling Controls. *14th International Teletraffic Congress*, Antibes Juan-les-Pins, June 1994, pp. 367-376.
- [257] HUBER, J. F.; MAIR, E.: Universelle Paketvermittlung durch flexible EWSP-Architektur. *telcom report*, 10. Jhrg., Nr. 1/87, Januar/Februar 1987, S. 12-18.
- [258] HUBER, M. N.; KIESEL, W. M.; KÜHN, P. J.; RAITH, T.; KERSCHNER, G. W.: Performance Modelling of a Highly Modularized Packet Switching Node. *8th International Conference on Computer Communications (ICCC'86)*, Munich, September 1986, pp. 600-605.
- [259] HUBIG, W.; WEBER, D.: Overload Control in ISDN PABXs. *14th International Teletraffic Congress*, Antibes Juan-les-Pins, June 1994, pp. 243-252.
- [260] HWANG, R.-H.; KUROSE, J. F.; TOWSLEY, D.: On-Call Processing Delay in High Speed Networks. *IEEE/ACM Transactions on Networking*, Vol. 3, No. 6, December 1995, pp. 628-639.
- [261] IDE, H.-D.: *Über ein Warte-Verlust-System mit dynamischer Zuteilung der Bedienungseinheiten zur Abwehr lokaler Überlastsituationen*. Fortschr.-Ber. VDI Reihe 8 Nr. 93, VDI-Verlag, Düsseldorf, 1985.
- [262] INSIGNIA: Network-Based Signalling Traffic Models and Performance Parameters. *ACTS Project INSIGNIA (IN and B-ISDN Signalling Integration on ATM Platforms)*, Deliverable AC068/UOT/114/DS/P/006/b0, August 1997.
- [263] IRLAND, E. A.; STAGG, U. K.: New Developments in Suburban and Rural ESS (No. 1 ESS and No. 2 ESS). *7th International Switching Symposium (ISS '74)*, München, September 1974, pp. 512/1-512/6.
- [264] ISHIBASHI, K.; YAMAMOTO, Y.; IAI, S.: Trends of Multimedia Applications from the Viewpoint of Quality of Service and Traffic Aspects. *NTT REVIEW*, Vol. 10, No. 2, March 1998, pp. 44-49.
- [265] ITOH, M.; NUNOTANI, Y.; UEDA, T.; OKADA, K.: Methods of Estimating Central Processing System Traffic Performances in SPC-Electronic Switching Systems. *8th International Teletraffic Congress*, Vol. 2, Melbourne, November 1976, Paper #621.
- [266] JACOBSON, V.: Congestion Avoidance and Control. *Proc. SIGCOMM'88 Symposium Communication, Architecture Protocols*, August 1988, pp. 314-329.
- [267] JAFFE, J. M.: Flow Control Power is Nondecentralizable. *IEEE Transactions on Communications*, Vol. 29, No. 9, September 1980, pp. 1301-1306.
- [268] JAIN, R.: A Timeout-Based Congestion Control Scheme for Window Flow-Controlled Networks. *IEEE Journal on Selected Areas in Communication*, Vol. 4, No. 7, October 1986, pp. 1162-1167.
- [269] JAIN, R.: A Delay-Based Approach for Congestion Avoidance in Computer Networks. *Computer Communications Review*, Vol. 19, No. 5, October 1989, pp. 56-71.
- [270] JAIN, R.: Myths about Congestion Control in High Speed Networks. *7th ITC Specialist*

- Seminar: Broadband Technologies*, Morriston, October 1990.  
Bzw.: Myths about Congestion Management in High Speed Networks. *Internetworking: Research and Experience*, Vol. 3, 1992, pp. 101-113.
- [271] JAIN, R.: Congestion Control in Computer Networks: Issues and Trends. *IEEE Network Magazine*, Vol. 4, No. 3, May 1990, pp. 24-30.
- [272] JEANES, D. L.: Evolution of the SL-10 Packet Switching System. *8th International Conference on Computer Communications (ICCC '86)*, Munich, September 1986, pp. 430-435.
- [273] JEONG, K.-J.; KIM, S.-G.: Effective Overload Prevention and Control for Traffic Management in Intelligent Network. *4th International Conference on Intelligence in Networks*, Bordeaux, November 1996, pp. 229-234.
- [274] JOEL, A. J.: The Classification and Unification of Switching System Functions. *6th International Switching Symposium (ISS '72)*, Cambridge, USA, June 1972, pp. 446-453.
- [275] JOEL, A. J.: Realization of the Advantages of Stored Program Control. *7th International Switching Symposium (ISS '74)*, München, September 1974, pp. 143/1-143/6.
- [276] JOEL, A. J.: ESS: Minimonster. *IEEE Spectrum*, Vol. 13, No. 8, August 1976, pp. 33-35.
- [277] JOHANNESSEN, J. D.: No. 1 ESS Local/Toll/Tandem. *7th International Switching Symposium (ISS '74)*, München, September 1974, pp. 526/1-526/4.
- [278] JONINS, G.; SEDOLS, J.: Overload Control in Teletraffic Systems with Busy Subscribers and Repeated Attempts. *2nd St. Petersburg Regional International Teletraffic Seminar*, St. Petersburg, June/July 1995, pp. 597-601.
- [279] JORDAN, S.; JIANG, H.: Connection Establishment in High-Speed Networks. *IEEE Journal on Selected Areas in Communication*, Vol. 13, No. 7, September 1995, pp. 1150-1161.
- [280] JORGENSEN, A. A.: The ESC-1 – An Electronic Switching System with Distributed Control. *6th International Switching Symposium (ISS '72)*, Cambridge, USA, June 1972, pp. 188-194.
- [281] KADERALI, F.; GIEBLER, H. J.: Characteristics of a Digital Exchange for ISDN. *11th International Switching Symposium (ISS '84)*, Florence, Italy, May 1984, Paper #22B4.
- [282] KALLENBERG, P.: Load Estimation for Overload Control. *12th International Teletraffic Congress*, Torino, June 1988, pp. 4.3A6.
- [283] KAMIGAKI, Y.; ET AL.: Broadband Switching System for Public Network. *FUJITSU Scientific & Technical Journal*, Vol. 32, No. 1, June 1996, pp. 3-12.
- [284] KAMINSKY, F. C.; RUMPF, D. L.: Simulating Nonstationary Poisson Processes: A Comparison of Alternatives including the Correct Approach. *SIMULATION, Technical Journal of the Society for Computer Simulation*, Vol. 29, No. 1, July 1977, pp. 17-20.
- [285] KAMOLPHIWONG, S.; KARBOWIAK, A. E.; MEHRPOUR, H.: Flow Control in ATM Networks: A Survey. *Computer Communications*, Vol. 21, No. 11, August 1998, pp. 951-968.
- [286] KANKARE, J.; KILKKI, K.: Recursive Estimators for Traffic Analysis and Prediction. *4th Int. Conference on Telecommunications*, Vol. 1, Melbourne, April 1997, pp. 131-136.
- [287] KANT, K.: Flow Control Mechanisms for SAAL Links. *International IFIP-IEEE Conference on Broadband Communications*, Canada, April 1996, pp. 173-184.

- [288] KANT, K.; ONG, L.: Signaling in Emerging Telecommunications and Data Networks. *Proceedings of IEEE*, Vol. 85, No. 10, October 1997, pp. 1612-1621.
- [289] KAPPEL, J. G.; STONE, R. C.: Digital Switching Systems Traffic Analysis. *12th International Teletraffic Congress, Proceedings in 'North-Holland Studies in Telecommunication'*, Vol. 12, Torino, June 1988, pp. 285-291.
- [290] KARAGIANNIS, G.; VAN BEIJNUM, B. J. F.; NIEMEGEREERS, I. G. M. M.: On the Integration of the UMTS and B-ISDN System. *IFIP TC6/WG6.4 7th International Conference on High Performance Networking, VII (HPN '97)*, New York, April/May 1997, pp. 39-56.
- [291] KARAGIANNIS, G.; NICOLA, V. F.; NIEMEGEREERS, I. G. M. M.: Quantitative Evaluation of Scalability in Broadband Intelligent Networks. *IFIP TC6/WG7.3 7th International Conference on Performance of Information and Communication Systems (PICS '98)*, Lund, Sweden, May 1998, pp. 65-82.
- [292] KARLANDER, B.: Control of Central Processor Load in a SPC System. *7th International Teletraffic Congress*, Stockholm, June 1973, Paper #436.
- [293] KATSCHNER, L.; VAN DEN BRANDE, F.: SYSTEM 12 – Konzeption und Realisierung der Software. *Elektrisches Nachrichtenwesen*, Band 56, Nr. 2/3, 1981, S. 173-183.
- [294] KAUFMAN, J. S.; KUMAR, A.: Traffic Overload Control in a Fully Distributed Switching Environment. *12th International Teletraffic Congress*, Torino, June 1988, pp. 4.3A3.
- [295] KAUSHAL, A.; ET AL.: Performance Benchmarking of ATM Signaling Software. *Proceedings of the OpenSig Workshop*, Columbia University, October 1996.
- [296] KAWAHARA, R.; ASAKA, T.; SUMITA, S.: Overload Control for Intelligent Networks and Its Analysis by Simulation. *IEICE Transactions on Communications*, Vol. E78-B, No. 4, April 1995, pp. 494-503.
- [297] KAWAHARA, R.; ASAKA, T.: Overload Control for Intelligent Networks based on an Estimation of Maximum Number of Calls in a Node. *5th IEEE Intelligent Network Workshop (IN '96)*, Melbourne, Vol. 1, April 1996, Paper #1.3.1.
- [298] KAWAI, A.; IWAKI, S.; KIYOKU, H.; KUSABA, K.: Next-Generation Switching Systems. *Hitachi Review*, Vol. 47, No. 2, April 1998.
- [299] KAWARASAKI, M.; JABBARI, B.: B-ISDN Architecture and Protocol. *IEEE Journal on Selected Areas in Communication*, Vol. 9, No. 9, December 1991, pp. 1405-1415.
- [300] KEISTER, W.; KETCHLEDGE, R. W.; VAUGHAN, H. E.: No. 1 ESS: System Organization and Objectives. *The Bell System Technical Journal*, Vol. XLIII, No. 5, Part 1, September 1964, pp. 1831-1844.
- [301] KEISTER, W.; KETCHLEDGE, R. W.; VAUGHAN, H. E.: A New Electronic Switching System. *IEEE Spectrum*, Vol. 2, Februar 1965, pp. 87-94.
- [302] KESHAV, S.: A Control-Theoretic Approach to Flow Control. *Proc. SIGCOMM'91*, September 1991, pp. 3-15.
- [303] VON KIENLIN, A.: Architecture of a High Performance Packet Switching System. *7th International Conference on Computer Communications (ICCC '84)*, Sydney, October/November 1984, pp. 481-484.
- [304] KIHIL, M.; NYBERG, C.: Transient and Stationary Investigations of Overload Control in Intelligent Networks. *12th International Conference on Computer Communication*, Seoul, August 1995, pp. 602-607.
- [305] KIHIL, M.; RUMSEWICZ, M.: Overload Control of Intelligent Peripherals. *ITC Mini-Se-*

- minar on Engineering and Congestion Control in Intelligent Networks*, Melbourne, Australia, April 1996.
- [306] KIHLE, M.; RUMSEWICZ, M.: Flow Model Analysis of an Intelligent Peripheral Overload Control Strategy. *10th ITC Specialists Seminar on Control in Communications*, Lund, Sweden, September 1996, pp. 103-114.
- [307] KIHLE, M.; RUMSEWICZ, M.: Analysis of Overload Control Strategies in Combined SSP-SCPs in the Intelligent Network. *15th International Teletraffic Congress*, Washington, June 1997, pp. 1209-1218.
- [308] KIM, S. K.: A Traffic Overload Management for SCP in Intelligent Network. *International Seminar on Teletraffic and Network (ISTN'97): „Teletraffic Issues and Technology on Broadband Telecommunication“*, Xi'an, China, November 1997, pp. 117-120.
- [309] KIM, Y. B.; LEE, S. S.; OH, C. H.; KIM, Y. S.; HAN, C.; YIM, C. H.: An Architecture of Scalable ATM Switching System and Its Call Processing Capacity Estimation. *ETRI Journal*, Vol. 18, No. 3, October 1996, pp. 107-125.
- [310] KLEINROCK, L.: On Flow Controls in Computer Networks. *14th IEEE International Conference on Communications (ICC '78)*, Vol. 2, Toronto, June 1978, Paper #27.2.
- [311] KLEINROCK, L.: Power and Deterministic Rules of Thumb for Probabilistic Problems in Computer Communications. *15th IEEE International Conference on Communications (ICC '79)*, Vol. 2, Boston, June 1979, Paper #43.1.
- [312] KLEINROCK, L.: Distributed Systems. *Communications of the ACM*, Vol. 28, No. 11, November 1985, pp. 1200-1213.
- [313] KLEINROCK, L.: The Latency/Bandwidth Tradeoff in Gigabit Networks. *IEEE Communications Magazine*, Vol. 30, No. 4, April 1992, pp. 36-40.
- [314] KO, K.-T.; MISHRA, P. P.; TRIPATHI, S. K.: Predictive Congestion Control in High-Speed Wide-Area Networks. *2nd International Workshop on Protocols for High-Speed Networks*, Palo Alto, November 1990, pp. 33-53.
- [315] KOCHER, H.: *Entwurf und Implementierung einer Simulationsbibliothek unter Anwendung objektorientierter Methoden*. 59. Bericht über verkehrstheoretische Arbeiten (Dissertationsschrift), Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1994.
- [316] KÖRNER, U.: Congestion Control in Packet Switching Computer Communication Networks. *10th Int. Teletraffic Congress*, Vol. 1, Montreal, June 1983, Paper 3.3-#5.
- [317] KÖRNER, U.: Overload Control of SPC Systems. *13th International Teletraffic Congress, Proceedings in 'North-Holland Studies in Telecommunication'*, Vol. 14, Copenhagen, June 1991, pp. 105-114.
- [318] KÖRNER, U.; NYBERG, C.: Overload Control in Communications Network. *IEEE GLOBECOM '91*, Vol. 2, Phoenix, December 1991, pp. 1331-1335.
- [319] KÖRNER, U.; WALLSTRÖM, B.; NYBERG, C.: The Impact of New Services and New Control Architectures on Overload Control. *14th International Teletraffic Congress*, Antibes Juan-les-Pins, June 1994, pp. 275-283.
- [320] KOIKE, A.; OZAWA, T.: Traffic Control Techniques in the Multimedia Era. *NTT REVIEW*, Vol. 10, No. 2, March 1998, pp. 55-57.
- [321] KOINUMA, T.; TAKAHASHI, T.: ATM Node System Technologies. *NTT REVIEW*, Vol. 6, No. 1, January 1994, pp. 43-48.

- [322] KOLYVAS, G. T.; POLYKALAS, S. E.; VENIERIS, I. S.: Performance Evaluation of Integrated IN/B-ISDN Signalling Platforms. *Computer Communications*, Vol. 21, No. 7, June 1998, pp. 606-623.
- [323] KOONO, Z.; SHODA, A.; TOKITA, Y.; WORKS, T.: A Distributed Control System for Electronic Switching Systems. *8th International Switching Symposium (ISS '76)*, Vol. 2, Kyoto, October 1976, Paper #412-3.
- [324] KOSTEN, L.: On the Influence of Repeated Calls in the Theory of Probabilities of Blocking. *De Ingenieur*, Vol. 59, 1947, pp. 1-25 (in Dutch).
- [325] KOTSIS, G.; KRITHIVASAN, K.; RAGHAVAN, S. V.: A Workload Characterization Methodology for WWW Applications. *5th International Conference on the Performance and Management of Complex Communication Networks*, Tsukuba, Japan, November 1997, pp. 145-159.
- [326] KRAIMECHE, B.; SCHWARTZ, M.: Analysis of Traffic Access Control Strategies in Integrated Service Networks. *IEEE Transactions on Communications*, Vol. 33, No. 10, October 1985, pp. 1085-1093.
- [327] KRÖNER, H.; RENGER, T.; KNOBLING, R.: Performance Modelling of an Adaptive CAC Strategy for ATM Networks. *14th International Teletraffic Congress*, Vol. 2, Antibes Juan-les-Pins, June 1994, pp. 1077-1088.
- [328] KRÖNER, H.: *Verkehrssteuerung in ATM-Netzen – Verfahren und verkehrstheoretische Analysen zur Zellpriorisierung und Verbindungsannahme*. 62. Bericht über verkehrstheoretische Arbeiten (Dissertationsschrift), Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1995.
- [329] KÜHN, P. J.: Verteilte Mikrorechner-Steuerungen in Nachrichtenvermittlungssystemen – Strukturen, Organisation und Verkehrsanalyse. In *Nachrichtensysteme – Dienstintegration in künftigen Kommunikationsnetzen: Vorträge des Nachrichtentechnischen Kolloquiums 1981 der Technischen Universität Braunschweig*, H. L. Hartmann (Hrsg.), B. G. Teubner Verlag, Stuttgart, 1982, S. 85-106.
- [330] KÜHN, P. J.: Modelling of Switching System Controls. *10th International Teletraffic Congress*, Vol. 1, Montreal, June 1983, Paper 1.3-#1.
- [331] KÜHN, P. J.: Modelling of New Services in Computer and Communication Networks. *International Seminar on Computer Networking and Performance Evaluation*, Tokyo, Japan, September 1985, Paper 7-1.
- [332] KÜHN, P. J.: Stochastische Modelle für Bedienungs- und Transportprozesse in Kommunikationsnetzen. *ITG-Fachtagung 'Stochastische Modelle und Methoden in der Informationstechnik'*, Nürnberg, April 1989. In: *ITG-Fachbericht*, Nr. 107, VDE-Verlag, Berlin, S. 79-94.
- [333] KÜHN, P. J., SCHOPP, M.: Signalling Networks for ISDN, IN and Mobile Networks – Modelling, Analysis, and Overload Control. *10th ITC Specialists Seminar on Control in Communications*, Lund, Sweden, September 1996, pp. 35-49.
- [334] KUMAR, A.: Adaptive Load Control of the Central Processor in a Distributed System with a Star Topology. *IEEE Transactions on Computers*, Vol. 38, No. 11, November 1989, pp. 1502-1512.
- [335] KÜMMERLE, K.: Ein Vorschlag zur Berechnung der Vertrauensintervalle bei Verkehrstests. *Archiv der elektrischen Übertragung (AEÜ)*, Bd. 23, H. 10, 1969, S. 507-510.
- [336] KUROSE, J. F.; CHIPALKATTI, R.: Load Sharing in Soft Real-Time Distributed Compu-

- ter Systems. *IEEE Transactions on Computer*, Vol. 36, No. 8, August 1987, pp. 993-1000.
- [337] KWIATKOWSKI, M.: Queue Length Congestion Control at an SCP. *Proceedings of the Australian Telecommunication Networks & Applications Conference*, Melbourne, December 1994, pp. 743-749.
- [338] KWIATKOWSKI, M.; NORTHCOTE, B.: Calculating Mean Delays in Intelligent Networks Under Overload. *Proceedings of the Australian Telecommunication Networks & Applications Conference*, Melbourne, December 1994, pp. 737-741.
- [339] LAM, S. S.; LIEN, Y. C. L.: Congestion Control of Packet Communication Networks by Input Buffer Limits – A Simulation Study. *IEEE Transactions on Computer*, Vol. 30, No. 10, October 1981, pp. 733-742.
- [340] LANG, F.: *Simulation von Überlastabwehrverfahren im Intelligenzen Netz*. Semesterarbeit Nr. 1464 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, Januar 1997.
- [341] LANGENBACH-BELZ, M.; MELIS, A.; VERHILLE, H.: SYSTEM 12 – Eine Einführung. *Elektrisches Nachrichtenwesen*, Band 56, Nr. 2/3, 1981, S. 114-125.
- [342] LANGENBACH-BELZ, M.: Digitale Fernsprechvermittlungstechnik — Einfluß der Technologie und der betrieblichen Anforderungen auf die Systemarchitektur. In *Nachrichtensysteme – Dienstintegration in künftigen Kommunikationsnetzen: Vorträge des Nachrichtentechnischen Kolloquiums 1981 der Technischen Universität Braunschweig*, H. L. Hartmann (Hrsg.), B. G. Teubner Verlag, Stuttgart, 1982, S. 63-84.
- [343] LA PORTA, T., F.; VEERARAGHAVAN, M.; BUSKENS, R. W.: Comparison of Signaling Loads for PCS Systems. *IEEE/ACM Transactions on Networking*, Vol. 4, No. 6, December 1996, pp. 840-856.
- [344] LAW, A. M.: Statistical Analysis of Simulation Output Data. *Operations Research*, Vol. 31, No. 6, 1983, pp. 983-1029.
- [345] LAWSON, D. A.: A New Software Architecture for Switching Systems. *IEEE Transactions on Communications*, Vol. 30, No. 6, June 1982, pp. 1281-1289.
- [346] LEE, Y.: Overload Control of SCP in Intelligent Network with Priority. *IEICE Transactions on Communications*, Vol. E80-B, No. 11, November 1997, pp. 1753-1755.
- [347] LE GALL, P.: Traffic Simulation and Traffic Modelling in the Control Plane. *10th ITC Specialists Seminar on Control in Communications*, Lund, Sweden, September 1996, pp. 131-142.
- [348] LEHTINEN, P.: Window Access Control in a Distributed Switching System. *1st St. Petersburg Regional International Teletraffic Seminar*, St. Petersburg, Russia, June 1993, pp. 189-198.
- [349] LEHTINEN, P.: Architecture of Service Traffic Measurements in Intelligent Networks. *2nd St. Petersburg Regional International Teletraffic Seminar*, St. Petersburg, Russia, June/July 1995, pp. 135-149.
- [350] LEINER, B. M.; CERF, V. G.; CLARK, D. D.; KAHN, R. E.; KLEINROCK, L.; LYNCH, D. C.; POSTEL, J.; ROBERTS, L. G.; WOLFF, S. S.: The Past and Future History of the Internet. *Communications of the ACM*, Vol. 40, No. 2, February 1997, pp. 102-108.
- [351] LEKKOU, M. E.; VENIERIS, I. S.: A Workload Model for Performance Evaluation of Multimedia Signalling Systems. *Computer Communications*, Vol. 20, No. 10, Septem-



- ber 1997, pp. 884-898.
- [352] LEMPPENAU, W.: *Umweltsimulator für den Funktions- und Lasttest von Nachrichtenvermittlungssystemen*. 47. Bericht über verkehrstheoretische Arbeiten (Dissertationschrift), Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1989.
- [353] LEVY, M.: Das Internet und die Vermittlung im Kernnetz. *Alcatel Telecom Rundschau*, 4. Quartal, 1997, S. 274-279.
- [354] LEVY, Y.; DURINOVIC-JOHRI, S.; MILITO, R. A.: Dynamic Network Call Distribution with Periodic Updates. *14th International Teletraffic Congress*, Antibes Juan-les-Pins, June 1994, pp. 85-94.
- [355] LEWIS, A.; LÉONARD, G.: Measurements of Repeated Call Attempts in the International Telephone Service. *10th Int. Teletraffic Congress*, Montreal, June 1983, Paper 2.4-#2.
- [356] LEWIS, P. A. W.; SHEDLER, G. S.: Simulation Nonhomogeneous Poisson Process by Thinning. *Naval Research Logistics*, Vol. 26, 1979, pp. 403-413.
- [357] LI, M.; YAN, J.: Dimensioning of Line-Access Systems Serving Internet Traffic. *16th Int. Switching Symposium (ISS '97)*, Vol. 2, Toronto, September 1997, pp. 67-73.
- [358] LIAO, K.-Q.; MASON, L.: Self-Optimizing Window Flow Control in High-Speed Data Networks. *Computer Communications*, Vol. 16, No. 11, November 1993, pp. 706-716.
- [359] LITTLE, J. D. C.: A Proof for the Queueing Formula:  $L = \lambda W$ . *Operations Research*, Vol. 9, 1961, pp. 383-387.
- [360] LIU, S. S.: Impacts of Signaling-Intensive Local Services (SILS) on B-ISDN Switching Systems: A Simulation Study. *Computer Networks and ISDN Systems*, Vol. 25, 1992, pp. 121-143.
- [361] LODGE, F.; CURRAN, T.: A Congestion Control Strategy for Combined IN and Non-IN Traffic Load at the Service Switching Point of an Intelligent Network. *7th International Network Planning Symposium*, Sydney, November 1996, pp. 83-88.
- [362] LOHRUM, T.: Verkehrs- und Überlaststeuerung in ATM-Netzen. *Deutsche Telekom Unterrichtsblätter*, 50. Jg., April 1997, S. 236-245.
- [363] LOVEGROVE, W. P.: *Methodology for Studying Transient Behavior in Computer Networks using Simulation*. Ph.D. Thesis, Clemson University, USA, December 1990.
- [364] LOVEGROVE, W. P.; HAMMOND, J. L.; TIPPER, D.: Simulation Methods for Studying Nonstationary Behavior of Computer Networks. *IEEE Journal on Selected Areas in Communication*, Vol. 8, No. 9, December 1990, pp. 1696-1708.
- [365] LUCKY, R. W.: New Communications Services – What Does Society Want? *Proceedings of IEEE*, Vol. 85, No. 10, October 1997, pp. 1536-1543.
- [366] LUDERER, G. W. R.: Architecture of the AT&T No. 1 Packet Switching System. *7th International Conference on Computer Communications (ICCC '84)*, Sydney, October/November 1984, pp. 485-490.
- [367] LYONS, M. H.; BURTON, F.; EGAN, B.; LYNCH, T.; SKELTON, S.: Dynamic Modeling of Present and Future Service Demand. *Proceedings of IEEE*, Vol. 85, No. 10, October 1997, pp. 1544-1555.
- [368] MAHER, A. T.: Mit CP113-Modularität Vermittlungsleistung flexibel erweitern. *telcom report*, 12. Jhrg., Nr. 1-2/89, Januar/Februar 1989, S. 10-13.

- [369] MAIR, E.; HAUSMANN, H.; NAEßL, R.: EWSP — A High-Performance Packet Switching System. *8th International Conference on Computer Communications (ICCC '86)*, Munich, September 1986, pp. 359-364.
- [370] MANFIELD, D.; DENIS, B.; BASU, K.; ROULEAU, G.: Overload Control in a Hierarchical Switching System. *11th International Teletraffic Congress*, Vol. 2, Kyoto, September 1985, Paper #5.1B-4.
- [371] MANFIELD, D. R.; MILLSTEED, G. K.; ZUKERMAN, M.: Congestion Controls in SS7 Signaling Networks. *IEEE Communications Magazine*, Vol. 31, No. 6, June 1993, pp. 50-57.
- [372] MARK, B. L.; ET AL.: Large Capacity Multiclass ATM Core Switch Architecture. *16th Int. Switching Symposium (ISS '97)*, Vol. 1, Toronto, September 1997, pp. 417-423.
- [373] MARTIKAINEN, O.; NAOUMOV, V.; NIKITINE, A.: Broadband Signalling for Multimedia Services. *St. Petersburg International Teletraffic Seminar*, Proceedings, St. Petersburg, Russia, June/July 1995, pp. 36-51.
- [374] MASE, K.; YAMAMOTO, H.: Quality and Traffic Management in Multimedia Era. *NTT REVIEW*, Vol. 10, No. 2, March 1998, pp. 34-39.
- [375] MATSUMOTO, T.; YAMADA, K.; FAN, R.; NAKAGAWA, T.: Evolution of ATM Switch Family for Diverse Networks and Services. *NEC Research & Development*, Vol. 36, No. 4, October 1995, pp. 502-512.
- [376] MATSUMURA, R.; YOSHINO, H.; HORIGOME, H.; MIWA, H.: Traffic Control for Server Overload and Network Congestion by Dynamic Multi-Server System. *NTT REVIEW*, Vol. 10, No. 2, March 1998, pp. 58-62.
- [377] MAZUMDAR, R.; MASON, L. G.; DOULIGERIS, C.: Fairness in Network Optimal Flow Control: Optimality of Product Forms. *IEEE Transactions on Communications*, Vol. 39, No. 5, May 1991, pp. 775-762.
- [378] McCulley, B.: Confusing Telephone System Overload Message. *ACM SIGSOFT, Software Engineering Notes*, Vol. 13, Issue 9, February 1992. Bzw.: <http://catless.ncl.ac.uk/Risks/13.09.html> (Forum on Risks to the Public in Computers and Related Systems).
- [379] McMILLAN, D.; RUMSEWICZ, M.: Analysis of Congestion Control for SCCP Traffic and the Impact on Intelligent Network Services. *5th IEEE Intelligent Network Workshop (IN '96)*, Melbourne, Vol. 1, April 1996, Paper #1.3.2.
- [380] MCQUILLAN, J. M.: Interactions between Routing and Congestion Control in Computer Networks. *International Symposium on Flow Control in Computer Networks*, Versailles, February 1979, pp. 63-75.
- [381] MEDINA, M.; TUDELA, R.; GONZÁLES SOTO, O.; BORJA, T.: Traffic Characterization for ISDN Planning Studies. *4th ITC Specialists Seminar on ISDN Traffic Issues*, Brussels, May 1986.
- [382] MILITO, R. A.: Direct Adaptive Job Distribution and Overload Control in a Decentralized System. *10th ITC Specialists Seminar on Control in Communications*, Lund, Sweden, September 1996, pp. 23-33.
- [383] MITROU, N. M.: Traffic Control in ATM: A Review, an Engineer's Critical View and a Novel Approach. *6th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks (ATM '98)*, Ilkley, UK, July 1998, Tutorial Paper #T08.

- [384] MIYAHO, N.; SHIOMOTO, K.; TAKENAKA, T.: Development of a Commercial ATM Node as the Backbone Node for High-Speed Data Communication Networks. *15th International Switching Symposium (ISS '95)*, Vol. 1, Berlin, April 1995, pp. 464-468.
- [385] MIYAHO, N.; ITOH, A.; SHIOMOTO, K.: ATM Node System Technology for Effective Maintainability. *IEICE Transactions on Communications*, Vol. E79-B, No. 12, December 1996, pp. 1873-1886.
- [386] MIYAHO, N.; MIURA, A.: Integrated Switching Architecture and Its Traffic Handling Capacity in Data Communication Networks. *IEICE Transactions on Communications*, Vol. E79-B, No. 12, December 1996, pp. 1887-1899.
- [387] MORALES-ANDRÉS, G.; VILLÉN-ALTAMIRANO, M.: System 12 Traffic Overload Control. *Electrical Communication*, Vol. 59, No. 1/2, 1985, pp. 74-79.
- [388] MORGAN, S.; DELANEY, M.: The Internet and the Local Telephone Network: Conflicts and Opportunities. *16th International Switching Symposium (ISS '97)*, Vol. 1, Toronto, September 1997, pp. 561-569. Bzw. in *IEEE Communications Magazine*, Vol. 36, No. 1, January 1998, pp. 42-48.
- [389] MOYLAN, P. J.: The Decoupling of Estimation and Control in Teletraffic Problems. *Computer Networks and ISDN Systems*, Vol. 20, No. 1-5 (Special Volume on the ITC Specialist Seminar, Adelaide, September 1989), December 1990, pp. 203-207.
- [390] MÜLLER, H.: Verfügbarkeit bei Kommunikationssystemen — Sicherheitstechnik bei Hicom 300 E von Siemens. *taschenbuch der telekom praxis 1998*, 35. Jhrg., Fachverlag Schiele & Schön, Berlin, 1998, S. 44-69.
- [391] MUMMERT, V. S.: Network Management and its Implementation on the No. 4 ESS. *8th Int. Switching Symposium (ISS '76)*, Vol. 1, Kyoto, Japan, October 1976, Paper #241-2.
- [392] MURAKAMI, T.; ET AL.: A Development Environment for Hierarchically Structured Switching Software. *16th International Switching Symposium (ISS '97)*, Vol. 2, Toronto, September 1997, pp. 511-517.
- [393] MURATA, M.: Requirements on ATM Switch Architectures for Quality-of-Service Guarantees. *IEICE Transactions on Communications (Special Issue on 'ATM Switching Systems for future B-ISDN')*, Vol. E81-B, No. 2, Februar 1998, pp. 138-151.
- [394] MYSKJA, A.; AAGESEN, F. A.: On the Interaction between Subscribers and a Telephone System. *8th Int. Teletraffic Congress*, Vol. 1, Melbourne, November 1976, Paper #322.
- [395] MYSKJA, A.: Modelling of Non-Stationary Traffic Processes. *9th International Teletraffic Congress*, Vol. 1, Torremolinos, October 1979.
- [396] NAGLE, J.: On Packet Switches with Infinite Storage. *IEEE Transactions on Communications*, Vol. 35, No. 4, April 1987, pp. 435-438.
- [397] NAKAMURA, H.; KAI, T.; TANAKA, H.: Software Architecture of the All Band Switching Node System for Efficient Processing of N/B-ISDN Calls. *15th International Switching Symposium (ISS '95)*, Vol. 1, Berlin, April 1995, pp. 98-102.
- [398] NEWMAN, P.; MINSHALL, G.; LYON, T. L.: IP Switching – ATM Under IP. *IEEE/ACM Transactions on Networking*, Vol. 6, No. 2, April 1998, pp. 117-129.
- [399] NI, L. M.; HWANG, K.: Optimal Load Balancing in a Multiple Processor System with Many Job Classes. *IEEE Transactions on Software Engineering*, Vol. 11, No. 5, May 1985, pp. 491-496.
- [400] NIEHAUS, D.; ET AL.: Performance Benchmarking of Signaling in ATM Networks.

- IEEE Communications Magazine*, Vol. 35, No. 8, August 1997, pp. 134-143.
- [401] NIU, Z.; TAKAHASHI, Y.; ENDO, N.: Performance Evaluation of SVC-Based IP-Over-ATM Networks. *IEICE Transactions on Communications (Special Issue on ,ATM Traffic Control and Performance Evaluation‘)*, Vol. E81-B, No. 5, May 1998, pp. 948-957.
- [402] NOGUCHI, K.; KAWASHIMA, Y.; NARITA, S.: A Highly Reliable Frame-Relay Switching Node Architecture. *16th International Switching Symposium (ISS ‘97)*, Vol. 1, Toronto, September 1997, pp. 341-349. Bzw.: A Highly Reliable Frame-Relay Switching Node Architecture based on ATM Switching Technology. *IEICE Transactions on Communications (Special Issue on ,ATM Switching Systems for future B-ISDN‘)*, Vol. E81-B, No. 2, Februar 1998, pp. 315-323.
- [403] NORTHCOTE, B.: Analysing SCP Congestion Control Strategies: A Study of Control Dynamics. *5th IEEE Intelligent Network Workshop (IN ‘96)*, Melbourne, Australia, Vol. 1, April 1996, Paper #1.3.3.
- [404] NORTHCOTE, B.; ATAI, A. H.: AIN Focused Overloads – A Review of USA CCS Network Failures and Lessons Learned. *ITC Mini-Seminar on Engineering and Congestion Control in Intelligent Networks*, Melbourne, Australia, April 1996.
- [405] NORTHCOTE, B.; SMITH, D. E.: Service Control Point Overload Rules to Protect Intelligent Network Services. *IEEE/ACM Transactions on Networking*, Vol. 6, No. 1, February 1998, pp. 71-81.
- [406] NYBERG, C.: *On Overload Control in Telecommunication Systems*. Ph.D. Thesis, Department of Communication Systems, Lund Institute of Technology, 1992.
- [407] NYBERG, C.; AHLFORS, U.: On Overload Control in Distributed Control Systems. *St. Petersburg International Teletraffic Seminar, Proceedings*, St. Petersburg, June/July 1995, pp. 169-181.
- [408] NYBERG, C.; AHLFORS, U.; KIHLE, M.: Modified PD Controller applied to Overload Control of a Multiprocessor System. *12th Nordic Teletraffic Seminar*, Espoo, Finland, August 1995, pp. 421-427.
- [409] NYBERG, H.; OLIN, B.: On Load Control of an SCP in the Intelligent Network. *Proceedings of the Australian Telecommunication Networks & Applications Conference*, Melbourne, December 1994, pp. 751-756.
- [410] OEHLERICH, J.; HELD, W.; WITTE, M.: Load Control and Load Balancing in Switching Systems with Distributed Processor Architecture. *MMB-Mitteilungen (GI/ITG-Interessengruppe ,Messung, Modellierung und Bewertung von Rechensystemen“)*, Nr. 32, Herbst 1997, S. 32-42.
- [411] OHNISHI, H.; ET AL.: All Band Switching Node Architecture for Flexible and Cost-Effective Evolution towards B-ISDN. *15th International Switching Symposium (ISS ‘95)*, Vol. 1, Berlin, April 1995, pp. 57-61.
- [412] OHNISHI, H.; NAGASAWA, M.: Using the NS8000 Series to Build Multimedia Networks. *NTT REVIEW*, Vol. 9, No. 1, January 1997, pp. 79-85.
- [413] OHNISHI, H.; MIYAKE, K.: Issues in ATM Network Service Development, Standardization and Deployment. *IEICE Transactions on Communications (Special Issue on ,ATM Switching Systems for future B-ISDN‘)*, Vol. E81-B, No. 2, Februar 1998, pp. 152-163.
- [414] OMIDYAR, C. G.; PUJOLLE, G.: Introduction to Flow and Congestion Control. *IEEE Communications Magazine, Special Issue on ,Flow and Congestion Control‘*, Guest Editorial, Vol. 34, No. 11, November 1996, pp. 30-32.

- [415] ORLAMÜNDER, H.; BLEUSE, P.: Handling of Internet Traffic in Telecommunication Networks. *16th International Switching Symposium (ISS '97)*, Vol. 1, Toronto, September 1997, pp. 579-586.
- [416] PACK, C. D.; WHITAKER, B. A.: Kalman Filter Models for Network Forecasting. *The Bell System Technical Journal*, Vol. 61, No. 1, January 1982, pp. 1-14.
- [417] PANCHA, P.: ATM Switch Routers for Combined Connection-Oriented and Connectionless Transport. *16th International Switching Symposium (ISS '97)*, Vol. 1, Toronto, September 1997, pp. 261-269.
- [418] PANTLEON, K.: *Untersuchung dynamischer Vorgänge in einem ATM-Vermittlungsrechner*. Diplomarbeit Nr. 1482 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, Februar 1997.
- [419] PARK, K.: Warp Control: A Dynamically Stable Congestion Protocol and its Analysis. *ACM SIGCOMM Computer Communication Review*, Vol. 23, No. 4, October 1993, pp. 137-147.
- [420] PARK, W.-G.; RHEE, J.-H.; LEE, S.-J.; LEE, S.-H.: A Novel Overload Control Strategy for Distributed Mobile Communication Systems. *IEICE Transactions on Communications*, Vol. E81-B, No. 11, November 1998, pp. 2131-2140.
- [421] PASCH, H.-L.: *Design and Analysis of a Load-Adaptive High-Speed Network*. Dissertation (Proefschrift), Universiteit Twente, Nederlands, 1993.
- [422] PATEROK, M.: *Warteschlangensysteme mit Rückkopplung und Prioritäten*. Dissertation, Arbeitsberichte des Instituts für Mathematische Maschinen und Datenverarbeitung (Informatik), Bd. 23, Nr. 12, Universität Erlangen-Nürnberg, 1990.
- [423] PEEREN, R.: IN in the Hybrid Broadband Network: The Intelligent Web. *5th IEEE Intelligent Network Workshop (IN '96)*, Melbourne, Vol. 2, April 1996, Paper #9.2.3.
- [424] PENNEY, B. K.; WILLIAMS, J. W. J.: The Software Architecture for a Large Telephone Switch. *IEEE Transactions on Communications*, Vol. 30, No. 6, June 1982, pp. 1369-1378.
- [425] PERROS, H. G.; ELSAYED, K. M.: Call Admission Control Schemes: A Review. *IEEE Communications Magazine*, Vol. 34, No. 11, November 1996, pp. 82-91.
- [426] PETERSSEN, B.; RUCK, H.; BARTEL, A.: A Model to Determine the Effects of Repeated Calls on the Answer Bid Ratio and Revenue. *10th ITC Specialists Seminar on Control in Communications*, Lund, Sweden, September 1996, pp. 143-150.
- [427] PHAM, X. H.: Control Loop for Traffic Management of Networks under Focussed Overloads. *13th International Teletraffic Congress, Proceedings in 'North-Holland Studies in Telecommunication'*, Vol. 14, Copenhagen, June 1991, pp. 315-322.
- [428] PHAM, X. H.; BETTS, R.: Congestion Control for Intelligent Networks. *1992 International Zurich Seminar on Digital Communications – Intelligent Networks and their Applications*, Zurich, March 1992, pp. 375-389.
- [429] PHAM, X. H.; BETTS, R.: Congestion Control for Intelligent Networks. *Computer Networks and ISDN Systems*, Vol. 26, No. 5, January 1994, pp. 511-524.
- [430] PITSILLIDES, A.: *Control Structures and Techniques for Broadband-ISDN Communication Systems*. Doctor Thesis, Swinburne University of Technology, Melbourne, Australia, June 1993.
- [431] PRASAD, R. R. N.: Overload Traffic Management of E10B Switching System. *15th In-*

- ternational Teletraffic Congress, Workshop „Traffic Engineering for Developing Countries“, Washington, June 1997, pp. 43-54.*
- [432] PTACNIK, E.: SSP103 — Ein Zentralsteuerwerk im Elektronischen Fernsprech-Wählsystem EWS. *telcom report*, 1. Jhrg., Nr. 2/78, April 1978, S. 131-136.
- [433] RAMALHOTO, M. F.; AMARAL, J. A.; TERESA COCHITO, M.: A Survey of J. Little's Formula. *International Statistical Review*, Vol. 51, 1983, pp. 255-278.
- [434] RAMAMURTHY, G.; DIGHE, R. S.: Analysis of Multilevel Hierarchical Congestion Controls in B-ISDN. *14th International Teletraffic Congress*, Antibes Juan-les-Pins, June 1994, pp. 1281-1290.
- [435] RATHGEB, E. P.; FISCHER, W.; HINTERBERGER, C.; WALLMEIER, E.; WILLE-FIER, R.: The MainStreetXpress Core Services Node – A Versatile ATM Switch Architecture for the Full Service Network. *IEEE Journal on Selected Areas in Communication*, Vol. 15, No. 5, June 1997, pp. 795-806.
- [436] RATHGEB, E. P.: Redundancy Concepts for a Large ATM Switching Node. *16th International Switching Symposium (ISS '97)*, Vol. 1, Toronto, September 1997, pp. 425-433.
- [437] RATHGEB, E. P.: Architecture of a Multigigabit ATM Core Switch for the B-ISDN. *IEICE Transactions on Communications (Special Issue on „ATM Switching Systems for future B-ISDN“)*, Vol. E81-B, No. 2, Februar 1998, pp. 251-257.
- [438] REIM, T.: *Überlastabwehr in „Intelligenten Netzen“*. Diplomarbeit Nr. 1237 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, Sep. 1993.
- [439] RICHARDS, P. C.: No. 2 ESS Call Processing Capacity: Estimation, Measurement and Control. *6th International Switching Symposium (ISS '72)*, Cambridge, USA, June 1972, pp. 285-292.
- [440] RICHARDS, P. C.; HERNDON, J. A.: No. 2 ESS: An Electronic Switching System for the Suburban Community. *Bell Laboratories Record*, Vol. 51, May 1973, pp. 130-135.
- [441] RIGAULT, C.; KOVÁCIKOVÁ, T.: A Signalling Scheme for B-ISDN and IN Integration. *IEEE ATM'97 Workshop*, Lisboa, Portugal, May 1997, pp. 551-559.
- [442] RITTER, M.; TRAN-GIA, P.: Mechanismen zur Steuerung und Verwaltung von ATM-Netzen – Teil 1: Grundlegende Prinzipien. *Informatik-Spektrum*, Bd. 20, Nr. 4, August 1997, S. 216-224.
- [443] RITTER, M.; TRAN-GIA, P.: Mechanismen zur Steuerung und Verwaltung von ATM-Netzen – Teil 2: Modellierung und Leistungsbewertung. *Informatik-Spektrum*, Bd. 20, Nr. 5, Oktober 1997, S. 276-285.
- [444] ROCKSTRÖM, A.; ZDEBEL, B.: A Network Strategy for Survival. *16th International Switching Symposium (ISS '97)*, Vol. 1, Toronto, September 1997, pp. 361-367. Bzw. in *IEEE Communications Magazine*, Vol. 36, No. 1, January 1998, pp. 36-40.
- [445] ROHRS, C. E.; BARRY, R. A.; O'HALEK, S. J.: A Control Engineer's Look at ATM Congestion Avoidance. *IEEE GLOBECOM '95*, Vol. 2, Singapore, November 1995, pp. 1089-1094.
- [446] ROMBACH, H.: A Distributed Object-Oriented Control Software for Private ATM Switching Systems. *15th International Switching Symposium (ISS '95)*, Vol. 1, Berlin, April 1995, pp. 93-97.
- [447] ROSS, K. W.; TSANG, D. H. K.: The Stochastic Knapsack Problem. *IEEE Transactions on Communications*, Vol. 37, No. 7, July 1989, pp. 740-747.

- [448] ROZMARYN, C.; TRELUT, J.; VIELLEVOYE, L.: The E11 Switching System. *7th International Switching Symposium (ISS '74)*, München, September 1974, pp. 134/1-134/8.
- [449] RUMSEWICZ, M.: On the Efficacy of Using the Transfer-Controlled Procedure During Periods of STP Processor Overload in SS7 Networks. *IEEE Journal on Selected Areas in Communication*, Vol. 12, No. 3, April 1994, pp. 415-423.
- [450] RUMSEWICZ, M. P.: Critical Congestion Control Issues in the Evolution of Common Channel Signaling Networks. *14th International Teletraffic Congress*, Anitbes Juan-les-Pins, June 1994, pp. 115-124.
- [451] RUMSEWICZ, M. P.; SMITH, D. E.: A Comparison of SS7 Congestion Control Options During Mass Call-In Situations. *IEEE/ACM Transactions on Networking*, Vol. 3, No. 1, February 1995, pp. 1-9.
- [452] RYDÉN, T.: Optimal Overload Control of Systems with Access Buffers – A Difficult Problem. *Computer Networks and ISDN Systems*, Vol. 24, No. 9-10, May 1998, pp. 999-1005.
- [453] SABOURIN, T.; FICHE, G.; LIGEOUR, M.: Overload Control in a Distributed System. *13th International Teletraffic Congress, Proceedings in 'North-Holland Studies in Telecommunication'*, Vol. 14, Copenhagen, June 1991, pp. 421-427.
- [454] SAITO, H.; SHIOMOTO, K.: Dynamic Call Admission Control in ATM Networks. *IEEE Journal on Selected Areas in Communication*, Vol. 9, No. 7, September 1991, pp. 982-981.
- [455] SAITO, H.: Signalling-Free ATM Core Network. *7th International Network Planning Symposium*, Sydney, November 1996, pp. 69-74.
- [456] SAITO, H.; Nakagawa, S.; Kato, Y.; Ohta, S.: Innovation of Circuit/Path Operations in ATM Networks – Self-Sizing Network. *NTT REVIEW*, Vol. 8, No. 1, January 1996, pp. 57-65.
- [457] SAITO, H.; ET AL.: Performance Issues in Public ABR Service. *IEEE Communications Magazine*, Vol. 34, No. 11, November 1996, pp. 82-91.
- [458] SAITO, H.: Dynamic Resource Allocation in ATM Networks. *IEEE Communications Magazine*, Vol. 35, No. 5, May 1997, pp. 146-153.
- [459] SAITO, H.: Congestion Control through Network Money. *IEICE Transactions on Communications*, Vol. E80-B, No. 7, July 1997, pp. 991-999.
- [460] SAKURAI, Y.; HIROSHIMA, S.; TANABE, S.: ATM Switching System for B-ISDN. *Hitachi Review*, Vol. 40, No. 3, 1991, pp. 193-198.
- [461] SCHMIDT, B.: Die Bestimmung von Konfidenzintervallen in der Simulation stochastischer zeitdiskreter Systeme. *Elektronische Rechenanlagen*, 24. Jhrg., Heft Nr. 3, 1982, pp. 118-124.
- [462] SCHNEPS-SCHNEPPE, M.; SCHWETZOV, J.: On Standard Call Models and Network GOS. *2nd St. Petersburg Regional International Teletraffic Seminar*, St. Petersburg, June/July 1995, pp. 578-596.
- [463] SCHOEN, U.; ET AL.: Convergence Between Public Switching and the Internet. *16th International Switching Symposium (ISS '97)*, Vol. 1, Toronto, September 1997, pp. 549-560. Bzw. in *IEEE Communications Magazine*, Vol. 36, No. 1, January 1998, pp. 50-65.
- [464] SCHOUTE, F. C.: Optimal Control and Call Acceptance in a SPC Exchange. *9th International Teletraffic Congress*, Vol. 2, Torremolinos, October 1979.

- [465] SCHOUTE, F. C.: Adaptive Overload Control Schemes for an SPC Exchange. *10th International Teletraffic Congress*, Montreal, June 1983, Paper 5.2-#3.
- [466] SCHRAMMEL, F. J.; VAN'T SLOT, A. W.: General Introduction to PRX 205. *6th International Switching Symposium (ISS '72)*, Cambridge, USA, June 1972, pp. 348-357.
- [467] SCHRÖDER, M.: Einführung in die kurzfristige Zeitreihenprognose und Vergleich der einzelnen Verfahren. In „*Prognoserechnung*“ (Herausgeber P. Mertens), Physica-Verlag, Würzburg, Wien, 1981, S. 23-33.
- [468] SCHRODI, K. J.; HEY, H.-J.: Integration of IP Packet Forwarding in an ATM Switch. *16th Int. Switching Symposium (ISS '97)*, Vol. 1, Toronto, September 1997, pp. 247-254.
- [469] SCHWARZ, A.; DOMSCHITZ, P.; TOTZKE, J.: Modelling and Nonstationary Simulations of a Control Platform for a Public B-ISDN Switching System. *4th International Conference on Telecommunications*, Vol. 3, Melbourne, April 1997, pp. 1181-1186.
- [470] SCHWARZ, A.; DOMSCHITZ, P.; TOTZKE, J.; VON DER STRATEN, G.; KREILINGER, R.: Overload Protection for Public B-ISDN Switching Systems. *15th International Teletraffic Congress*, Vol. 2, Washington, June 1997, pp. 921-932.
- [471] SCHWARZ, A.: Overload Indicators and Strategies for a B-ISDN/ATM Switching System. *6th Open Workshop on High Speed Networks*, Stuttgart, October 1997, pp. 125-132.
- [472] SCHWARZ, A.; DOMSCHITZ, P.; VAN DER VEKENS, A.: Performance Modelling and Simulations Studies of a Broadband Service Switching Point. *5th Inter. Conference on the Performance and Management of Complex Communication Networks, Workshop 3: „Queueing Theorie and its Applications“*, Tsukuba, Japan, November 1997, pp. 46-57.
- [473] SCHWARZ, A.: Methodik der Instationären Simulation – Einführung, Werkzeugunterstützung und bisherige Arbeiten am IND. Universität Stuttgart, Institut für Nachrichtenvermittlung und Datenverarbeitung, *Interner Bericht*, Nr. 23, April 1998.
- [474] SCHWARZ, A.: Instationäre Simulation – Teil 2 (Beispiel für ein einfaches Simulationsmodell; G/G/1/5-System und Little'sches Gesetz). Universität Stuttgart, Institut für Nachrichtenvermittlung und Datenverarbeitung, *Interner Bericht*, Nr. 24, Mai 1998.
- [475] SCHWARTZ, M.; SAAD, S.: Analysis of Congestion Control Techniques in Computer Communication Networks. *International Symposium on Flow Control in Computer Networks*, Versailles, February 1979, pp. 113-130.
- [476] SCHWERTFEGER, H.-J.: Strukturierung von Steuerungen für Vermittlungssysteme mit Zeitteilung nach dem Durchschalteprinzip. *Nachrichtentechnik Elektronik*, 34. Jhrg., Nr. 8, VEB Verlag Technik, Berlin-Ost, 1984, S. 306-309.
- [477] SCHWERTFEGER, H.-J.: Kommunikationseinrichtungen in verteilten Strukturen zur Steuerung von Vermittlungsprozessen. *Nachrichtentechnik Elektronik*, 34. Jhrg., Nr. 10, VEB Verlag Technik, Berlin-Ost, 1984, S. 368-371.
- [478] SELGA, J. M.: New Flow Control Power is Decentralizable and Fair. *3rd IEEE Conference on Computer Communications (INFOCOM '84)*, San Francisco, 1984, pp. 87-94.
- [479] SENOH, A.; MIZUNO, T.: Multiprocessor Architecture for Large-Capacity ATM Switching System. *15th International Switching Symposium (ISS '95)*, Vol. 1, Berlin, April 1995, pp. 263-267.
- [480] SERAJ, J.: An Analysis of Processor Load Control in SPC Systems. *11th International Teletraffic Congress*, Kyoto, September 1985, Paper #4.3B-5.



- [481] SIEGELE, L.: Handel mit heißer Luft: Der amerikanische Forscher Bernardo Huberman bringt Computersystemen die Grundregeln der Marktwirtschaft bei – mit verblüffendem Erfolg. *DIE ZEIT*, Nr. 27, Juni 1997.
- [482] SIEMENS AG: Voice and Leased Line Services Over ATM: An Architectural Overview. *Siemens, Public Communication Networks*, Whitepaper, December 1997.
- [483] SINGER, J.; KÖHN, I.; SCHINK, H.: Narrowband Services over ATM Networks: Evaluation of Trunking Methods. *16th International Switching Symposium (ISS '97)*, Vol. 2, Toronto, September 1997, pp. 17-25.
- [484] SHARMA, V.; MAZUMDAR, R.: Estimating Traffic Parameters in Queueing Systems with Local Information. *Performance Evaluation*, Vol. 32, No. 13, April 1998, pp. 217-231.
- [485] SHENKER, S.: Fundamental Design Issues for the Future Internet. *IEEE Journal on Selected Areas in Communications*, Vol. 13, No. 7, September 1995, pp. 1176-1188.
- [486] SHIN, K. G.; CHANG, Y.-C.: Load Sharing in Distributed Real-Time Systems with State-Change Broadcasts. *IEEE Transactions on Computer*, Vol. 38, No. 8, August 1989, pp. 1124-1142.
- [487] SHIODA, S.; SAITO, H.: Real-Time Cell Loss Ratio Estimation and Its Application to ATM Traffic Controls. *16th IEEE Conference on Computer Communications (INFOCOM '97)*, Vol. 3, Kobe, April 1997, pp. 1072-1079.
- [488] SHIRAZI, B. A.; HURSON, A. R.; KAVI, K. M.: *Scheduling and Load Balancing in Parallel and Distributed Systems*. IEEE Computer Society Press, Los Alamitos, 1995.
- [489] SHIVARATRI, N. G.; KRUEGER, P.; SINGHAL, M.: Load Distributing for Locally Distributed Systems. *IEEE Computer*, Vol. 25, No. 12, December 1992, pp. 33-44.
- [490] SIU, K.-Y.; TZENG, H.-Y.: Congestion Control for Multicast Service in ATM Networks. *IEEE GLOBECOM '95*, Singapore, November 1995, pp. 310-314.
- [491] SKLIROS, A.: IP Switching over ATM Networks. *6th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks (ATM '98)*, Ilkley, UK, July 1998, Tutorial Paper #T01.
- [492] SMITH, D. E.: Effects of Feedback Delay on the Performance of the Transfer-Controlled Procedure in Controlling CCS Network Overloads. *IEEE Journal on Selected Areas in Communication*, Vol. 12, No. 3, April 1994, pp. 424-432.
- [493] SMITH, D. E.: Preventing Release Message Avalanches in Common Channel Signaling Networks. *14th International Teletraffic Congress*, Antibes Juan-les-Pins, June 1994, pp. 105-114.
- [494] SMITH, D. E.: Ensuring Robust Call Throughput and Fairness for SCP Overload Controls. *IEEE/ACM Transactions on Networking*, Vol. 3, No. 5, October 1995, pp. 538-548.
- [495] SOMOZA, P.; GUERRERO, A.: Dynamic Processor Overload Control and its Implementation in Certain Single-Processor and Multiprocessor SPC Systems. *9th International Teletraffic Congress*, Vol. 2, Torremolinos, October 1979.
- [496] SOMOZA, P.; GUERRERO, A.: Dynamic Processor Overload Control for SPC Switching Systems. *Electrical Communication*, Vol. 55, No. 1, 1980, pp. 37-45. Bzw.: Dynamische Überlastabwehrstrategie für speicherprogrammierte Vermittlungssysteme. *Elektrisches Nachrichtenwesen*, Bd. 55, Nr. 1, 1980, S. 37-45.
- [497] DE SOUZA, J. M.; ET AL.: Window Based Overload Control in Fully Distributed Sy-

- stems. *15th International Teletraffic Congress, Workshop „Traffic Engineering for Developing Countries“*, Washington, June 1997, pp. 83-92.
- [498] SPENCER, A. E.; VIGILANTE, F. S.: No. 2 ESS – System Organization and Objectives. *The Bell System Technical Journal*, Vol. 48, No. 8, October 1948, pp. 2607-2618.
- [499] STANKOVIC, J. A.: A Perspective on Distributed Computer Systems. *IEEE Transactions on Computers*, Vol. 33, No. 12, December 1984, pp. 1102-1115.
- [500] STIDHAM, S.: Optimal Control of Admission to a Queueing System. *IEEE Transactions on Automatic Control*, Vol. 30, No. 8, August 1985, pp. 705-713.
- [501] STONE, H. S.: Critical Load Factors in Two-Processor Distributed Systems. *IEEE Transactions on Software Engineering*, Vol. 4, No. 3, May 1978, pp. 254-258.
- [502] STOREY, T. F.: Design of a Microprogram Control for a Processor in an Electronic Switching System. *The Bell System Technical Journal*, Vol. 55, No. 2, Februar 1976.
- [503] SUCKFÜLL, H.: Architektur einer neuen Linie digitaler öffentlicher Fernsprechvermittlungen. *telcom report*, 2. Jhrg., Nr. 3/79, 1979, S. 174-183.
- [504] SUMITA, S.; NAKAGAWA, S.: Performance Specification, Call-Processing Capacity, and Overload Control for SPC Systems with Extremely High Peak Traffic. *14th International Teletraffic Congress*, Antibes Juan-les-Pins, June 1994, pp. 253-262.
- [505] SUN, H.; HUANG, K.; LI, L.: Supporting IP on the ATM Networks: An Overview. *Computer Communications*, Vol. 21, No. 11, August 1998, pp. 1020-1021.
- [506] SUZUKI, K.; ET AL.: An ATM Switching System – Development and Evaluation. *NEC Research & Development*, Vol. 31, April 1991, pp. 242-251.
- [507] SUZUKI, S.; ISHIKAWA, H.: The NS 8000 Series and Multimedia-era Networks. *NTT REVIEW*, Vol. 9, No. 1, January 1997, pp. 56-66.
- [508] SZÉKELY, S.; MOLDOVÁN, I.; SIMON, C.: Overload Generated by Signalling Message Flows in ATM Networks. *IFIP TC6/WG7.3 7th International Conference on Performance of Information and Communication Systems (PICS '98)*, Lund, Sweden, May 1998, pp. 51-64.
- [509] SZÉKELY, S.; SIMON, C.; SZÜCS, G.: Performance Analysis of Switched Virtual Connections in ATM Networks. *6th IFIP Workshop on Performance Modelling and Evaluation of ATM Networks (ATM '98)*, Ilkley, UK, July 1998, Paper #102.
- [510] TAKANO, M.; SAITO, I.: Input Regulation Control of Communication Systems – A Fuzzy Logic Approach –. *13th Int. Teletraffic Congress, Workshop, Proceedings in 'North-Holland Studies in Telecommunication'*, Vol. 17, Copenhagen, June 1991, pp. 411-416.
- [511] TAKATORI, M.; ET AL.: A Multi-Functional Large-Scale ATM Switch Architecture. *15th International Switching Symposium (ISS '95)*, Vol. 1, Berlin, April 1995, pp. 489-493.
- [512] TAKATORI, M.; ET AL.: Hierarchical System Architecture for Multi-Service and Multi-Rate Traffic in Broadband ISDN. *16th International Switching Symposium (ISS '97)*, Vol. 1, Toronto, September 1997, pp. 231-237.
- [513] TAKEI, T.; MAKI, H.: Damage from the Great Hanshin/Awaji Earthquake and NTT's Efforts to Recover from It. *NTT REVIEW*, Vol. 7, No. 5, September 1995, pp. 8-16.
- [514] TANG, H.; SIMULA, O.: The Adaptive Resource Assignment and Optimal Utilization of Multi-Service SCP. *4th International Conference on Intelligence in Networks*, Bordeaux, November 1996, pp. 235-240.

- [515] THEIMER, T.: *Vergleichende Untersuchungen an ATM-Koppelnetzstrukturen*. 56. Bericht über verkehrstheoretische Arbeiten (Dissertationsschrift), Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1994.
- [516] THEIS, R.: *Instationäre Simulation von Protokollen der Fertigungsautomatisierung*. Semesterarbeit Nr. 973 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, Januar 1990.
- [517] THOMPSON, K.; MILLER, G. J.; WILDER, R.: Wide-Area Internet Traffic Patterns and Characteristics. *IEEE Network Magazine*, Vol. 11, No. 6, Nov./Dec. 1997, pp. 10-23.
- [518] TIPPER, D.; SUNDARESHAN, M. K.: Numerical Methods for Modeling Computer Networks Under Nonstationary Conditions. *IEEE Journal on Selected Areas in Communication*, Vol. 8, No. 9, December 1990, pp. 1682-1695.
- [519] TOKUNAGA, H.; KAWANO, H.; NAKAJIMA, S.: Traffic Congestion Control System. *REVIEW of the Electrical Communications Laboratories (NTT)*, Vol. 36, No. 2, March 1988, pp. 189-193.
- [520] TOMANIK, J.: *Instationäre Simulation von Warteschlangenmodellen*. Semesterarbeit Nr. 1021 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, August 1990.
- [521] TRAN-GIA, P.; VAN HOORN, M. H.: Dependency of Service Time on Waiting Time in Switching Systems – A Queueing Analysis with Aspects of Overload Controls. *Forschungsbericht*, Universität Siegen - Freie Universität Amsterdam, 1982.
- [522] TRAN-GIA, P.: *Überlastprobleme in rechnergesteuerten Fernsprechvermittlungssystemen – Modellbildung und Analyse*. 36. Bericht über verkehrstheoretische Arbeiten (Dissertationsschrift), Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1982.
- [523] TRAN-GIA, P.: Subcall-Oriented Modelling of Overload Control in SPC Switching Systems. *10th International Teletraffic Congress*, Montreal, June 1983, Paper 5.2-#1.
- [524] TRAN-GIA, P.: Simulation of Instationary Processes for Performance Evaluations of Switching Systems. *1st European Simulation Congress*, Aachen, September 1983, in *Informatik-Fachberichte*, Bd. 71, Springer-Verlag, pp. 362-367.
- [525] TRAN-GIA, P.: Analysis of a Load-Driven Overload Control Mechanism in Discrete-Time Domain. *12th International Teletraffic Congress*, Torino, June 1988, pp. 4.3A2.
- [526] TRAN-GIA, P.: A Class of Renewal Interrupted Poisson Processes and Applications to Queueing Systems. *Operations Research*, Vol. 32, 1988, pp. 231-250.
- [527] TRAN-GIA, P.; RATHGEB, E.: Performance Analysis of Load-Balancing Semidynamic Scheduling Mechanisms in Distributed Systems. *Archiv für Elektronik und Übertragungstechnik (AEÜ)*, Bd. 43, H. 1, Januar/Februar 1989, S. 38-45.
- [528] TROUP, H.: Radio Contest Overloads Phone System. *ACM SIGSOFT, Software Engineering Notes*, Vol. 15, Issue 70, March 1994. Bzw.: <http://catless.ncl.ac.uk/Risks/15.70.html> (Forum on Risks to the Public in Computers and Related Systems).
- [529] TSOLAS, N.; ABDO, G.; BOTTHEIM, R.: Performance and Overload Considerations when Introducing IN into an Existing Networks. *Int. Zurich Seminar on Digital Communications – Intelligent Networks and their Applications*, Zurich, March 1992, pp. 407-414.
- [530] TURNER, J.; YAMANAKA, N.: Architectural Choices in Large Scale ATM Switches.

- IEICE Transactions on Communications (Special Issue on 'ATM Switching Systems for future B-ISDN')*, Vol. E81-B, No. 2, Februar 1998, pp. 120-137.
- [531] TURNER, P. M. D.; KEY, P. B.: A New Call Gapping Algorithm for Network Traffic Management. *13th International Teletraffic Congress, Proceedings in 'North-Holland Studies in Telecommunication'*, Vol. 14, Copenhagen, June 1991, pp. 121-126.
- [532] ULMER, J.: *Routingverfahren mit Ressourcenmanagement für Breitband-ISDN-Systeme*. Fortschr.-Ber. VDI Reihe 10, Nr. 450, VDI Verlag, Düsseldorf, 1996.
- [533] UNDERWOOD, D. J.: Using ATM-based Tandem Switches to Consolidate Narrowband Trunking. *16th International Switching Symposium (ISS'97)*, Vol. 2, Toronto, September 1997, pp. 27-34.
- [534] VAN BRUSSEL, M.; CAMPOS FLORES, A.: SYSTEM 12 – Ortsvermittlungsstellen. *Elektrisches Nachrichtenwesen*, Band 56, Nr. 2/3, 1981, S. 218-234.
- [535] VANDENHOUTE, M.: IP-Verkehr in Netzen unterschiedlicher Größe. *Alcatel Telecom Rundschau*, 4. Quartal, 1997, S. 294-299.
- [536] VAN DER VEKENS, A.; URBAN, J., HUSSMANN, H.; FABRELLAS, A.; VENIERIS, I.; ZIZZA, F.: Object-Oriented Realization of a Broadband Intelligent Network Architecture. *16th Int. Switching Symposium (ISS'97)*, Vol. 1, Toronto, September 1997, pp. 97-104.
- [537] VAN DER VEKENS, A.; SCHWARZ, A.: INSIGNIA – A New Approach towards Broadband Intelligent Networks. *6th Open Workshop on High Speed Networks*, Stuttgart, October 1997, pp. 271-277.
- [538] VASILAKOS, A. V.; PAXIMADIS, C. T.; MARITSAS, D. G.: Congestion Avoidance Scheme for ISDN Frame Relay Networks. *Computer Communications*, Vol. 16, No. 11, November 1993, pp. 689-694.
- [539] VEERARAGHAVAN, M.: Connection Control in ATM Networks. *Bell Labs Technical Journal*, Vol. 2, No. 1, Winter 1997, pp. 48-64.
- [540] VIGILANTE, F. S.: General Description of No. 2 Electronic Switching System. *IEE Conference on Switching Techniques for Telecommunications Networks*, London, April 1969, pp. 479-482.
- [541] VILLAR DE VILLACIAN, J. E.: Traffic Calculations in SPC Systems. *8th International Teletraffic Congress*, Melbourne, November 1976, Paper #611. Bzw.: Verkehrstheoretische Berechnungen von SPC-Systemen. *Elektrisches Nachrichtenwesen*, Bd. 52, Nr. 3, 1977, S. 260-267.
- [542] VILLAR DE VILLACIAN, J. E.; SOTO, O. G.: Simulation Methods for Estimating Traffic Capacity of SPC Systems. *2th ITC Seminar on Traffic Measurement and Traffic Simulation and their Role in Telecommunications System Planning*, Budapest, Hungary, October 1978, pp. 131-152.
- [543] VILLEN-ALTAMIRANO, M.; MORALES-ANDRES, G.; BERMEJO-SAEZ, L.: An Overload Control Strategy for Distributed Control Systems. *11th International Teletraffic Congress*, Kyoto, September 1985, Paper #4.4B-5.
- [544] WALLSTRÖM, B.: A Feedback Queue with Overload Control. *10th International Teletraffic Congress*, Montreal, June 1983, Paper 1.3-#4.
- [545] WALLSTRÖM, B.; VOIGT, H.: Transient Behaviour of Simple Overload Strategies for SPC Switching Systems – An Analytic Approach. *12th International Teletraffic Congress*, Torino, June 1988, pp. 4.3A1.

- [546] WALLSTRÖM, B.; NYBERG, C.: Transient Model of Overload Control and Priority Service in SPC-Systems. *13th International Teletraffic Congress, Proceedings in 'North-Holland Studies in Telecommunication'*, Vol. 14, Copenhagen, June 1991, pp. 429-434.
- [547] WANG, Y.-T.; MORRIS, R. J. T.: Load Sharing in Distributed Systems. *IEEE Transactions on Computers*, Vol. 34, No. 3, March 1985, pp. 204-217.
- [548] WEIGEL, I.: *Modelle für Entwurf und Bewertung eines dynamischen Lastverbunds lose gekoppelter Rechensysteme*. Dissertation am Institut für Mathematische Maschinen und Datenverarbeitung, Universität Erlangen-Nürnberg, Juli 1994.
- [549] WEISSCHUH, H.: *Entwicklung der Steuersoftware für eine rechnergesteuerte Vermittlungsstelle*. 24. Bericht über verkehrstheoretische Arbeiten (Dissertationsschrift), Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1977.
- [550] WELCH, P.: The Statistical Analysis of Simulation Results. In *Computer Performance Modeling Handbook*, Edited by S. S. Lavenberg, Academic Press, New York, 1983, pp. 268-329.
- [551] WELLER, A.: *Instationäre Simulation von FDDI*. Diplomarbeit Nr. 1129 am Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, Januar 1992.
- [552] WHITE, P. E.: The Changing Role of Switching Systems in the Telecommunications Network. *IEEE Communications Magazine*, Vol. 31, No. 1, January 1993, pp. 12-13.
- [553] WILDLING, K.; KARLSTEDT, T.: Call Handling and Control of Processor Load in a SPC – A Simulation Study. *9th Int. Teletraffic Congress*, Vol. 2, Torremolinos, October 1979.
- [554] WILE, G. E.; GOWAN, D. S.: The Architecture of the DPN Data Networking System. *8th International Conference on Computer Communications (ICCC '86)*, Munich, September 1986, pp. 365-369.
- [555] WILLMANN, G.; KÜHN, P. J.: Performance Modeling of Signaling System No. 7. *IEEE Communications Magazine*, Vol. 28, No. 7, July 1990, pp. 44-56.
- [556] WIZGALL, M.: *Über Architektur, Betriebsweise und Verkehrsverhalten der Steuerung einer rechnergesteuerten Vermittlungsstelle*. 27. Bericht über verkehrstheoretische Arbeiten (Dissertationsschrift), Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1980.
- [557] WOLFF, R. W.: Poisson Arrivals See Time Averages. *Operations Research*, Vol. 30, No. 2, March-April 1982, pp. 223-231.
- [558] WOODRUFF, G. M.; ROGER, R. G. H.; RICHARDS, P. S.: A Congestion Control Framework for High Speed Integrated Packetized Transport. *IEEE GLOBECOM '88*, Hollywood, Florida, November/December 1988, pp. 203-207.
- [559] WU, C.-S.; JIAU, J.-C.; CHEN, K.-J.; CHOY, M.: Minimizing Call Setup Delay in ATM Networks via Optimal Processing Capacity Allocation. *IEEE Communications Letters*, Vol. 2, No. 4, April 1998, pp. 110-112.
- [560] YAJIMA, T.; SUDA, K.: Switching System Software Design for ISDN. *IEEE Journal on Selected Areas in Communication*, Vol. 4, No. 8, November 1986, pp. 1222-1229.
- [561] YANG, C.-Q.; REDDY, A. V. S.: A Taxonomy for Congestion Control Algorithms in Packet Switching Networks. *IEEE Network Magazine*, Vol. 9, No. 5, July/August 1995, pp. 34-45.
- [562] YANOFF, L.: Die Vermittlungsplattform Alcatel 1100 HSS. *Alcatel Telecom Rundschau*, 4. Quartal, 1996, S. 242-256.

- [563] YANOFF, L.: Das Internet der nächsten Generation. *Alcatel Telecom Rundschau*, 4. Quartal, 1997, S. 246-255.
- [564] YASHIRO, Z.; TANAKA, T.; DOI, Y.: The Next-Generation ATM Switching Architecture for Multimedia Communications. *IEICE Transactions on Communications (Special Issue on ,ATM Switching Systems for future B-ISDN')*, Vol. E81-B, No. 2, Februar 1998, pp. 209-214.
- [565] YEO, H. G.; SONG, K. S.; RO, S. H.; KI, J. G.: Performance Evaluation of the Call Connection Processor in Distributed ATM Exchange System. *4th International Conference on Telecommunications*, Vol. 1, Melbourne, April 1997, pp. 161-166.
- [566] YIM, C. H.; HARTMANN; H. L.: Throughput Behavior of Switching Systems under Heavy Load Conditions. *11th International Teletraffic Congress*, Kyoto, September 1985, Paper #4.3B-2.
- [567] YOKOTANI, T.; ICHIHASHI, T.; TONAMI, S.: New Signaling Procedures in ATM Networks for Advance Resource Reservation. *11th ITC Specialist Seminar on Multimedia and Nomadic Communications*, Yokohama, October 1998, pp. 25-32.
- [568] YONESHIGE, T.; IGARASHI, K.: Great Hanshin/Awaji Earthquake Results in Revised Contingency Plans. *NTT REVIEW*, Vol. 8, No. 1, January 1996, pp. 79-84.
- [569] YUM, T.: Measuring the Utilization of a Synchronous Data Link: An Application of Busy-Period Analysis. *The Bell System Technical Journal*, Vol. 59, No. 5, May-June 1980, pp. 731-744.
- [570] YUMIBA, H.: Towards Intelligent and Flexible Telecommunications. *4th International Conference on Intelligence in Networks (ICIN'96)*, Bordeaux, November 1996.
- [571] ZEGURA, E. W.: Architectures for ATM Switching Systems. *IEEE Communications Magazine*, Vol. 31, No. 2, February 1993, pp. 28-37.
- [572] ZEPF, J.; WILLMANN, G.: Transient Analysis of Congestion and Flow Control Mechanisms in Common Channel Signalling Networks. *13th International Teletraffic Congress*, Copenhagen, June 1991, pp. 413-419.
- [573] ZEPF, J.; HUBER, M. N.; DASCH, B.: Overload Control Strategies for Packet Switching. *Telecommunications*, International Edition, Vol. 26, No. 5, May 1992, pp. 47, 83-88.
- [574] ZEPF, J.; RUFA, G.: Congestion and Flow Control in Signalling System No. 7 – Impacts of Intelligent Networks and New Services. *IEEE Journal on Selected Areas in Communication*, Vol. 12, No. 3, April 1994, pp. 501-509.
- [575] ZEPF, J.: *Modellierung und Bewertung von Überlastabwehrmechanismen in Signaliernetzen*. 61. Bericht über verkehrstheoretische Arbeiten (Dissertationsschrift), Institut für Nachrichtenvermittlung und Datenverarbeitung, Universität Stuttgart, 1995.
- [576] ZHOU, S.: A Trace-Driven Simulation Study of Dynamic Load Balancing. *IEEE Transactions on Software Engineering*, Vol. 14, No. 9, September 1988, pp. 1327-1341.

## Patente

Alle wichtigen Patente auf dem Gebiet der Vermittlungstechnik, und insbesondere auch für Lastregelalgorithmen für Vermittlungssteuerungen, werden auch in den USA angemeldet. Die durchgeführte Recherche umfasste deshalb in erster Linie U.S.-Patente.

- [577] ALBAL, N. A.; GOYAL, P. K.: Arbitration in Multiprocessor Communication Node. *U.S. Patent No. 4949338*, August 1990.
- [578] ARDON, M. T.; ET AL.: Call Processing Method for Distributed Switching. *U.S. Patent No. 5119366*, June 1992.
- [579] BAHR, J. E.; ET AL.: Process for Dispatching Tasks among Multiple Information Processor. *U.S. Patent No. 5109512*, April 1992.
- [580] BODNAR, B. L.: Method for Overload Control in a Packet Switch that Processes Packet Streams having different Priority Levels. *U.S. Patent No. 5539729*, July 1996.
- [581] CLARK, M. W.; SIEGMUND, W.; WUNDERLICH, H.-D.: Communication System having a Multiprocessor System serving the purpose of Central Control. *U.S. Patent No. 5327419*, July 1994.
- [582] CYR, B. L.; KAUFMAN, J. S.; LEE, T. P.: Load Balancing and Overload Control in a Distributed Processing Telecommunications System. *U.S. Patent No. 4974256*, November 1990.
- [583] DEBRULER, D. L.; ET AL.: Twin-Threshold Load-Sharing System with each Processor in a Multiprocessor Ring adjusting its own assigned Task List based on Workload Threshold. *U.S. Patent No. 5155858*, October 1992.
- [584] DIAZ, A. H.; MILITO, R. A.: System and Method for Integrated Overload Control and Message Distribution. *U.S. Patent No. 5754786*, May 1998.
- [585] EGUCHI, H.; ET AL.: Method and Apparatus for Controlling Job Transfer between Computer Systems. *U.S. Patent No. 4839798*, June 1989.
- [586] GAO, Z.; ET AL.: Overload Control for a Central Processor in the Switching Network of a Mobile Communications System. *U.S. Patent No. 5548533*, August 1996.
- [587] GREWAL, K.; ET AL.: System for Load Balancing between Message Processors by Routing all Queued Messages to a Particular Processor selected by a Deterministic Rule. *U.S. Patent No. 5592672*, January 1997.
- [588] HANSELKA, P.; OEHLERICH, J.: Circuit Arrangement for Telecommunications Switching Systems, particularly Telephone Switching Systems with Information Processing Switching Device and Processing Load Limiting. *U.S. Patent No. 4790005*, December 1988.
- [589] HESSE, D.: Digital Telecommunications Switching System. *U.S. Patent No. 5434851*, July 1995.
- [590] HIDAKA, T.; KANEKO, Y.: Load Control System for Controlling a Plurality of Processes and Exchange System having Load Control System. *U.S. Patent No. 5425086*, June 1995.
- [591] JONSSON, B. E. R.; AVSAN, O.: Multi-Processing Stored Program Controlled Telecommunication Establishment. *U.S. Patent No. 4731817*, March 1988.
- [592] KRISHNAN, K. R.: Routing an Incoming Data Stream to Parallel Processing Stations. *U.S. Patent No. 5155851*, October 1992.
- [593] McMILLEN, R. J.: Load Balancing for Packet Switching Nodes. *U.S. Patent No. 4621359*, November 1986.
- [594] MORALES, J.; ET AL.: Software Overload Control Method. *U.S. Patent No. 5440741*, August 1995.

- [595] NAGANUMA, J.; OGURA, T.: Multiprocessor System and a Method of Load Balancing thereof. *U.S. Patent No. 5241677*, August 1993.
- [596] SAMUELSON, D. A.: System for Regulating Arrivals of Customers to Servers. *U.S. Patent No. 4858120*, August 1989.
- [597] SAND, D. S.: Architecture for Distributed Control Telecommunication Systems. *U.S. Patent No. 4720854*, January 1988.
- [598] SCHOUTE, F. C.: Method of Overload Control for the Central Control of a Telecommunication System and Arrangement for Implementing the Method. *U.S. Patent No. 4809318*, February 1989.
- [599] SERVI, L. D.; KEILSON, J.: Control Process for Allocating Services in Communications Systems. *U.S. Patent No. 5381546*, January 1995.
- [600] STOCKDALE, R.: Overload Detection and Control System for a Telecommunications Exchange. *U.S. Patent No. 4511762*, April 1985.
- [601] WARTY, P.; ET AL.: Call Control of a Distributed Processing Communications Switching System. *U.S. Patent No. 4827499*, May 1989.
- [602] WEBER, A; CLARK, M.: Method for Load Distribution among the Central Processors of a Multiprocessor Central Control Unit of a Switching System. *U.S. Patent No. 4807281*, 1989.
- [603] WITTE, M.; OEHLERICH, J.; HELD, W.: Verfahren zum Lastausgleich in einem Multiprozessorsystem. *Europäische Patentanmeldung Nr. 0645702*, September 1993.
- [604] WITTE, M.; OEHLERICH, J.; HELD, W.: Method for Load Balancing in a Multi-Processor System where arising Jobs are Processed by a Plurality of Processors under Real-Time Conditions. *U.S. Patent No. 5655120*, August 1997.
- [605] YOO, C. H.; ET AL.: Method for Controlling Overload of Distributed Processors of Full Electronic Switching System. *U.S. Patent No. 5513255*, April 1996.
- [606] YOO, C. H.; ET AL.: Method for Controlling Overload in Centralized and Distributed Operating Hybrid Switching System. *U.S. Patent No. 5513257*, April 1996.
- [607] YOO, C. H.; ET AL.: Method for Controlling Overload of Main Processor of Distributed Switching System with Hierarchy Structure. *U.S. Patent No. 5574770*, November 1996.

## **Empfehlungen, Normen, Standards**

### **ATM-Forum<sup>1</sup>:**

- [608] ATM FORUM, TECHNICAL COMMITTEE: *BISDN Inter Carrier Interface (B-ICI) V2.0*. ATM Forum, December 1995.
- [609] ATM FORUM, TECHNICAL COMMITTEE: *Interim Inter-Switch Signaling Protocol (IISP) Specification Version V1.0*. ATM Forum, December 1994.
- [610] ATM FORUM, TECHNICAL COMMITTEE: *Private Network-Network Interface V1.0*. ATM Forum, March 1996.

---

<sup>1</sup> Das ATM-Forum wurde im September 1991 von CISCO, NET/Adaptive, Northern Telecom und US-Sprint gegründet. Mittlerweile (Stand: Februar 1998) gehören im mehr als 700 stimmberechtigte Mitglieder an.



- [611] ATM FORUM, TECHNICAL COMMITTEE: *ATM User-Network Interface Signalling V4.0*. ATM Forum, June 1996.
- [612] ATM FORUM, TECHNICAL COMMITTEE: *Traffic Management Specification V4.0*. ATM Forum, April 1996.

**Nationale Gremien:**

- [613] INFORMATIONSTECHNISCHE GESELLSCHAFT (ITG) IM VDE<sup>2</sup>: *ISDN-Begriffe*. ITG-Empfehlung 1.6-01, 1987.
- [614] INFORMATIONSTECHNISCHE GESELLSCHAFT (ITG) IM VDE: *Architektur und Verfahren der Vermittlungstechnik*. Entwurf der ITG-Empfehlung 5.2-01, April 1996.
- [615] INFORMATIONSTECHNISCHE GESELLSCHAFT (ITG) IM VDE: *Systeme der Vermittlungstechnik*. Entwurf der ITG-Empfehlung 5.2-02, April 1996.
- [616] INFORMATIONSTECHNISCHE GESELLSCHAFT (ITG) IM VDE: *Begriffe der Verkehrstheorie*. Entwurf der ITG-Empfehlung 5.2-03, Januar 1997.
- [617] NACHRICHTENTECHNISCHE GESELLSCHAFT (NTG) IM VDE: *Nachrichtenverkehrstheorie*. NTG-Empfehlung 0903, 1984.

**Internationale Gremien:**

- [618] ISO 7498: *Information Processing Systems – Open Systems Interconnection – Basic Reference Model*. International Standard, 1984.

**International Telecommunication Union<sup>3</sup> – Telecommunication Standardization Sector:**

- [619] ITU-T, RECOMMENDATION E.177: *B-ISDN Routing*. International Telecommunication Union, Geneva, October 1996.
- [620] ITU-T, RECOMMENDATION E.507: *Models for Forecasting International Traffic*. International Telecommunication Union, Blue Book, Fascicle II.3, Geneva, 1989.
- [621] ITU-T, RECOMMENDATION E.711: *User Demand Modelling*. International Telecommunication Union, Geneva, October 1992.
- [622] ITU-T, RECOMMENDATION E.712: *User Plane Traffic Modelling*. International Telecommunication Union, Geneva, October 1992.
- [623] ITU-T, RECOMMENDATION E.713: *Control Plane Traffic Modelling*. International Telecommunication Union, Geneva, October 1992.
- [624] ITU-T, RECOMMENDATION E.716: *User Demand Modelling in B-ISDN*. International Telecommunication Union, Geneva, October 1996.
- [625] ITU-T, RECOMMENDATION E.731: *Methods for Dimensioning Resources Operating in Circuit Switched Mode*. Int. Telecommunication Union, Geneva, October 1992.
- [626] ITU-T, RECOMMENDATION E.734: *Methods for Allocating and Dimensioning Intelligent Network Resources*. International Telecommunication Union, Geneva, October 1996.
- [627] ITU-T, RECOMMENDATION E.735: *Framework for Traffic Control and Dimensioning in B-ISDN*. International Telecommunication Union, Geneva, May 1997.
- [628] ITU-T, RECOMMENDATION E.736: *Methods for Cell Level Traffic Control in B-ISDN*.

---

<sup>2</sup> VDE ... Verband Deutscher Elektrotechniker

<sup>3</sup> Die *International Telecommunication Union* wurde 1932 unter dem Namen CCITT (*Comité Consultatif International Téléphonique et Télégraphique*) gegründet und 1993 in ITU umbenannt.

- International Telecommunication Union, Geneva, May 1997.
- [629] ITU-T, RECOMMENDATION E.737: *Dimensioning Methods for B-ISDN*. International Telecommunication Union, Geneva, May 1997.
  - [630] ITU-T, RECOMMENDATION E.744: *Traffic and Congestion Control Requirements for SS No. 7 and IN-Structured Networks*. International Telecommunication Union, Geneva, October 1996.
  - [631] ITU-T, RECOMMENDATION E.800: *Terms and Definitions related to the Quality of Telecommunication Services*. International Telecommunication Union, Blue Book, Fascicle II.3, Geneva, 1989.
  - [632] ITU-T, RECOMMENDATION I.120: *Integrated Services Digital Networks (ISDNs)*. International Telecommunication Union, Blue Book, Fascicle III.7, Geneva, 1989.
  - [633] ITU-T, RECOMMENDATION I.121: *Broadband Aspects of ISDN*. International Telecommunication Union, Geneva, April 1991.
  - [634] ITU-T, RECOMMENDATION I.140: *Attribute Technique for the Characterization of Telecommunication Services supported by an ISDN and Network Capabilities of an ISDN*. International Telecommunication Union, Blue Book, Fascicle III.7, Geneva, 1989.
  - [635] ITU-T, RECOMMENDATION I.150: *B-ISDN Asynchronous Transfer Mode Functional Characteristics*. International Telecommunication Union, Geneva, November 1995.
  - [636] ITU-T, RECOMMENDATION I.211: *B-ISDN Service Aspects*. International Telecommunication Union, Geneva, March 1993.
  - [637] ITU-T, RECOMMENDATION I.311: *B-ISDN General Network Aspects*. International Telecommunication Union, Geneva, August 1996.
  - [638] ITU-T, RECOMMENDATION I.352: *Network Performance Objectives for Connection Processing Delays in an ISDN*. International Telecommunication Union, Blue Book, Fascicle III.8, Geneva, 1989.
  - [639] ITU-T, RECOMMENDATION I.361: *B-ISDN ATM Layer Specification*. International Telecommunication Union, Geneva, November 1995.
  - [640] ITU-T, RECOMMENDATION I.362: *B-ISDN ATM Adaptation Layer (AAL) Functional Description*. International Telecommunication Union, Geneva, March 1993.
  - [641] ITU-T, RECOMMENDATION I.363.2: *B-ISDN ATM Adaptation Layer Specification – Type 2 AAL*. International Telecommunication Union, Geneva, September 1997.
  - [642] ITU-T, RECOMMENDATION I.363.5: *B-ISDN ATM Adaptation Layer Specification – Type 5 AAL*. International Telecommunication Union, Geneva, August 1996.
  - [643] ITU-T, RECOMMENDATION I.370: *Congestion Management for the ISDN Frame Relaying Bearer Service*. International Telecommunication Union, Geneva, October 1991.
  - [644] ITU-T, RECOMMENDATION I.371: *Traffic and Congestion Control in B-ISDN*. International Telecommunication Union, Geneva, August 1996.
  - [645] ITU-T, RECOMMENDATION M.3010: *Principles for a Telecommunications Management Network*. International Telecommunication Union, Geneva, May 1996.
  - [646] ITU-T, RECOMMENDATION Q.521: *Exchange Functions*. International Telecommunication Union, Blue Book, Fascicle VI.5, Geneva, 1989.
  - [647] ITU-T, RECOMMENDATION Q.541: *Digital Exchange Design Objectives – General*. International Telecommunication Union, Blue Book, Fascicle VI.5, Geneva, 1989.

- [648] ITU-T, RECOMMENDATION Q.543: *Digital Exchange Performance Design Objectives*. International Telecommunication Union, Blue Book, Fascicle VI.5, Geneva, 1989.
- [649] ITU-T, RECOMMENDATION Q.544: *Digital Exchange Measurements*. International Telecommunication Union, Blue Book, Fascicle VI.5, Geneva, 1989.
- [650] ITU-T, RECOMMENDATION Q.700: *Introduction to CCITT Signalling System No. 7*. International Telecommunication Union, Geneva, March 1993.
- [651] ITU-T, RECOMMENDATION Q.701: *Functional Description of the Message Transfer Part (MTP) of Signalling System No. 7*. International Telecommunication Union, Geneva, March 1993.
- [652] ITU-T, RECOMMENDATION Q.702: *Signalling Data Link*. International Telecommunication Union, Geneva, November 1988.
- [653] ITU-T, RECOMMENDATION Q.703: *Signalling Link*. International Telecommunication Union, Geneva, July 1996.
- [654] ITU-T, RECOMMENDATION Q.704: *Signalling Network Functions and Messages*. International Telecommunication Union, Geneva, July 1996.
- [655] ITU-T, RECOMMENDATION Q.931: *Digital Subscriber Signalling No. 1 (DSS1) – ISDN User-Network Interface Layer 3 Specification for Basic Call Control*. International Telecommunication Union, Geneva, March 1993.
- [656] ITU-T, RECOMMENDATION Q.932: *Digital Subscriber Signalling No. 1 (DSS1) – Generic Procedures for the Control of ISDN Supplementary Services*. International Telecommunication Union, Geneva, March 1993.
- [657] ITU-T, RECOMMENDATION Q.1200: *Q-Series Intelligent Network Recommendation Structure*. International Telecommunication Union, Geneva, March 1993.
- [658] ITU-T, RECOMMENDATION Q.2010: *Broadband Integrated Services Digital Network Overview – Signalling Capability Set 1, Release 1*. International Telecommunication Union, Geneva, Februar 1995.
- [659] ITU-T, RECOMMENDATION Q.2100: *B-ISDN Signalling ATM Adaptation Layer (SAAL) Overview Description*. International Telecommunication Union, Geneva, July 1994.
- [660] ITU-T, RECOMMENDATION Q.2110: *B-ISDN ATM Adaptation Layer – Service Specific Connection Oriented Protocol (SSCOP)*. International Telecommunication Union, Geneva, July 1994.
- [661] ITU-T, RECOMMENDATION Q.2120: *B-ISDN Meta-Signalling Protocol*. International Telecommunication Union, Geneva, Februar 1995.
- [662] ITU-T, RECOMMENDATION Q.2130: *B-ISDN Signalling ATM Adaptation Layer – Service Specific Coordination Function for Support of Signalling at the User Network Interface (SSCF at UNI)*. International Telecommunication Union, Geneva, July 1994.
- [663] ITU-T, RECOMMENDATION Q.2140: *B-ISDN ATM Adaptation Layer – Service Specific Coordination Function for Signalling at the Network Node Interface (SSCF at NNI)*. International Telecommunication Union, Geneva, Februar 1995.
- [664] ITU-T, RECOMMENDATION Q.2721.1: *B-ISDN User Part – Overview of the B-ISDN Network Node Interface Signalling Capability Set 2, Step 1*. International Telecommunication Union, Geneva, July 1996.
- [665] ITU-T, RECOMMENDATION Q.2726.2: *B-ISDN User Part – Call Priority*. International Telecommunication Union, Geneva, July 1996.

- [666] ITU-T, RECOMMENDATION Q.2761: *Broadband Integrated Services Digital Network (B-ISDN) – Functional Description of the B-ISDN User Part (B-ISUP) of Signalling System No. 7*. International Telecommunication Union, Geneva, Februar 1995.
- [667] ITU-T, RECOMMENDATION Q.2762: *Broadband Integrated Services Digital Network (B-ISDN) – General Functions of Messages and Signals of the B-ISDN User Part (B-ISUP) of Signalling System No. 7*. Int. Telecommunication Union, Geneva, Februar 1995.
- [668] ITU-T, RECOMMENDATION Q.2763: *Broadband Integrated Services Digital Network (B-ISDN) – Signalling System No. 7 B-ISDN User Part (B-ISUP) – Formats and Codes*. International Telecommunication Union, Geneva, Februar 1995.
- [669] ITU-T, RECOMMENDATION Q.2764: *Broadband Integrated Services Digital Network (B-ISDN) – Signalling System No. 7 B-ISDN User Part (B-ISUP) – Basic Call Procedures*. International Telecommunication Union, Geneva, Februar 1995.
- [670] ITU-T, RECOMMENDATION Q.2931: *Broadband Integrated Services Digital Network (B-ISDN) – Digital Subscriber Signalling No. 2 (DSS2) – User-Network Interface (UNI): Layer 3 Specification for Basic Call/Connection Control*. International Telecommunication Union, Geneva, Februar 1995.
- [671] ITU-T, RECOMMENDATION Q.2951: *Stage 3 Description for Number Identification Supplementary Services Using B-ISDN Digital Subscriber Signalling No. 2 (DSS2) – Basic Call*. International Telecommunication Union, Geneva, Februar 1995.
- [672] ITU-T, RECOMMENDATION Q.2957: *Stage 3 Description for Additional Information Transfer Supplementary Services Using B-ISDN Digital Subscriber Signalling No. 2 (DSS2) – Basic Call*. International Telecommunication Union, Geneva, Februar 1995.
- [673] ITU-T, RECOMMENDATION Q.2957.1: *Stage 3 Description for Additional Information Transfer Supplementary Services Using B-ISDN Digital Subscriber Signalling No. 2 (DSS2) – Basic Call: Clause 1 – User-to-User Signalling (UUS)*. International Telecommunication Union, Geneva, Februar 1995.
- [674] ITU-T, RECOMMENDATION Q.2959: *Digital Subscriber Signalling No. 2 (DSS2) – Call Priority*. International Telecommunication Union, Geneva, July 1996.
- [675] ITU-T, RECOMMENDATION Q.2961: *Broadband Integrated Services Digital Network (B-ISDN) – Digital Subscriber Signalling No. 2 (DSS2) – Additional Traffic Parameters*. International Telecommunication Union, Geneva, October 1995.
- [676] ITU-T, RECOMMENDATION Q.2961.2: *Digital Subscriber Signalling No. 2 (DSS2) – Additional Traffic Parameters: Support of ATM Transfer Capability in the Broadband Bearer Capability Information Element*. International Telecommunication Union, Geneva, June 1997.
- [677] ITU-T, RECOMMENDATION Q.2961.3: *Digital Subscriber Signalling No. 2 (DSS2) – Signalling Capabilities to Support Traffic Parameters for the Available Bit Rate (ABR) ATM Transfer Capability*. Int. Telecommunication Union, Geneva, September 1997.
- [678] ITU-T, RECOMMENDATION Q.2962: *Digital Subscriber Signalling No. 2 (DSS2) – Connection Characteristics Negotiation during Call/Connection Establishment Phase*. International Telecommunication Union, Geneva, May 1998.
- [679] ITU-T, RECOMMENDATION Q.2963: *Digital Subscriber Signalling No. 2 (DSS2) – Connection Modification*. International Telecommunication Union, Geneva.
- [680] ITU-T, RECOMMENDATION Q.2963.1: *Peak Cell Rate Modification by the Connection*

*Owner*: International Telecommunication Union, Geneva, July 1996.

- [681] ITU-T, RECOMMENDATION Q.2963.2: *Modification Procedures for Sustainable Cell Rate Parameters*. International Telecommunication Union, Geneva, September 1997.
- [682] ITU-T, RECOMMENDATION Q.2963.3: *ATM Traffic Descriptor Modification with Negotiation by the Connection Owner*. Int. Telecommunication Union, Geneva, May 1998.
- [683] ITU-T, RECOMMENDATION Q.2964.1: *Digital Subscriber Signalling No. 2 (DSS2) – Basic Look-Ahead*. International Telecommunication Union, Geneva, July 1996.
- [684] ITU-T, RECOMMENDATION Q.2971: *Broadband Integrated Services Digital Network (B-ISDN) – Digital Subscriber Signalling No. 2 (DSS2) – User-Network Interface Layer 3 Specification for Point-to-Multipoint Call/Connection Control*. International Telecommunication Union, Geneva, October 1995.
- [685] ITU-T, DRAFT RECOMMENDATION Q.298X: *Broadband Integrated Services Digital Network (B-ISDN) – Digital Subscriber Signalling No. 2 (DSS2) – User-Network Interface Layer 3 Specification for Point-to-Point Multiconnection Call Control*. International Telecommunication Union, Melbourne, February 1995.
- [686] ITU-T, RECOMMENDATION X.2: *International Data Transmission Services and Optional User Facilities in Public Data Networks and ISDNs*. International Telecommunication Union, Geneva, October 1996.
- [687] ITU-T, RECOMMENDATION X.200: *Open Systems Interconnection – Basic Reference Model: The Basic Model*. International Telecommunication Union, Geneva, July 1994.
- [688] ITU-T, RECOMMENDATION Z.200: *CCITT High Level Language (CHILL)*. International Telecommunication Union, Geneva, October 1996.

#### **Internet Engineering Task Force:**

- [689] IETF, REQUEST FOR COMMENTS: RFC 1755: *ATM Signaling Support for IP over ATM*. Internet Engineering Task Force, November 1995.

## **Datenblätter, Produktschriften<sup>4</sup>**

Nachfolgende Informationen können alle über die WWW-Seiten der jeweiligen Institutionen abgerufen werden.

- [690] 3COM: *CoreBuilder 9000 Enterprise Switch*. Product Overview, 3Com Corporation, 1998. → <http://www.3com.com/solutions/documents/guides/400346.html>
- [691] ALCATEL: *1100 HSS Switch*. Product Overview, Alcatel Data Products, 1998. → <http://wwwusa.adn.alcatel.com/dataprod/hssatmt.htm>
- [692] ALCATEL: *1000 BBX – Broadband Transit and Local Multiservice Multifunctional Exchange*. Product Overview, Alcatel Products, 1998. → <http://www.alcatel.com/telecom/mbd/products/products/opg/products/1000bbx.htm>
- [693] ASCOM: *ATMLightRing 1250*. Product Documentation, Ascom Transmission, BroadNet Division, June 1998. → <http://www.broadnet.ascom.ch/product.html>
- [694] BAY NETWORKS: *Centillion 1000 Multiservice ATM Switch*. Data Sheet, Bay Networks,

---

<sup>4</sup> Die URL-Angaben der WWW-Seiten entsprechen dem Stand ‚Oktober 1998‘.

1998. → <http://business5.baynetworks.com>
- [695] CASCADE: *CBX 500 Multiservice ATM Switch*. Product Description, Cascade Communication Corporation bzw. Ascend Communications, 1998.  
→ <http://www.casc.com/products/atm/specifications.html>
- [696] CISCO: *LightStream 1010 Multiservice ATM Switch*. Product Overview, Cisco Systems, 1998. → [http://www.cisco.com/warp/public/730/LS1010/ls10m\\_ds.htm](http://www.cisco.com/warp/public/730/LS1010/ls10m_ds.htm)
- [697] DATACOMM: *APEX-IMX Multiservice ATM Switch*. Data Sheet, General DataComm, 1997. → <http://www.gdc.com/>
- [698] DIGITAL: *GIGAswitch/ATM System*. Product Description, Digital Products, Compaq Corporation, 1998. → <http://www.networks.digital.com/dr/npdg/dagfm-mn.html>
- [699] ERISSEN: *AXD 301 ATM Switch*. Product Overview, Ericsson Telecom AB, 1998.  
→ [http://www.ericsson.com/datacom/axd\\_301.htm](http://www.ericsson.com/datacom/axd_301.htm)
- [700] FORE: *ForeRunner ASX-4000 – ATM Backbone Switch*. Data Sheet, FORE Systems, September 1998. → <http://www.fore.com/products/swtch/asx4000.html>
- [701] FUJITSU: *FETEX-150 ATM ESP Broadband Switching System*. Product Description, Fujitsu Network Communications, 1997.  
→ <http://www.fujitsu.com/FNC/products/fetex150.html>
- [702] HUGHES: *BX5000 Dual-Service (ATM/FR) Broadband Switch*. Product Overview, Hughes Network Systems, 1998.  
→ <http://www.hns.com/products/ens/bx5000/bx5000.htm>
- [703] IBM: *IBM 2230 Nways ATM Switch Models 600 and 650*. Product Overview, International Business Machines Corporation, 1998.  
→ <http://www.networking.ibm.com/223/223over.html>
- [704] LUCENT: *5ESS<sup>®</sup>-2000 Switch*. Product Overview, Lucent Technologies, 1997.  
→ <http://www.lucent.com/netsys/5ESS/5esswtch.html>
- [705] LUCENT: *GlobeView<sup>®</sup>-2000 Switch*. Product Preview, Lucent Technologies, 1997.  
→ <http://www.lucent.com/netsys/supercomm/atmglobeview.html>
- [706] MOTOROLA: *MCM69C432 Content Addressable Memory (16K × 64 CAM)*. Data Sheet, Motorola, Inc., March 1998.
- [707] MUSIC: *MU9C4320L ATMCAM (4K × 32 CAM)*. Data Sheet, MUSIC Semiconductors, January 1998.
- [708] NEC: *ATOMNET/M20 Backbone Switch*. Product Overview, NEC Corporation, 1997.  
→ [http://www1e.meshnet.or.jp/CNPWORLD/english/product/main\\_e.html](http://www1e.meshnet.or.jp/CNPWORLD/english/product/main_e.html)
- [709] NORTEL: *Magellan Concorde*. Product Overview, Northern Telecom, 1996.  
→ [http://www.nortel.com/broadband/Carrier\\_Data\\_Networks/concorde.html](http://www.nortel.com/broadband/Carrier_Data_Networks/concorde.html)
- [710] QUALITY: *QS762470 Content Addressable Memory (2K × 64 CAM)*. Data Sheet, Quality Semiconductor, Inc., October 1997.
- [711] SIEMENS/NEWBRIDGE: *MainStreetXpress 36190 Core Services Switch*. Data Sheet, Siemens Public Communication Networks, September 1996.  
→ <http://www.newbridge.com/products/index.html>

## Anhang A

# Steuerungsstrukturen ausgewählter ATM-Netzknoten

Die Problematik der Laststeuerung und Überlastabwehr wurde mit Einzug der rechnergestützten Steuerung für Vermittlungssysteme in den Vordergrund gerückt. Viele Elemente von Lastregelkreisen (z. B. Lastsensoren, Regelalgorithmen) wurden in diesem Zusammenhang in den letzten 30 bis 40 Jahren entworfen, weiterentwickelt oder aus anderen Disziplinen adaptiert. Für die Vermittlungssysteme der verschiedenen Netztypen wurden spezifische bzw. teilweise unterschiedliche Überlastabwehrstrategien konzipiert. Die einzelnen Verfahren sind kaum übertragbar und machen nur im Kontext des jeweiligen Anwendungsbereiches Sinn. Wenn man jedoch die Entwicklung der Überlastabwehr über die Jahre betrachtet, dann wurden für neue Systemarchitekturen und Netzgenerationen keine grundsätzlich neuen Lastregelverfahren entwickelt. Viele Elemente und Einzelmechanismen können und werden deshalb auch zukünftig noch eingesetzt werden.

Eine Untersuchung von Überlastabwehrstrategien für B-ISDN/ATM-Netzknoten muss deshalb die Verfahren vergangener System- und Netzgenerationen berücksichtigen. Für das Verständnis der einzelnen Überlastabwehrverfahren (siehe auch Abschnitt 3.4.9) sind beispielsweise in [4, 31, 49] typische Knoten zusammengestellt. Der Schwerpunkt liegt dort im wesentlichen auf Systemen mit ausgeprägten Leistungs- und Architekturmerkmalen.

Bevor in diesem Kapitel Steuerungsstrukturen gängiger ATM-Netzknoten vorgestellt werden, soll zuerst ein kurzer Rückblick auf die Entwicklungsschritte im *Kontrollpfad* eines Vermittlungssystems erfolgen.

## A.1 Entwicklungstendenzen in der Steuerungsarchitektur

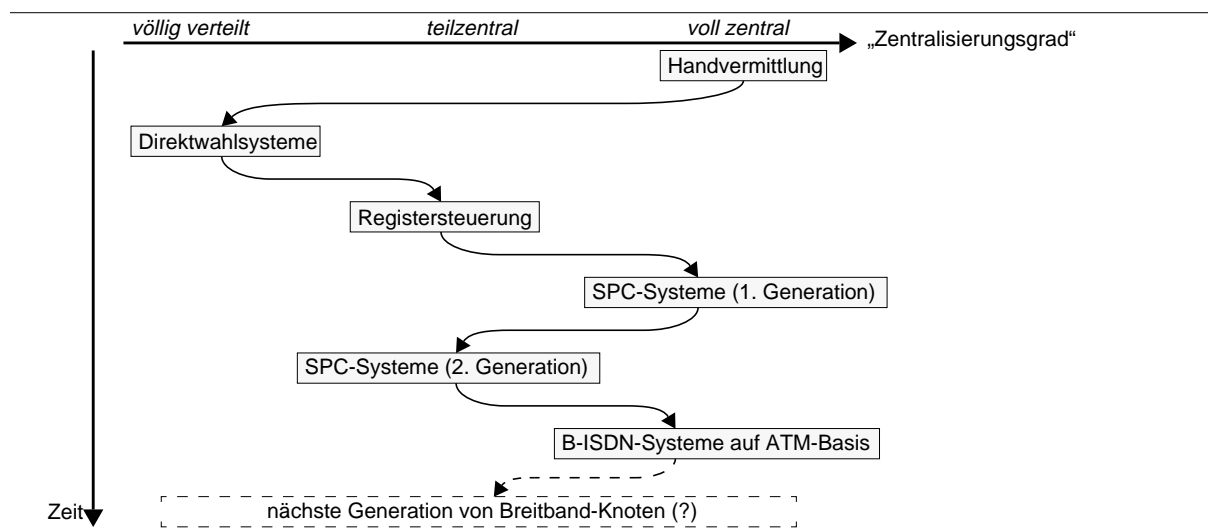
In Deutschland wurde 1877 die erste öffentliche Telegraphenvermittlung mit Fernsprechbetrieb handvermittelt in Betrieb genommen. Betrachtet man die Vermittlungssteuerung, dann lassen sich aus Sicht der technischen und historischen Entwicklung seit diesem Zeitpunkt verschiedene Generationen von Vermittlungssystemen unterscheiden. Bevor in den 60er Jahren die Epoche der rechnergesteuerten Vermittlungssysteme begann, waren davor die *Handvermittlung*, die *Direktwahlssysteme* bzw. Systeme mit indirekter Wahl (*Registersteuerung*) die wichtigsten Entwicklungsschritte [13, 88, 109, 114, 139, 203].

Für die rechnergesteuerten Vermittlungssysteme kann man rückblickend folgende Generationen identifizieren [4, 163, 342]: Mit der Einführung der Koordinatenschaltersysteme (*Crossbar*) ging der Weg von den Direktwahl- zu den Indirektwahlssystemen hin, mit zunächst elektromechanischer, danach elektronischer Steuerung und schließlich rechnerbasierter Programmsteuerung (*Stored-Program Control, SPC*). Bezüglich der Steuerung war diese erste Generation von SPC-Vermittlungssystemen durch analoge Durchschaltung im Sprechweg und eine zentralisierte Rechner-Steuerung gekennzeichnet. Durch das Aufkommen preiswerter, leistungsfähiger Mikroprozessoren in den 70er Jahren konnte ein Teil der zunächst zentralisierten Steuerungsfunktionen wieder ganz oder teilweise in den peripheren Bereich ausgelagert wer-

den (SPC-Systeme der zweiten Generation) [245]. Einzelne Steuerungen konnten im zentralen wie auch im peripheren Bereich durch mehrere kleinere Teilsteuerungen, die nach dem Last- und/oder Funktionsteilungsprinzip arbeiten, ersetzt werden [83,217]. Durch die Auslagerung von Funktionen in Prozessoren der Anschlussbaugruppen entstanden hierarchische Steuerungsstrukturen. DENZEL [146] gibt hierzu eine übersichtliche Darstellung prinzipieller Steuerungsarchitekturen.

Dieser Ansatz führte dann zu rechnergesteuerten Vermittlungssystemen mit einer Vielzahl von Mikroprozessoren und einem umfangreichen, leistungsfähigen Kommunikationssystem für den Steuerdatenaustausch zwischen den Prozessoren [476,477]. Die teilzentralen Steuerungsstrukturen haben ihre großen Vorteile bei der digitalen PSTN- und der ISDN-Technik. Durch den zugrundeliegenden synchronen Transfermodus<sup>1</sup> musste die Signalisierinformation hier in den Anschlussgruppen ausgekoppelt werden. Es lag daher nahe, die Protokollbearbeitung für Teilnehmer- und Netzsignalisierung bereits dezentral in diesen peripheren Einheiten durch Mikroprozessoren vorzunehmen. Übergeordnete Aufgaben wie beispielsweise die Koppelnetz-durchschaltung werden hier von zentralen Steuerungseinheiten vorgenommen. In [49] werden eine repräsentative Auswahl rechnergesteuerter Nicht-ATM-Vermittlungssysteme zusammengestellt und deren Steuerungsmerkmale diskutiert.

Wie bereits in Abschnitt 2.6 ausgeführt wurde, zeichnet sich bei der gegenwärtigen Generation von ATM-Netzknotten (siehe auch Abschnitt A.2) wiederum ein Trend in Richtung höherer Zentralisierung ab. Der Begriff der *Zentralisierung* soll dabei hier im Sinne der *räumlichen Verteilung* von Verarbeitungseinheiten verstanden werden.<sup>2</sup> Betrachtet man diesen „Zentralisierungsgrad“ der Steuerungsfunktionen in einem Vermittlungssystem, dann lässt sich grob folgende Entwicklung festmachen (Bild A-1):



**Bild A-1:** Zentralisierung von Steuerungsfunktionen

Neben dem Funktions- und Lastteilungsgrad (siehe Abschnitt 2.3.5) einer Vermittlungssteuerung sind aus Sicht der Überlastabwehr aber auch grundlegende interne Organisationsmerkmale von Interesse.

<sup>1</sup> Durchschaltvermittlung im synchronen Zeitmultiplex mit Inband- oder Außerband-Signalisierung.

<sup>2</sup> Zur *Systematik von Steuerungsarchitekturen* wurde von KÜHN [329] diesbezüglich ein verfeinertes Klassifizierungsschema eingeführt. In einer *zentralen Steuerung* werden dabei die Hauptfunktionen durch ein oder wenige gleichartige Prozessoren realisiert, wohingegen bei *dezentraler Steuerung* sich in der Regel eine Vielzahl funktions-spezifischer Teilsteuerungen die Aufgaben teilen. *Konzentrierte Steuerung* liegt vor, wenn die Steuerungsfunktionen eines Vermittlungssystems in einer Steuerungseinheit zusammengefasst sind; bei *verteilter Steuerung* werden dagegen Teile der Steuerungsaufgaben auf örtlich getrennte (Peripherie-)Steuerwerke aufgeteilt.



## A.2 Auswahl repräsentativer Systeme

Wie bereits einführend erwähnt, soll hier nur eine Auswahl von ATM-Vermittlungssystemen vorgestellt werden. Diese kann und will auch keinen repräsentativen Charakter und Vollständigkeitsanspruch besitzen. Die Vielzahl kleinerer bis mittlerer *ATM Switches* besitzt einen zentralen Prozessorknoten als Vermittlungssteuerung. Nachfolgend sind nun solche Systeme dokumentiert, die eine aufwendigere Steuerungsstruktur aufweisen und damit prinzipiell für höhere Leistungen ausgelegt sind. Die neun dokumentierten Systeme wurden in zwei Gruppen eingeteilt:

- *Zentraler Vermittlungsrechner*

Merkmale: ein ATM-Koppelnetz, zentrale Vermittlungssteuerung, alle ATM-Zellen mit Signalisierinformation werden an diese Einheit weitergeleitet. Der Vermittlungsrechner selbst kann als Multiprozessorsystem mit den unterschiedlichsten Architekturmerkmalen realisiert sein.

- *Verteilte Vermittlungssteuerungen*

Merkmale: i. d. R. zweistufige Koppelnetze<sup>3</sup>, Koppelnetzmodule der ersten Stufe besitzen eine lokale zentrale Steuerung. Diese stellen für das Gesamtsystem eine verteilte Vermittlungssteuerung dar.

In der gegenwärtigen Entwicklungsphase (Stand: Ende 1998) von ATM-Systemen sind Informationen zu detaillierten technischen Systemmerkmalen, aus verständlichen Gründen, öffentlich nicht zugänglich. Dies trifft umso mehr für Hardwarestruktur, Softwarearchitektur und Funktionsverteilung der jeweiligen Vermittlungssteuerungen zu. Aus diesem Grund können auch im folgenden keine oder nur sehr ungenaue Angaben zur Überlastabwehr bzw. Lastregelverfahren gemacht werden.

### A.2.1 Zentraler Vermittlungsrechner

#### A.2.1.1 A1000 Multimedia Multiservice Switch (Alcatel)

Dieses Breitband-Vermittlungssystem von *Alcatel* vereint Architekturmerkmale von den Systemen S 12 und E 10 [431]. Der Kern des Datenpfades wird durch ein mehrpfadfähiges, selbst-routendes Koppelnetz gebildet. Die zentrale Vermittlungssteuerung besteht aus einer Multiprozessorarchitektur.

#### A.2.1.2 AXD 301 (Ericsson)

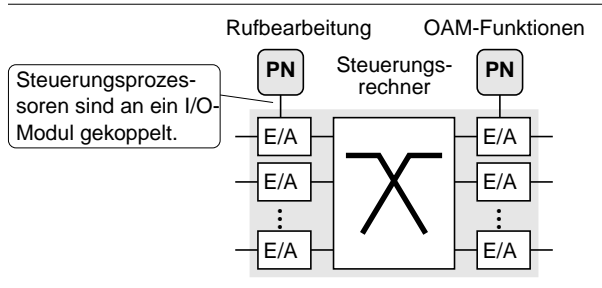
Das Koppelnetz des *AXD 301* von *Ericsson* ist skalierbar bis zu einem Gesamtdurchsatz von 2 Tbit/s [96]. Im Grundausbau besteht der Vermittlungsrechner aus zwei Prozessorknoten. Im Normalbetrieb bearbeitet einer als zentrale Instanz alle Rufe und der andere Prozessor ist zuständig für OAM-Aufgaben. Dabei stellt jeder die Stand-By-Einheit des Gegenüber dar. Bei Ausfall eines Prozessors werden dann die jeweiligen Funktionen vom anderen übernommen. Sichergestellt wird dies durch ein verteiltes, echtzeitfähiges Datenbanksystem. Die Rufver-

---

<sup>3</sup> Jedes ATM-Koppelnetz selbst stellt wiederum eine mehrstufige Koppelanordnung dar.

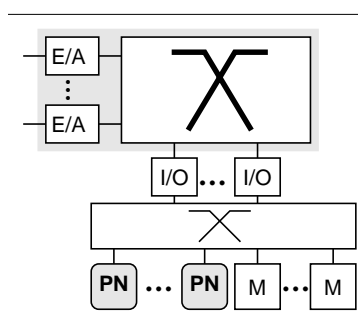
mittlungsleistung kann durch zusätzliche Steuerungsprozessoren erhöht werden. Jeder Prozessorknoten ist dann zuständig für eine bestimmte Gruppe von I/O-Modulen. Es liegt somit eine völlig verteilte Steuerung in Lastteilung vor.

Jeder Prozessorknoten ist mit einer Laststeuerung, die nach der LIFO-Strategie (siehe auch Abschnitt 3.4.9.2) arbeitet, ausgestattet. Neue Anforderungen kommen in eine Warteschlange mit LIFO-Abfertigungsdisziplin. Die Bedienrate, mit welcher die Aufträge an die *Call Control* weitergereicht werden, ist umgekehrt proportional zur CPU-Auslastung. Übersteigt die Wartezeit einer Anforderung einen gewissen Grenzwert, dann wird sie zurückgewiesen.



**Bild A-2:** AXD 301 von Ericsson

### A.2.1.3 Magellan Concorde (Nortel)



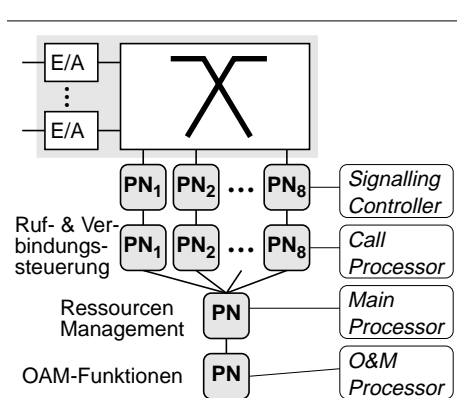
**Bild A-3:** Magellan Concorde (Nortel)

Das System *Magellan Concorde* von Nortel ist als *Backbone Switch* konzipiert [709]. Die Vermittlungssteuerung ist als verteiltes Multiprozessorsystem ausgelegt [16] und trägt die Bezeichnung *XA-Core*. Die Steuerungsstruktur entspricht einer Parallel-Processorarchitektur mit gemeinsamem Speicher.<sup>4</sup> Die einzelnen Prozessorelemente kommunizieren über ein Verbindungsnetzwerk mit den Speichermodulen (Bild A-3). Dieses eng gekoppelte Steuerungssystem ist über mehrere I/O-Module mit dem Koppelnetz verbunden. Ein Call wird in verschiedene *Subcalls* unterteilt, die parallel auf verschiedenen Prozessoren bearbeitet werden können [156]. Durch den gemeinsamen Speicher und damit der gemeinsamen Haltung der Rufdaten können hierbei Zugriffskonflikte und Konsistenzprobleme auftreten. Insgesamt wächst deshalb die Vermittlungsleistung nicht ganz proportional zur Anzahl von Prozessorknoten.

### A.2.1.4 FETEX-150 ATM ESP B-160 (Fujitsu)

Das Breitbandvermittlungssystem *FETEX-150 ATM ESP* von Fujitsu ist als *Core Switch* für öffentliche Netzbetreiber konzipiert [16].<sup>5</sup> Es besitzt eine hierarchische Steuerungsstruktur mit Last- und Funktionsteilungsprinzip. Da Koppelnetz und Kontrollplattform lose gekoppelt sind, wird unabhängige Skalierbarkeit der Leistung in Nutzer- und Steuerungs-Ebene garantiert [283].

Bild A-4 zeigt die verschiedenen Funktionsebenen der Multiprozessorarchitektur. Die Schicht-3-Signalisierung sowie die Ruf- und Verbindungssteuerung sind in den *Call Prozessoren* lokalisiert. Die Prozesse für ankommende und abgehende Rufrichtung (*originating* bzw. *termina-*



**Bild A-4:** FETEX-150 ESP

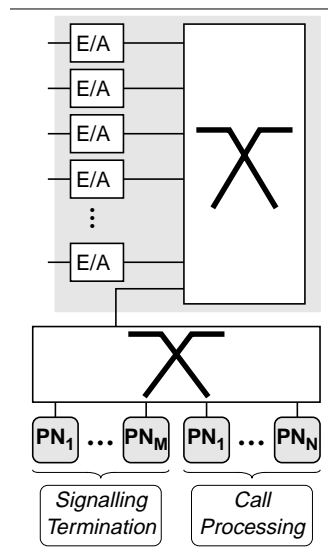
<sup>4</sup> Shared-Memory Multi-Processor Call Control

<sup>5</sup> FETEX ... Fujitsu Electronic Telecommunication Exchange, ESP ... Enhanced Switching Platform.

ting call) können auf verschiedene Prozessoren verteilt werden. Diese flexible Möglichkeit der Verteilung sogenannter *Half Calls* auf verschiedene Verarbeitungseinheiten verbessert die Skalierbarkeit. Die Vergabe und Verwaltung von Ressourcen (z.B. Bandbreite, VPI/VCI) wird von einer zentralen Instanz, dem sogenannten *Main Processor*, vorgenommen.

Die Leistung der aktuellen Systemversion wird mit 1800000 BHCAs („*typical POTS mix with an average feature penetration*“) angegeben [701].

### A.2.1.5 MainstreetXpress (Siemens/Newbridge)



Das leistungsfähigste Breitband-Vermittlungssystem von *Siemens/Newbridge* ist der *MainstreetXpress 36190*. Die Leistung ist skalierbar bis zu einem Durchsatz von 1 Tbit/s und einer Vermittlungskapazität von 4MBHCA [711]. Dabei können gleichzeitig bis zu 2000000 VPCs und VCCs gehalten werden. Ermöglicht wird dies durch einen zentralen Vermittlungsrechner in einer Multiprozessorarchitektur. Die Steuerungsaufgaben sind somit verteilt auf ein Cluster von Prozessoren [253].

Die Steuerungsstruktur wird gebildet durch zwei Funktionsebenen (genannt *Signalling* und *Call Processing*). In jeder Funktionsebene arbeiten mehrere Prozessorknoten parallel in Lastteilung. Alle Prozessoren sind vom selben Typ. Ihre Funktionsebene wird ihnen beim Neustart des Systems durch die hochgeladene Software zugeteilt.

**Bild A-5:** *Mainstreet-Xpress*

Die Prozessorknoten kommunizieren untereinander mittels eines separaten ATM-Koppelnetzes (Multiplexer) [435]. Für die Kommunikation mit Leitungsmodulen und Koppelnetz werden ebenfalls ATM-Kanäle verwendet. Ankommende Signalisierzellen werden in den Signalisier-Processor-knoten terminiert. An der Bearbeitung eines Rufes sind in der Regel mehrere Prozessoren beteiligt.

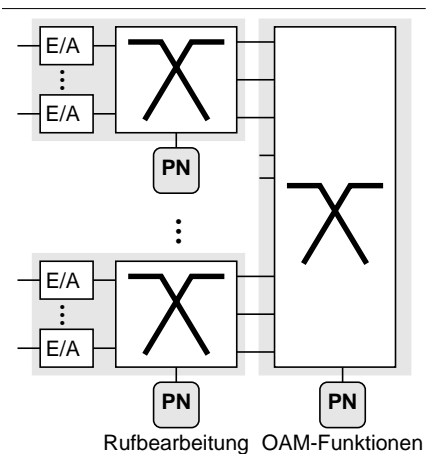
Das System besitzt einen ausgeklügelten Lastverteialgorithmus [410,604] (siehe auch Abschnitt A.3.1).

## A.2.2 Verteilte Vermittlungssteuerungen

### A.2.2.1 TDX-ATM (ETRI)

Ein ATM-System mit völlig verteilter Steuerungsarchitektur wurde vom ETRI entwickelt [565]. Sogenannte Zugangsmodule sind mit einem ATM-Koppelfeld gekoppelt (Struktur siehe Bild A-6). Jedes Zugangsmodule besitzt einen lokalen Steuerungsprozessor. In ihrer Gesamtheit stellen diese ein verteiltes Steuerungssystem dar. Zusätzlich steht ein zentraler OAM-Prozessor zur Verfügung.

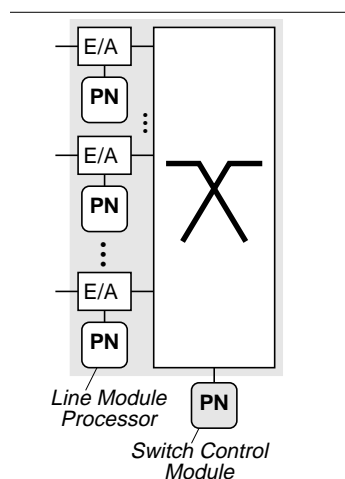
Die Bearbeitung von Rufes mit Verbindungen über mehrere Zugangsmodule bedingt eine umfangreiche Interprozessor-



**Bild A-6:** *TDX-ATM* (ETRI)

kommunikation zwischen den beteiligten Verarbeitungsinstanzen [309]. Je nachdem wie die Funktionen Adressumsetzung, Routing, CAC und Ressourcenverwaltung hier verteilt werden, können verschiedene Steuerstrategien (*fully-distributed*, *partially-distributed* oder *partially centralized*) implementiert werden. Für die Datenkonsistenz ist ein verteiltes, echtzeitfähiges Datenmanagement-System verantwortlich [230].

### A.2.2.2 GlobeView-2000 (Lucent)



**Bild A-7:** *GlobeView-2000* (Lucent)

Große PSTN/N-ISDN-Systeme besitzen eine Rufkapazität von bis zu 2 MBHCA. Die Steuerungsleistung aktueller ATM-Anlagen liegt z.Zt. noch weit darunter. Zukünftig werden aber für große ATM-Vermittlungen mehr als 10000 cps (= 36 MBHCA) erwartet [128]. Diese Steuerungskapazitäten sind nur noch mit Mehrprozessorarchitekturen handhabbar. Aus diesem Grund wurden in den *Bell Labs* Konzepte für verteilte Vermittlungssteuerungen ausgearbeitet [128]. Als die rechenzeitintensivsten Funktionen bei der Rufbearbeitung wurden Signalisierung, Rufsteuerung und Routing identifiziert. Die vorgestellten Steuerungskonzepte sehen deshalb eine Verteilung und parallele Ausführung dieser Aufgaben vor.

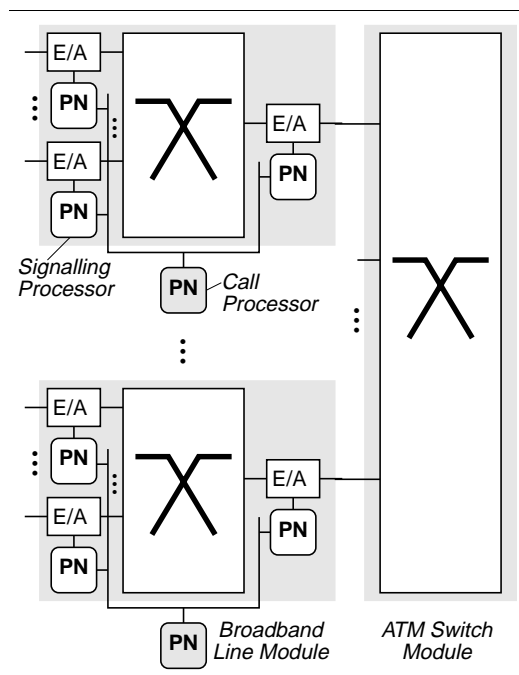
Für die erste Version ihres Vermittlungssystems, *GlobeView*<sup>®</sup>-2000 Release 1, wurde die Signalisierung in die Leitungsmodule ausgelagert und damit parallelisiert [128,238] (*Distributed Signaling Call Processing Architecture*). Rufsteuerung und Routing sind im *Switch Control Module* zentralisiert (Bild A-7).

Die nächste Generation des *ATM Core Switch* soll eine Rufleistung von 3000 cps (gleich 10.8 MBHCA) erreichen [61].

### A.2.2.3 ATOMNET/M20 (NEC)

Die Familie von Breitband-Systemen *ATOMNET* von NEC ist für öffentliche Netzbetreiber ausgelegt [375,708]. Das Grundelement stellt das *Broadband Line Module* dar (Bild A-8). Dieses besitzt eine zweistufige hierarchische Steuerungsstruktur. Jedes Leitungsmodul besitzt einen *Signalling Processor*, denen ein zentraler *Call Processor* übergeordnet ist [506]. Die Interprozessorkommunikation wird hier nicht über das ATM-Koppelfeld abgewickelt, sondern über ein separates Bussystem.

Aus mehreren *Broadband Line Modules* und einem zentralen Koppelnetz können leistungsfähigere Systeme konfiguriert werden. Das *ATM Switch Module* besitzt keinen Prozessorknoten mit Steuerungsfunktionen. Das zweistufige Breitband-Vermittlungssystem besitzt damit eine verteilte Vermittlungssteuerung.



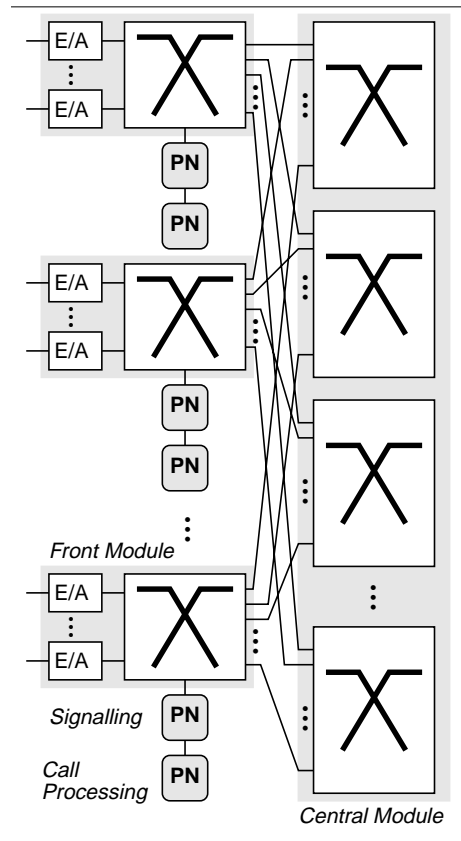
**Bild A-8:** *ATOMNET/M20* (NEC)

#### A.2.2.4 *Hyper-Distributed System (Hitachi)*

Bereits 1991 wurde von *Hitachi* ein völlig verteiltes B-ISDN-System (genannt „*Hyper-Distributed*“) vorgeschlagen [460]. Dieser Ansatz ist dadurch motiviert, Skalierbarkeit sowohl in der Nutzer- als auch in der Steuerungsebene garantieren zu können.<sup>6</sup>

Sogenannte *Front Modules* stellen autonome ATM Switches mit einer zentralen Steuerung (Dualprozessor-Struktur) dar. Diese ATM-Knoten sind über ein zentrales Verbindungsnetzwerk (*Central Module*), das aus mehreren selbstständig arbeitenden Koppelfeldern besteht, gekoppelt (siehe Bild A-9). Zwischen zwei *Front Modules* existieren dabei immer mehrere Verbindungsmöglichkeiten. Beide Ebenen bestehen aus einer flexiblen Anzahl von verteilten Einheiten.

Aufgaben der Steuerungsebene sind nur in den *Front Modules* lokalisiert. Aus Sicht der Vermittlungssteuerung ist somit das *Central Module* komplett passiv. Die einzelnen Verbindungen eines Rufes können über mehrere *Front Modules* gehen. Diese Ruftypen bedingen eine hohe Interprozessorkommunikation (siehe auch ETRI-Switch, Abschnitt A.2.2.1). Insgesamt besitzt damit dieses B-ISDN-Vermittlungssystem eine vollständig verteilte Vermittlungssteuerung.



**Bild A-9:** Konzept von *Hitachi*

### A.2.3 Weitere Informationen zu ATM-Systemen

Nachfolgend soll noch auf weitere ATM-Vermittlungssysteme hingewiesen werden. Die Auflistungen entsprechen dabei dem Stand von Ende 1998.<sup>7</sup>

#### A.2.3.1 Bekannte Steuerungsstruktur

Anhand der in Tabelle A-1 aufgelisteten ATM-Systeme soll eine Übersicht derzeit aktueller Steuerungsstrukturen gegeben werden. Die *Klassifizierung* basiert dabei auf dem in Abschnitt 2.6.5 eingeführten *einfachen Schema*. Alle Informationen sind den aufgeführten Literaturstellen entnommen. Ein sogenannter Kernnetzknotten (*Core Switch*) besitzt vornehmlich eine auf den Backbone-Bereich zugeschnittene hohe Zellvermittlungsleistung. Bei einem *Multiservice Switch* liegt der Schwerpunkt mehr auf der Unterstützung und dem Interworking verschiedener Netzdienste. Abschließend soll noch angemerkt werden, dass sowohl die Begriffsverwendungen als auch die Schreibweisen in den verschiedenen Produktschriften alles andere als einheitlich sind.

<sup>6</sup> Diese Anforderung wird in [460] durch die Darstellung einer Funktion  $Leistung_{UserPlane} = f(Leistung_{ControlPlane})$ , die Angabe erfolgt in bit/s bzw. BHCAs, unterlegt. Dabei wird zwischen B- und N-ISDN sowie verschiedenen Diensten unterschieden.

<sup>7</sup> Die Marktlandschaft der Anbieter von ATM-Systemen ist derzeit durch einen starken Wandel gekennzeichnet, wobei Firmenübernahmen und die Bildung strategischer Allianzen eine wesentliche Rolle spielen.

Struktur	Produkt- bzw. Prototypenbezeichnung	Entwickler/Hersteller	Referenzen	Bemerkungen
(1)	NS10A	NTT	[321, 384, 397, 412]	Erster kommerzieller ATM-Knoten
(1)	NS8000 Familie	NTT	[392, 402, 402, 411, 507, 564]	Multiservice Knoten ( <i>MHN, Multimedia Handling Node</i> )
(1)	Konzept einer Steuerungsarchitektur	CSELT	[80]	
(1)	Centillion 1000	Bay Networks	[694]	Multiservice ATM Switch
(1)	2230 Nways 600 & 650	IBM	[703]	ATM Switch
(1)	E-7550	Fujitsu	[69]	ATM TK-Anlage
(1)	LightStream 1010 Multiservice ATM Switch	Cisco/Stratacom	[696]	MIPS R4600 $\mu$ P als ATM Switch Processor
(1)	ATOMNET/M20	NEC	[372, 708]	
(1, 1)		Hitachi	[511, 512]	
(1, 1)	TDX-ATM Switching System	ETRI	[309, 565, 605-607]	
(1, 1)	EWSB, ACC bzw. EWSXpress	Siemens	[181, 468, 536]	Prototypensysteme
(N)	AXD 301	Ericsson	[76, 96, 115, 699]	
(N)	Konzept eines B-ISDN-Vermittlungssystems	Alcatel	[85, 144]	Erweiterung von S12 [140] und E10
(N)	1000 BBX 1100 HSS	Alcatel	[176, 415, 692] [562, 691]	ATM Core Switch Multi-Service Switch
(N)	Konzept eines ATM Multiservice Switch	Bosch Telecom	[98]	Busarchitektur
(N)	CBX 500 Multiservice ATM Switch	Cascade	[695]	Verteilte Signalisier- und Rufbearbeitung <sup>a</sup>
(N)	Magellan Concorde	Nortel	[709]	
(N)	Konzept eines B-ISDN-Vermittlungssystems	NTT	[479]	Verteilte Signalisier- und Rufbearbeitung
(N, M)	MainstreetXpress 36190	Siemens/Newbridge	[54, 436, 437, 604, 711]	Core Services Switch
(N, M, 1, 1)	FETEX-150 ATM ESP	Fujitsu	[16, 283, 701]	Core Switch

a. Jedes I/O-Modul ist mit einem Vermittlungsprozessor (i960  $\mu$ P) bestückt. Der CBX500 ist derzeit für maximal 14 Module konzipiert, d.h. im Vollausbau arbeiten 14 Prozessoren in Lastteilung.

**Tabelle A-1:** Typische Steuerungsstrukturen gängiger ATM-Vermittlungssysteme

### A.2.3.2 Unbekannte Steuerungsstruktur

Im Rahmen dieser Arbeit wurden Informationen zu weiteren ATM-Vermittlungssystemen eingeholt (siehe Tabelle A-2). Dabei konnten die zugrundeliegenden Steuerungsstrukturen allerdings nicht identifiziert werden. Neben diesen Systemen sind derzeit noch eine Vielzahl von

ATM-Netzknoten anderer Hersteller, beispielsweise von *GPT, Ipsilon Networks, Mitsubishi, NOKIA, OKI, Philips, ZeitNet, TranSwitch* oder *Cabletron Systems*, auf dem Markt.

Produktbezeichnung	Entwickler/ Hersteller	Referenz	Bemerkungen
CoreBuilder 9000	3COM	[690]	Kombinierbarer ATM-Gigabit-Ethernet Core Switch
ATMLightRing 1250	ASCOM	[693]	Ringförmige Vermaschung von Add-Drop-Multiplexern bilden das ATM-Vermittlungssystem
APEX-IMX	DataCom	[697]	Multiservice ATM Switch
GIGAswitch/ATM	digital	[698]	Keine B-ISDN-Signalisierung
ForeRunner ASX-4000	FORE	[700]	ATM Backbone Switch
BX5000	Hughes	[702]	Dual-Service (ATM/FR) Broadband Switch

**Tabelle A-2:** Auswahl von ATM-Vermittlungssystemen mit unbekannter Steuerungsarchitektur

## A.3 Beispiele konkreter Lastregelverfahren

Spezifische Verfahren zur Überlastabwehr von B-ISDN-Vermittlungsrechner sind zum jetzigen Zeitpunkt wenig publiziert worden. Nachfolgend werden zwei Beispiele kurz erläutert, die hauptsächlich die *Lastverteilung in einer Mehrprozessorsteuerungsstruktur* zum Ziel haben, aber weniger dienstspezifische Elemente berücksichtigen.

### A.3.1 Load Balancing – Siemens/Newbridge

Die Vermittlungssteuerung basiert auf einer zweistufigen, skalierbaren und flexibel konfigurierbaren Mehrprozessorarchitektur (siehe auch Abschnitt A.2.1.5).<sup>8</sup> Die beiden Funktionsebenen sind in Bild A-10 dargestellt. Alle Prozessorknoten sind im Aufbau identisch. Beim Systemstart wird durch die geladene Software der Funktionstyp festgelegt. Für die Dienstvermittlung und -steuerung sind die Signalisier- (*SLT*) und Rufprozessoren (*CALLP*) zuständig.<sup>9</sup> Es gibt keine 1:1-Zuordnung zwischen diesen Funktionstypen (d.h. allgemein ist  $M \neq N$ ). Die Zuordnung ankommender Signalisierkanäle basiert auf der Unterscheidung von UNI und NNI.

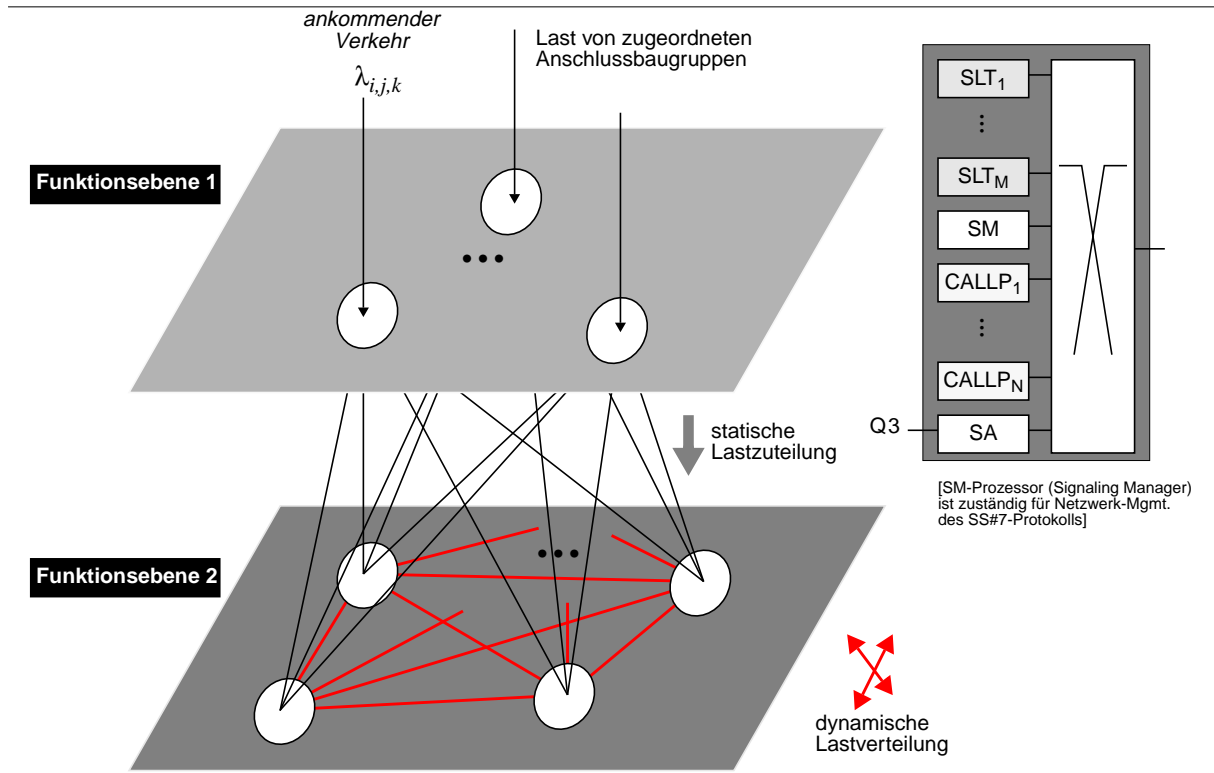
Von der Signalisierenebene nach der Rufebebene findet keine dynamische Lastzuteilung statt. Aus diesem Grund ist eine dynamische Lastweiterleitung (*Load Balancing*) in der Rufebebene vorgesehen (Bild A-11). Neben dieser globalen Lastverteilung sind die Prozessorknoten dieser Funktionsebene zusätzlich mit einer lokalen Überlastabwehr ausgestattet. Nachfolgend wird die Arbeitsweise dieser beiden kooperierenden Regelschleifen kurz erläutert.

<sup>8</sup> Alle Informationen zu diesem System entstammen [181,410,435,436,437]. Der Algorithmus zur Lastverteilung ist patentrechtlich geschützt [604]. Weitere Patente sollen folgen.

<sup>9</sup> Aufgabenzuordnung:

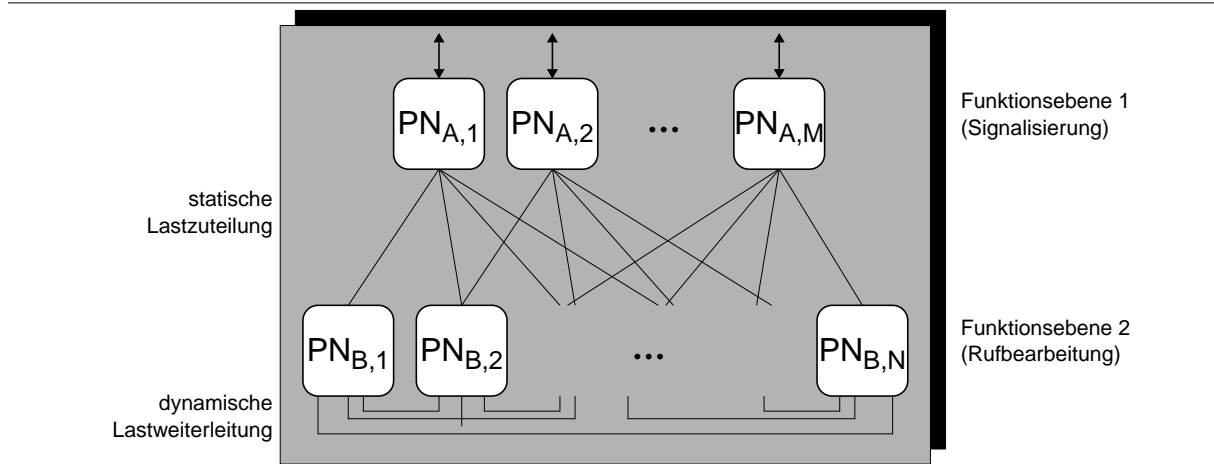
Ebene 1 – *SLTs* (*Signaling Link Termination*) für Schicht-2-Zeichengabe (SAAL für UNI und Breitband-NNI, MTP für Schmalband-CCS7). Die Zuordnung von Signalisierkanälen zu einer SLT wird vom Management vorgenommen. Zeigen Leistungsmessungen eine dauernde Überlastung einer SLT an, so kann während des Betriebs eine Umkonfigurierung vorgenommen werden.

Ebene 2 – *CALLPs* (*Call Processing*) für Schicht-3-Zeichengabe (z.B. Q.931, Q.2931, B-ISUP, N-ISUP) und Rufbearbeitung.



**Bild A-10:** Lastzu- und Lastweiterleitung (*Load Balancing*)

Jeder Prozessorknoten berechnet, auf Basis der eigenen CPU-Auslastung, den eigenen Lastzustand. Aus dieser Information u. a. wird für die lokale Überlastabwehr eine Rufzurückweisungswahrscheinlichkeit abgeleitet.



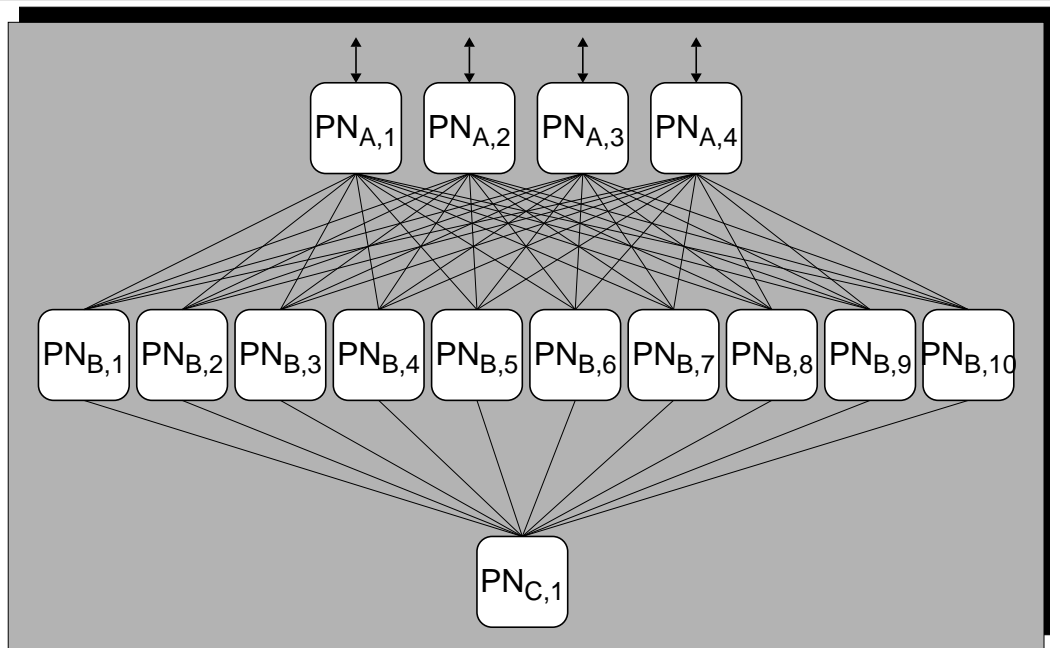
**Bild A-11:** Abstrahierte Steuerungsstruktur (zwei Funktionsebenen)

Weiterhin wird eine Akzeptanzwahrscheinlichkeit für indirekte Neurufe, d.h. Rufe die von anderen Prozessorknoten weitergeleitet wurden, berechnet. Diese wird prinzipiell erhöht, je kleiner die lokale Belastung ist und umgekehrt. Aus diesen Parametern wird abgeleitet, wieviel Prozent neuankommender Rufe lokal bearbeitet werden und wieviel Prozent jeweils an die anderen Prozessorknoten gehen. Ein nicht weitergeleiteter Ruf kann dennoch von der lokalen Überlastabwehr zurückgewiesen werden. In den Mechanismus zur Lastweiterleitung ist eine zentrale Instanz involviert, so dass kein völlig verteilter Algorithmus vorliegt.



### A.3.2 Mehrstufige Regelkreisstruktur – Lucent/AT & T

Von *Lucent* wurde nachfolgendes Konzept für die Laststeuerung in einer hierarchischen, dreistufigen Steuerungsstruktur vorgeschlagen.<sup>10</sup> Die Aufgabenbereiche von *A*- bzw. *B*-Ebene sind Signalisierung bzw. Rufbearbeitung (Bild A-12). Ebene *C* ist für zentrale Systemfunktionen verantwortlich. In jeder Ebene arbeiten mehrere Prozessorknoten in Lastteilung.



**Bild A-12:** Steuerungsstruktur (drei Funktionsebenen, die Anzahl der den einzelnen Ebenen zugeordneten *PN* entspricht der Simulation von [79])

Die Lastregelung im Vermittlungsrechner erfolgt durch verschiedene kooperierende Regelkreise mit unterschiedlichem räumlich und zeitlich begrenztem Wirkungsbereich. Zentrale Strategie ist die Lastverteilung auf *B*-Prozessoren, um eine gleichmäßige Belastung dieser Funktionsebene zu erreichen. Zunächst erfolgt eine einfache Zugangskontrolle (*Adaptive Throttling*) durch eine lokale Überlastabwehr in den  $PN_{A,i}$ .

Ein  $PN_{A,i}$  entscheidet, welchem  $PN_{B,j}$  eine neue Rufanforderung zugewiesen wird.<sup>11</sup> Alle zu einem angenommenen Ruf gehörigen Folgemeldungen werden zum selben  $PN_{B,j}$  geroutet (sogenanntes *Call Routing*). Alle Algorithmen zur Lastregelung sollen ohne zentrale Instanz auskommen.

<sup>10</sup> Alle Informationen zu diesem Konzept entstammen [79, 354, 382]. Der Algorithmus ist patentiert [584].

<sup>11</sup> Lastzuteilung auf Basis von Berechtigungen (*Token*). Jeder *A*-Prozessor besitzt für jeden *B*-Prozessor eine „*Token-Bank*“ sowie eine weitere für abgewiesene Rufe („*wastebasket*“). Bei jeder neuen Rufankunft wird jeder Tokenbank ein bestimmter Creditanteil aufaddiert und von der Tokenbank, dessen *B*-Prozessor der Ruf zugewiesen wird (= Prozessor mit den meisten Credits), ein Credit subtrahiert. Der Creditanteil berechnet sich aus der Rufzurückweisungswahrscheinlichkeit für einen *A*-Prozessor („*Throttling*“) und einem Zuweisungsfaktor für den  $PN_{B,j}$ , der auf Basis eines Lastkennwertes (*Queue Length*) ermittelt wird.

Das vorgeschlagene Verfahren besitzt folgende Vorteile:

- ⊕ Es findet keine Lastverschiebung (*Load Forwarding*) zwischen *B*-Prozessoren statt.
- ⊕ Jeder *B*-Prozessor meldet einen Lastkennwert an alle *A*-Prozessoren (einfache Bestimmung der Systembelastung).
- ⊕ Es können unterschiedlich leistungsfähige Prozessoren in der *B*-Ebene eingesetzt werden (Inhomogenität).
- ⊕ Robustes Verfahren für diese Form von Steuerungsarchitektur.

## Anhang B

# Leistungsverhalten einfacher Warteschlangensysteme

Im Kapitel 3 wurde der Funktionsverlauf exemplarischer Last- und Leistungskenngrößen einfacher Systeme diskutiert. Für leistungsoptimierende Verfahren ist dabei wichtig, welche Extrema solche Funktionen prinzipiell besitzen und in welchem Lastbereich diese liegen. Bei Systemen mit mehreren Verkehrsklassen liegen mehrdimensionale Funktionen für die Leistungsmaße vor. Als vorteilhafte Größe hat sich hier das Maß *Power-Produkt* (siehe Abschnitt 3.1.3.2) mit seinem globalen Maximum bewährt. Ein mathematisch fundierter Beweis für dieses Extremum existiert noch nicht und erscheint auch sehr aufwendig. Durch eine Abschätzung soll aber hier der „Lastraum“ dieses Maximums näherungsweise lokalisiert werden (Abschnitt B.4).

Vorbereitend werden die bekannten Leistungsgrößen *Durchsatz* und *Antwortzeit* an relativ einfachen Modellen<sup>1</sup> diskutiert sowie das prinzipielle Verhalten der *Power*-Kenngröße aufgezeigt.

### B.1 M/M/N-System – Multiprozessorstruktur

Das *M/M/N*-Modell ist charakterisiert durch eine endliche Anzahl von  $N$  Prozessoren. Bei kleinem Angebot und großem  $N$  ist die Wahrscheinlichkeit eines Ressourcenengpasses sehr klein und jeder Auftrag wird sofort bedient. Der mittlere Durchsatz  $\phi$  entspricht dann der mittleren Ankunftsrate  $\lambda$  und die mittlere Antwortzeit  $\tau$  der mittleren Bedienzeit  $h$ . Im Zusammenhang mit Vermittlungsrechnern kann dieses System als Modell im Niedriglast-Arbeitsbereich für entsprechende Steuerungsstrukturen verwendet werden.

### B.2 M/M/1/∞-System – Begrenzte Ressource ‚CPU‘

Beim *M/M/1/∞*-Modell ist nur eine Verarbeitungseinheit vorhanden (*Single Server*). Während somit die Ressource *Prozessor* limitiert ist, steht dagegen von der Ressource *Speicher* eine unbegrenzte Kapazität zur Verfügung. Die Bedieneinheit bildet somit den Systemengpass.

#### B.2.1 Grundgleichungen

Es treten keine Verluste auf. Aus den *Zustandswahrscheinlichkeiten* [42]

---

<sup>1</sup> Poisson-Ankunftsprozess, negativ-exponentiell verteilte Bediendauern, FIFO-Abfertigungsdisziplin sowie nicht-unterbrechende Prioritäten.

$$p_0 = 1 - \rho \quad \text{und} \quad p_K = (1 - \rho)\rho^K \quad \text{für} \quad \rho = \frac{\lambda}{\mu} < 1, \quad K = 1, 2, \dots \quad (\text{B-1})$$

und der *mittleren Systembelegung*

$$\bar{X} = E[X] = \frac{\rho}{1 - \rho} \quad (\text{B-2})$$

ergibt sich nach *Little* [359] folgende *mittlere Aufenthaltsdauer*:

$$\tau(\lambda) = h \cdot \frac{1}{1 - \rho} = \frac{h}{1 - \lambda h}. \quad (\text{B-3})$$

Für die Leistungsgröße *Power*  $\Lambda_P(\lambda)$  (siehe Gl. (3-9), S. 53) ergibt sich folgender Zusammenhang für den *Normallastbereich* ( $\rho < 1$ ):

$$\Lambda_P(\lambda) = \frac{\phi(\lambda)}{\tau(\lambda)} = \frac{\phi^2(\lambda)}{\bar{X}(\lambda)} = \lambda \cdot (\mu - \lambda) = -\lambda^2 + \frac{1}{h}\lambda. \quad (\text{B-4})$$

Dieses Leistungsmaß folgt somit einer Parabelfunktion und besitzt sein *Maximum*  $\Lambda_{P,max}$ , gleich  $1/(4 \cdot h^2)$ , bei halber Systemauslastung ( $A = \rho = 0.5$ , d.h.  $\lambda = 1/(2 \cdot h)$ ).<sup>2</sup> Für ein Angebot von  $A \geq 1$  geht die Bedieneinheit in Sättigung und das System weist ein instationäres Verhalten auf. Die Berechnung einer mittleren Antwortzeit ist deshalb dann nicht mehr möglich. Ein System, bei welchem überhaupt keine Überlast entstehen kann, wird nachfolgend beschrieben.

## B.2.2 System mit Entmutigung

Bei Terminal-Host-Systemen wurde u.a. die Beobachtung gemacht, dass bei ansteigenden Antwortzeiten der Ankunftsstrom zurückgeht. Dieses Verhalten wird als Teilnehmerentmutigung bezeichnet (*Discouraged Arrivals*, [34]). Bei diesem Modell wird somit eine Rückwirkung auf den Teilnehmer miteinbezogen. Die Ankunftsrate wird bei zunehmender Systembelegung kleiner. Durch eine *zustandsabhängige Ankunftsrate*  $\lambda_i$

$$\lambda_i = \lambda^* \cdot \frac{1}{i + 1} \quad \text{für} \quad i = 0, 1, 2, \dots \quad (\text{B-5})$$

kann dieser Effekt modelliert werden [34]. Im Mittel befinden sich hier

$$\bar{X} = \frac{\lambda^*}{\mu} = \lambda^* \cdot h \quad (\text{B-6})$$

<sup>2</sup> Mittlerer Durchsatz und mittlere Antwortzeit werden so ausbalanciert, dass das System den *halben Maximaldurchsatz*  $\phi_{opt} = 1/2 \cdot \phi_{max} = 1/(2 \cdot h)$  bei *doppelter Minimalantwortzeit*  $\tau_{opt} = 2 \cdot \tau_{min} = 2 \cdot h$  in diesem Arbeitspunkt erzielt! In diesem Lastpunkt gilt weiterhin für die *Steigungen* von Durchsatzfunktion und Antwortzeitfunktion folgender Zusammenhang:

$$\phi(\lambda) \frac{d}{d\lambda} \tau(\lambda) = \tau(\lambda) \frac{d}{d\lambda} \phi(\lambda) \quad \text{bzw.} \quad \frac{d\tau(\lambda)}{\tau(\lambda)} = \frac{d\phi(\lambda)}{\phi(\lambda)}.$$

D.h., der relative Zuwachs an Antwortzeit ist gleich dem relativen Zuwachs an Durchsatz. KLEINROCK hat in [198] das *Power-Maß* für einfache, unbegrenzte Systeme ausführlich diskutiert. Für eine große Klasse von Warteschlangensystemen gilt dabei, dass für eine optimale Leistung  $\Lambda_{P,max}$  (d.h. großer Durchsatz bei kleinem Delay) jede Resource *Bedieneinheit* im Mittel so belastet werden sollte, so dass *immer gerade ein Auftrag in Bearbeitung* ist.

Aufträge im System [34]. Als Durchsatz ergibt sich eine *stationäre Rate*  $\phi$  von

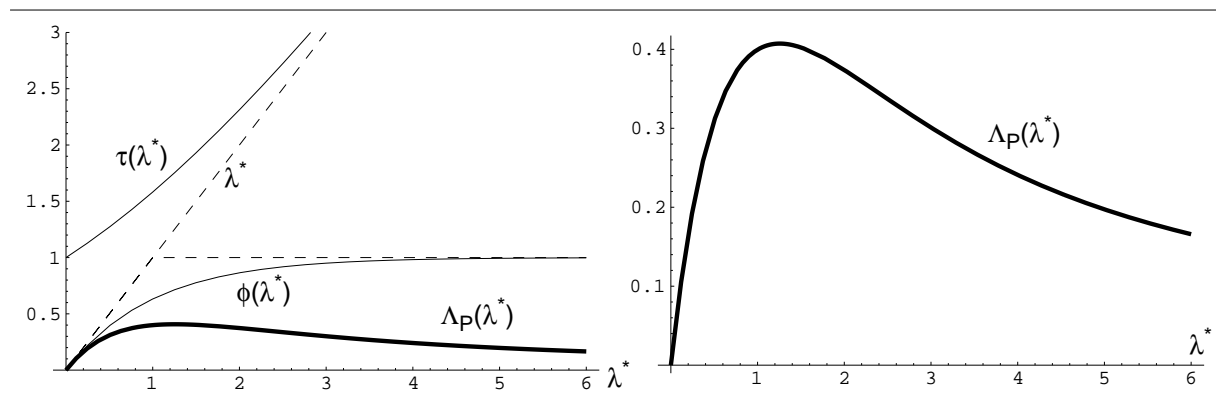
$$\phi = \lambda = \mu(1 - e^{-\lambda^*/\mu}) = \frac{1 - e^{-\lambda^*h}}{h}. \quad (\text{B-7})$$

Daraus resultiert eine *mittlere Aufenthaltsdauer*  $\tau$  von:

$$\tau = \frac{\bar{X}}{\phi} = \frac{\lambda^*h^2}{1 - e^{-\lambda^*h}}. \quad (\text{B-8})$$

Für das Leistungsmaß *Power*  $\Lambda_P(\lambda)$  ergibt sich folgende Exponentialfunktion:

$$\Lambda_P(\lambda) = \frac{\phi}{\tau} = \frac{1}{\lambda^*h^3} \cdot (1 - e^{-\lambda^*h})^2. \quad (\text{B-9})$$



**Bild B-1:** System mit Entmutigung – Funktionsverläufe für  $h = 1$  (Abszisse: Angebot  $\lambda^*$  bei leerem System).<sup>3</sup> Links: Antwortzeit  $\tau_C(\lambda^*)$ , Durchsatz  $\phi(\lambda^*)$  und Power  $\Lambda_P(\lambda^*)$ ; Rechts: Ausschnittsvergrößerung von  $\Lambda_P(\lambda^*)$ .

Die Funktionsverläufe für eine mittlere Bedienzeit von  $h = 1$  sind in Bild B-1 dargestellt.<sup>4</sup> Die Leistungskenngröße *Power* weist für diesen Systemtyp ebenfalls ein globales Maximum auf.

### B.3 M/M/1/K-System – Begrenzte Ressource ‚Speicher‘

Neben der begrenzten Ressource *CPU-Zeit* werden jetzt Systeme betrachtet, bei denen nun auch nur endlich viele Warteplätze ( $s = K - 1$ ) zur Verfügung stehen. Ab einer gewissen Last gehen nun zwangsläufig Aufträge verloren. In diesem Unterkapitel wird zunächst der Funktionsverlauf der *Power-Kenngröße* für den gesamten Lastbereich ermittelt. Anschließend wird das System zusätzlich mit einer speziellen Überlastabwehrstrategie ausgestattet.

<sup>3</sup> Zur Beschriftung der Schaubilder in diesem Kapitel: Alle Ankunftsrate und Wartezeiten sind auf die jeweiligen Bedienzeiten (i. d. R.  $h = 1$ ) normiert. Dann sind i. a. normierte Ankunftsrate und Angebot identisch.

<sup>4</sup> Die Veränderliche ist hier nicht das wirkliche Verkehrsangebot  $A$ , sondern das nur bei leerem System auftretende Angebot  $\lambda^* \cdot h$ . Das mittlere Angebot  $A$  bleibt hier immer kleiner eins. Für den Zusammenhang  $\lambda^*(A)$  gilt nach Gleichung (B-7):  $\lambda^*(A) = (1/h) \ln[1/(1 - A)]$ .

### B.3.1 Ohne Überlastabwehr

Zunächst soll das System keine expliziten Lastregelmechanismen vornehmen. Eine neue Anforderung, die auf ein volles System trifft, geht verloren. Diese „natürliche Überlastabwehr“, genannt auch *FIFO-Blocking* [152], macht für bestimmte Systeme Sinn (z.B. [204,205]) und kann auch relativ einfach realisiert werden. Die *Zustandswahrscheinlichkeiten*  $p_i$  berechnen sich nach [25] wie folgt:

$$p_i = \begin{cases} p_0 \cdot \rho^i & \text{für } i \leq K \\ 0 & \text{für } i > K \end{cases} \quad (\text{B-10})$$

mit

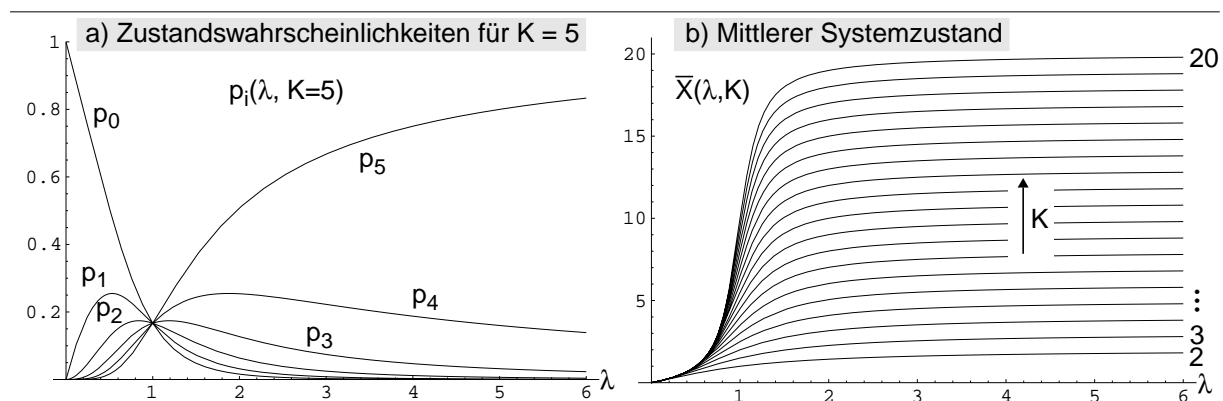
$$p_0 = \frac{1 - \rho}{1 - \rho^{K+1}} \text{ und } \rho = \frac{\lambda}{\mu} = \lambda h. \quad (\text{B-11})$$

Bei diesem Systemtyp darf das Verhältnis der Übergangsraten  $\rho \geq 1$  sein. Das *Verkehrsangebot*  $A$  entspricht diesem Verhältnis  $\rho$  ( $A = \rho = \lambda h$ ). Für den *Grenzfall*  $A = 1$  besitzen alle Zustandswahrscheinlichkeiten denselben Wert (siehe Bild B-2 für  $K = 5$  und  $h = 1$ ):

$$\lim_{A \rightarrow 1} p_i = \frac{1}{K+1}. \quad (\text{B-12})$$

Die *mittlere Systembelegung*  $\bar{X}$  ergibt sich allgemein zu (siehe auch Bild B-2):

$$\bar{X}(\lambda, K) = \frac{\lambda h [1 - (K+1)(\lambda h)^K + K(\lambda h)^{K+1}]}{(1 - \lambda h)(1 - (\lambda h)^{K+1})}. \quad (\text{B-13})$$



**Bild B-2:** Zustandswahrscheinlichkeiten  $p_i(\lambda)$  für  $K = 5$  und mittlere Systembelegung  $\bar{X}(\lambda, K)$  mit  $K \in [2, 20]$  ( $h = 1$ )

Trifft eine neue Anforderung auf eine vollbelegte Warteschlange, dann geht sie ohne Inanspruchnahme von Prozessorzeit verloren. Die Verlustwahrscheinlichkeit ist gleich  $p_K$  (entspricht  $p_5$  in Bild B-2). Der ankommende Verkehr  $\lambda$  teilt sich somit in die zwei Ströme  $\phi_{AC}$  und  $\phi_{BC}$  von angenommenen und blockierten Rufen auf. Die mittleren Raten ergeben sich wie folgt (siehe Bild B-3):

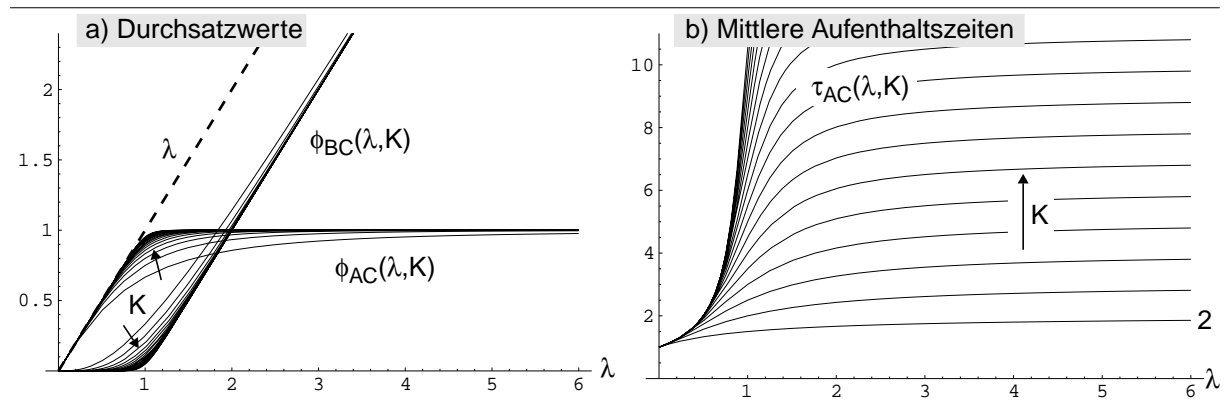
$$\phi_{AC}(\lambda, K) = \lambda \cdot (1 - p_K) = \begin{cases} \lambda \cdot \frac{1 - (\lambda h)^K}{1 - (\lambda h)^{K+1}} & \text{für } \rho \neq 1 \\ \lambda \cdot \frac{K}{1 + K} & \text{für } \rho = 1 \end{cases} \quad (\text{B-14})$$

$$\phi_{BC}(\lambda, K) = \lambda \cdot p_K = \begin{cases} \lambda \cdot \frac{(\lambda h)^K [1 - (\lambda h)]}{1 - (\lambda h)^{K+1}} & \text{für } \rho \neq 1 \\ \lambda \cdot \frac{1}{1 + K} & \text{für } \rho = 1 \end{cases}$$

Nach LITTLE [359] lassen sich die *mittleren Aufenthaltszeiten*  $\tau_{AC}$  für die akzeptierten Rufe berechnen (siehe Bild B-3):

$$\tau_{AC}(\lambda, K) = \frac{\bar{X}}{\phi_{AC}} = h \cdot \frac{1 - (K + 1)(\lambda h)^K + K(\lambda h)^{K+1}}{1 - \lambda h - (\lambda h)^K + (\lambda h)^{K+1}}. \quad (\text{B-15})$$

Die Aufenthaltszeit  $\tau_{BC}$  einer blockierten Meldung macht hier keinen Sinn.



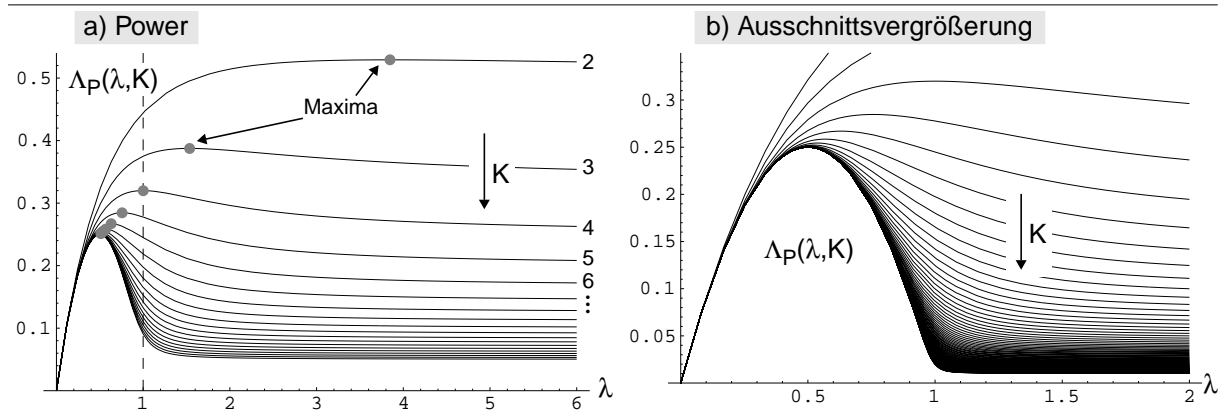
**Bild B-3:** Angebot  $\lambda$  und Durchsätze  $\phi_{AC}(\lambda, K)$  bzw.  $\phi_{BC}(\lambda, K)$  sowie die mittleren Aufenthaltszeiten  $\tau_{AC}(\lambda, K)$  für  $h = 1$

Aus Gleichungen (B-14, B-15) ergibt sich für das Leistungsmaß *Power*  $\Lambda_P(\lambda, K)$  folgender Zusammenhang:

$$\Lambda_P(\lambda, K) = \frac{\phi_{AC}}{\tau_{AC}} = \frac{\lambda h \cdot (1 - \lambda h) \cdot (1 - (\lambda h)^K)^2}{(1 - (\lambda h)^{K+1}) \cdot [1 - (K + 1)(\lambda h)^K + K(\lambda h)^{K+1}]}. \quad (\text{B-16})$$

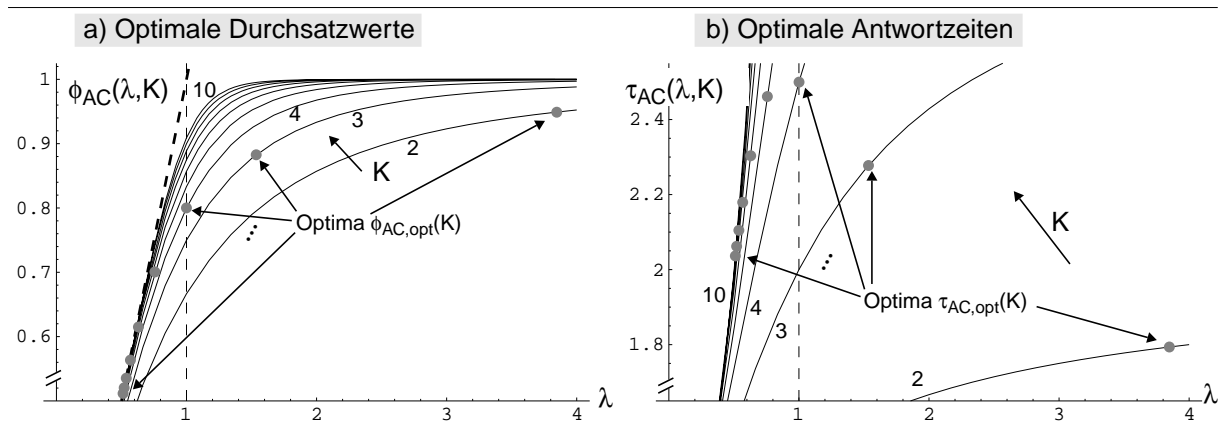
Diese Funktion besitzt ein globales Maximum:

$$\frac{\partial}{\partial \lambda} \Lambda_P(\lambda, K) \stackrel{!}{=} 0. \quad (\text{B-17})$$



**Bild B-4:** Funktionsverlauf von  $\Lambda_P(\lambda, K)$  für  $h = 1$  und  $K = 2, 3, \dots$

Die numerisch<sup>5</sup> berechneten Maxima sowie der Funktionsverlauf von  $\Lambda_P(\lambda, K)$  für  $h = 1$  sind in Bild B-4 dargestellt. Die gestrichelte Linie trennt den Arbeitsbereich in den *Normal-* ( $A < 1$ ) und *Überlastbereich* ( $A \geq 1$ ). Wie sich zeigen lässt, befindet sich beim *M/M/1/K*-System das globale Maximum für  $K > 4$  im Normallastbereich.<sup>6</sup>



**Bild B-5:** Optimale Durchsatzwerte und Antwortzeiten für die *Power-Maxima*  $\Lambda_{P,max}(K)$  für  $K = 2, 3, \dots, 10$ .

In Bild B-5 sind die den maximalen *Power*-Werten  $\Lambda_{P,max}(K)$  zugehörigen Durchsatz- und Antwortzeitwerte dargestellt. Die Funktion  $\phi_{AC,opt}(K)$  ist monoton fallend und  $\tau_{AC,opt}(K)$  besitzt ein Maximum für  $K = 4$ . Für einen *unbegrenzten Warteraum* konvergiert der optimale Durchsatz gegen  $\phi_{AC,opt,K \rightarrow \infty} = 1/2$  und die optimale Antwortzeit gegen  $\tau_{AC,opt,K \rightarrow \infty} = 2$  (siehe auch Abschnitt B.2.1).

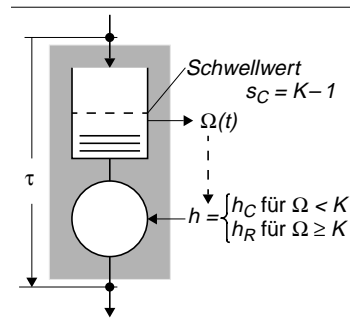
<sup>5</sup> Alle Berechnungen dieses Kapitels wurden mit dem Werkzeug MATHEMATICA 3.0 durchgeführt.

<sup>6</sup> Für  $K = 4$  liegt das Maximum auf der Grenze zwischen Normal- und Überlastbereich, d.h.,  $\Lambda_{P,max} = 8/25 = 0.32$  für  $A = 1$  (mit  $\phi_{opt} = 4/5$  und  $\tau_{opt} = 5/2$ ).

### B.3.2 Überlastabwehr auf Basis der Warteschlangenbelegung

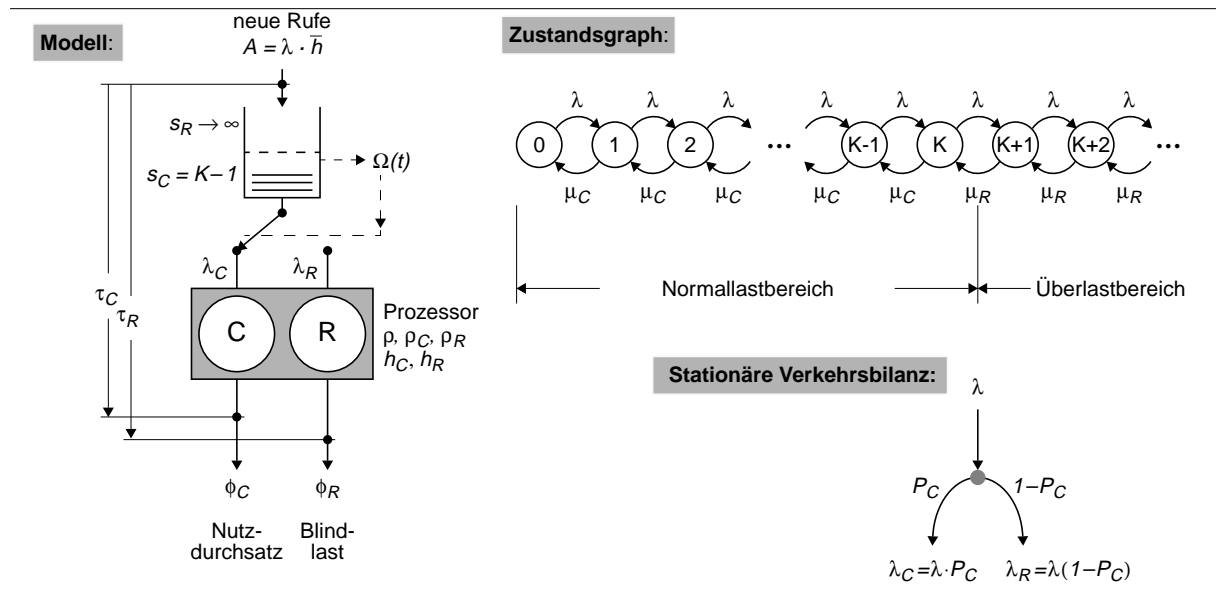
Neben der im vorherigen Abschnitt betrachteten impliziten Überlastabwehr („endlicher Warteraum“), soll nun das  $M/M/1/K$ -System um ein typisches Lastregelverfahren erweitert werden. In einem Vermittlungsrechner sei folgender einfacher *Überlastabwehrmechanismus* realisiert:

- Der Pufferbereich für *neue Rufe* kann maximal  $K-1$  Meldungen aufnehmen (siehe Bild B-6).
- Trifft ein neuer Ruf auf einen vollen Puffer, dann wird der älteste Ruf *zurückgewiesen*.
- Die notwendige (mittlere) Bedienzeit für eine Rufzurückweisung  $h_R$  sei wesentlich kleiner als die (mittlere) Bearbeitungszeit  $h_C$  für einen angenommenen Ruf ( $h_C \gg h_R$ ).



**Bild B-6:** Modell

Das stationäre Verhalten dieses Monoprocessorsystems lässt sich auf ein  $M/M_1M_2/1/K, \infty$ -Warteschlangenmodell<sup>7</sup> abstrahieren (siehe Bild B-7 mit Parameterbeschreibung). Während in Bild B-6 der Steuermechanismus in ein *Single-Server-System* eingezeichnet wurde, soll zusätzlich zum besseren Verständnis derselbe Mechanismus in Bild B-7 als zweiphasiges Bediensystem dargestellt werden.



**Bild B-7:**  $M/M_1M_2/1/K, \infty$ -Warteschlangensystem mit Zustandsgraph

Die Abfertigungsdisziplin ist FIFO. Jeweils zu Bedienbeginn beeinflusst der Pufferfüllstand die Entscheidung, ob die Rufanforderung auf dem vordersten Platz des FIFO angenommen oder ob sie schnell abgefertigt und abgewiesen wird. Diese Aussage ist eine Interpretation der Markoff-Kette. Der Zustandsgraph beschreibt letztlich die exakte Arbeitsweise dieses Systems.

<sup>7</sup> Die Bezeichnung „ $M/M_1M_2/1/K, \infty$ “ soll verdeutlichen, dass die verschiedenen Rufotypen unterschiedliche mittlere Bedienzeiten besitzen, und dass der Warteraum in zwei logische Bereiche mit unterschiedlicher Kapazität ( $s_C = K-1$  bzw.  $s_R \rightarrow \infty$ ) aufgeteilt ist.



### B.3.2.1 Stationäres Verhalten

Im stationären Zustand gilt für die Meldungsraten (siehe Bild B-7)  $\lambda_C = \phi_C$  bzw.  $\lambda_R = \phi_R$ . Die *Zustandswahrscheinlichkeiten* berechnen sich nach [34] wie folgt:

$$p_i(\lambda, K) = \begin{cases} p_0 \cdot \tilde{\rho}_C^i & \text{für } 0 \leq i \leq K \\ p_0 \cdot \tilde{\rho}_C^K \cdot \tilde{\rho}_R^{i-K} & \text{für } i \geq K + 1 \end{cases} \quad (\text{B-18})$$

mit

$$p_0(\lambda, K) = \frac{1}{1 + \sum_{i=1}^K \tilde{\rho}_C^i + \tilde{\rho}_C^K \cdot \sum_{i=K+1}^{\infty} \tilde{\rho}_R^{i-K}} = \frac{(1 - \tilde{\rho}_R)(1 - \tilde{\rho}_C)}{1 - \tilde{\rho}_R + (\tilde{\rho}_R - \tilde{\rho}_C)\tilde{\rho}_C^K} \quad (\text{B-19})$$

für die Randbedingung

$$\tilde{\rho}_R < 1 \quad (\text{Anmerkung: } \tilde{\rho}_C < 1 \text{ ist nicht erforderlich}). \quad (\text{B-20})$$

Dabei repräsentieren die  $\tilde{\rho}_R$  und  $\tilde{\rho}_C$  die *Verhältnisse der Übergangsraten* (siehe Zustandsgraph)

$$\tilde{\rho}_C = \frac{\lambda}{\mu_C} = \lambda h_C \quad \text{bzw.} \quad \tilde{\rho}_R = \frac{\lambda}{\mu_R} = \lambda h_R, \quad (\text{B-21})$$

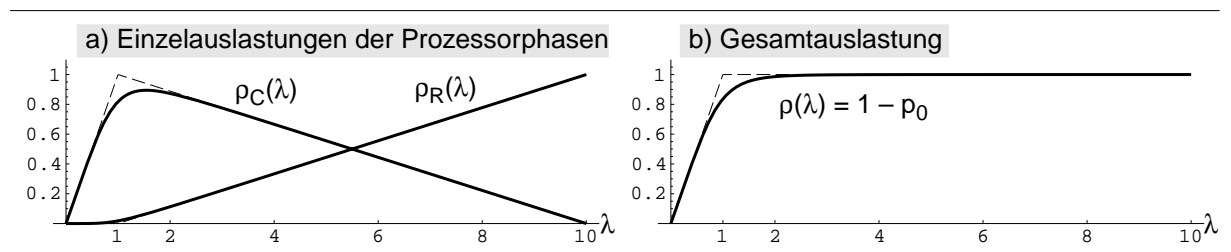
d.h.  $\rho_i$  die *Auslastungen der Prozessorphasen*. Diese entsprechen folgenden Zustandsräumen:<sup>8</sup>

$$\rho_C = P\{X \in [1, K]\} \quad \text{bzw.} \quad \rho_R = P\{X \in [K + 1, \infty)\}. \quad (\text{B-22})$$

Mit Gl. (B-18, B-19) ergeben sich hierfür folgende Funktionsverläufe (siehe auch Bild B-8):

$$\rho_C(\lambda, K) = p_0 \sum_{i=1}^K \tilde{\rho}_C^i = \tilde{\rho}_C \cdot \frac{(1 - \tilde{\rho}_R)(1 - \tilde{\rho}_C^K)}{1 - \tilde{\rho}_R + (\tilde{\rho}_R - \tilde{\rho}_C)\tilde{\rho}_C^K} \quad (\text{B-23})$$

$$\rho_R(\lambda, K) = \sum_{i=K+1}^{\infty} p_i = p_0 \tilde{\rho}_C^K \sum_{i=K+1}^{\infty} \tilde{\rho}_R^{i-K} = \tilde{\rho}_R \cdot \frac{(1 - \tilde{\rho}_C)\tilde{\rho}_C^K}{1 - \tilde{\rho}_R + (\tilde{\rho}_R - \tilde{\rho}_C)\tilde{\rho}_C^K} \quad (\text{B-24})$$

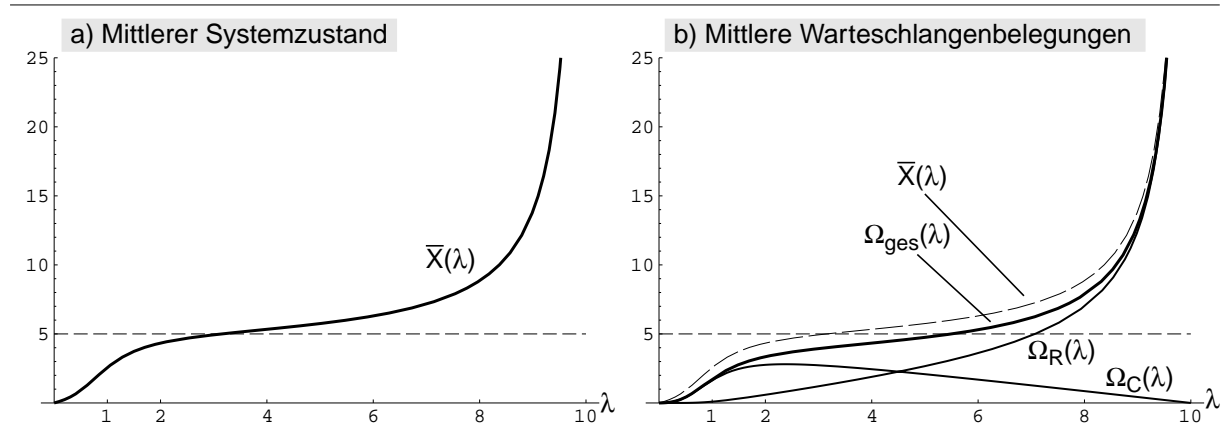


**Bild B-8:** Funktionsverläufe für  $K = 5$ ,  $h_C = 1$  und  $h_R = 0.1 h_C$ : a) Teilauslastungen  $\rho_C(\lambda)$  bzw.  $\rho_R(\lambda)$ ; b) Gesamtauslastung  $\rho(\lambda)$ .

<sup>8</sup> Die Wahrscheinlichkeit für ein leeres System ist  $p_0 = 1 - \rho_C - \rho_R$ .

Der Funktionsverlauf der Auslastung der Komplettierungsphase  $\rho_C(\lambda)$  entspricht einer klassischen *Überlastkurve*. Die Gesamtauslastung  $\rho(\lambda)$  konvergiert relativ schnell gegen eins. Die *mittlere Gesamtsystembelegung*  $\bar{X}$  (siehe Bild B-9) ergibt sich zu:

$$\bar{X}(\lambda, K) = \sum_{i=0}^{\infty} (i \cdot p_i) = K - 1 + \frac{1}{1 - \tilde{\rho}_C} + \frac{1}{1 - \tilde{\rho}_R} - \frac{1 + K(1 - \tilde{\rho}_R)}{1 - \tilde{\rho}_R + (\tilde{\rho}_R - \tilde{\rho}_C)\tilde{\rho}_C^K}. \quad (\text{B-25})$$



**Bild B-9:** Funktionsverläufe (Parameter siehe Bild B-8): a) mittlere Systembelegung  $\bar{X}(\lambda)$ ; b) mittlere Warteschlangenlängen  $\Omega_C(\lambda)$ ,  $\Omega_R(\lambda)$  und  $\Omega_{ges}(\lambda) = \Omega_C(\lambda) + \Omega_R(\lambda)$ .

Für das System lassen sich die folgenden *drei Arbeitsbereiche* definieren:

- *Normallast*, d.h. der stationäre Zustandsbereich ist  $X \leq K$ ,
- *Überlast*, d.h. der stationäre Zustandsbereich ist  $X > K$  und  $\tilde{\rho}_R < 1$  und
- *Instabilität*, d.h. das System verhält sich instationär wegen  $\tilde{\rho}_R \geq 1$ .

Da kein Verlustsystem vorliegt, gilt für  $\tilde{\rho}_R < 1$  der *Erhaltungssatz*, d.h.  $\lambda = \phi_C + \phi_R$ , als stationäre Verkehrsbilanz, wobei sich dann *Nutzdurchsatz* und *Ablehnungsrate* nach Little berechnen lassen:

$$\phi_C = \frac{\rho_C}{h_C} \text{ bzw. } \phi_R = \frac{\rho_R}{h_R} \text{ für } \lambda < \frac{1}{h_R}. \quad (\text{B-26})$$

Ausgeschrieben ergibt sich für den *Nutzdurchsatz*  $\phi_C$  (Bild B-10)

$$\phi_C(\lambda, h_C, h_R, K) = \lambda \frac{(1 - h_R \lambda)(1 - h_C^K \lambda^K)}{1 - h_R \lambda + (h_R - h_C)h_C^K \lambda^{K+1}}. \quad (\text{B-27})$$

Damit lassen sich nun die *Komplettierungswahrscheinlichkeit*  $P_C$  und die *Ablehnungswahrscheinlichkeit*  $P_R$  für den stabilen Arbeitsbereich berechnen:

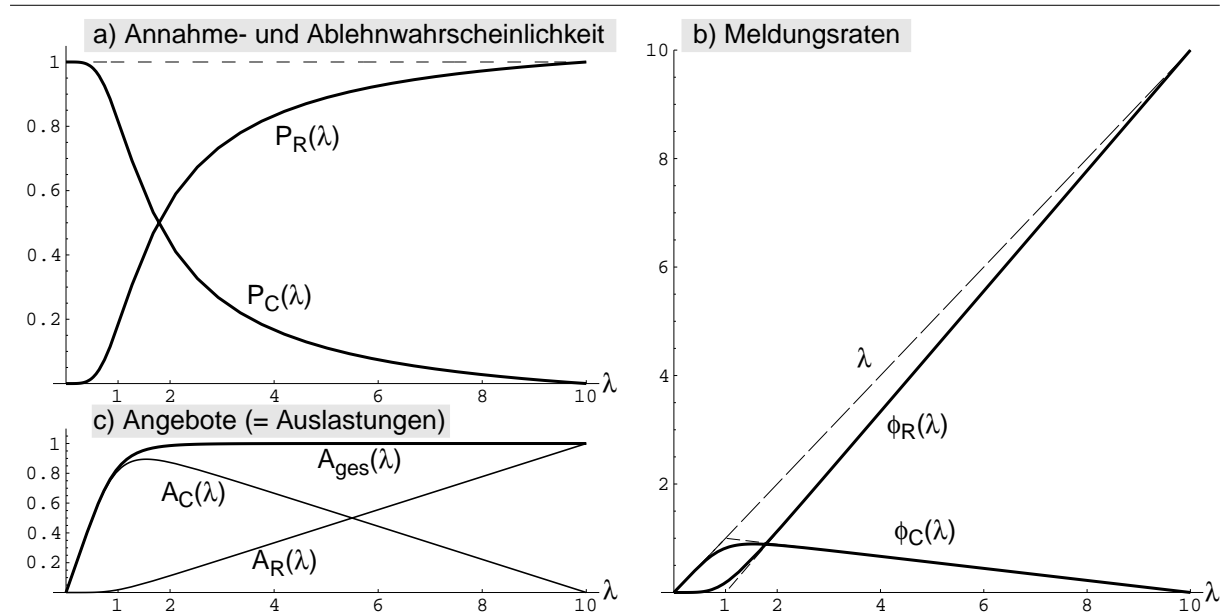
$$P_C = \frac{\phi_C}{\lambda} = \frac{\rho_C}{\lambda h_C} = \frac{\rho_C}{\tilde{\rho}_C} \text{ bzw. } P_R = \frac{\phi_R}{\lambda} = \frac{\rho_R}{\lambda h_R} = \frac{\rho_R}{\tilde{\rho}_R} \text{ mit } P_R = 1 - P_C. \quad (\text{B-28})$$

Man beachte, dass in allen Diagrammen (von Bild B-10) die einzelnen Kenngrößen hier als Funktion der Ankunftsrate  $\lambda$ , und nicht wie üblicherweise als Funktion des Gesamtangebotes  $A_{ges}$ , aufgetragen sind. Folgende stationären *Angebote*  $A_i$  stellen sich ein:

$$A_C(\lambda, K) = \rho_C = \phi_C h_C, A_R(\lambda, K) = \rho_R = \phi_R h_R \text{ und} \quad (\text{B-29})$$

$$A_{ges}(\lambda, K) = A_C + A_R = 1 - p_0(\lambda, K). \quad (\text{B-30})$$

Üblicherweise wird  $A_C$  auch als *Nutz-* und  $A_R$  als *Blindlast* bezeichnet. Unter Überlast ergibt sich im stabilen Arbeitsbereich ein Gesamtangebot von nahezu eins (Bild B-10).

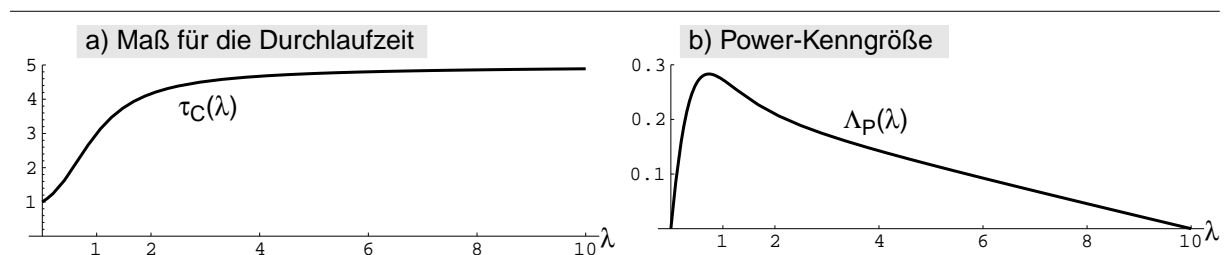


**Bild B-10:** Funktionsverläufe (Parameter siehe Bild B-8): a) Komplettierungswahrscheinlichkeit  $P_C(\lambda)$  und Ablehnwahrscheinlichkeit  $P_R(\lambda)$ ; b) Ankunftsrate  $\lambda$ , Nutzrate  $\phi_C(\lambda)$  und Ablehnungsrate  $\phi_R(\lambda)$ ; c) Gesamt-  $A_{ges}(\lambda)$  und Teilangebote  $A_C(\lambda)$  bzw.  $A_R(\lambda)$ .

Für die mittlere Durchlaufzeit *kompletierter Rufe*  $\tau_C$  ergibt sich hier:

$$\tau_C(\lambda, K) = \frac{\rho_C + \Omega_C}{\phi_C} = h_C \left( 1 + \frac{\Omega_C}{\rho_C} \right) = h_C \left[ K + \frac{1}{1 - h_C \lambda} - \frac{K}{1 - h_C^K \lambda^K} \right]. \quad (\text{B-31})$$

Wie Bild B-11 zu entnehmen ist, konvergiert die *Durchlaufzeit*  $\tau_C$  hier gegen  $K \cdot h_C = 5$  im stabilen Arbeitsbereich.



**Bild B-11:** Funktionsverläufe (Parameter siehe Bild B-8): a) mittlere Durchlaufzeit  $\tau_C(\lambda)$ ; b) *Power*-Maß  $\Lambda_P(\lambda)$ , wobei das Maximum hier im Normallastbereich liegt.

Die Power-Kenngröße  $\Lambda_P$  berechnet sich hier wie folgt (siehe auch Bild B-11):

$$\Lambda_P(\lambda, K) = \frac{\phi_C}{\tau_C} = \frac{1}{h_C^2} \cdot \frac{\rho_C^2}{\rho_C + \Omega_C}. \quad (\text{B-32})$$

Das Power-Maß  $\Lambda_P$  besitzt den typischen Funktionsverlauf mit einem *globalen Maximum*.

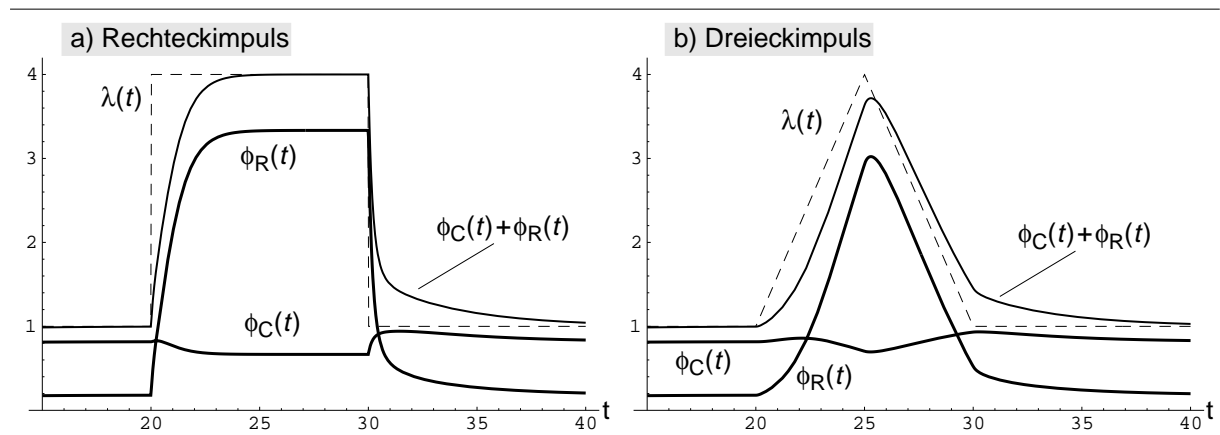
### B.3.2.2 Dynamisches Verhalten

Es soll der Zeitverlauf von *Komplettierungs-*  $\phi_C(t)$  und *Ablehnungsrate*  $\phi_R(t)$  als Reaktion auf eine Überlast ermittelt werden. Für *Geburts-/Sterbeprozesse* berechnen sich die *Zustandswahrscheinlichkeiten*  $p_i(t)$  nach den *Kolmogoroff'schen Vorwärtsgleichungen* [42]. Für das vorliegende System ergibt sich folgendes Differentialgleichungssystem:

$$\begin{aligned} \frac{d}{dt}p_0(t) &= -\lambda(t)p_0(t) + \mu_C p_1(t) && \text{für } i = 1 \\ \frac{d}{dt}p_i(t) &= -[\lambda(t) + \mu_C]p_i(t) + \lambda(t)p_{i-1}(t) + \mu_C p_{i+1}(t) && \text{für } 1 \leq i \leq K - 1 \\ \frac{d}{dt}p_K(t) &= -[\lambda(t) + \mu_C]p_K(t) + \lambda(t)p_{K-1}(t) + \mu_R p_{K+1}(t) && \text{für } i = K \\ \frac{d}{dt}p_i(t) &= -[\lambda(t) + \mu_R]p_i(t) + \lambda(t)p_{i-1}(t) + \mu_R p_{i+1}(t) && \text{für } i \geq K + 1 \end{aligned} \quad (\text{B-33})$$

mit der Normierungsbedingung  $\sum_{i=0}^{\infty} p_i(t) = 1$  für alle  $t$ .

Die Reaktion auf einen *Rechteck-* und einen *Dreieckimpuls* sind in Bild B-12 dargestellt.



**Bild B-12:** Funktionsverläufe<sup>9</sup>  $\phi_C(t)$ ,  $\phi_R(t)$  und  $\lambda(t)$  für  $K = 5$ ,  $h_C = 1$  und  $h_R = 0.1 h_C$  (Abszisse: Zeit  $t$  [ZE]): a) Rechteckimpuls; b) Dreieckimpuls.

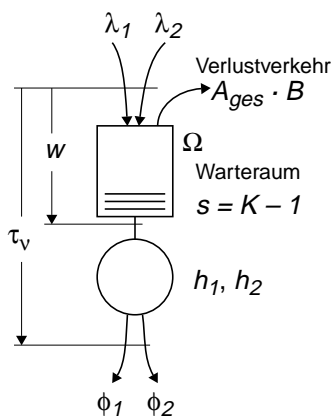
Nach einem Lastanstieg steigt die Kompletierungsrate  $\phi_C(t)$  kurzfristig minimal an bevor sie wieder abfällt. Unter Überlast wird hier der Großteil der Anforderungen abgelehnt. Der stationäre Durchsatz  $\phi_C$  unter Überlast ist kleiner als unter Normallast (siehe Rechteckimpuls).

<sup>9</sup> Für die numerisch berechneten Lösungen wurde der gesamte Warteraum auf 19 Plätze begrenzt, d.h. bei  $K = 5$  ist  $s_C = 4$  und  $s_R = 15$ . Es gehen also hier Anforderungen mit der Wahrscheinlichkeit  $p_{20}(t)$  verloren, die allerdings hier für die zugrundeliegenden Parameter vernachlässigbar klein ist.

## B.4 Stationäre Leistungsmaße des $M_1M_2/M_1M_2/1/K$ -Systems

Im folgenden soll der *stationäre Verlauf* von bestimmten Leistungsmaßen für ein einfaches System mit *mehr als einer Verkehrsklasse* (auch Auftragsklassen genannt) gezeigt werden. Im Mittelpunkt stehen dabei die in Abschnitt 3.1.3.2 eingeführten *Power*-Maße. Für das sogenannte *Power-Produkt* soll gezeigt werden, dass dieses ein globales Maximum besitzt.

Aus Gründen der Einfachheit wurde das  $M_1M_2/M_1M_2/1/K$ -Modell<sup>10</sup> gewählt, d.h. es werden zwei Auftragsklassen betrachtet. Für dieses Warteschlangensystem lassen sich mathematisch geschlossene Ausdrücke für bestimmte Zeitmittelwerte angeben, oder es können zumindest relativ einfach approximative Lösungen abgeschätzt werden. Bild B-13 zeigt das entsprechende Strukturbild und die im folgenden betrachteten Leistungsmaße.



### Leistungsmaße (Mittelwerte)

Einfache (direkte) Größen:

- *Durchsatz*  $\phi_v$  für Verkehrsklasse  $v$  bzw.  $\phi_{ges}$  und
- *Antwortzeiten*  $\tau_v$  für Verkehrsklasse  $v$  bzw.  $\tau_{ges}$ .

Abgeleitete (indirekte) Größen:

- *Power*  $\Lambda_{P,v} = \phi_v / \tau_v$  für Verkehrsklasse  $v$  bzw.  $\Lambda_{P,ges}$ ,
- *Power-Summe*  $\Lambda_{P,\Sigma} = \sum_v \Lambda_{P,v}$  und
- *Power-Produkt*  $\Lambda_{P,\Pi} = \prod_v \Lambda_{P,v}$ .

**Bild B-13:**  $M_1M_2/M_1M_2/1/K$ -System und interessierende Leistungsmaße

Zunächst soll eine Unterteilung der *Lastebene* in verschiedene *Arbeitsbereiche* vorgenommen werden. Da hier *stationäre* Größen betrachtet werden sollen, ist der jeweilige Arbeitspunkt  $\{\lambda_1, \lambda_2\}$  (oder Lastvektor) *zeitunabhängig*. Die *mittleren Angebote* für die beiden Verkehrsklassen sind

$$A_1(\lambda_1) = \lambda_1 \cdot h_1 \text{ und } A_2(\lambda_2) = \lambda_2 \cdot h_2. \quad (\text{B-34})$$

Ein *unbegrenztes System* ( $K \rightarrow \infty$ ) würde für

$$A_{ges}(\lambda_1, \lambda_2) = A_1 + A_2 < 1 \quad (\text{B-35})$$

in einen *stationären Zustand* übergehen. Da jede Meldung – egal welcher Verkehrsklasse – dieselbe *Wartezeit*  $w(\lambda_1, \lambda_2)$  in einem bestimmten Arbeitspunkt besitzt, kann dieser Kennwert als „gemeinsame Kopplungsgröße“ aller Verkehrsklassen  $v$  aufgefasst werden.

<sup>10</sup> Die Schreibweise  $M_1M_2\dots/M_1M_2\dots/1$  für ein System mit mehreren Verkehrsklassen (*multi-class systems*) wurde von [3,297] übernommen.

## Warte- und Durchlaufzeiten

Für ein *reines Wartesystem* ( $M_1M_2/M_1M_2/1$ ) ergibt sich für die *mittlere Wartezeit* z.B. nach PATEROK [422]<sup>11</sup> folgende Gleichung:

$$w_{WS}(\lambda_1, \lambda_2) = \frac{A_1 \cdot h_1 + A_2 \cdot h_2}{1 - (A_1 + A_2)} = \frac{\lambda_1 \cdot h_1^2 + \lambda_2 \cdot h_2^2}{1 - (\lambda_1 \cdot h_1 + \lambda_2 \cdot h_2)} \text{ für } A_{ges} < 1. \quad (\text{B-36})$$

Da aber von einer *begrenzten Warteschlange* ( $s = K - 1$  Pufferplätze) ausgegangen wird, ist für  $A_{ges} \geq 1$  das System nahe der Sättigungsgrenze und die Blockierungsrate wird mit zunehmendem Angebot größer. Im Warteraum befinden sich dann für eine Verkehrsklasse im Mittel

$$X_v^Q(\lambda_1, \lambda_2) \approx (K - 1) \cdot \frac{\lambda_v}{\lambda_1 + \lambda_2} \quad (\text{B-37})$$

Anforderungen. Die folgenden beiden Faktoren (siehe Gl. (B-38, B-39)) geben die *mittlere Belegung von Systemressourcen* einer Verkehrsklasse an:

- Der Faktor  $\xi_v^Q$  gibt an, welcher Anteil der *Warteplätze* durch eine Verkehrsklasse im Mittel belegt wird:

$$\xi_v^Q(\lambda_1, \lambda_2) = \frac{\lambda_v}{\lambda_1 + \lambda_2} \quad (\text{B-38})$$

- und Faktor  $\xi_v^P$  gibt an, welcher Anteil der *CPU-Zeit* durch eine Verkehrsklasse im Mittel in Anspruch genommen werden:

$$\xi_v^P(\lambda_1, \lambda_2) = \frac{A_v}{A_{ges}}. \quad (\text{B-39})$$

Weiterhin ist die *Blockierwahrscheinlichkeit*  $B$  für beide Verkehrsklassen gleich groß:

$$B(\lambda_1, \lambda_2) = 1 - \frac{1}{A_{ges}} = p_K \text{ für } A_{ges} \geq 1. \quad (\text{B-40})$$

Die *Wartezeit* für ein nahezu gesättigtes System lässt sich mit Gl. (B-37, B-38) durch folgenden Ausdruck approximieren:

$$w_{VS}(\lambda_1, \lambda_2) \approx (K - 1) \cdot \sum_v \xi_v^Q \cdot h_v = (K - 1) \cdot \frac{\lambda_1 \cdot h_1 + \lambda_2 \cdot h_2}{\lambda_1 + \lambda_2} \text{ für } A_{ges} \geq 1. \quad (\text{B-41})$$

Die beiden Gl. (B-36, B-41) bieten eine erste Abschätzung der mittleren Wartezeit für den gesamten Lastbereich. Da für  $A_{ges}(\lambda_1, \lambda_2) \rightarrow 1$  die Wartezeitfunktion unstetig wäre – es gilt

$$\lim_{A_{ges} \rightarrow 1} w_{WS}(A_{ges}) > w_{VS}(A_{ges} = 1) \quad (\text{B-42})$$

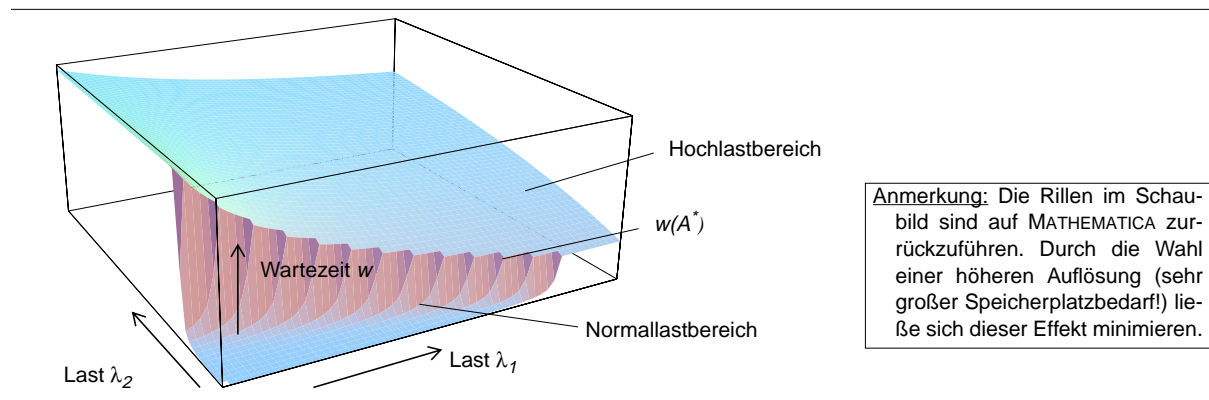
<sup>11</sup> Das vorliegende Modell repräsentiert einen Spezialfall des *M/G/1-Prioritätensystems mit Rückkopplungen*. Für *ungesättigte Systeme* diesen Typs wurden für die Berechnung von Zeitmittelwerten von PATEROK [422] detaillierte Untersuchungen durchgeführt. Seine Berechnungen stützen sich auf die *Momentenmethode*, die selbst wiederum auf dem *Little'schen Gesetz* [359], dem *PASTA-Theorem* [557] (*Poisson Arrivals See Time Averages*) und dem Ergebnis für die mittlere Vorwärts-Rekurrenzeit aus der Erneuerungstheorie basiert [42]. In der Arbeit von BODAMER [97] wird ein verallgemeinertes Verfahren für *gesättigte Systeme* präsentiert.

– wird als Trennlinie der beiden Arbeitsbereiche *Normal-* und *Hochlast*<sup>12</sup>, die Schnittlinie  $A^*(\lambda_1, \lambda_2)$  der beiden Flächen verwendet:

$$w_{WS}(A^*) = w_{VS}(A^*). \quad (\text{B-43})$$

$A^*(\lambda_1, \lambda_2)$  muss kleiner eins sein und repräsentiert einen elliptischen Verlauf in der Lastebene. Zusammengefasst wird für die nachfolgenden Berechnungen die mittlere Wartezeit durch folgenden Verlauf approximiert:

$$w(\lambda_1, \lambda_2) \approx \begin{cases} \frac{\lambda_1 \cdot h_1^2 + \lambda_2 \cdot h_2^2}{1 - (\lambda_1 \cdot h_1 + \lambda_2 \cdot h_2)} & \text{für } A_{ges} \leq A^* \\ (K - 1) \cdot \frac{\lambda_1 \cdot h_1 + \lambda_2 \cdot h_2}{\lambda_1 + \lambda_2} & \text{für } A_{ges} > A^* \end{cases}. \quad (\text{B-44})$$



**Bild B-14:** Qualitativer Verlauf der *Wartezeitfunktion*  $w(\lambda_1, \lambda_2)$   
( $h_2 = 4 \cdot h_1, K = 100$ )

Bild B-14 gibt exemplarisch den Verlauf dieser Funktion wieder. Wichtige, beobachtbare Leistungskenngrößen sind die *mittleren Antwortzeiten*  $\tau_v$ . Für die beiden Verkehrsklassen lassen sich diese mit Gl. (B-44) wie folgt berechnen:

$$\tau_v(\lambda_1, \lambda_2) = h_v + w(\lambda_1, \lambda_2) \quad (\text{B-45})$$

Damit ergeben sich für den gesamten Arbeitsbereich nach Gl. (B-44) und (B-45) für die einzelnen Verkehrsklassen folgende Funktionsverläufe:

$$\tau_v(\lambda_1, \lambda_2) = \begin{cases} h_v + \frac{\lambda_1 \cdot h_1^2 + \lambda_2 \cdot h_2^2}{1 - (\lambda_1 \cdot h_1 + \lambda_2 \cdot h_2)} & \text{für } A_{ges} \leq A^* \\ h_v + (K - 1) \cdot \frac{\lambda_1 \cdot h_1 + \lambda_2 \cdot h_2}{\lambda_1 + \lambda_2} & \text{für } A_{ges} > A^* \end{cases}. \quad (\text{B-46})$$

Mit der *mittleren Bedienzeit*  $h$

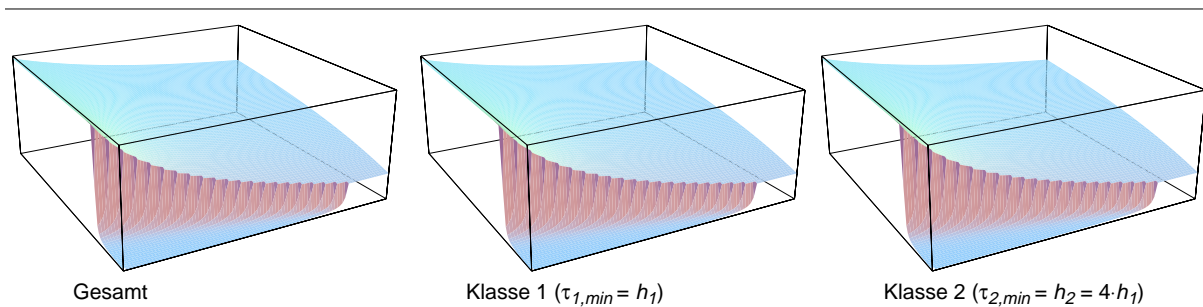
$$h(\lambda_1, \lambda_2) = \frac{\lambda_1 \cdot h_1 + \lambda_2 \cdot h_2}{\lambda_1 + \lambda_2} = \frac{1}{\phi_{ges, max}(\lambda_1, \lambda_2)} \quad (\text{B-47})$$

<sup>12</sup> Später wird in der *Hochlastfläche* ein weiterer Arbeitsbereich – genannt *Überlast* – definiert.

ergibt sich für die *mittlere Gesamtverzögerungszeit*  $\tau_{ges}(\lambda_1, \lambda_2)$  folgende Näherung:

$$\tau_{ges}(\lambda_1, \lambda_2) = \begin{cases} \frac{\lambda_1 \cdot h_1 + \lambda_2 \cdot h_2}{\lambda_1 + \lambda_2} + \frac{\lambda_1 \cdot h_1^2 + \lambda_2 \cdot h_2^2}{1 - (\lambda_1 \cdot h_1 + \lambda_2 \cdot h_2)} & \text{für } A_{ges} \leq A^* \\ K \cdot \frac{\lambda_1 \cdot h_1 + \lambda_2 \cdot h_2}{\lambda_1 + \lambda_2} & \text{für } A_{ges} > A^* \end{cases} \quad (\text{B-48})$$

In Bild B-15 ist für  $h_1 < h_2$  beispielhaft der Verlauf dieser Größen gezeichnet. Die vorliegende Parameterwahl von  $h_2 = 4 \cdot h_1$  und  $K = 100$  gilt auch für alle nachfolgenden Diagramme. Für diesen Fall ergibt sich als maximale Systemzeit  $\tau_{max} \approx K \cdot h_2$ .



**Bild B-15:** Mittlere Antwortzeiten (links:  $\tau_{ges}(\lambda_1, \lambda_2)$ , Mitte:  $\tau_1(\lambda_1, \lambda_2)$ , rechts:  $\tau_2(\lambda_1, \lambda_2)$ ); Beschriftung x- bzw. y-Achse s. Bild B-14.

Liegt keine Last an, d.h.  $A_{ges} = 0$ , dann sind die Durchsätze gleich null und demzufolge können auch keine Antwortzeiten gemessen werden. Für diesen Lastpunkt  $(\lambda_1, \lambda_2) = (0, 0)$  kann aber mit folgenden Grenzwerten gerechnet werden:

$$\tau_{v, min} = \lim_{A_{ges} \rightarrow 0} \tau_v(\lambda_1, \lambda_2). \quad (\text{B-49})$$

## Durchsätze

Der zweite wichtige Typus von Leistungskenngrößen sind die Nutzdurchsätze. Für eine Abschätzung der Durchsatzfunktionen soll wiederum eine Trennung in zwei Lastbereiche erfolgen. Hier erfolgt nun eine Aufteilung durch die Gerade  $\lambda_1 \cdot h_1 + \lambda_2 \cdot h_2 = 1$  in zwei Arbeitsbereiche. Ist  $A_{ges} \geq 1$ , wird hier von *Überlast* gesprochen. Für  $A_{ges}(\lambda_1, \lambda_2) < 1$  seien die Verluste vernachlässigt und für die *mittleren Durchsätze*  $\phi_v$  ergibt sich demzufolge:<sup>13</sup>

$$\phi_v \approx \lambda_v. \quad (\text{B-50})$$

Im Überlastbereich ergibt sich mit Gl. (B-40):

$$\phi_v = \lambda_v \cdot (1 - B) = \frac{\lambda_v}{A_{ges}}. \quad (\text{B-51})$$

Damit ergeben sich für den *gesamten Arbeitsbereich* nach Gl. (B-50) und (B-51) folgende *Durchsatzfunktionen*:

<sup>13</sup> Für  $A_{ges} < 1$  gehen aber bekanntlich auch Meldungen verloren da  $B = p_K = f(K, \lambda_1, \lambda_2, h_1, h_2)$  größer null ist.

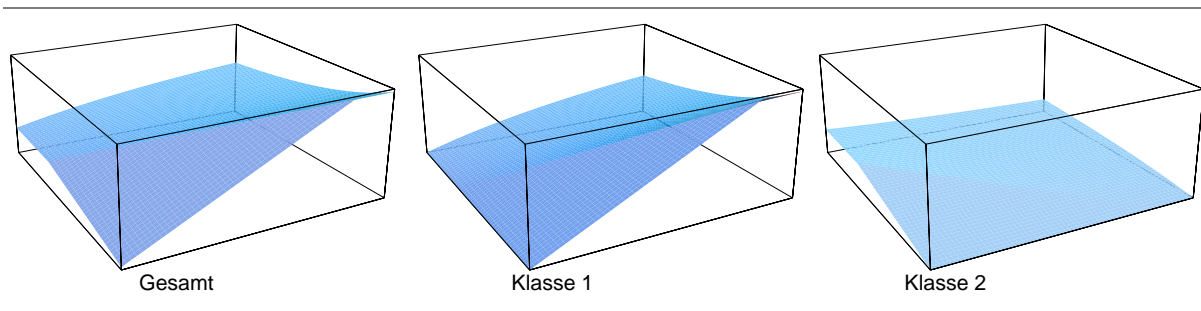


$$\phi_v(\lambda_1, \lambda_2) = \begin{cases} \lambda_v & \text{für } A_{ges} < 1 \\ \frac{\lambda_v}{\lambda_1 \cdot h_1 + \lambda_2 \cdot h_2} & \text{für } A_{ges} \geq 1 \end{cases} \quad (\text{B-52})$$

und

$$\phi_{ges}(\lambda_1, \lambda_2) = \phi_1 + \phi_2 = \begin{cases} \lambda_1 + \lambda_2 & \text{für } A_{ges} < 1 \\ \frac{\lambda_1 + \lambda_2}{\lambda_1 \cdot h_1 + \lambda_2 \cdot h_2} & \text{für } A_{ges} \geq 1 \end{cases} \quad (\text{B-53})$$

In Bild B-16 ist für  $h_1 < h_2$  beispielhaft der Verlauf der Durchsatzfunktionen gezeichnet. Für diesen Fall ergibt sich als maximaler Durchsatz  $\phi_{ges,max} = 1/h_1$ .



**Bild B-16:** Verlauf der *Durchsatzfunktionen* (li.:  $\phi_{ges} = \phi_1 + \phi_2$ , Mitte:  $\phi_1$ , re.:  $\phi_2$ ). Koordinatensystem und Wertebereiche: x-Achse  $\rightarrow \lambda_1 \in [0, 6]$ , y-Achse  $\rightarrow \lambda_2 \in [0, 2]$ , z-Achse  $\rightarrow \phi_i \in [0, 5.1]$ .

## Fehlerbetrachtung

Insgesamt wurde nun die Lastebene in die folgenden drei Flächen unterteilt:

- *Normallast* für  $0 < A_{ges}(\lambda_1, \lambda_2) \leq A^*(\lambda_1, \lambda_2)$
- *Hochlast* für  $A^*(\lambda_1, \lambda_2) < A_{ges}(\lambda_1, \lambda_2) \leq 1$
- *Überlast* für  $1 < A_{ges}(\lambda_1, \lambda_2) < \infty$

Für Normallast wird mit einem reinen Warte-, für Überlast mit einem Verlustsystem gerechnet. Da im Normal- und Hochlastbereich bei den Durchsatzfunktionen die Verluste nicht berücksichtigt wurden, sind die geschätzten Durchsatzwerte in diesem Arbeitsbereich etwas zu hoch. Dieselbe Betrachtung gilt auch für die Systemzeiten, die tatsächlich kleiner sein sollten als sie hier approximiert wurden. Für das im nachfolgenden betrachtete Leistungsmaß *Power* wirken sich diese Fehler unterschiedlich aus: der Durchsatz geht zu optimistisch, die Antwortzeit eher zu pessimistisch in die Rechnung ein. Auf der anderen Seite ist der sich einstellende Fehler im Hochlastbereich am größten. Für ein System mit sehr großem Speicher (d.h.  $K$  groß) geht  $A^*(\lambda_1, \lambda_2)$  gegen eins und der Hochlastbereich kann vernachlässigt werden. Da bei nachfolgenden Berechnungen nur die Extrema der *Power*-Funktionen, und weniger der exakte Funktionsverlauf<sup>14</sup>, im Mittelpunkt stehen, ist diese Annahme hier ausreichend und es werden nur noch Normal- und Überlastbereich betrachtet.

<sup>14</sup> Durch den hier erzielten Fehler sind die *Power*-Funktionen an den Grenzlinien zwischen den einzelnen Arbeitsbereichen nicht stetig. Dies ist jedoch hier von untergeordneter Bedeutung.

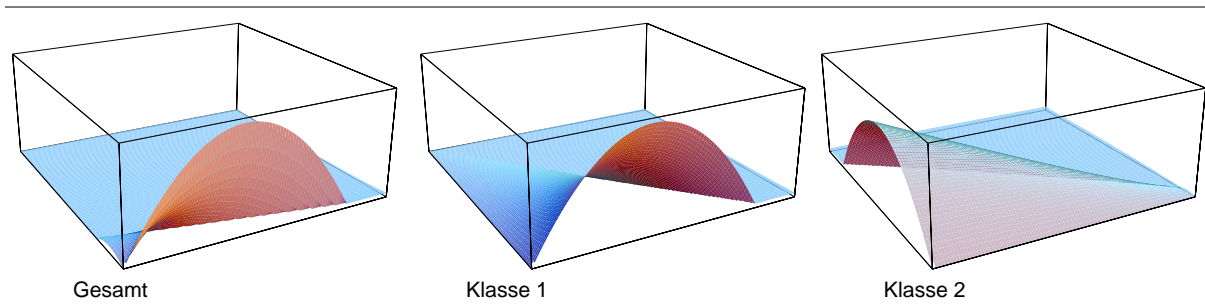
## Leistungsmaße vom Typ Power

Im folgenden werden nun die abgeleiteten Leistungsmaße genauer betrachtet. Für die Größe *Power* (siehe auch Gl. (3-9)) ergeben sich folgende Zusammenhänge:

$$\Lambda_{P,v}(\lambda_1, \lambda_2) = \frac{\phi_v(\lambda_1, \lambda_2)}{\tau_v(\lambda_1, \lambda_2)} \text{ bzw. } \Lambda_{P,ges}(\lambda_1, \lambda_2) = \frac{\phi_{ges}(\lambda_1, \lambda_2)}{\tau_{ges}(\lambda_1, \lambda_2)}. \quad (\text{B-54})$$

Mit Gl. (B-46, B-52, B-54) ergibt sich z.B. für *Verkehrsklasse 1*:

$$\Lambda_{P,1}(\lambda_1, \lambda_2) \approx \begin{cases} \lambda_1 \cdot \frac{1 - \lambda_1 h_1 - \lambda_2 h_2}{h_1 + (h_2 - h_1) \lambda_2 h_2} & \text{für } A_{ges} < 1 \\ \frac{\lambda_1 \cdot (\lambda_1 + \lambda_2)}{(\lambda_1 h_1 + \lambda_2 h_2) [(K-1) \lambda_2 h_2 + (K \lambda_1 + \lambda_2) h_1]} & \text{für } A_{ges} \geq 1 \end{cases}. \quad (\text{B-55})$$



**Bild B-17:** Leistungsmaß *Power* für  $h_2 = 4 \cdot h_1$  (links:  $\Lambda_{P,ges}(\lambda_1, \lambda_2)$ , Mitte:  $\Lambda_{P,1}(\lambda_1, \lambda_2)$ , rechts:  $\Lambda_{P,2}(\lambda_1, \lambda_2)$ ). Anmerkung: z-Achse hat unterschiedliche Maßstäbe in den einzelnen Schaubildern.

In Bild B-17 sind die sich ergebenden *Power*-Funktionen für den Normal- und Überlastbereich zusammengestellt. In den Diagrammen sind einzelne Maxima zu sehen. Diese liegen jeweils im Normallastbereich. Für  $\lambda_v = 0$  und  $\lambda_i \rightarrow \infty$  ( $i \neq v$ ) ist  $\Lambda_{P,v} = 0$  bzw. für  $\lambda_v \rightarrow \infty$  ist  $\Lambda_{P,v} = 1 / (K h_v^2)$ . Da Durchsatz und Antwortzeit nicht negativ sein können, muss *Power* ein *globales Maximum*  $\Lambda_{P,max,v}$  besitzen. Im folgenden werden die Extrema der einzelnen Lastbereiche gesucht. Die Berechnungen werden hier für Verkehrsklasse 1 durchgeführt.<sup>15</sup> Im Überlastbereich ist das System gesättigt und  $\Lambda_{P,1}$  erreicht sein Maximum für  $\lambda_2 = 0$ :

$$\Lambda_{P,max,1}^{ÜL}(\lambda_1, \lambda_2 = 0) = \frac{1}{K h_1^2}. \quad (\text{B-56})$$

Im Normallastbereich ist

$$\frac{\partial}{\partial \lambda_1} \Lambda_{P,1}(\lambda_1, \lambda_2) \approx \frac{1 - 2\lambda_1 h_1 - \lambda_2 h_2}{h_1 + (h_2 - h_1) \lambda_2 h_2} \stackrel{!}{=} 0 \text{ für } \lambda_1 = \frac{1 - \lambda_2 h_2}{2 h_1} \quad (\text{B-57})$$

mit dem Maximum

$$\Lambda_{P,max,1}^{NL}(\lambda_2) = \frac{(1 - \lambda_2 h_2)^2}{4 h_1 [h_1 + (h_2 - h_1) \lambda_2 h_2]}, \quad (\text{B-58})$$

<sup>15</sup> Das *Power*-Maß  $\Lambda_{P,ges}$  hat für  $h_1 \neq h_2$  mindestens noch ein lokales Maximum. Diese Kenngröße wird aber im Rahmen dieser Arbeit sowieso nicht weiter betrachtet, da ja gerade auf die spezifischen Einflüsse der einzelnen Verkehrsklassen reagiert werden soll.

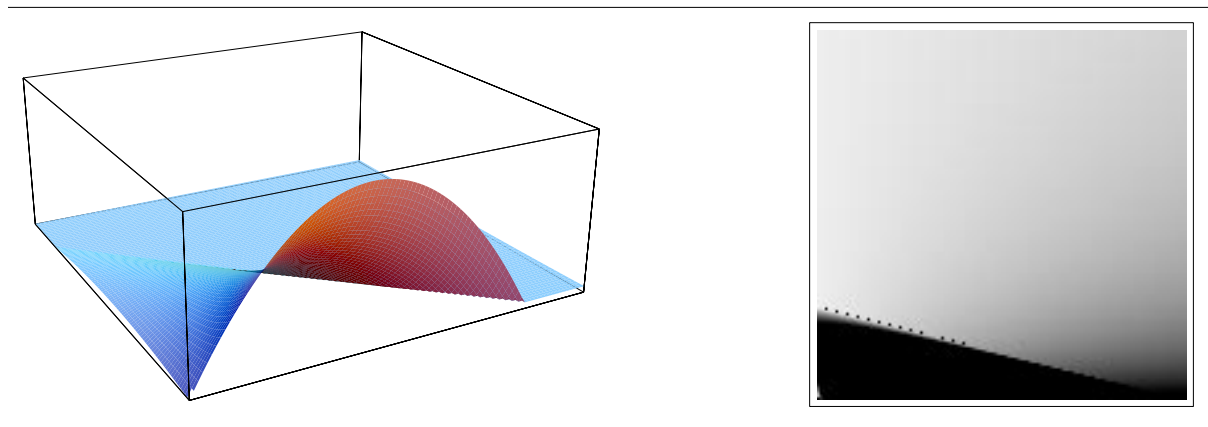
das auch für  $\lambda_2 = 0$  seinen Extremwert

$$\Lambda_{P,max,1}^{NL}(\lambda_1, \lambda_2 = 0) = \frac{1}{4h_1^2} \quad (\text{B-59})$$

annimmt. D.h. für  $K > 4$  liegt hier das globale Maximum von  $\Lambda_{P,v}$  im Normallastbereich. Aus den *Power*-Maßen der einzelnen Verkehrsklassen  $\Lambda_{P,v}$  können weitere Größen definiert werden. Interessant für Mehrklassensysteme sind insbesondere die beiden folgenden Kenngrößen. Die *Power-Summe* wird gebildet aus der Addition aller Einzel-*Power*-Maße:

$$\Lambda_{P,\Sigma} = \sum_v \Lambda_{P,v} = \Lambda_{P,1} + \Lambda_{P,2}. \quad (\text{B-60})$$

In Bild B-18 ist diese Funktion  $\Lambda_{P,\Sigma}(\lambda_1, \lambda_2)$  und ihr Dichtediagramm<sup>16</sup> abgebildet. Aufgrund vorliegender Parameterwahl ( $h_2 = 4 \cdot h_1$ ) dominiert *Power*  $\Lambda_{P,1}$  und das Maximum von  $\Lambda_{P,2}$  tritt hier kaum in Erscheinung.



**Bild B-18:** Leistungsmaß *Power-Summe* (li.:  $\Lambda_{P,\Sigma}(\lambda_1, \lambda_2)$ , re.: Dichtediagramm). Koordinatensystem und Wertebereiche: x,y-Ebene  $\rightarrow \lambda_1 \in [0, 6], \lambda_2 \in [0, 1.5]$ ; z-Achse  $\rightarrow \Lambda_{P,\Sigma} \in [0, 6.5]$

Durch zusätzliche Faktoren können die Einzel-*Power*-Maße unterschiedlich gewichtet werden (siehe auch Gl. (3-12)). Durch die lokalen Maxima ist diese Funktion jedoch weniger als Optimierungskriterium für eine Leistungsregelung geeignet. Diese Anforderung wird aber vom sogenannten *Power-Produkt* (siehe Gl. (3-11)) erfüllt, das aus der Multiplikation aller Einzel-*Power*-Maße gebildet wird:

$$\Lambda_{P^{\Pi}} = \prod_v \Lambda_{P,v} = \Lambda_{P,1} \cdot \Lambda_{P,2} = \frac{\phi_1 \cdot \phi_2}{\tau_1 \cdot \tau_2}. \quad (\text{B-61})$$

Für die beiden Lastbereiche ergeben sich hier folgende Funktionen:

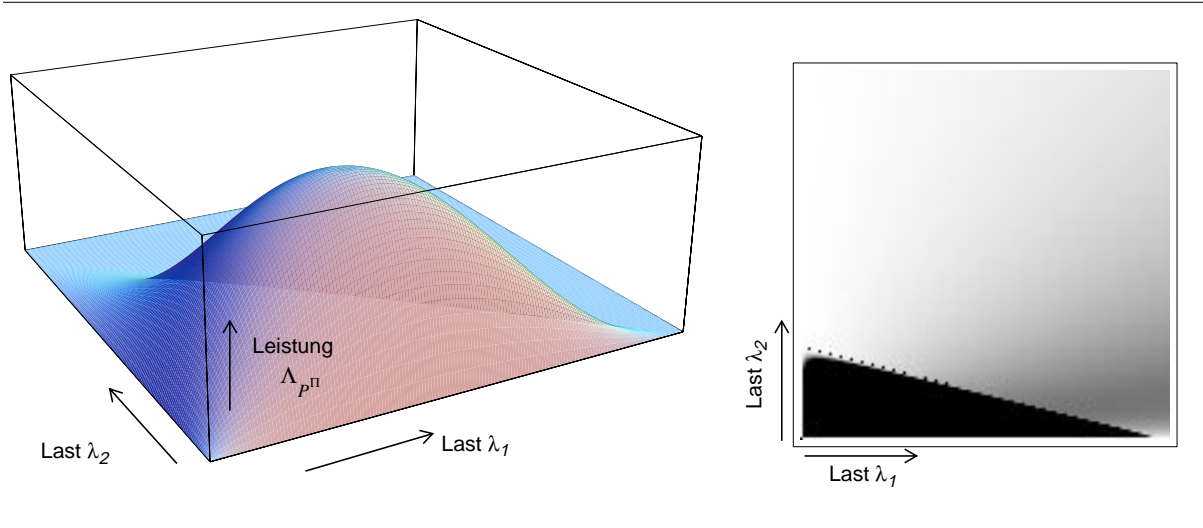
$$\Lambda_{P^{\Pi}}^{NL}(\lambda_1, \lambda_2) = \frac{\lambda_1 \lambda_2 \cdot (1 - \lambda_1 h_1 - \lambda_2 h_2)^2}{[h_2 + (h_1 - h_2) \lambda_1 h_1] \cdot [h_1 + (h_2 - h_1) \lambda_2 h_2]} \quad (\text{B-62})$$

<sup>16</sup> Darstellung eines dreidimensionalen Funktionsverlaufs als zweidimensionales Schaubild, in dem die Funktionsgröße als Dichte interpretiert wird, dessen Größe beispielsweise proportional zum Grauwert ist.

für den *Normallastbereich* ( $A_{ges} < 1$ ) und

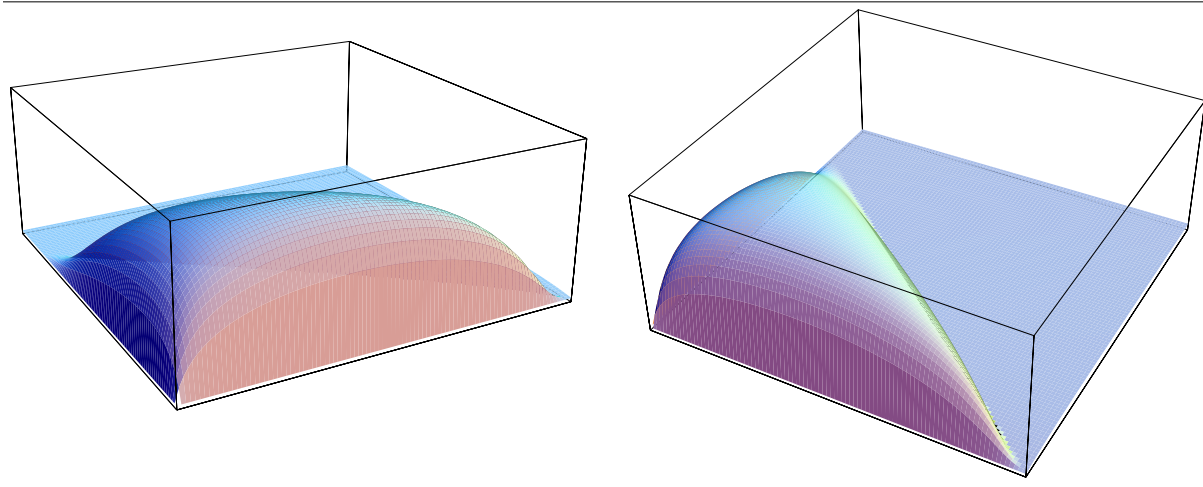
$$\Lambda_{P\Pi}^{ÜL}(\lambda_1, \lambda_2) = \frac{\lambda_1 \lambda_2 \cdot (\lambda_1 + \lambda_2)^2}{A_{ges}^2 [(K-1)\lambda_1 h_1 + h_2(K\lambda_2 + \lambda_1)][(K-1)\lambda_2 h_2 + h_1(K\lambda_1 + \lambda_2)]} \quad (\text{B-63})$$

für den *Überlastbereich* ( $A_{ges} \geq 1$ ). In Bild B-19 ist diese Funktion  $\Lambda_{P\Pi}$  und ihr Dichtediagramm abgebildet.



**Bild B-19:** Leistungsmaß *Power-Produkt* für  $h_2 = 4 \cdot h_1$  (li.:  $\Lambda_{P\Pi}(\lambda_1, \lambda_2)$ , re.: Dichtediagramm). Koordinatensystem und Wertebereiche: x,y-Ebene  $\rightarrow \lambda_1 \in [0, 5], \lambda_2 \in [0, 1.3]$ ; z-Achse  $\rightarrow \Lambda_{P\Pi}$ . Dichtediagramm unverzerrt in x-y-Richtung.

Bei sehr vielen Verkehrsklassen kann durch die multiplikative Bildung des *Power-Produktes*  $\Lambda_{P\Pi}$  dessen Wertebereich sehr viele Größenordnungen umfassen. Hier kann es dann ratsam sein, das *logarithmische Power-Produkt* zu verwenden. In Bild B-20 ist dieses Leistungsmaß für das vorliegende Zwei-Klassensystem dargestellt.



**Bild B-20:** Logarithmische Darstellung von *Power-Produkt*, d. h.  $\log_{10}\Lambda_{P\Pi}$  (dargestellt sind zwei verschiedene Perspektiven).

## Globales Maximum

Aus dem Dichtediagramm wird deutlich, dass bei vorliegender Parameterwahl die Leistungswerte im Normallastbereich am größten sind. Nach Gl. (B-62) kann hier  $\Lambda_{P,\Pi}$  in der Nähe seines Maximums (für  $h_1 \neq h_2$ ) durch einen elliptischen Paraboloiden approximiert werden. D.h., „wandert“ der Lastvektor in Richtung Maximum, dann wächst die Leistung quadratisch an. Umgekehrt besitzt  $\Lambda_{P,\Pi}$  im *Überlastbereich* ein hyperbolisches Verhalten (siehe Gl. (B-63)). Die Leistung fällt deshalb in diesem Bereich umgekehrt proportional zur Last.

Gesucht ist nun eine Abschätzung des Maximalwertes. Für  $K = 1$  wird  $\Lambda_{P,\Pi}$  unter Überlast maximal (siehe Gl. (B-63)) und der Wert beträgt

$$\Lambda_{P^{\Pi},max}^{\ddot{U}L}(\lambda_1, \lambda_2, K = 1) = \frac{\lambda_1 \lambda_2}{h_1 h_2 \cdot (\lambda_1 h_1 + \lambda_2 h_2)^2}. \quad (\text{B-64})$$

Da hier  $A_{ges} \geq 1$  sein muss, liegt das Maximum von Gl. (B-64) bei  $A_{ges} = 1$  und beträgt:

$$\Lambda_{P^{\Pi},max}^{\ddot{U}L}\left(\lambda_1, \lambda_2 = \frac{1 - \lambda_1 h_1}{h_2}, K = 1\right) = \frac{\lambda_1 (1 - \lambda_1 h_1)}{h_1 h_2^2}. \quad (\text{B-65})$$

Damit ergibt sich für den Überlastbereich folgender Maximalwert:

$$\Lambda_{P^{\Pi},max}^{\ddot{U}L}\left(\lambda_1 = \frac{1}{2h_1}, \lambda_2 = \frac{1}{2h_2}, K = 1\right) = \frac{1}{4h_1^2 h_2^2}. \quad (\text{B-66})$$

Das reine Verlustsystem ( $K = 1$ ) ist im Rahmen dieser Arbeit weniger interessant, so dass hier das Extremum als Funktion von  $K$  gesucht ist. Aus Gl. (B-63) ergibt sich für große  $K$  folgende Abschätzung:

$$\tilde{\Lambda}_{P^{\Pi}}^{\ddot{U}L}(\lambda_1, \lambda_2) = \frac{\lambda_1 \lambda_2 \cdot (\lambda_1 + \lambda_2)^2}{K^2 \cdot (\lambda_1 h_1 + \lambda_2 h_2)^4} < \Lambda_{P^{\Pi}}^{\ddot{U}L}(\lambda_1, \lambda_2) \quad (\text{B-67})$$

Wiederum wird diese hyperbolische Funktion maximal für  $A_{ges} = 1$ :

$$\tilde{\Lambda}_{P^{\Pi},max}^{\ddot{U}L}\left(\lambda_1, \lambda_2 = \frac{1 - \lambda_1 h_1}{h_2}, K\right) = \frac{\lambda_1 (1 - \lambda_1 h_1) [1 + (h_2 - h_1) \lambda_1]^2}{K^2 h_2^3} \quad (\text{B-68})$$

mit dem Maximalwert

$$\tilde{\Lambda}_{P^{\Pi},max}^{\ddot{U}L}\left(\lambda_1 = \frac{2}{5h_1 - 3h_2 + \sqrt{9h_1^2 - 14h_1 h_2 + 9h_2^2}}, \lambda_2 = \frac{1 - \lambda_1 h_1}{h_2}, K\right) = \frac{2(3h_1 - 3h_2 + \sqrt{9h_1^2 - 14h_1 h_2 + 9h_2^2})(3h_1 - h_2 + \sqrt{9h_1^2 - 14h_1 h_2 + 9h_2^2})^2}{K^2 h_2^3 \cdot (5h_1 - 3h_2 + \sqrt{9h_1^2 - 14h_1 h_2 + 9h_2^2})^4} \quad (\text{B-69})$$

Aufgrund des paraboloiden Verlaufs von  $\Lambda_{P,\Pi}$  im *Normallastbereich* muss der Gradient im Maximum gleich null sein:<sup>17</sup>

<sup>17</sup> Als Lösung für die Bedingung, dass eine partielle Ableitung null sein muss, ergeben sich hier jeweils Parabelfunktionen. Der Gradient ist dann null, wenn sich beide Parabeln schneiden. Als Lösung ergeben sich hier vier Schnittpunkte, wobei sowohl für  $h_1 > h_2$  als auch für  $h_1 < h_2$  immer nur ein Schnittpunkt im ersten Quadranten und damit im Arbeitsbereich liegt.

$$\nabla \Lambda_{P^{\Pi}, max}^{NL}(\lambda_1, \lambda_2) = \begin{pmatrix} \frac{\partial \Lambda_{P^{\Pi}}^{NL}}{\partial \lambda_1} \\ \frac{\partial \Lambda_{P^{\Pi}}^{NL}}{\partial \lambda_2} \end{pmatrix} = 0 \text{ für } \begin{pmatrix} \tilde{\lambda}_1 \\ \tilde{\lambda}_2 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{h_2}}{2h_1(\sqrt{h_1} + \sqrt{h_2})} \\ \frac{\sqrt{h_1}}{2h_2(\sqrt{h_1} + \sqrt{h_2})} \end{pmatrix}. \quad (\text{B-70})$$

Damit ergibt sich für den Normallastbereich folgender Maximalwert:

$$\Lambda_{P^{\Pi}, max}^{NL}(\tilde{\lambda}_1, \tilde{\lambda}_2) = \frac{1}{4h_1h_2(\sqrt{h_1} + \sqrt{h_2})^4}. \quad (\text{B-71})$$

Soll das Maximum im Normallastbereich liegen, muss nach Gl. (B-69, B-71) die Ungleichung

$$\Lambda_{P^{\Pi}, max}^{NL}(h_1, h_2) > \tilde{\Lambda}_{P^{\Pi}, max}^{\ddot{U}L}(h_1, h_2, K) \quad (\text{B-72})$$

erfüllt sein. Dies trifft für große Warteschlangenlängen zu, wobei eine Abschätzung für den Grenzwert  $K^* = f(h_1, h_2)$  aus den Gl. (B-69, B-71) berechnet werden kann.<sup>18</sup>

## Zusammenfassung

Es wurden für die Leistungsregelung wichtige Kenngrößen für das  $M_1M_2/M_1M_2/1/K$ -Modell abgeschätzt und diskutiert. Weiterhin wurde plausibel argumentiert, dass das *Power-Produkt*  $\Lambda_{P,\Pi}$  ein *globales Maximum* besitzt. Ab einer gewissen Warteschlangenlänge  $K^*$  liegt dieses im Normallastbereich. Die diskutierten Betrachtungen lassen sich prinzipiell auch auf Systeme mit mehr als zwei Verkehrsklassen erweitern. Ein strenger Beweis im mathematischen Sinne ist die hier dargestellte Vorgehensweise nicht.

Für den Entwurf und die Dimensionierung einer konkreten Leistungsregelung ist jedoch, aufgrund des vorhandenen Wissens über das System, dieser Beweis nicht unbedingt erforderlich. Durch die nun bekannten Systemparameter (z.B. mittlere Bearbeitungszeiten, Puffergrößen) lässt sich dann mittels obiger Vorgehensweise das Maximum lokalisieren. Für eine Abschätzung sind deshalb die hergeleiteten Resultate ausreichend.

<sup>18</sup> Beispielsweise für  $h_2 = 4 \cdot h_1$  und  $h_1 = 0.2$  beträgt  $K^* = 7$  Pufferplätze.

# Anhang C

## Dienstszzenarien und Simulationsmodelle

Die Feinheit der Modellierung für die durchgeführten Untersuchungen übertrifft hier den sonst üblichen Detaillierungsgrad für derartige Systemanalysen. Die den Messungen zugrunde liegenden Simulationsmodelle und Verkehrsabläufe sollen deshalb ausführlich vorgestellt werden.

### C.1 Verkehrsmodelle

Als Modelle für einen zentralen Vermittlungsrechner wurden zwei Mehrprozessorsysteme entwickelt und implementiert. Beide Modelle basieren auf der in Abschnitt 5.4.4.2 beschriebenen Elementarzelle eines generischen Prozessorknotens. Dabei wurde das Simulationswerkzeug so flexibel gehalten, dass dynamisch beliebig aufwendige Steuerungsstrukturen erzeugt werden können. Dies betrifft auch die interne Konfiguration nahezu aller Eigenschaften eines Prozessorknotens (z. B. Prioritäten, Abfertigungsdisziplinen, Bedienzeiten).

#### C.1.1 Modell für einen Dualprozessor

Ein Dualprozessor als zentrale Vermittlungssteuerung besitzt zwei Funktionsebenen (Bild C-1). Die Steuerungseinheit  $PN_A$  ist hauptsächlich für die Signalisierung zuständig.  $PN_B$  bearbeitet alle höheren Funktionen (siehe Abschnitt 2.6.5). Prinzipiell ist jeder Prozessorknoten mit einer lokalen Last-/Leistungsregelung ausgestattet. Beiden Regelkreisen kann eine globale Regelung überlagert werden, wodurch eine hierarchische Regelstruktur entsteht. Der Prozess dieser zentralen Regelung kann auf einem der beiden Prozessoren lokalisiert werden. Betrachtet man den Ressourcenbedarf der beiden Funktionsebenen, dann wird diese Steuerungsfunktion sinnvollerweise im  $PN_A$  installiert.

Dieses Verkehrsmodell bildet die Obermenge des für die Untersuchung der *NRAC*- und *PPC*-Algorithmen (siehe Abschnitte 4.4 bzw. 4.5) verwendeten Simulationsmodells. Je nach Verfahren werden im Kontrollpfad eines Simulationsmodells bestimmte Komponenten aktiviert bzw. deaktiviert.

Der Weg einer Signalisiernachricht einer bestimmten Verkehrsklasse durch die Vermittlungssteuerung lässt sich am Meldungsszenario verfolgen. Für die Dualprozessorstruktur ist dieses in Abschnitt C.3.3.1 erläutert.

Ein Kontrollpfad besteht aus der Sensorik, Aktorik und einer Regel- oder Übertragungsfunktion. Als Sensorelemente werden beispielsweise die Lastindikatoren *CPU-Auslastung*  $\rho$  (gesamt sowie einzelne Bedienphasen), *Pufferbelegung*  $\Omega$  und *Ankunftsrate*  $\lambda$  bestimmter Meldungstypen sowie die direkten Leistungskenngrößen *Nutzdurchsatz*  $\phi_{out}$  und *Aufenthaltszeiten*  $\tau$  den Regelkreisen zur Verfügung gestellt. Die Aktorik beinhaltet die Lastvertagung sowie Möglichkeiten der Zurückweisung oder Verwerfung von bestimmten Meldungen (hier nur  $PN_B$ ).

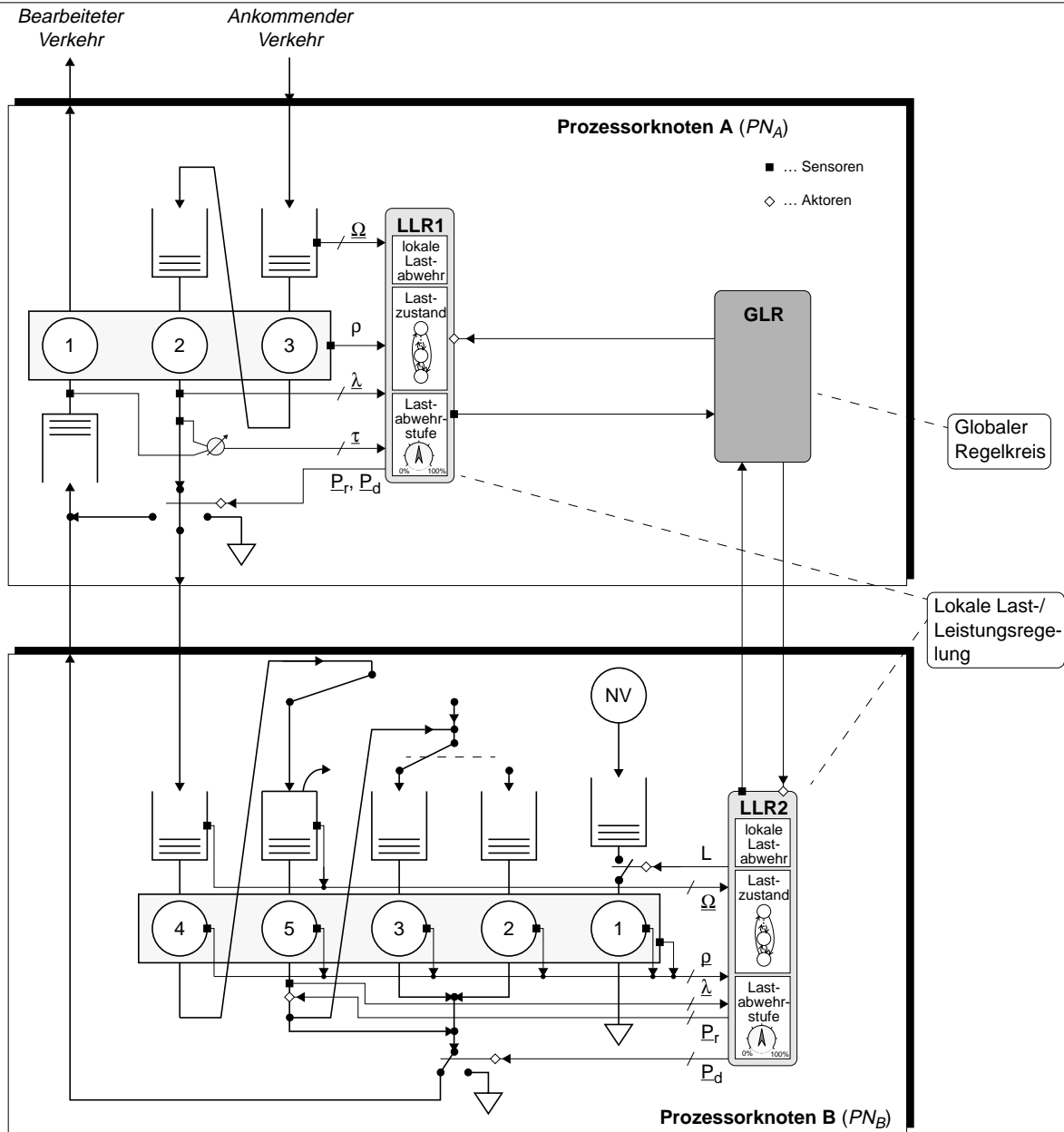


Bild C-1: Vermittlungssteuerung als Dualprozessorstruktur

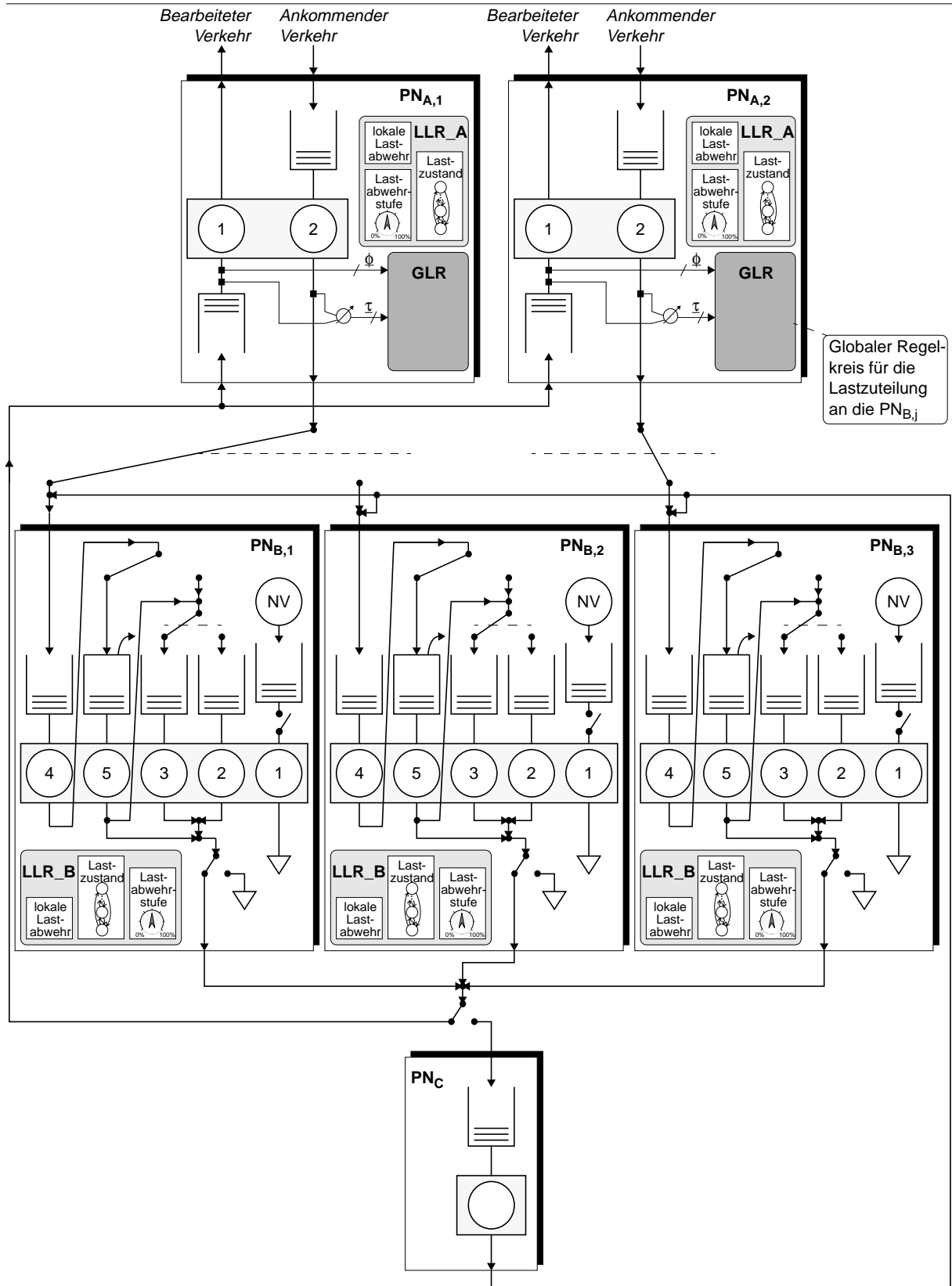
## C.1.2 Multiprozessor mit Last- und Funktionsteilung

Für die Untersuchung der Lastzuteilungsverfahren wurde das Dualprozessormodell um eine zusätzliche Funktionsebene sowie um parallele Prozessorknoten auf A- und B-Ebene erweitert. Das konkrete Modell mit zwei  $PN_{A,i}$  und drei  $PN_{B,j}$  wurde für die Messungen verwendet (Bild C-2). Bis auf  $PN_C$  ist jeder Knoten grundsätzlich mit einer lokalen Lastregelung ausgestattet.<sup>1</sup>

Wie in Abschnitt 4.6 beschrieben wurde, befinden sich die Instanzen zur Lastzuteilung in jedem  $PN_{A,i}$ . Hierdurch entsteht ein vollständig verteilter Lastverteilmechanismus, der hier der globalen Regelung entspricht.

<sup>1</sup> Bei den Untersuchungen zum DLB-Algorithmus sind die lokalen Lastregelkreise der  $PN_{A,i}$  deaktiviert.





**Bild C-2:** Modell für verteiltes Lastzuteilungsverfahren

Die internen Verkehrsabläufe sind durch die Szenariodiagramme in Abschnitt C.3.3.2 spezifiziert. Prinzipiell durchläuft jede ankommende Nachricht mehrere Bedienphasen, bevor sie die Vermittlungssteuerung wieder verlässt.

## C.2 Simulationsmodelle

Als Werkzeug wurde eine in C++ geschriebene objektorientierte Simulationsbibliothek eingesetzt [315]. Die im vorherigen Abschnitt beschriebenen Warteschlangenmodelle müssen hierbei aus entsprechenden Bibliothekskomponenten aufgebaut werden. Da die Bibliothek in ihrer Grundversion vornehmlich nur Klassen für Datenpfadelemente (z.B. Phase, FIFO-Puffer) besitzt, musste eine Vielzahl von Komponenten für die Implementierung der Kontrollpfade, d.h. Sensoren, Aktoren, Regelkreiselemente usw., implementiert werden. Diese sind in einer separaten Bibliothek zusammengefasst worden [63, 174, 418].

Um die Eigenschaften von Verkehrsklassen, Signalisierabläufen, Prioritäten, Rufwiederholungen u.a. simulieren zu können, war es notwendig, die Simulationsnachrichten um zusätzliche Felder zu erweitern. Auf Basis dieser Parameter erfolgt das Routing einer Nachricht innerhalb einer Prozessoreinheit – und auch zwischen den *PN* – durch Multiplexer-/Demultiplexerkomponenten. Diese Informationen werden auch für die Messung spezifischer Kenngrößen verwendet.

## C.3 Signalisierszenarien

### C.3.1 Anforderungen für die Modellierung

Von Beginn bis Ende einer Kommunikationsbeziehung (z.B. eines erfolgreichen Rufes) wird Steuerinformation zwischen Endgeräten und Netzelementen ausgetauscht. Der zeitliche Ablauf für einen bestimmten Dienst wird häufig mit einem Signalisierszenario dokumentiert. Aufgrund der Zustandsräume von Signalisierprotokollautomaten kann solch ein Ablaufdiagramm nur ein mögliches Szenario widerspiegeln.<sup>2</sup>

Die an einer Vermittlungssteuerung ankommenden Signalisiernachrichten bilden den zu bearbeitenden Verkehr und damit die Steuerungslast. Ein realer Signalisierablauf wird für eine Simulation auf ein entsprechendes Meldungsszenario abgebildet. Aufgrund der Vielfalt ist eine Abstraktion unerlässlich. Um diesen Schritt zu verdeutlichen, werden in Abschnitt C.3.2.2 ein reales Szenario für einen komplexeren Dienst, und in Abschnitt C.3.3 exemplarische Szenarien aufgezeigt. Sie bilden die Grundlage für die Simulationen in dieser Arbeit.

Weitere Aspekte (z.B. Rufwiederholungen) und Randbedingungen wurden bereits in den Abschnitten 5.2.5 und 5.2.6 diskutiert.

### C.3.2 Signalisierabläufe für reale Dienste

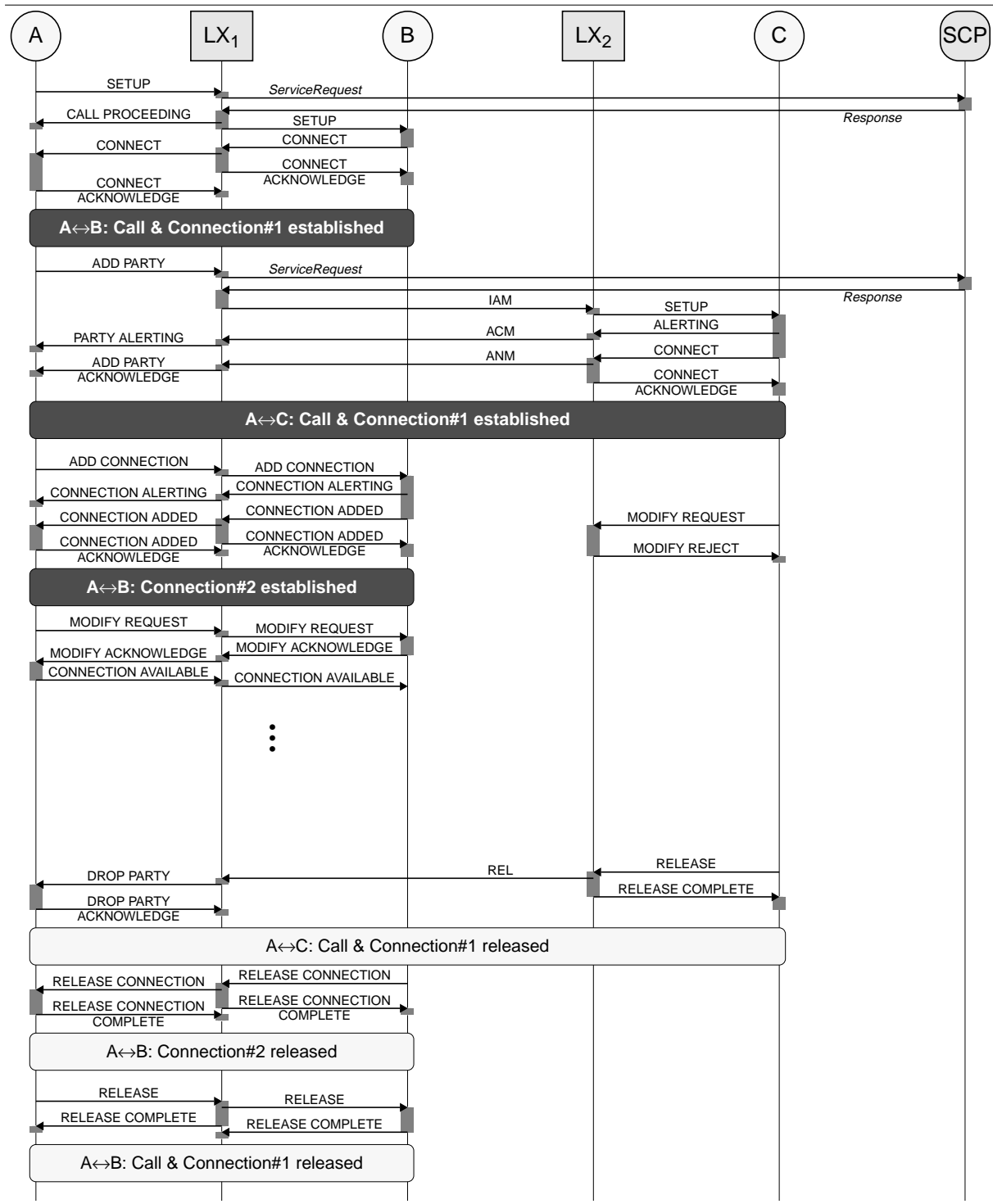
#### C.3.2.1 Basisruf

Der klassische Basisruf entspricht einem Ruf mit einer Punkt-zu-Punkt-Verbindung. Für einen abgehenden Ruf an einer Teilnehmervermittlungsstelle, d.h. UNI und NNI sind involviert, sind

---

<sup>2</sup> Erfolgreiche Dienstbearbeitungen lassen sich oft durch wenige Szenarien dokumentieren. Die Vielzahl resultiert vielmehr aus der Abdeckung aller potenzieller Fehlerfälle (z.B. mehrere hundert für Q.2931).

mindestens vier Signalisier Nachrichten (SETUP, CONNECT, CONNECT ACKNOWLEDGE und RELEASE) für eine erfolgreiche Rufkompletierung zu bearbeiten. In der Regel ist aber mit mehr Nachrichten zu rechnen. Ein typisches Szenario für den Basisruf ist beispielsweise in [54] zu finden.



**Bild C-3:** IN-Dienst „Videokonferenz mit drei Teilnehmern“ – Ausschnitt aus einem vereinfachten Signalisierszenario. Zwischen den verschiedenen Teilnehmern werden zusätzliche Verbindungen auf- und wieder abgebaut bzw. Verbindungsparameter temporär modifiziert.

### C.3.2.2 IN-Dienst

Eine Videokonferenz ist ein Beispiel für einen komplexeren Dienst. Bild C-3 zeigt den Ausschnitt aus einem entsprechenden Signalisierdiagramm, wobei folgende Randbedingungen angenommen wurden:

Die Konferenz besteht aus drei Teilnehmern, die an zwei Netzknoten angeschlossen sind. Durch einen zusätzlichen IN-Dienst sind alle Endgeräteadressen Teile eines virtuellen privaten Netzes (VPN). Für eine Adressauflösung ist somit eine SCP-Anfrage notwendig. Von jedem Teilnehmer besteht jeweils eine Punkt-zu-Mehrpunkt-Rufbeziehung zu den anderen Konferenzteilnehmern. Die zugehörigen Verbindungen sind unidirektional. Weitere Punkt-zu-Punkt-Verbindungen zwischen zwei Partnern sollen temporär auf- und nach einer gewissen Zeit wieder abgebaut werden (*Multiconnection Call*).

In Bild C-3 sind die wichtigsten Signalisierabschnitte aufgeführt. Weiterhin sind erfolgreiche und nicht erfolgreiche Anfragen nach Parameteränderungen einer Verbindung dargestellt. Betrachtet man beispielsweise Netzknoten  $LX_j$ , dann kann man sich leicht vorstellen, dass für größere Konferenzen während der gesamten Dienstbearbeitung mehr als hundert zu bearbeitende Signalisiernachrichten ankommen können.

## C.3.3 Exemplarische Szenarien für die Leistungsuntersuchung

Die nachfolgend vorgestellten Szenarien sind nicht direkt Abbilder realer Signalisierabläufe bestimmter Dienste. Aus Sicht des Vermittlungsrechners spielt die räumliche Trennung in einzelne Teilnehmer keine Rolle. Alle Verkehrsquellen und -senken sind deshalb in einem einzigen Generatormodell zusammengefasst.<sup>3</sup> Das Simulationsmodell entspricht damit, auf einer höheren Abstraktionsebene, einer *Client-Server*-Struktur. Eine Quelle/Senke repräsentiert dabei einen Client und initiiert einen Dialog durch eine Anfrage an den Vermittlungsrechner als Server. Dieses generische Quellmodell kann deshalb auch für andere Anwendungen verwendet werden, beispielsweise für die Modellierung von Signalisieraktivitäten im IN zwischen SSPs und einem SCP.

Nachfolgend werden die Signalisierszenarien und internen Meldungsketten von Verkehrsklasse 1 für die beiden Systemmodelle erläutert.

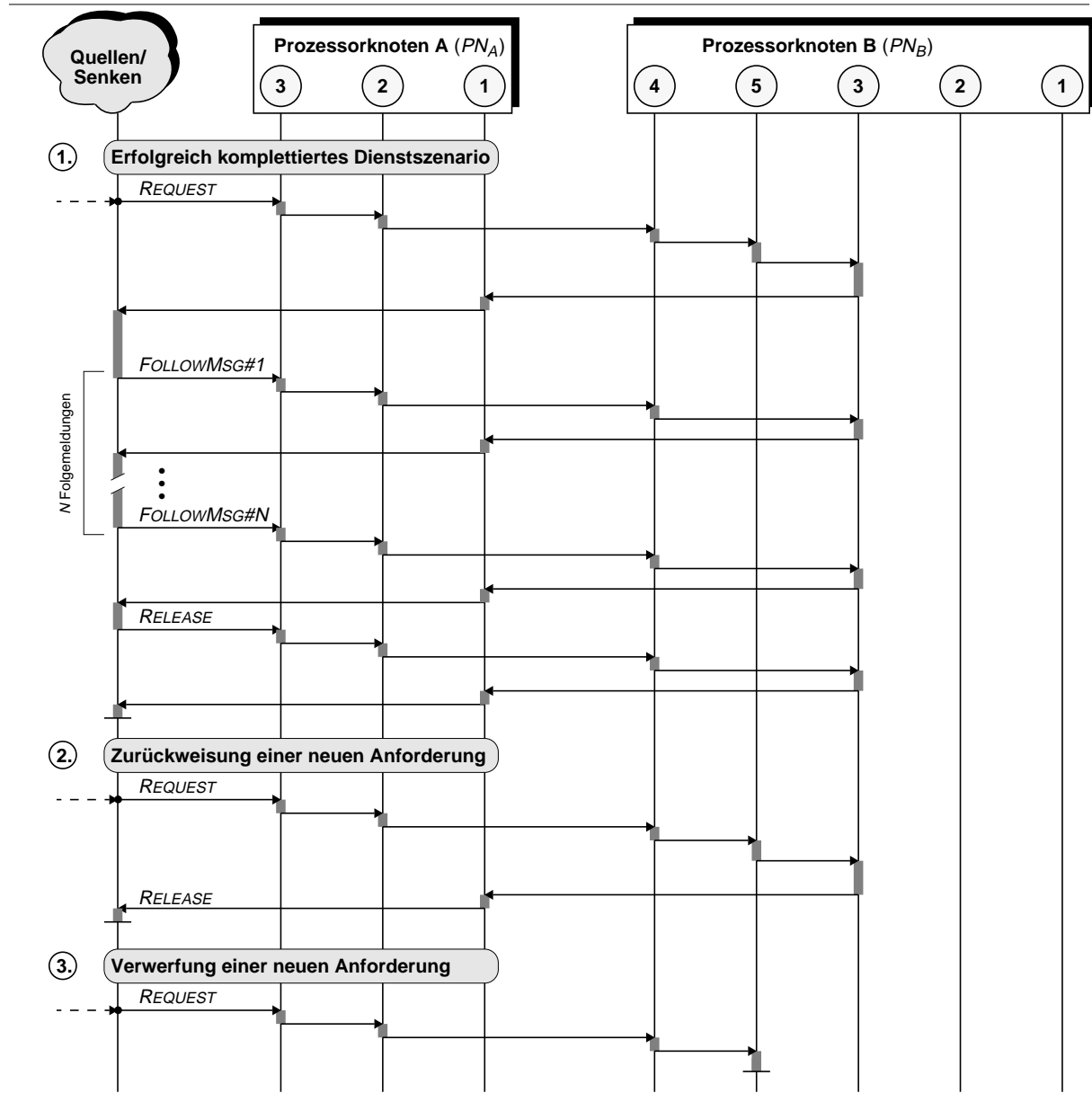
### C.3.3.1 Dualprozessorstruktur

Jedes Szenario beginnt mit einer Anforderung für einen bestimmten Dienst. Wie bereits erwähnt wurde, sind hier verschiedene Dienste zu Gruppen oder Verkehrsklassen zusammengefasst. Die Verkehrsklasseninformation ist in einem bestimmten Feld der REQUEST-Meldung codiert.

Beim Dualprozessormodell durchläuft jede ankommende Nachricht zunächst  $PN_A$ , bevor sie in  $PN_B$  von Phase 2 analysiert wird (Bild C-4). Der weitere interne Ablauf ist in Abschnitt 5.4.4.2 beschrieben. Die Dienstbearbeitung für diese Verkehrsklasse findet in Phase 3 statt. Erfolgreich bearbeitete Meldungen gehen via  $PN_A$  wieder an den Teilnehmer zurück. Nach einer gewissen Verzögerungszeit (Typ und Parameter der Verteilungsfunktion kann frei vorgegeben werden) wird eine Folgemeldung (FOLLOWMESSAGE) an den Vermittlungsrechner gesendet.

<sup>3</sup> Innerhalb des Generators existiert für jede Verkehrsklasse ein separates Teilnehmermodell.

Die Anzahl der Folgemeldungen ist für Verkehrsklasse 1 eine Zufallszahl, die einer diskreten Gleichverteilung gehorcht. Die variable Anzahl an Folgemeldungen steht auch für die Zusammenfassung mehrerer Dienste in eine Verkehrsklasse. Abgeschlossen wird ein Dienstszenario durch eine RELEASE-Meldung. Damit werden für Verkehrsklasse 1 im Mittel sieben Meldungen innerhalb eines Dialoges für eine erfolgreiche Komplettierung ausgetauscht.



**Bild C-4:** Die drei Szenarien für Verkehrsklasse 1. Die grauen Balken entsprechen den Bedienzeiten in den Prozessorknoten bzw. Antwortzeiten beim Teilnehmer.

Weiterhin sind in Bild C-4 die Szenarien für die Zurückweisung und die Verwerfung einer neuen Anforderung dargestellt. Dabei soll angedeutet werden, dass beide Überlastabwehrstrategien einen unterschiedlichen Blindlastanteil verursachen. Die Szenarien für *Repeated Call Attempts* sind nicht eingezeichnet (Reaktion auf Ablehnung einer Anforderung oder Ablauf eines Protokolltimers).

Die Signalisierszenarien für die einzelnen Verkehrsklassen unterscheiden sich in der Anzahl von Folgemeldungen, in dem Antwortzeitverhalten des Teilnehmers und damit der Dauer einer Kommunikationsbeziehung sowie dem Ressourcenbedarf an CPU-Zeit in  $PN_B$ . Die größten Unterschiede treten dabei in den Phasen drei bzw. vier auf. Auch kostet die Zurückweisung weniger an Ressourcen als die Annahme einer Anforderung.

### C.3.3.2 (2, 3, 1)-Steuerungsstruktur

Die Signalisierszenarien der einzelnen Verkehrsklassen bleiben für dieses Multiprozessormodell dieselben. Was sich geringfügig ändert sind die Meldungsketten innerhalb des Vermittlungsrechners. Wie Bild C-2 entnommen werden kann, besitzt ein  $PN_{A,i}$  hier zwei Bedienphasen. Weiterhin wird für jede neue und akzeptierte Dienstanforderung eine Anfrage an  $PN_C$  gesendet (siehe Bild C-5).

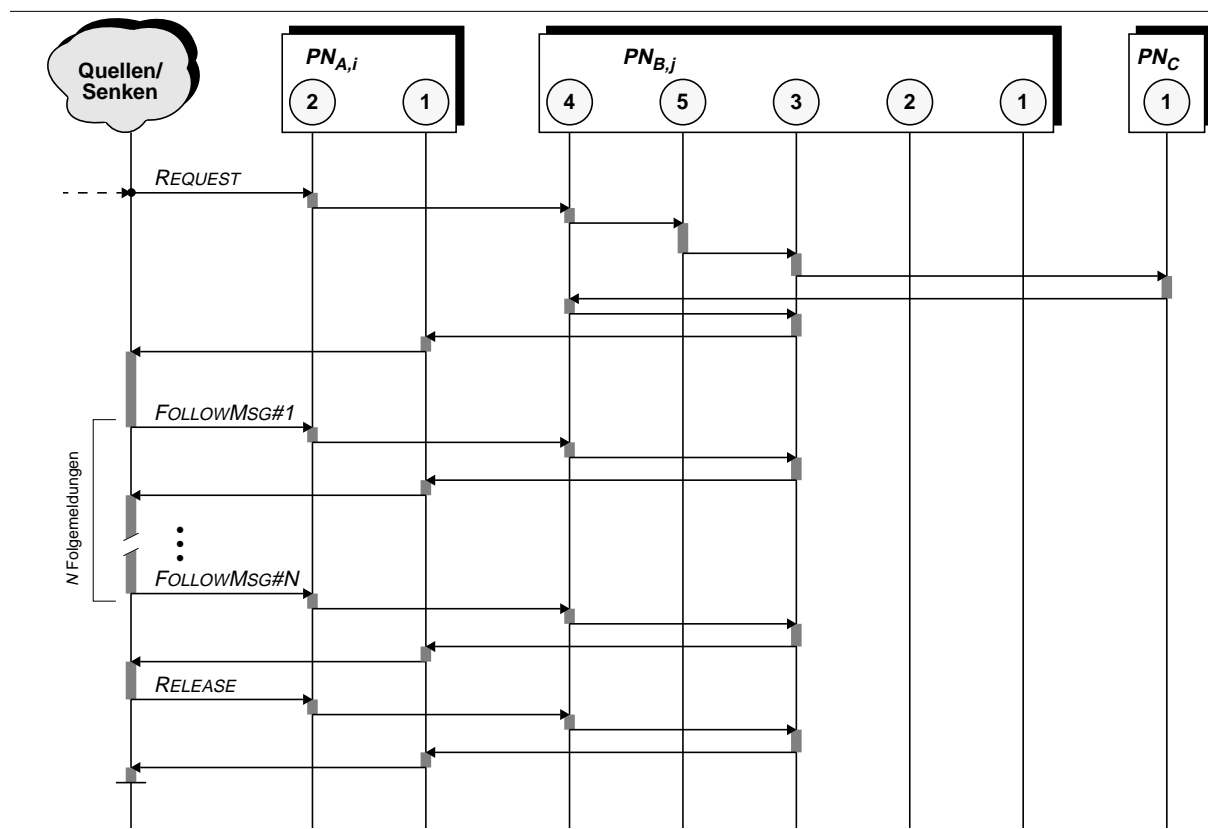


Bild C-5: Erfolgreiches Szenario für Verkehrsklasse 1

## C.4 Gewählte Bediendauerverteilung

Für alle Messungen und Fallstudien im Rahmen dieser Arbeit wurden die Bedienzeiten als negativ-exponentiell verteilt angenommen. Begründet wird dies u. a. durch die Tatsache, dass hier eine Vielzahl „ähnlicher“ Dienste durch eine Dienstgruppe bzw. Verkehrsklasse repräsentiert wird.

## Anhang D

# Exemplarische Messdaten und Ergebnisse der *DLB*-Algorithmen

In Kapitel 8 wurden Messungen und Parameterstudien für die Leistungsregelung nach dem Verfahren des *Distributed Load Balancing* (DLB) vorgestellt. Aufgrund der Vielzahl an Prozessorknoten, Sensortypen, Messgrößen, definierten Leistungsmaßen und den drei verschiedenen Regelalgorithmen fallen in einer Fallstudie große Mengen an Statistikdaten an. Für die Bewertung eines bestimmten Regelverfahrens ist es aber nicht ausreichend, nur den Zeitverlauf eines ausgewählten Leistungsmaßes für die Gesamtleistung zu betrachten. Für die Beurteilung von z.B. Stabilitäts- oder Reagibilitätskriterien muss auch die Aktorik bewertet werden. Mit aus diesen Gründen sind in diesem Anhangskapitel exemplarische Simulationsdaten zusammengestellt.

Als Messreihe wurde die Fallstudie mit *internem Störverkehr* und der *Power-Produkt-Methode* ausgewählt. Siehe hierzu auch Abschnitt 8.4.3 für die Beschreibung der zugehörigen Verkehrsparameter und des zeitlichen Lastprofils.

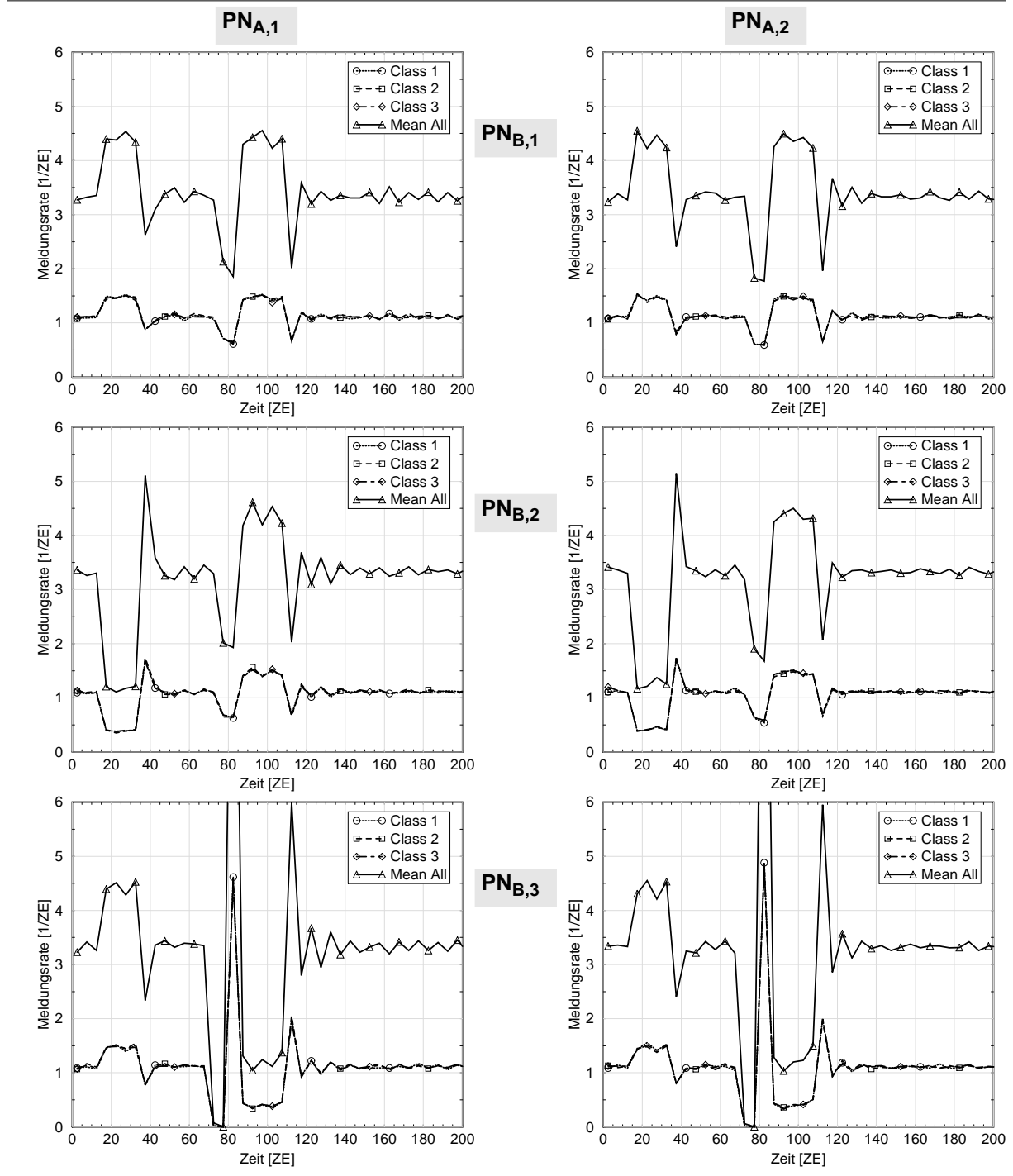
### D.1 Sensorik beim Power-Produkt-Verfahren

Wie in Abschnitt 4.6.4.1 ausgeführt wurde, erfolgt die Messung der Leistungskenngrößen in der *A*-Ebene. Durch das verteilte Verfahren kann jeder  $PN_{A,i}$  nur Messungen an dem von ihm zu bearbeiteten Verkehrsstrom vornehmen. Weiterhin sind in jedem  $PN_{A,i}$  pro Kenngröße bei  $K$  Verkehrsklassen und  $M$  *B*-Prozessorknoten  $K \cdot M$  Statistiken zu führen.

In Bild D-1 sind die gemessenen Annahmeraten  $\hat{\Phi}_{out, RM, A_i, B_j}(k)$  sowie die berechneten Summenmittelwerte  $\hat{\Phi}_{out, RM, A_i, B_j}(k)$  dargestellt. Durch die externe homogene Verkehrsmischung – die mittleren Ankunftsraten  $\hat{\Phi}_{in, RM, A_i, v}$  sind stationär und für alle Klassen und für jeden  $PN_{A,i}$  gleich groß (siehe Tabelle 8-1, Seite 165) – sind die gemessenen Sensorwerte zum einen für beide  $PN_{A,i}$ , und zum anderen für alle drei Verkehrsklassen nahezu identisch. Durch die internen Lastschwankungen unterscheiden sich natürlich die Messwerte für die einzelnen  $PN_{B,j}$ .

Aufgrund der stochastischen Prozesse ist den Signalen ein gewisses Rauschen überlagert. Dieser Anteil wird durch den Einsatz von Filter- und Schätzverfahren erheblich reduziert. Die in den gemessenen Scharmittelwerten sichtbaren Schwankungen sind durch die hier relativ groß gewählte Mess- und Regelintervalldauer bedingt ( $T_{MI} = T_{CI} = 5$  Zeiteinheiten). Die Überschwinger sind die Folge der idealisierten und relativ starken Lastsprünge durch die nichtvermittlungstechnischen Prozesse. Folgende zwei interne Lastimpulse liegen den Messungen hier zugrunde:

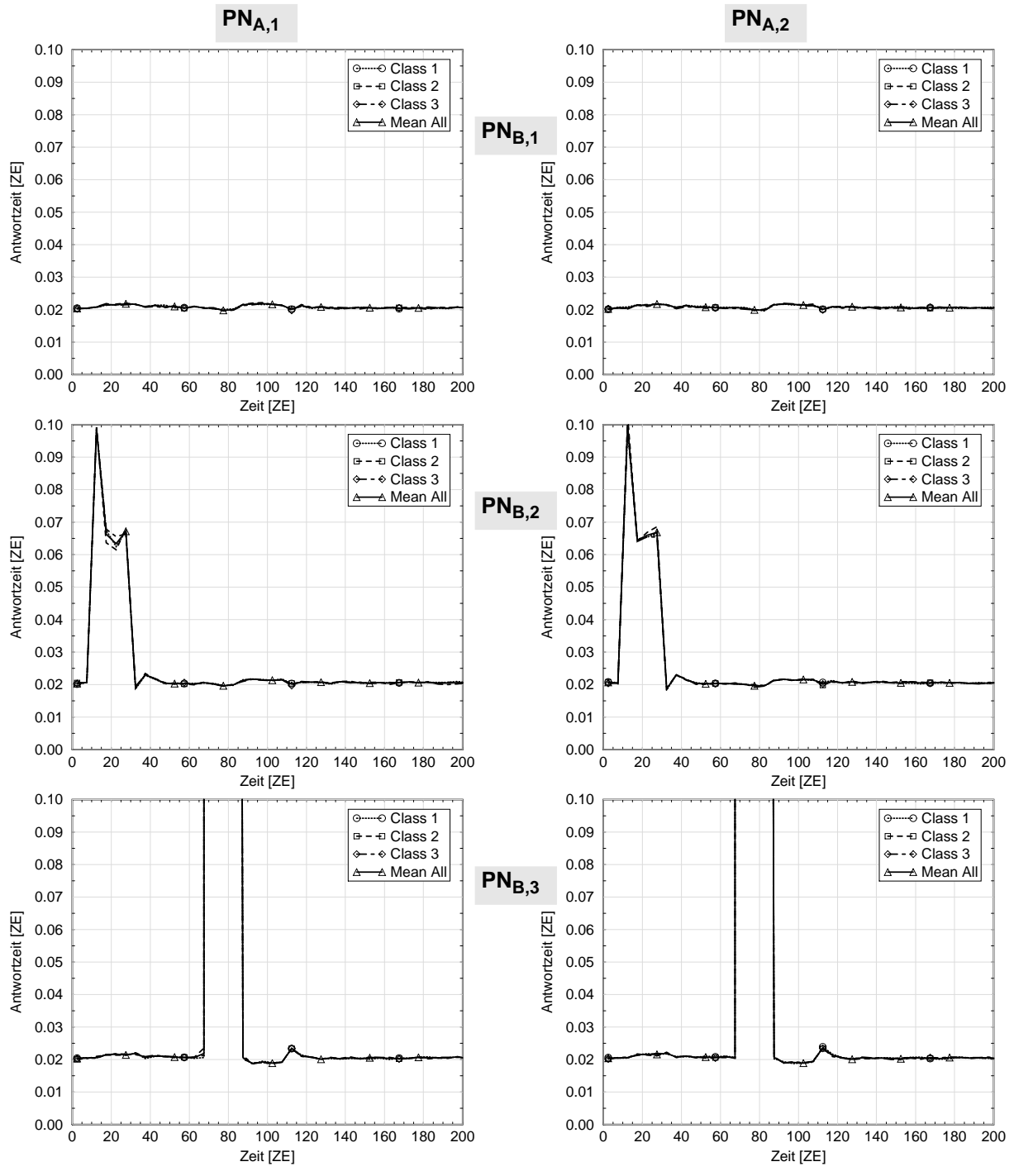
- $PN_{B,2}$ :  $A_{BL, B_2}(t) = 62.5\%$  (d.h. Lastanstieg um Faktor 12.5) für  $t \in [10, 30]$
- $PN_{B,3}$ :  $A_{BL, B_3}(t) = 100\%$  (d.h. Lastanstieg um Faktor 20) für  $t \in [70, 80]$



**Bild D-1:**  $\widehat{\phi}_{out, RM, A_p, B_j}(k)$  und  $\widehat{\phi}_{out, RM, A_p, B_j}(k)$ . Jede Zeile repräsentiert einen  $PN_{B,j}$ , jede Spalte die Sicht eines  $PN_{A,i}$ .

In Bild D-2 sind die entsprechenden Messwerte des zweiten Kenngrößentyps – der *Anwortzeiten*  $\widehat{\tau}_{RM, A_p, B_j}(k)$  bzw. Mittelwerte  $\widehat{\tau}_{RM, A_p, B_j}(k)$  – dargestellt. Aufgrund der homogenen Verkehrsverteilung und der relativ niedrigen Grundlast sind nahezu keine Klassenunterschiede bei den Antwortzeiten messbar.





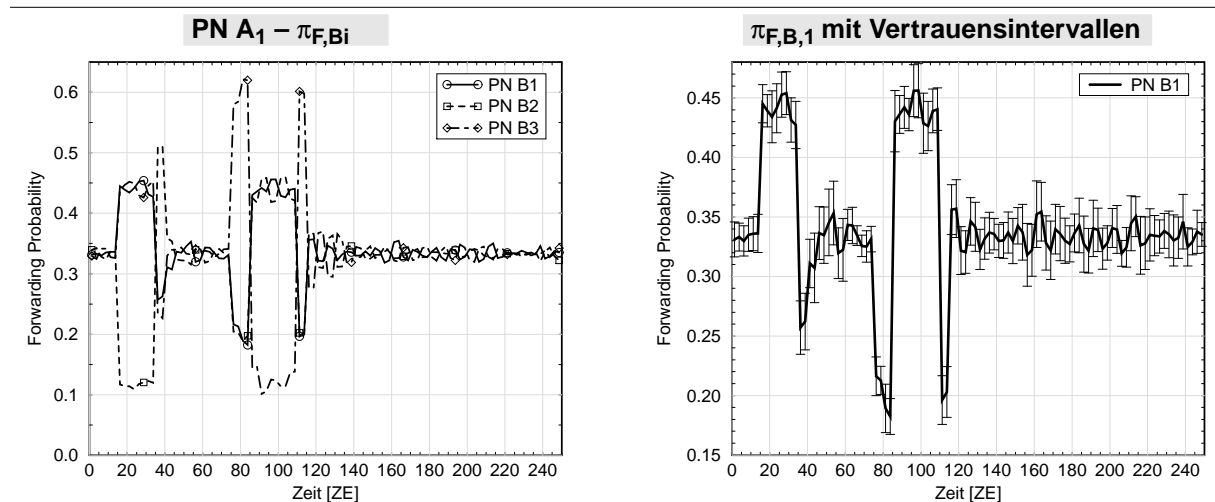
**Bild D-2:** Sensorwerte  $\overrightarrow{\hat{\tau}_{RM, A_i, B_j}(k)}$  und  $\overleftarrow{\hat{\tau}_{RM, A_i, B_j}(k)}$ . Jede Zeile repräsentiert einen  $PN_{B,j}$ , jede Spalte die Sicht eines  $PN_{A,i}$ .

## D.2 Aktorik und Vertrauensintervalle

In fast allen präsentierten Ergebnisdiagrammen wurden die berechneten Vertrauensintervalle aus Gründen der Übersichtlichkeit weggelassen. Dies beruht zum einen auf der Vielzahl von

Funktionsverläufen in einem Schaubild, und zum anderen von der gewählten Messintervalldauer und damit der Vielzahl von Messzeitpunkten in den dargestellten Zeitbereichen.

Aus diesen Gründen wurde in Bild D-3 stellvertretend der gemessene zeitliche Funktionsverlauf der Zuteilungswahrscheinlichkeit  $\pi_{F, A_1, B_1}(k)$  mit zugehörigem berechneten Konfidenzbereich dargestellt. Dieses Schaubild, als auch alle nachfolgenden Simulationsdaten, basieren auf 600 Einzelmessungen, die in zehn Teiltests zu je 60 Elementartests aufgeteilt wurden. Somit gehen in die Berechnung jedes Scharmittelwertes und die Schätzung des zugehörigen Vertrauensbereiches 600 statistisch unabhängige Ereignisse (Stichproben) ein.<sup>1</sup>



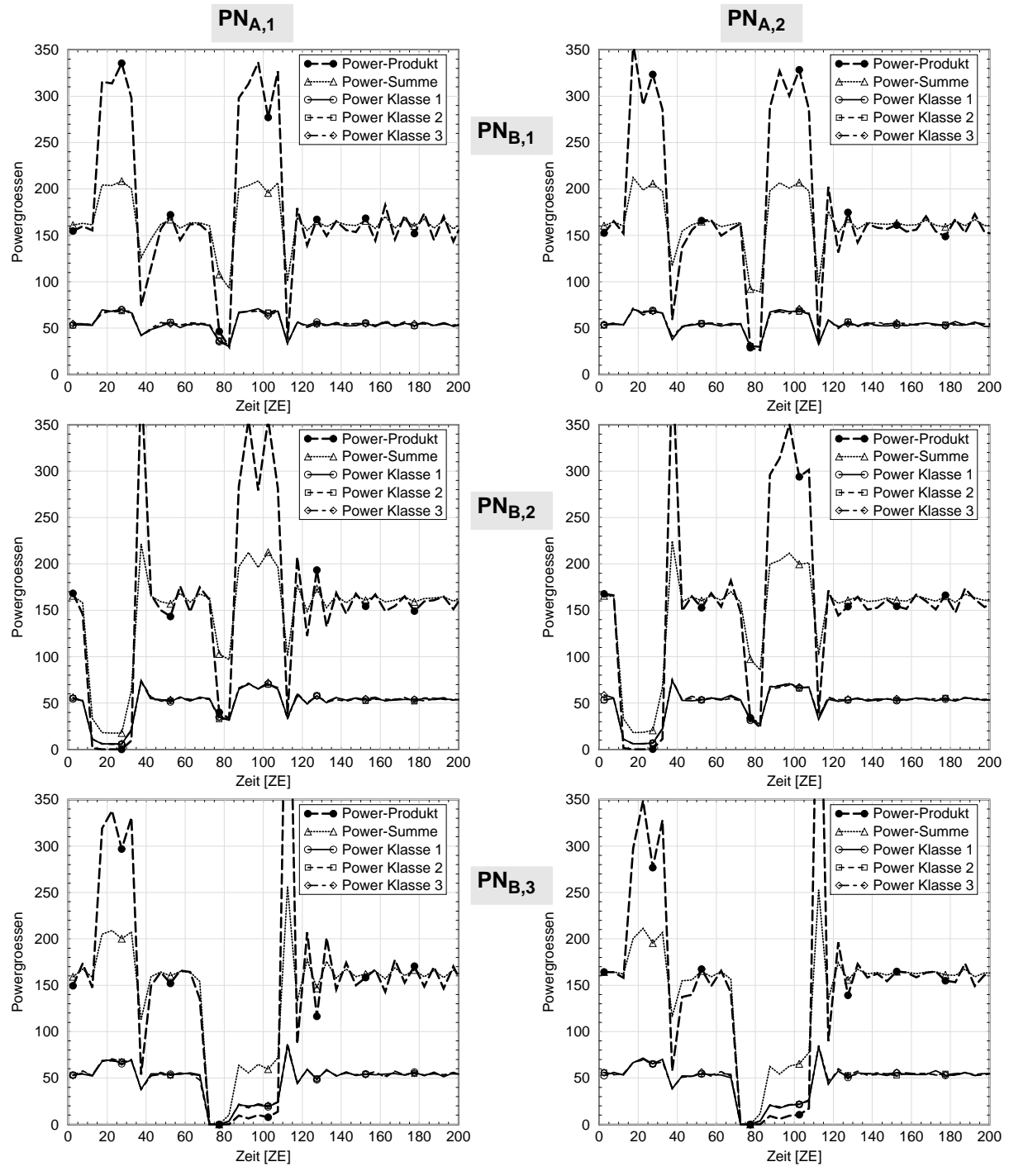
**Bild D-3:** Lastzuteilungsvektor  $\overrightarrow{\pi_{F, A_1}}(k)$  von  $PN_{A, I}$  sowie  $\pi_{F, A_1, B_1}(k)$  mit Konfidenzintervallen.

Der Lastzuteilungsvektor  $\overrightarrow{\pi_{F, A_1}}(k)$  ist Bestandteil der Aktorik. Da diese besonders sensitiv ist wurde ihre Kenngrößen für dieses Beispiel ausgewählt. So sind auch die Konfidenzintervalle aller anderen Systemvariablen im Vergleich hierzu durchweg kleiner.

### D.3 Leistungswerte beim Power-Produkt-Verfahren

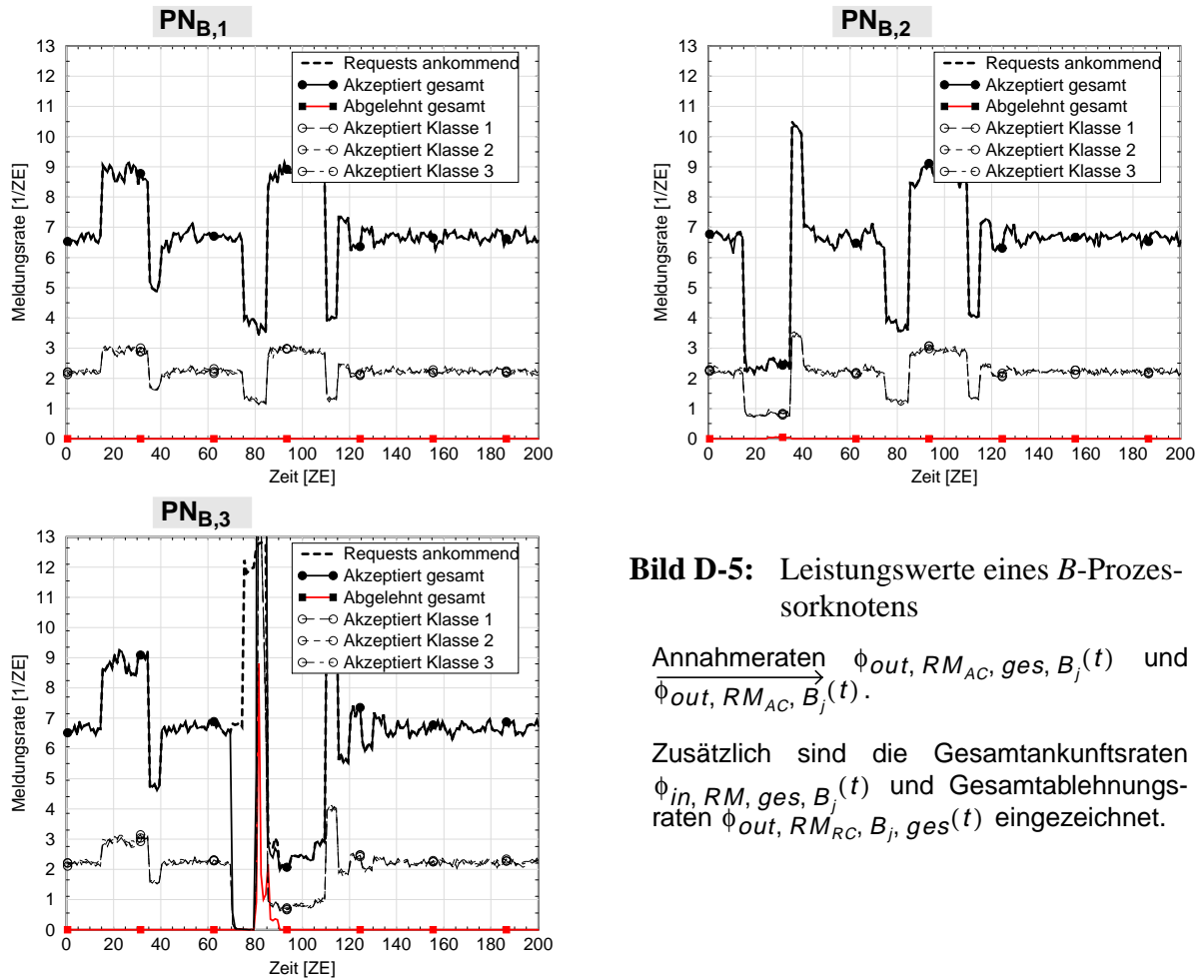
Die Lastzuteilung durch die A-Knoten erfolgt auf Basis der berechneten *Power-Produkte* für die einzelnen  $PN_{B, j}$ . Diese Kennwerte sind deshalb auch direkte Maßzahlen für die Leistung eines einzelnen B-Prozessorknotens. In Bild D-4 sind die sich ergebenden Zeitverläufe für diesen Kenngrößentyp – die Einzelmaße  $\Lambda_{P, RM, A_i, B_j, v}(k)$ , die Summe  $\Lambda_{P^\Sigma, RM, A_i, B_j}(k)$  sowie das Power-Produkt  $\Lambda_{P^\Pi, RM, A_i, B_j}(k)$  – zusammengestellt.

<sup>1</sup> Im Vergleich zu stationären Simulationen ist hier die Anzahl konkreter Stichproben relativ klein [473]. Dies begründet sich neben der instationären Simulationstechnik beim DLB-Verfahren aus dem sehr komplexen Systemmodell sowie der Vielzahl an Messpunkten und der zu führenden Statistiken. So benötigte vorliegende Messung ca. drei Wochen an reiner Prozessorzeit. Wirtschaftliche Gründe verboten deshalb genauere Messungen durchzuführen.



**Bild D-4:** Leistungskennwerte  $\Lambda_{P^{\Pi}, RM, A_i, B_j}(k)$  (mit Faktor  $10^{-3}$  skaliert),  $\Lambda_{P^{\Sigma}, RM, A_i, B_j}(k)$  und die  $\Lambda_{P, RM, A_i, B_j, v}(k)$ . Jede Zeile repräsentiert einen  $PN_{B,j}$ , jede Spalte die Sicht eines  $PN_{A,i}$ .

Die *Power*-Berechnungen basieren auf den Annahmeraten und entsprechenden Antwortzeiten. Im Vergleich zu Bild D-4 sind in Bild D-5 die gemessenen *Annahmeraten*  $\phi_{out, RM_{AC}, ges, B_j}(t)$  und  $\phi_{out, RM_{AC}, B_j}(t)$  dargestellt. Diese instationären Statistiken wurden direkt in den  $PN_{B,j}$  erfasst. Zusätzlich sind die Gesamtankunftsrate  $\phi_{in, RM, ges, B_j}(t)$  und Gesamtablehnungsraten  $\phi_{out, RM_{RC}, B_j, ges}(t)$  eingezeichnet. Bei den Parametereinstellungen in dieser Fallstudie werden nahezu keine neuen Anforderungen von den lokalen Überlastregelungen der  $PN_{B,j}$  abgelehnt.



**Bild D-5:** Leistungswerte eines *B*-Prozessorknotens

Annahmeraten  $\phi_{out, RM_{AC}, ges, B_j(t)}$  und  $\overrightarrow{\phi_{out, RM_{AC}, B_j(t)}}$ .

Zusätzlich sind die Gesamtankunftsrate  $\phi_{in, RM, ges, B_j(t)}$  und Gesamtablehnungsraten  $\phi_{out, RM_{RC}, B_j, ges(t)}$  eingezeichnet.

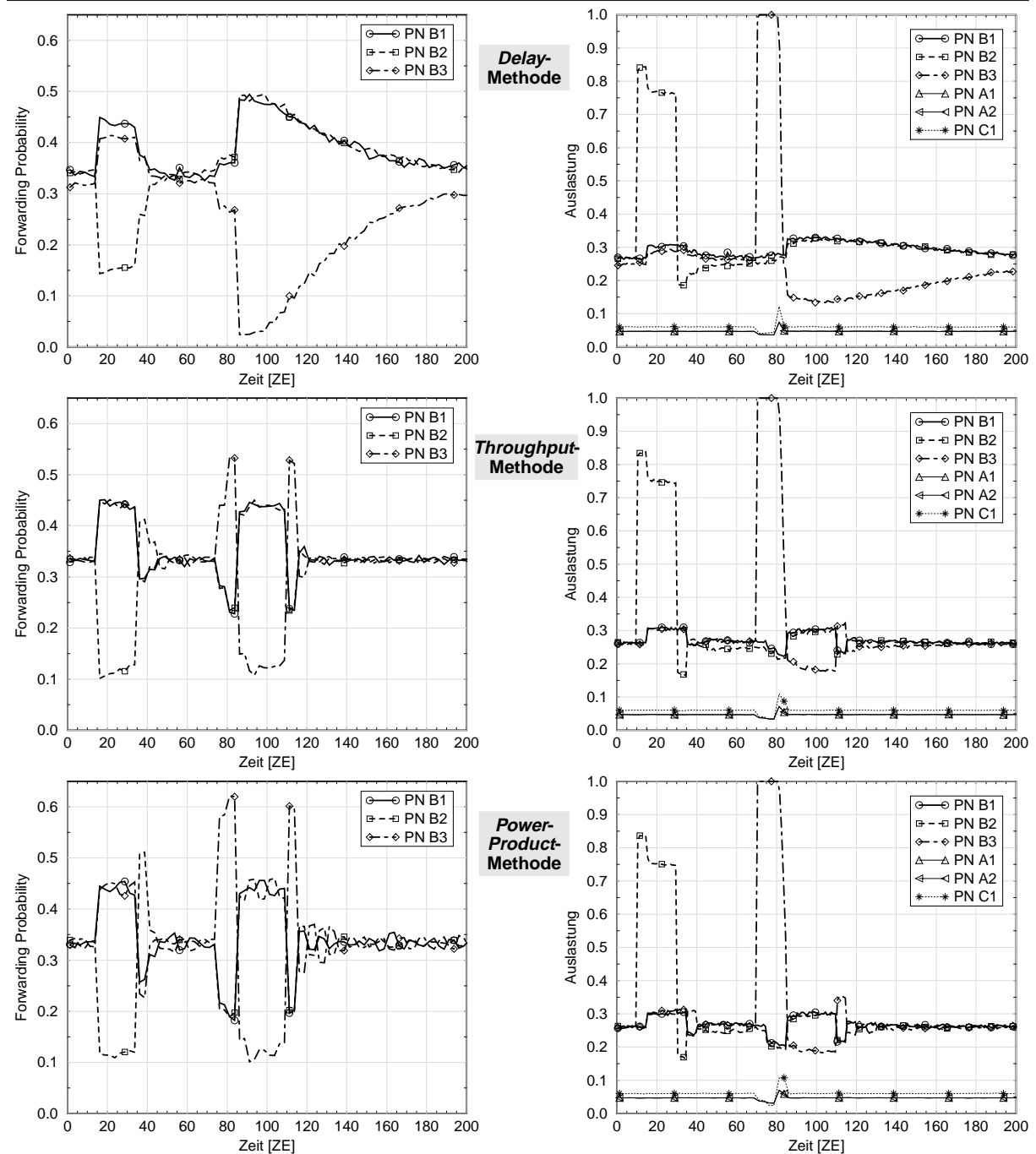
## D.4 Bewertung der drei verschiedenen Algorithmen

Nachdem im vorangegangenen Abschnitt nur Ergebnisse der *Power-Produkt-Methode* vorgestellt wurden, sollen nun diese mit den beiden anderen Algorithmen verglichen werden. Die Bewertung der drei Verfahren soll anhand der Lastzuteilungswahrscheinlichkeiten  $\overrightarrow{\pi_{F, A_i}}$ , der Nutzdurchsätze  $\overrightarrow{\phi_{out, RM_{AC}, B_j}}$ , der knotenindividuellen *Power*-Werte (z.B.  $\Delta P^{II, RM, A_i, B_j}$ ) sowie verschiedener Gesamtleistungsdaten erfolgen.

In Bild D-6 werden die Lastzuteilungswahrscheinlichkeiten verglichen. Aus Platzgründen wurde jeweils nur der Vektor  $\overrightarrow{\pi_{F, A_i}(k)}$  von  $PN_{A, I}$  dargestellt.<sup>2</sup> Anhand der Aktorik lässt sich gut das Reagibilitätsverhalten vergleichen. Dabei weisen das *Power-Produkt*- und das Durchsatzverfahren kürzere Übergangsphasen auf als die *Delay-Methode*. Die Ursache hierfür liegt nicht im Algorithmus sondern vielmehr am Kenngrößentyp *Antwortzeit*, die zwangsläufig eine größere Regeltotzeit bedingt. Deutlich wird dies vor allem am Ausschwingverhalten da hier aufgrund großer Pufferfüllstände relativ lange Wartezeiten zu verzeichnen sind.

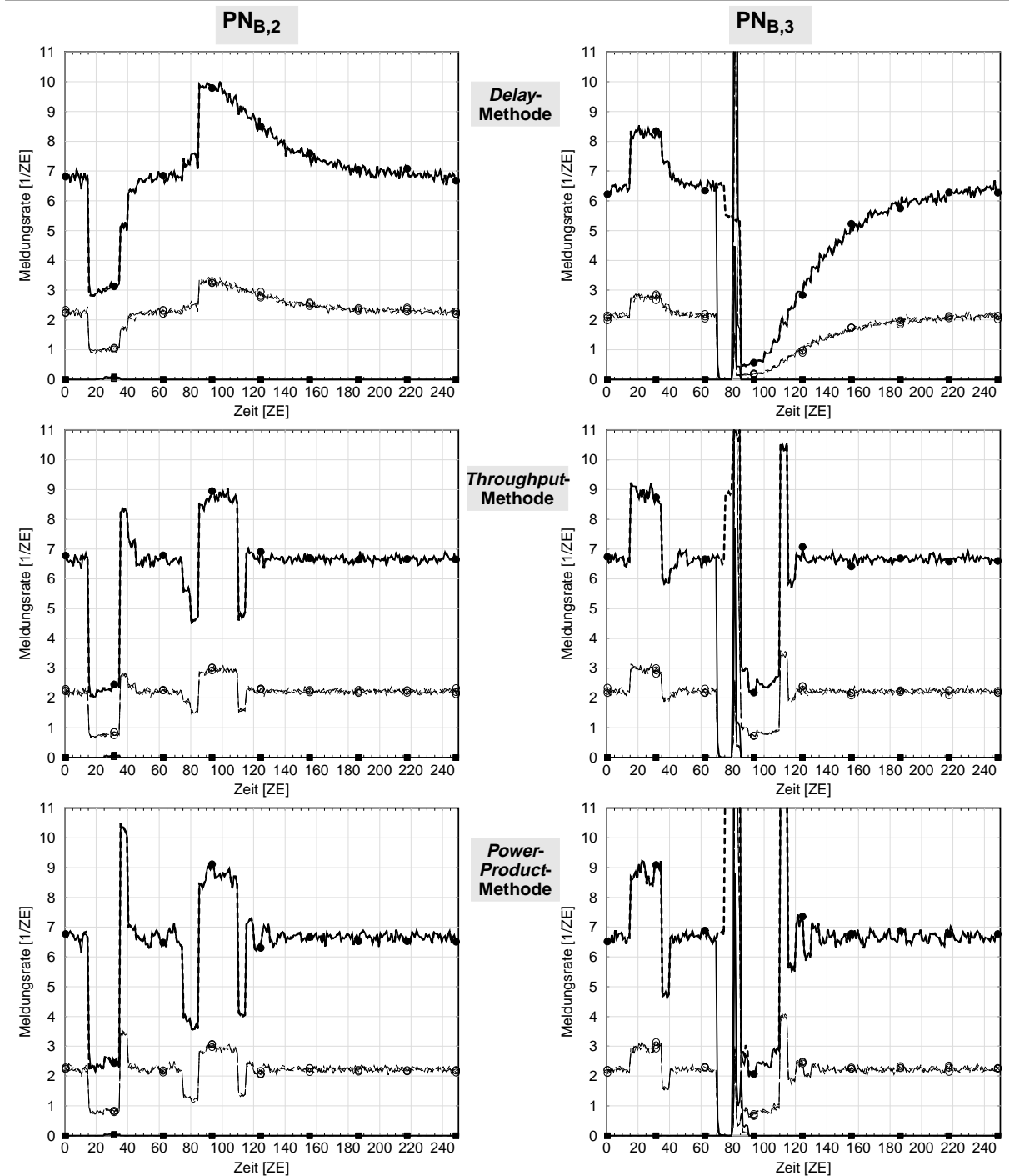
<sup>2</sup> Da für die A-Ebene in dieser Fallstudie symmetrische Verhältnisse vorliegen sind auch die Messwerte für beide A-Prozessorknoten nahezu identisch.

Auf der anderen Seite wird durch diese „inhärente Dämpfungseigenschaft“ auch die Neigung zu Überschwingern kleiner. Gerade hinsichtlich dieser dynamischen Eigenschaft weist das *Power-Produkt*-Verfahren eine größeres Schwankungsverhalten auf als die beiden anderen Algorithmen. Zurückzuführen ist dies auf die Tatsache, dass in die *Power*-Berechnung beide Sensortypen eingehen.



**Bild D-6:** Lastzuteilungsvektor  $\overrightarrow{\pi_{F, A_1}}(k)$  von  $PN_{A, I}$  und CPU-Auslastung  $\rho_{PN_i}(t)$  aller PN für die drei verschiedenen Methoden.

Bild D-6 zeigt auch CPU-Auslastung  $\rho_{PN_i}(t)$  aller sechs Prozessorknoten. So wird durch diese Messdaten auch deutlich, dass sowohl A- als auch C-Ebene bei diesem Vermittlungsrechner nicht den Systemengpass darstellen.

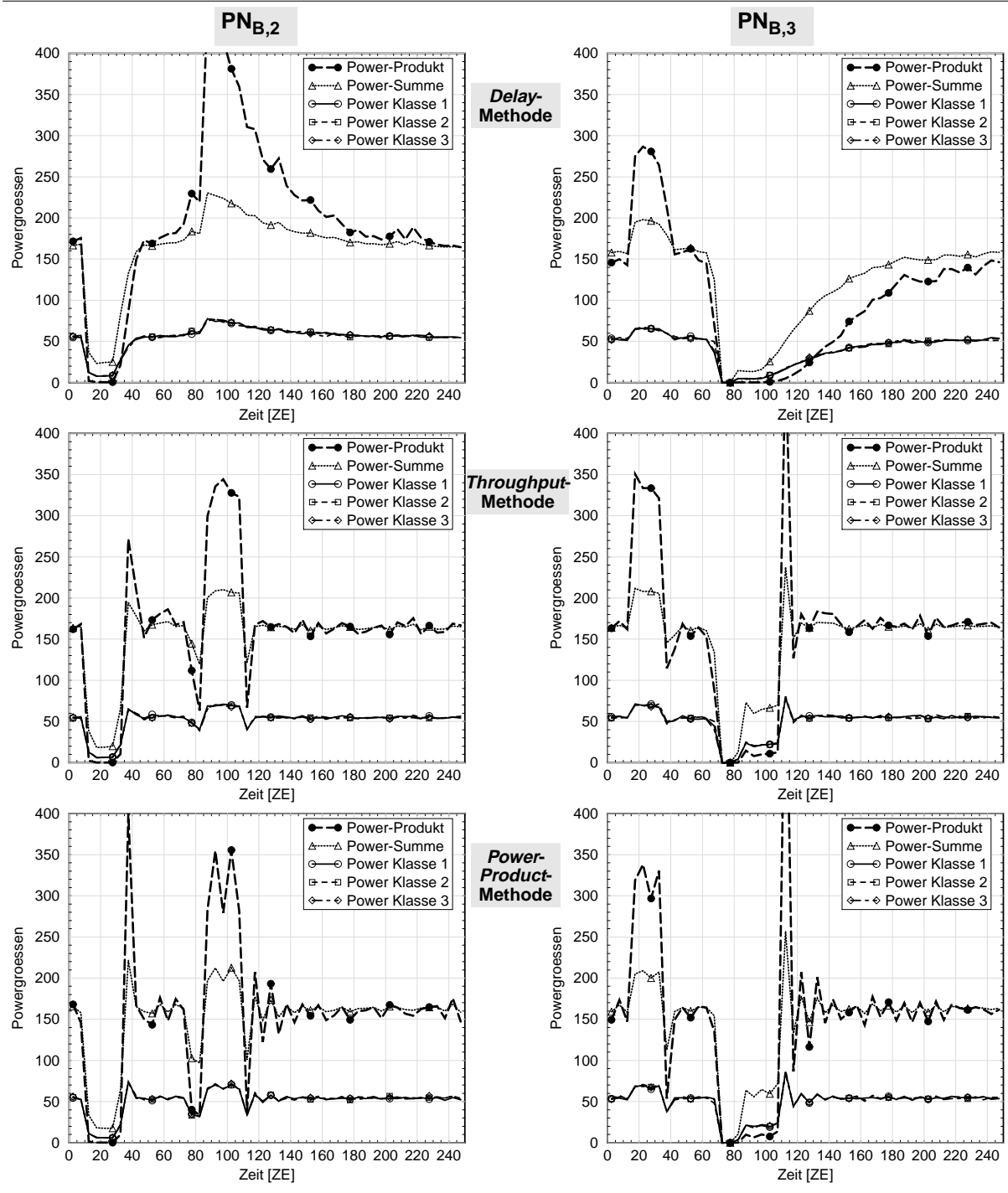


**Bild D-7:** *Meldungsraten*  $\phi_{in, RM, B_j, ges}(t)$ ,  $\phi_{out, RM_{AC}, B_j, ges}(t)$  und  $\phi_{out, RM_{RC}, B_j, ges}(t)$ ; sowie individuell  $\phi_{out, RM_{AC}, B_j, v}(t)$  und  $\phi_{out, RM_{RC}, B_j, v}(t)$  von  $PN_{B,2}$  und  $PN_{B,3}$  für die drei verschiedenen Methoden. (Legende siehe Bild D-5.)

In Bild D-7 können die gemessenen *Annahmeraten*  $\phi_{out, RM_{AC}, ges, B_j}(t)$  und  $\overrightarrow{\phi_{out, RM_{AC}, B_j}}(t)$  für  $PN_{B,2}$  und  $PN_{B,3}$  verglichen werden. Aufgrund der Zuteilungswahrscheinlichkeiten weisen obige Kurven ähnliche Charakteristiken wie diese auf.

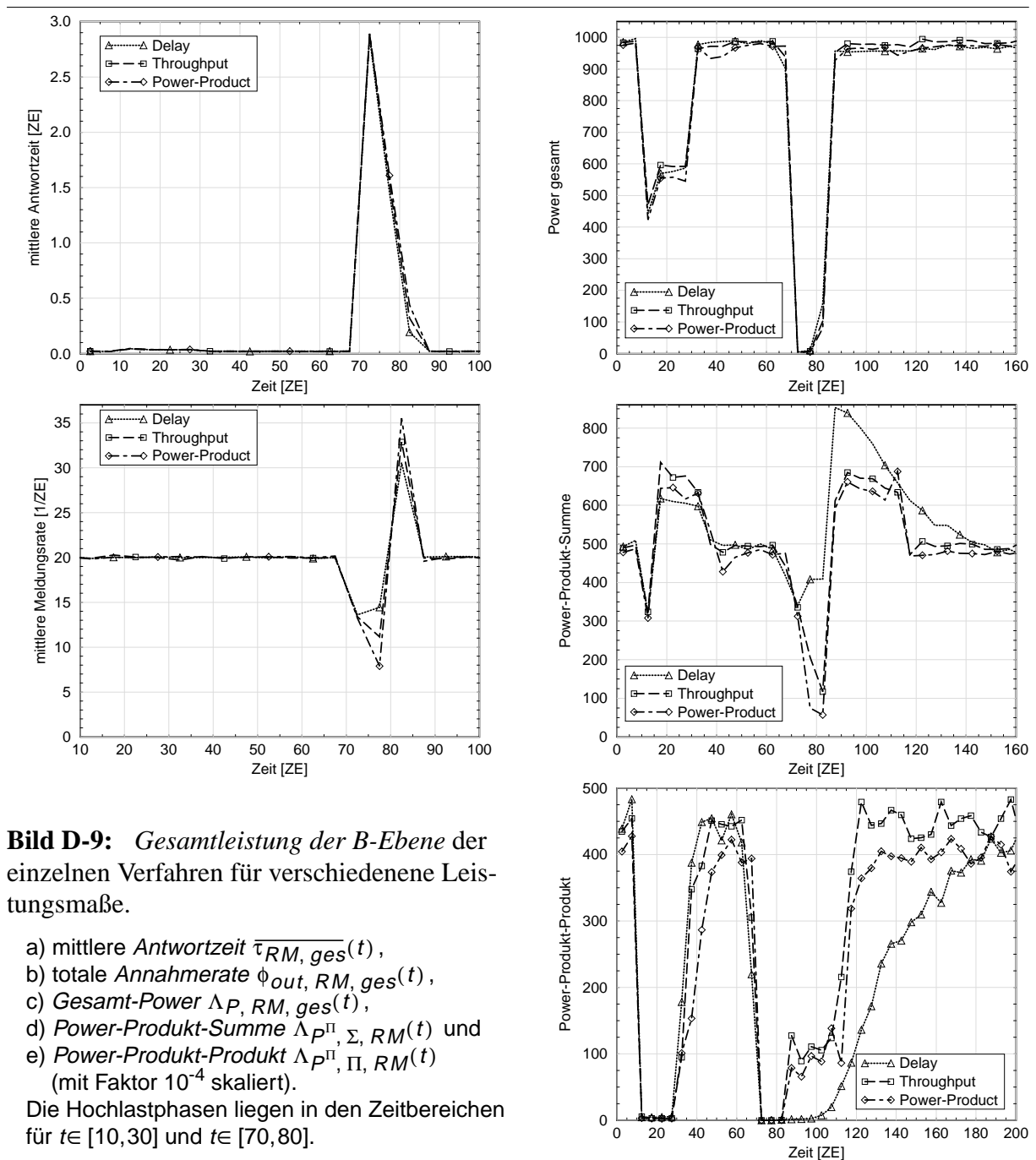
Das *Power-Produkt*  $\Lambda_{P^I, RM}$  ist zum einen die zentrale Regelgröße beim gleichnamigen Verfahren, zum anderen aber auch eine wichtige Leistungs Kenngröße. Aus diesem Grund wurden

in Bild D-8 diese Leistungswerte der beiden Knoten  $PN_{B,2}$  und  $PN_{B,3}$  für die drei verschiedenen Algorithmen gegenübergestellt. Deutlich wird auch hier wieder die relativ lange Ausschwingdauer beim Delay-Verfahren. Weiterhin wurden die *klassenindividuellen Power-Daten*  $\Lambda_{P, RM, A_i, B_j, v}$  als auch die *Power-Summe*  $\Lambda_{P^S, RM, A_i, B_j}$  aufgeführt.



**Bild D-8:** Leistungskennwerte  $\Lambda_{P^H, RM, A_i, B_j}(k)$  (mit Faktor  $10^{-3}$  skaliert),  $\Lambda_{P^S, RM, A_i, B_j}(k)$  und die  $\Lambda_{P, RM, A_i, B_j, v}(k)$  für die drei verschiedenen Methoden *aus Sicht des  $PN_{A,1}$* . Von der B-Ebene sind nur die Daten der  $PN_{B,2}$  und  $PN_{B,3}$  dargestellt.

Verschiedene Maßzahlen zur *Gesamtleistung der B-Ebene* sind in Bild D-9 zusammengestellt.



Betrachtet man die mittlere Antwortzeit  $\overline{\tau_{RM, ges}}$  und die totale Annahmerate  $\phi_{out, RM, ges}$  dann hat der erste Hochlastimpuls kaum sichtbare Auswirkungen auf diese Leistungsmaße. Deutlich werden diese allerdings im *Gesamt-Power-Maß*  $\Lambda_{P, RM, ges}$  (Berechnung nach Gl. (8-8)). Für obige drei Kenngrößen lassen sich kaum deutliche Abweichungen zwischen den verschiedenen Algorithmen erkennen.

Unterschiede lassen sich aber hinsichtlich der *Power-Produkt-Summe*  $\Lambda_{P^{\Pi}, \Sigma, RM}$  als auch des *Power-Produkt-Produkts*  $\Lambda_{P^{\Pi}, \Pi, RM}$  feststellen. So wird hier die stärkere Dämpfung beim Delay-Verfahren besonders deutlich.