

Institut für Nachrichtenvermittlung und Datenverarbeitung
Prof. Dr.-Ing. habil. P. J. Kühn

62. Bericht über verkehrstheoretische Arbeiten

Verkehrssteuerung in ATM-Netzen Verfahren und verkehrstheoretische Analysen zur
Zellpriorisierung und Verbindungsannahme

von

Hans Kröner

1995



## **University of Stuttgart**

Institute of Communications Switching and Data Technics
Prof. Dr.-Ing. habil. P. J. Kühn

**62th Report on Studies in Congestion Theory** 

Traffic Control in ATM Networks 
Mechanisms and Teletraffic Analysis for Cell

Prioritization and Connection Admission

by
Hans Kröner
1995

## Summary

The Asynchronous Transfer Mode (ATM) is a transmission, multiplexing and switching technique which has been developed to meet the growing demand for broadband communications in the wide area as well as in the local area. ATM networks are based on a packet-and connection-oriented information transfer using fixed size packets called cells. Several connections share the network resources according to their actual demand by using the asynchronous time division multiplexing principle. This provides a flexible bit rate allocation which is needed to support a wide variety of different communication services with its distinct bit rate and quality of service requirements.

In particular, variable bit rate (VBR) services may share transmission capacity in a statistical way, i.e., the mutual compensation of the bit rate requirements of different connections may be exploited to reserve less than peak bit rate for VBR connections. This statistical multiplexing scheme provides an economic utilization of the network resources, but requires a proper control of the network load in order to guarantee network performance objectives in terms of cell loss and delay. The international standardization bodies have specified a general traffic and congestion control framework, which is based on a comprehensive set of control functions working on different time scales.

This thesis is concerned with two key components of this traffic control framework — the cell loss priority and the connection admission control. The cell loss priority control provides two different bearer services with distinct cell loss probabilities, while the connection admission control function decides whether a new connection request can be accepted or not. The first part of the thesis is devoted to a systematic investigation of the relationship among traffic, system and performance characteristics. This provides a solid basis for a comparative study of different cell loss priority and connection admission control algorithms, which is the aim of the second part of the thesis.

Chapter 1 motivates the need for broadband networks and describes the fundamental characteristics of the Asynchronous Transfer Mode.

The following chapter presents some background information concerning the services, the architecture and the protocols of the Broadband Integrated Services Digital Network (B-ISDN). Section 2.1 discusses the bit rate, delay and bit error rate requirements of some typical communication services. The second section contains a brief description of the main features of the Asynchronous Transfer Mode, covering the multiplexing and switching principle, the concept of virtual paths and virtual connections and the cell format. The subsequent section addresses the B-ISDN protocol reference model, which is structured into the physical layer, the ATM layer and the ATM adaptation layer. The basic functionality of the different layers and the signalling principles are briefly described.

Section 2.4 concentrates on the traffic management framework, which has to provide a satisfactory network performance as well as an economic use of the network resources. It consists of two main components that are based on different control principles. Traffic control aims at the prevention of a congestion, while congestion control reacts to a congestion situation and limits its intensity, spread and duration. The whole set of traffic and congestion control functions, their main properties and their interrelations are discussed in detail. It should be emphasized that the control mechanisms work on different time scales, which results in a hierarchical control structure.

The definition of traffic control mechanisms has to be based on adequate performance models and analytical solution techniques. Chapter 3 introduces a hierarchical traffic modelling framework, which is frequently used for the performance modelling of ATM networks. This model characterizes the traffic source behaviour at different levels, namely the cell, burst and connection level. The corresponding network models describing an ATM multiplexer or a complete reference connection are defined in Section 3.2. The following section deals with the event-by-event simulation method which will be used to validate analytical performance results. Analytical methods are preferred because traditional simulation techniques are not able to measure realistic cell loss probabilities. The analytical methods which will be used in the subsequent chapters are derived in Section 3.4.

The focus of Chapter 4 is the performance analysis of generic ATM network components. Section 4.1 is concerned with the analysis of a statistical multiplexer, which combines the traffic streams of different connections. The multiplexer can be modelled by a single server queueing system with constant service time  $\Delta t$  and finite buffer size S. The traffic characteristics of the different connections are described by the sporadic source model. A sporadic source alternates between burst and silence phases and is a realistic model for data and packetized voice traffic. A burst consists of a geometrically distributed number of cells with deterministic interarrival times and the duration of a silence phase is geometrically distributed. This model has been extensively used for the performance analysis of ATM networks because it captures the essential traffic and system characteristics. Section 4.1 presents a survey of the numerous approximations that have been proposed within the literature to solve this queueing model.

Using a discrete-time Markov chain approach, we have obtained an exact, but numerically complex, analytical solution. The computational burden is significantly smaller if the sum of the peak bit rates is lower than the transmission bit rate, i.e., a burst level overload is avoided. Assuming this peak bit rate allocation scheme, it is shown that the queueing behaviour becomes independent of the burstiness. Hence, it can be evaluated from the superposition of constant bit rate traffic streams with equivalent arrival rates. Furthermore, it is proven that the discrete-time M/D/1/(S+1)-system yields an upper bound for the delay and loss characteristics of this cell level queueing model. This can be explained by the fact that the

arrival process approaches a discrete-time Poisson arrival process as the number of traffic streams increases.

On the other hand, the burst level queueing behaviour can be characterized by a fluid flow model, which ignores cell level congestion by assuming continuous arrival and departure flows instead of discrete cell arrivals and departures. Appendix A is dedicated to the analysis of this model and contains novel analytical results for general Markov-modulated arrival processes.

The overall queueing model can be decomposed into the complementary M/D/1/(S+1) and fluid flow queueing models which describe the cell and burst level queueing behaviour, respectively. The results which are obtained from this simple decomposition method are in good agreement with the exact analysis. Moreover, Section 4.2 derives an appropriate analytical method to aggregate the performance results of different connection links in order to evaluate the end-to-end cell delay distribution and loss probability of a complete reference connection. The comprehensive analytical studies, which are described in Chapter 5, confirm the following conclusions with respect to ATM traffic control:

- A cell level congestion can be compensated by relatively small buffers, while a burst level congestion requires a buffer size which is larger than the mean burst size in order to reduce the cell loss probability significantly.
- The correlation of consecutive cell losses increases if the buffer becomes larger.
- The traffic characteristics are only slightly altered if the network buffers are dimensioned for cell level congestion. Therefore, the different network links can be considered separately.
- The cell loss probability is a crucial network performance parameter, while the cell
  delay variation is almost negligible for small network buffers.

Consequently, an ATM network supports selective cell discarding because the different services have diverse cell loss probability requirements. Furthermore, cell loss priorities can be used to isolate different traffic flows. The comparison of various cell loss priority mechanisms is the subject of Chapter 5. First, a taxonomy and a qualitative discussion of the properties of various loss priority schemes is used to identify the most attractive priority mechanisms. The simplest approach uses separate network routes to achieve different cell loss probability objectives. This scheme is simple to implement, because no special buffer management hardware is required. The second approach uses different buffer filling thresholds to decide whether a new cell may enter the buffer. This threshold mechanism can be implemented using a table which contains the corresponding threshold values, a counter which stores the filling level and a comparator which compares both values. Finally, a push-out priority scheme may be used to discard a low priority cell when a high priority cell meets a completely filled buffer. This priority scheme requires a relatively complex buffer management scheme in order to guarantee cell sequence integrity. A possible implementation is proposed in Section 5.2.

Section 5.3 deals with the performance evaluation of the different priority mechanisms considering cell and burst level queueing models. Analytical solutions are derived for Poisson arrival processes, which reflect the cell level queueing behaviour. The burst level queueing behaviour is assessed by simulation assuming sporadic input traffic streams. Additionally, a fluid flow analysis is developed for the threshold mechanism. Detailed performance studies confirm that the admissible traffic load can be increased significantly by using cell loss priorities. As expected, there is a tradeoff between the performance and the implementation complexity of the different cell loss priority schemes, i.e., the push-out scheme performs slightly better than the threshold mechanism while the performance of the route separation scheme is much worse.

Chapter 6 addresses the connection admission control (CAC) function which plays a key role within the previously mentioned traffic control framework, because this functions decides whether a new connection request can be accepted or should be rejected. The different CAC mechanisms can be classified as follows:

- Direct CAC mechanisms apply analytical performance models to estimate the relevant network performance parameters directly from the traffic and system characteristics.
   Various mechanisms will be considered which are based on a convolution algorithm, a normal approximation, a large deviation approximation or a cell loss probability upper bound.
- Indirect CAC mechanisms define different traffic classes, which comprise connections with similar traffic characteristics. Each connection reserves an effective, equivalent or virtual bit rate which depends on the traffic characteristics of the corresponding traffic class. Detailed studies will be carried out for the linear approach, the class related rule and the algorithm used in the plaNET test network. Another class of indirect CAC algorithms is based on neural networks, which learn the relationship between the actual connection mix and the resulting network performance. The connection mixes are classified into "good" patterns, where network performance objectives are fulfilled, and "bad" patterns.

Section 6.2 is concerned with a detailed description and a qualitative discussion of the properties of the different CAC algorithms, which reveals their major advantages and drawbacks. The subsequent section presents a comparison of the connection admission boundaries of the different CAC schemes. The connection admission boundary identifies the connection mixes which are admitted by the corresponding CAC algorithm. An ideal CAC algorithm can be defined such that all connection mixes inside of the connection admission boundary fulfill the network performance requirements, while all connection mixes outside of the connection admission boundary violate the network performance objectives. Furthermore, we compare the maximum cell loss probabilities and the call loss probabilities of the different algorithms.

The results indicate a tradeoff between the performance and the implementation complexity of the considered algorithms.

Based on these results, a new two-level CAC approach is suggested in Section 6.4. The first level uses a simple indirect CAC algorithm in order to achieve a real-time CAC decision. The second level performs a periodical adaptation of the effective bit rates used by the first level algorithm, i.e., a convolution algorithm is used to adapt the effective bit rate to the actual traffic composition. A simulation study confirms that this promising CAC scheme is able to guarantee network performance objectives and achieves a satisfactory network load.

Finally, Section 6.5 discusses various methods to specify and control the traffic parameters of a connection. The prediction and control of statistical traffic parameters like the mean bit rate seems to be impossible for most services. An algorithmic description allows a proper control of the traffic parameters but it is still difficult to predict the traffic parameters. Furthermore, it is shown that there is no unique worst case traffic source which provides an upper bound for the cell loss probability for a given set of traffic parameters. Therefore, an additional measurement of the traffic parameters seems to be a promising approach for traffic characterization and control.

# Inhaltsverzeichnis

A	bkür	zungei	n .	XI		
Fo	rme	lzeiche	en	XIII		
1	Ein	ıleitung				
	1.1	Diens	teintegrierende Breitband-Kommunikation	1		
	1.2	Übers	icht über die Arbeit	2		
2	Die	nste, A	Architektur und Protokolle des Breitband-ISDN	4		
	2.1	Chara	kterisierung zukünftiger Dienste	4		
		2.1.1	Klassifizierung der Dienste	4		
		2.1.2	Verkehrscharakteristik ausgewählter Dienste	6		
		2.1.3	Dienstgüteanforderungen ausgewählter Dienste	9		
	2.2	Der A	synchrone Transfer-Modus	11		
		2.2.1	Vermittlungsprinzip	11		
		2.2.2	Konzept der virtuellen Pfade	13		
		2.2.3	Zellformat	15		
	2.3	Protol	kollarchitektur des Breitband-ISDN	17		
		2.3.1	Protokoll-Referenzmodell	17		
		2.3.2	Bitübertragungsschicht	19		
		2.3.3	ATM-Schicht	20		
		2.3.4	ATM-Anpassungsschicht	20		
		2.3.5	Signalisierungsprinzipien	22		
	2.4	Verkel	hrsmanagement für ATM-Netze	23		

		2.4.1	Übersich	t und Zusammenwirken der Steuerungsfunktionen	23
		2.4.2	Funktion	nen zur Verkehrssteuerung	27
			2.4.2.1	Dienstgüteklassen und Prioritätsmechanismen	27
			2.4.2.2	Quellflußkontrolle	28
			2.4.2.3	Formung des Quellverkehrs	30
			2.4.2.4	Schnelle Ressourcenzuteilung	31
			2.4.2.5	Verbindungsannahme und Verkehrslenkung	32
			2.4.2.6	Verwaltung virtueller Pfade	33
		2.4.3	Funktion	nen zur Überlastbehandlung	34
			2.4.3.1	Selektives Verwerfen von Zellen	34
			2.4.3.2	Explizite Anzeige einer Überlast	35
3	Met	hoden	zur Mo	dellierung und Leistungsbewertung von ATM-Netzen	36
	3.1	Model	lierung vo	n ATM-Verkehrsströmen	37
		3.1.1	Zufallsva	riablen, Verteilungen, Momente und Transformationsmethoden	37
		3.1.2	Hierarch	ische Modellierung des Quellverkehrs	41
			3.1.2.1	Modellierung der Verbindungsebene	43
			3.1.2.2	Modellierung der Büschelebene	44
			3.1.2.3	Modellierung der Zellebene	44
	3.2	Modell	lierung de	s Kommunikationsnetzes	46
		3.2.1	Modellie	rung eines Verbindungsabschnitts	46
		3.2.2	Netzmod	ellierung	48
	3.3	Simula	tionstech	nik	49
	3.4	Grund	lagen der	verkehrstheoretischen Analyse	51
4	Leis	tungsu	ıntersucl	nung generischer ATM-Systemmodelle	55
	4.1	Leistur	ngsanalyse	e eines statistischen Multiplexers	55
		4.1.1	Exakte A	Analyse	58
			4.1.1.1	Modellierungsaspekte	58
			4.1.1.2	Analytischer Ansatz	59
			4.1.1.3	Charakteristische Kenngrößen	62

			4.1.1.4 Ergebnisse	63
		4.1.2	Leistungsuntersuchung auf Zellebene	67
			4.1.2.1 Überlagerung homogener sporadischer Quellen $\dots$	67
			4.1.2.2 Heterogene Verkehrszusammensetzung	69
			4.1.2.3 Einfache Näherungsmodelle	70
			4.1.2.4 Ergebnisse	71
		4.1.3	Leistungsuntersuchung auf Büschelebene	73
	4.2	Leistu	ngsanalyse einer Referenzverbindung	81
		4.2.1	Approximative Analyse	82
			4.2.1.1 Analyse einer einzelnen Bedienstation	82
			4.2.1.2 Analyse der Referenzverbindung	85
		4.2.2	Ergebnisse	89
	4.3	Schluß	Sfolgerungen für die Verkehrssteuerung in ATM-Netzen	93
5	Unt	ersuch	ung von Verlustprioritätsmechanismen	95
5	<b>Unt</b> 5.1		nung von Verlustprioritätsmechanismen	<b>95</b>
5		Anwer		
5	5.1	Anwer	ndungsbeispiele und Anforderungen	95
5	5.1	Anwer	reibung, Modellierung und Implementierung der Mechanismen	95 97
5	5.1	Anwer Beschi	reibung, Modellierung und Implementierung der Mechanismen  Getrennte Wegeführung  Selektives Verwerfen am Puffereingang	95 97 99
5	5.1	Anwer Beschi 5.2.1 5.2.2	ndungsbeispiele und Anforderungen reibung, Modellierung und Implementierung der Mechanismen Getrennte Wegeführung Selektives Verwerfen am Puffereingang Verdrängende Prioritäten	95 97 99 100
5	5.1	Anwer Beschr 5.2.1 5.2.2 5.2.3 5.2.4	reibung, Modellierung und Implementierung der Mechanismen  Getrennte Wegeführung  Selektives Verwerfen am Puffereingang  Verdrängende Prioritäten  Selektives Verwerfen am Pufferausgang	95 97 99 100
5	5.1	Anwer Beschr 5.2.1 5.2.2 5.2.3 5.2.4	reibung, Modellierung und Implementierung der Mechanismen  Getrennte Wegeführung  Selektives Verwerfen am Puffereingang  Verdrängende Prioritäten  Selektives Verwerfen am Pufferausgang	95 97 99 100 102
5	5.1	Anwer Beschi 5.2.1 5.2.2 5.2.3 5.2.4 Vergle	reibung, Modellierung und Implementierung der Mechanismen  Getrennte Wegeführung  Selektives Verwerfen am Puffereingang  Verdrängende Prioritäten  Selektives Verwerfen am Pufferausgang  ichende Leistungsbewertung  Leistungsuntersuchung auf Zellebene	95 97 99 100 102 105
5	5.1	Anwer Beschi 5.2.1 5.2.2 5.2.3 5.2.4 Vergle	reibung, Modellierung und Implementierung der Mechanismen  Getrennte Wegeführung  Selektives Verwerfen am Puffereingang  Verdrängende Prioritäten  Selektives Verwerfen am Pufferausgang  ichende Leistungsbewertung  Leistungsuntersuchung auf Zellebene  5.3.1.1 Analyse des Schwellwertverfahrens	95 97 99 100 102 105 106
5	5.1	Anwer Beschi 5.2.1 5.2.2 5.2.3 5.2.4 Vergle	reibung, Modellierung und Implementierung der Mechanismen  Getrennte Wegeführung  Selektives Verwerfen am Puffereingang  Verdrängende Prioritäten  Selektives Verwerfen am Pufferausgang  ichende Leistungsbewertung  Leistungsuntersuchung auf Zellebene  5.3.1.1 Analyse des Schwellwertverfahrens	95 97 99 100 102 105 106

6	Unt	tersuchung von Verbindungsannahmestrategien 12		
	6.1	Anford	derungen und Vergleichskriterien	123
	6.2	Ansätz	ze zur Verbindungsannahme	125
		6.2.1	Klassifizierung der Verbindungsannahmealgorithmen	125
		6.2.2	Direkte Verfahren	126
			6.2.2.1 Faltungsalgorithmus	126
			6.2.2.2 Normalverteilungsapproximation	130
			6.2.2.3 "Large Deviation"-Approximation	131
			6.2.2.4 Obere Schranke der Zellverlustwahrscheinlichkeit	133
		6.2.3	Indirekte Verfahren	135
			6.2.3.1 Linearer Ansatz	135
			6.2.3.2 "Class Related Rule"	136
			6.2.3.3 "plaNET"-Verbindungsannahmealgorithmus	137
			6.2.3.4 Verfahren basierend auf neuronalen Netzen	139
	6.3	Vergle	ichende Leistungsuntersuchung der Annahmealgorithmen	143
		6.3.1	Bestimmung der Verbindungsannahme-Grenzkurven	143
		6.3.2	Leistungsvergleich auf Verbindungsebene	152
		6.3.3	Diskussion der Ergebnisse	153
	6.4	Zweist	ufiger Annahmealgorithmus	154
		6.4.1	Beschreibung	155
		6.4.2	Leistungsuntersuchung	159
	6.5	Festleg	gung der Verkehrsparameter	162
		6.5.1	Statistische Beschreibung des Verkehrs	162
		6.5.2	Algorithmische Beschreibung des Verkehrs	163
		6.5.3	Messung des Verkehrs	164
7	<b>Z</b> 118	ammei	nfassung und Ausblick	166
	_ ~~			

A	"Fluid Flow"-Analyse für allgemeine Markoff-modulierte Ratenprozesse 1		
	A.1	System- und Quellcharakterisierung	187
	A.2	Grundlegender Lösungsansatz	190
	A.3	Allgemeine Lösung des Differentialgleichungssystems	193
	A.4	Lösung des Eigenwertproblems	195
	A.5	Lösung des Randwertproblems	197
	A.6	Charakteristische Leistungskenngrößen	198
	A.7	Einfache Näherungsmodelle	201
	A.8	Untersuchung des Schwellwertverfahrens	203

## Abkürzungen

AAL ATM Adaptation Layer

ACM Association for Computing Machinery

ADPCM Adaptive Differentielle Pulscodemodulation

AF Arrival First

ATM Asynchroner Transfer-Modus

BIN Binomial-Verteilung

B-ISDN Broadband Integrated Services Digital Network

CAC Connection Admission Control

CBR. Constant Bit Rate

CCITT Comité Consultatif International Téléphonique et Télégraphique

CDV Cell Delay Variation
CLP Cell Loss Priority

CMOS Complementary Metal Oxide Silicon

CS Convergence Sublayer

D Deterministische Verteilung
DCT Discrete Cosine Transform

DF Departure First

DPCM Differentielle Pulscodemodulation

EF Erzeugende Funktion

EFCI Explicit Forward Congestion Indication

 $\mathbf{E}_k$  Erlang-k-Verteilung FIFD First In First Drop FIFO First In First Out

FRM Fast Resource Management

G General (allgemeiner Zufallsprozeß)

GEO<sub>1</sub> Um 1 Stelle verschobene geometrische Verteilung

GFC Generic Flow Control

GI General Independent (allgemeine und unabhängige Verteilung)

GLOBECOM Global Telecommunications Conference
GMDP Generally Modulated Deterministic Process

H High

HDTV High Definition Television
HEC Header Error Control

H<sub>2</sub> Hyperexponentielle Verteilung 2. Ordnung

IBCN&S International Conference on Integrated Broadband Communication

Networks and Services

ICC International Conference on Communications

ICCC International Conference on Computer Communications

IEEE Institute of Electrical and Electronics Engineers

IEICE The Institute of Electronics, Information and Communication

Engineers

IFIP International Federation for Information Processing

IND Institut für Nachrichtenvermittlung und Datenverarbeitung

INFOCOM The Conference on Computer Communications

ISDN Integrated Services Digital Network

ISO International Organization for Standardization

ISS International Switching Symposium ITC International Teletraffic Congress

ITU International Telecommunication Union

JSAC Journal on Selected Areas in Communications

L Low

LIFD Last In First Drop
LT Laplace-Transformierte

M Markoffsche (negativ-exponentielle) Verteilung

MAP Markovian Arrival Process

MMBP Markoff-Modulierter Bernoulli-Prozeß
MMPP Markoff-Modulierter Poisson-Prozeß
MPEG Motion Pictures Expert Group

NNI Network-Node Interface

NPC Network Parameter Control

NRM Network Resource Management

NTG Nachrichtentechnische Gesellschaft (heutige Informationstechnische

Gesellschaft)

OAM Operation and Maintenance
OSI Open Systems Interconnection
PASTA Poisson Arrivals See Time Averages

PCM Pulscodemodulation PHY Physical Layer

plaNET Packetized Lightweight Architecture Protocol

PM Physical Medium (Sublayer)

POIS Poisson-Verteilung
PT Payload Type

RACE Research and Development in Advanced Communications

Technology in Europe

RAM Random Access Memory

S-AAL Signalling ATM Adaptation Layer

SAR	Segmentation and Reassembly (Sublayer)
SDH	Synchronous Digital Hierarchy
SDL	Specification and Description Language
SMDS	Switched Multi-megabit Data Service
SREJ	Selective Reject
TC	Transmission Convergence (Sublayer)
TSS	Telecommunication Standardization Sector
UAS	Uniform Arrival and Service
UNI	User Network Interface
UPC	Usage Parameter Control
VBR	Variable Bit Rate
VC	Virtual Channel
VCC	Virtual Channel Connection
VCI	Virtual Channel Identifier
VP	Virtual Path
VPC	Virtual Path Connection
VPI	Virtual Path Identifier

## Formelzeichen

Die typische Verwendung der Formelzeichen wird für die Zellebene dargestellt. Für die Büschel- und Verbindungsebene wird der Index Z durch die Indizes B bzw. V ersetzt. Zusätzliche Indizes k und n stehen für die Verkehrsklasse k bzw. die Verbindung n. Ebenso zeigen die Indizes H und L die Zugehörigkeit zur hohen bzw. niederen Prioritätsklasse an.

$A_Z$	Verkehrsangebot
$\widehat{A}_{Z}$	Maximal zulässiges Verkehrsangebot
$B_Z$	Verlustwahrscheinlichkeit
$\widehat{B}_{Z}$	Maximal zulässige Verlustwahrscheinlichkeit
$B_Z(l)$	Verlustwahrscheinlichkeit im Quellzustand $l$
$c_Z(l)$	Variationskoeffizient der Verweildauer im Quellzustand $l$
$c_{Z,A}$	Variationskoeffizient des Ankunftsabstands
$c_{Z,H}$	Variationskoeffizient der Bediendauer
$c_{Z,S}$	Variationskoeffizient der Ruhedauer
$f_Z(t,l)$	Verteilungsdichtefunktion der Verweildauer im Quellzustand $l$
$f_{Z,A}(t)$	Verteilungsdichtefunktion des Ankunftsabstands
$f_{Z,H}(t)$	Verteilungsdichtefunktion der Bediendauer
$f_{Z,S}(t)$	Verteilungsdichtefunktion der Ruhedauer

$T_Z(l)$	Zufallsvariable für die Verweildauer im Quellzustand $l$
$T_{Z,A}$	Zufallsvariable für den Ankunftsabstand
$T_{Z,H}$	Zufallsvariable für die Bediendauer
$T_{Z,S}$	Zufallsvariable für die Ruhedauer
$t_Z(l)$	Mittlere Verweildauer im Quellzustand $l$
$t_{Z,A}$	Mittlerer Ankunftsabstand
$t_{Z,H}$	Mittlere Bediendauer
$t_{Z,S}$	Mittlere Ruhedauer
$w_Z$	Mittlere Wartezeit aller angenommenen Anforderungen
$Y_Z$	Verkehrswert
$\lambda_Z$	Mittlere Ankunftsrate
$\Omega_Z$	Mittlere Warteschlangenlänge

#### Weitere Formelzeichen:

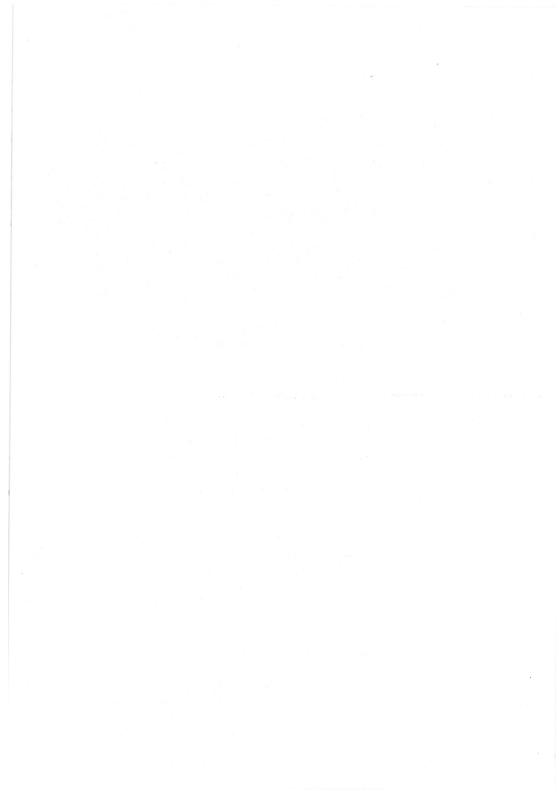
Zusätzliche Indizes k, n und j stehen für die Verkehrsklasse k, die Verbindung n und die Bedienstation j.

$A(t_i)$	Zufallsvariable für die Anzahl der Ankünfte zum Zeitpunkt $t_i$
$a_i$	Anzahl der Ankünfte zum Zeitpunkt $t_i$
$A_t$	Zufallsvariable für die Anzahl der Ankünfte innerhalb von $t$
	Zeitschlitzen
$a_t$	Anzahl der Ankünfte innerhalb von t Zeitschlitzen
b	Mittlere Anzahl von Zellen pro Sendephase
$B_L$	Verlustwahrscheinlichkeit innerhalb einer Verlustphase
$B_{\mathrm{M/D/1/(S+1)}}$	Verlustwahrscheinlichkeit des M/D/1/(S+1)-Systems
C	Bedienrate
$C_E$	Effektive Bedienrate
$C_k$	Effektive Gesamtbedienrate der Verkehrsklasse k
$C_{R,k}$	Freigegebene effektive Bedienrate für einen Verbindungsabbau aus
	Verkehrsklasse $k$
$C_{S,k}$	Zugewiesene effektive Bedienrate für einen Verbindungsaufbau aus
	Verkehrsklasse $k$
$C_W$	Effektive Gesamtbedienrate der anspruchsvollsten Verkehrsklasse
$c_X$	Variationskoeffizient der Zufallsvariablen X
$C_I,C_{II}$	Effektive Bedienraten des "plaNET"-Algorithmus
d	Minimaler, auf die Bedienzeit normierter Zellabstand einer Quelle
d(x, y, a)	Verbundwahrscheinlichkeit, daß eine niederprioritäre Zelle auf den
	Warteplatz x trifft und bis zur Warteposition y vorrücken kann,
	während $a$ hochprioritäre Zellen ankommen

$\mathrm{E}[X]$	Mittelwert oder erstes gewöhnliches Moment
$E[X^2]$	Zweites gewöhnliches Moment
F(l,x)	Verbundwahrscheinlichkeit für den Quellzustand $l$ und einen
	Pufferfüllstand kleiner gleich x
F(x)	Verteilungsfunktion
f(x)	Verteilungsdichtefunktion
$f_A(x,u)$	Verbundwahrscheinlichkeit für den Systemzustand $x$ und das
	Bedienalter $[u, u+du)$
$F_R(r)$	Verteilungsfunktion der Ankunftsrate
$f_R(r)$	Verteilungsdichtefunktion der Ankunftsrate
$f_R(x,t)$	Verbundwahrscheinlichkeit für den Systemzustand $x$ und die
	Restbedienzeit $[t, t+dt)$
G(z)	Erzeugende Funktion
$G_R(z)$	Erzeugende Funktion der Ankunftsrate
i	Ganzzahlige Zählvariable
J	Anzahl der Bedienstationen
j	Ganzzahlige Zählvariable bzw. Nummer der Bedienstation
K	Anzahl der Verkehrsklassen
k	Nummer der Verkehrsklasse bzw. Ordnung der Erlang- $k$ -Verteilung
L	Anzahl der Quellzustände
l	Nummer des Quellzustands
$l_A$	Anzahl der an einer Bedienstation hinzukommenden aktiven
-	Verbindungen
$l_D$	Anzahl der nach einer Bedienstation abzweigenden aktiven
	Verbindungen
m	Nummer des Quellzustands
N	Anzahl von Verbindungen
n	Nummer der Verbindung
$\widehat{N}$	Zulässige Anzahl von Verbindungen
$N_A$	Anzahl der an einer Bedienstation hinzukommenden Verbindungen
$N_D$	Anzahl der nach einer Bedienstation abzweigenden Verbindungen
P = (p(1), p(2),)	Vektor der Zustandswahrscheinlichkeiten
p	Wahrscheinlichkeit
P(X = x)	Wahrscheinlichkeit, daß die Zufallsvariable $X$ den Wert $x$ annimmt
p(x)	Wahrscheinlichkeit für den Zustand x
$p_A(a_i,, a_{i-d+1})$	Verbundwahrscheinlichkeit zur Beschreibung der Zellankünfte
$p_B(l)$	Wahrscheinlichkeit, daß sich $l$ Quellen im aktiven Zustand befinden
$p_C$	Überlastwahrscheinlichkeit
$p_{D,B}(x l)$	Auf den Quellzustand $l$ bedingte Durchlaufzeitverteilung des

	"Fluid-Flow"-Modells
$p_{D,j,J}(x l_j)$	Auf den Quellzustand $l_j$ bedingte Durchlaufzeitverteilung
	durch die Bedienstationen $jJ$
$p_{D,M/D/1/(S+1)}(x)$	Durchlaufzeitverteilung des $M/D/1/(S+1)$ -Systems
$p_{D,Z}(x l)$	Auf den Quellzustand l bedingte Durchlaufzeitverteilung einer
	Bedienstation
$p_{D,1,J}(x)$	Durchlaufzeitverteilung durch alle Bedienstationen
$p_G(x_i, a_i,, a_{i-d+1})$	Verbundwahrscheinlichkeit zur Beschreibung der Zellankünfte und
	des Systemzustands
$p_L$	Blockierungswahrscheinlichkeit
$p_L(l_j l_{j-1})$	Bedingte Wahrscheinlichkeit für $l_j$ aktive Verbindungen an
	Bedienstation $j$ , falls $l_{j-1}$ Verbindungen an der vorhergehenden
	Bedienstation $j-1$ aktiv sind
$p_R(i)$	Verteilung der Ankunftsrate
$p_{R,-n}(i)$	Verteilung der Ankunftsrate ohne Verbindung n
$p_R^*(i)$	Transformierte Verteilung der Ankunftsrate
$p_Z(x)$	Wahrscheinlichkeit, daß sich zu einem beliebigen Zeitpunkt x
	Zellen im System befinden
Q = [q(y x)]	Generator (Übergangsmatrix)
q(y x)	Übergangswahrscheinlichkeit vom Zustand $x$ in den Zustand $y$
$q_A(a_i a_{i-d},N-A_i)$	Wahrscheinlichkeit für $a_i$ Ankünfte zum Zeitpunkt $t_i$ falls $a_{i-d}$
	Ankünfte zum Zeitpunkt $t_{i-d}$ und $A_i$ Ankünfte im Zeitintervall
	$[t_{i-d}, t_{i-1}]$
$q_1(a)$	Wahrscheinlichkeit, daß a Zellen während einer Bedienzeit
	eintreffen
$q_2(a)$	Wahrscheinlichkeit, daß a hochprioritäre Zellen während einer
	Bedienzeit eintreffen
$q_3(a_1,a)$	Wahrscheinlichkeit, daß a Zellen während einer Bedienzeit
	eintreffen, falls die Ankunftsrate nach a1 eingetroffenen Zellen
	auf die Ankunftsrate der hohen Priorität reduziert wird
R	Zufallsvariable für die Ankunftsrate
r	Ankunftsrate
$\hat{r}$	Maximale Ankunftsrate
$\overline{r}$	Mittlere Ankunftsrate
$\overline{r}^*$	Mittelwert der transformierten Verteilung $p_R^*(i)$
r(l)	Ankunftsrate im Quellzustand $l$
S	Anzahl der Warteplätze
T	Zeitabstand zweier Parameteranpassungen
t	Zeit

$T_I$	Theoretischer Zellabstand
$t_i$	Beobachtungszeitpunkte
$T_L$	Zufallsvariable für die Verlustdauer
$t_L$	Mittlere Verlustdauer
$T_N$	Zufallsvariable für die verlustfreie Zeit
$t_N$	Mittlere verlustfreie Zeit
$T_{\mathcal{O}}$	Zufallsvariable für die Überlastdauer
$t_{O}$	Mittlere Überlastdauer
$T_U$	Zufallsvariable für die Unterlastdauer
$t_U$	Mittlere Unterlastdauer
VAR[X]	Varianz
$V_k$	Menge aller Verbindungszustände, in denen Verbindungen aus
	Verkehrsklasse k abgewiesen werden
X	Zufallsvariable
$\boldsymbol{x}$	Allgemeine Variable
[x]	Kleinste ganze Zahl größer oder gleich x
	Größte ganze Zahl kleiner oder gleich $x$
$\overline{x}$	Mittelwert
$X(t_i)$	Zufallsvariable für den Systemzustand zum Zeitpunkt $t_i$
$x_i$	Systemzustand zum Zeitpunkt $t_i$
Z	Büschelfaktor
α	Übergangsrate vom Ruhezustand in den aktiven Zustand
β	Übergangsrate vom aktiven Zustand in den Ruhezustand
$\Delta C_E$	Differenz zwischen den durch die erste und zweite Stufe
_	berechneten effektiven Bedienraten
$\Delta r$	Diskretisierungseinheit der Ankunftsrate
$\Delta t$	Zellübertragungsdauer
$\delta(x)$	Dirac-Distribution
$\Phi(s)$	Laplace-Transformierte
$\lambda(m l)$	Übergangsrate vom Quellzustand $l$ in den Quellzustand $m$
$\nu, \nu'$	Zählerstände des generischen Zellratenalgorithmus
Ω	Menge aller zulässigen Verbindungszustände
$\pi_Z = (\pi_Z(0), \pi_Z(1),)$	Vektor der Zustandswahrscheinlichkeiten zum Bedienende
$\pi_Z(x)$	Wahrscheinlichkeit für den Zustand x zum Bedienende
$\sigma$	Standardabweichung
$\sigma_R$	Standardabweichung der Ankunftsrate
$\sigma_R^*$	Standardabweichung der transformierten Verteilung $p_R^*(i)$
$\sigma_X$	Standardabweichung der Zufallsvariablen $X$
au	${\bf Maximale\ Abweichung\ vom\ theoretischen\ Ankunftszeitpunkt}$



## Kapitel 1

## **Einleitung**

## 1.1 Diensteintegrierende Breitband-Kommunikation

Das gesellschaftliche und industrielle Umfeld wird in zunehmendem Maße durch die Informationstechnik, d.h. die Verknüpfung von Informationsverarbeitung und Informationsübermittlung, geprägt. In diesem Zusammenhang wird auch immer häufiger das Schlagwort vom Produktionsfaktor Information gebraucht, dem für die Zukunft eine wachsende Bedeutung beigemessen wird. Ferner führt die Ausweitung des Dienstleistungssektors und die zunehmende internationale Verflechtung im privaten wie auch im geschäftlichen Bereich zu einem deutlichen Anwachsen der öffentlichen Telekommunikation. Neben dieser quantitativen Zunahme zeichnen sich auch qualitative Verbesserungen durch die Einführung neuer Kommunikationsdienste ab. Einerseits bedingt die Zunahme der Prozessorgeschwindigkeit und der Speichergröße von Rechnern ein exponentielles Wachstum der Übertragungsgeschwindigkeiten [197], andererseits werden Multimedia-Anwendungen, die verschiedene Kommunikationselemente wie z.B. Sprache, Bilder und Daten integrieren, immer wichtiger.

Nachdem der rasche Fortschritt der Mikro- und Optoelektronik die technischen Möglichkeiten geschaffen hat, war die Entwicklung eines diensteintegrierenden Digitalnetzes (Integrated Services Digital Network, ISDN) und dessen Fortführung zum Breitband-ISDN die logische Konsequenz dieser sich seit längerer Zeit abzeichnenden Tendenzen. Dieses Netz soll eine universelle Kommunikations-Infrastruktur bieten, auf der alle heutigen und zukünftigen Kommunikationsdienste und -anwendungen aufgesetzt werden können. Die dadurch entstehende Vielfalt der Dienste und Anwendungen führt zu stark streuenden Anforderungen in Bezug auf die benötigte Übertragungsgeschwindigkeit und die einzuhaltende Dienstgüte. Daher kann die Zukunftssicherheit und Wirtschaftlichkeit eines solchen weltumspannenden Kommunikationsnetzes nur dann gewährleistet werden, wenn die eigentliche Informationsvermittlung durch ein einheitliches, flexibles und schnelles Vermittlungsverfahren abgewickelt wird.

Da weder die Durchschalte- noch die herkömmliche Paketvermittlungstechnik diesen Anforderungen gerecht wird, wurde für das Breitband-ISDN ein neues Übertragungs-, Multiplexund Vermittlungsprinzip, der Asynchrone Transfer-Modus (ATM), entwickelt. Dieses Konzept basiert auf der asynchronen Zeitmultiplextechnik, wobei Nutz- und Signalisierinformationen in Pakete konstanter Länge (sog. Zellen) zerlegt und mittels virtueller Verbindungen übermittelt werden. Die Vermittlung dieser Zellen kann in einheitlichen Koppelnetzen erfolgen, welche die an einer Eingangsleitung eintreffenden Zellen auf freie Zeitschlitze der entsprechenden Ausgangsleitung umsetzen. Die asynchrone Belegung der Leitungen führt zu Zugriffskonflikten, die durch eine Zwischenpufferung von Zellen aufgelöst werden können.

Demgegenüber ermöglicht die asynchrone Übermittlung der Information eine äußerst flexible Zuteilung der Übertragungskapazität an die einzelnen Verbindungen. Im Gegensatz zur Durchschaltevermittlungstechnik können daher beliebige Bitraten bis hin zur Übertragungsbitrate mittels einheitlicher Koppelnetzstrukturen unterstützt werden. Ferner kann für Dienste mit zeitlich schwankendem Bitratenbedarf ein statistischer Ausgleich zwischen den Bitratenanforderungen einzelner Verbindungen stattfinden. Damit ist eine kostengünstige Nutzung des Netzes möglich, da die innerhalb des Netzes reservierte Bitrate unter der Spitzenbitrate der Verbindung liegen kann.

Die mit diesem Prinzip des statistischen Multiplexens verknüpften dynamischen Netzbelastungen können jedoch zu Überlastsituationen führen, die sich in Zellverlusten bzw. unzulässig großen Zellverzögerungen äußern. Für die Akzeptanz des B-ISDN spielt neben den Kosten die Qualität der Informationsübermittlung eine wichtige Rolle. Die Definition adäquater Verkehrssteuerungsmechanismen, die zwischen diesen gegenläufigen Interessen vermitteln, ist daher Gegenstand intensiver Forschungs-, Entwicklungs- und Standardisierungsbemühungen. Ein erster Entwurf für ein entsprechendes Verkehrsmanagement, das allerdings nur die Zuweisung der Spitzenbitrate vorsieht, wurde vom Telecommunication Standardization Sector (TSS) der International Telecommunication Union (ITU) innerhalb der Empfehlung I.371 [121] verabschiedet. Der Schwerpunkt der vorliegenden Arbeit liegt auf dem weitergehenden Prinzip des statistischen Multiplexens und den damit zusammenhängenden Verkehrssteuerungsverfahren zur Prioritätsbehandlung und Verbindungsannahme.

## 1.2 Übersicht über die Arbeit

Im folgenden Kapitel werden die Grundlagen des Breitband-ISDN beschrieben, soweit sie zum Verständnis dieser Arbeit notwendig sind. Einführend werden die in diesem Netz zusammengefaßten Dienste und deren Anforderungen in Bezug auf die Übertragungsgeschwindigkeit und Dienstgüte diskutiert. Im Anschluß wird auf die wichtigsten Aspekte des Asynchronen Transfer-Modus und dessen Einbettung in die Protokollarchitektur des Breitband-ISDN

eingegangen. Den Hauptteil des Kapitels nimmt die Beschreibung des Verkehrsmanagements ein. Anhand der Zeitkonstanten der verschiedenen Steuerungsfunktionen wird deren Zusammenwirken und die hierarchische Strukturierung des Verkehrsmanagements aufgezeigt. Für die einzelnen Funktionen werden die grundsätzlichen Anforderungen und die wichtigsten Lösungsansätze erläutert.

Das dritte Kapitel gibt eine Einführung in die für die nachfolgenden Untersuchungen benötigten verkehrstheoretischen Modelle und Methoden. Von besonderer Bedeutung ist dabei die wirklichkeitsgetreue Nachbildung der Verkehrsflüsse, was durch hierarchisch strukturierte Quellmodelle geschieht. Für die unterschiedlichen Hierarchiestufen — man spricht von der Zell-, Büschel- und Verbindungsebene — ergeben sich verschiedene Systemmodelle, die einen einzelnen Verbindungsabschnitt bzw. eine ganze Referenzverbindung nachbilden. Abschließend werden die Methoden zur simulativen und analytischen Leistungsuntersuchung erörtert.

Kapitel 4 dient der Klärung der grundlegenden Zusammenhänge zwischen der Netzbelastung und den daraus resultierenden Leistungskenngrößen auf der Zell- und Büschelebene. Ausgehend von einem exakten Lösungsansatz für die Überlagerung verschiedener Verkehrsströme wird eine Zerlegung der Leistungsanalyse motiviert, welche die Abläufe auf der Zell- und Büschelebene mit Hilfe getrennter Teilmodelle erfaßt. Auf der Basis dieser Teilmodelle wird der Einfluß der verschiedenen Verkehrsparameter auf die Leistungskenngrößen eines statistischen Multiplexers sowohl bei homogener als auch bei heterogener Verkehrszusammensetzung systematisch untersucht. Darauf aufbauend wird ein approximativer analytischer Lösungsansatz für eine ganze Referenzverbindung vorgestellt.

Der Vergleich unterschiedlicher Verlustprioritätsmechanismen, die das selektive Verwerfen von Zellen ermöglichen, ist Gegenstand von Kapitel 5. Zunächst werden die Einsatzmöglichkeiten und die daraus erwachsenden Anforderungen an dieses neuartige Prioritätskonzept behandelt. Nach einer Klassifizierung der verschiedenen Mechanismen wird die Funktionsweise und Realisierung der wichtigsten Vertreter ausführlich besprochen. Abschließend werden diese Verfahren im Hinblick auf ihre Leistungsfähigkeit verglichen. Dazu werden die aus dem vorhergehenden Kapitel bekannten analytischen Lösungsansätze entsprechend erweitert.

Kapitel 6 beschäftigt sich mit der Verbindungsannahmefunktion, die — abhängig von der momentanen Netzlast — die Entscheidung über die Annahme bzw. Abweisung eines neuen Verbindungswunschs trifft. Zu Beginn des Kapitels werden die Anforderungen an diese Verkehrssteuerungsfunktion erläutert und entsprechende Vergleichskriterien für die im Anschluß beschriebenen Verbindungsannahmealgorithmen abgeleitet. Die aus dem Vergleich gezogenen Schlußfolgerungen dienen dem Entwurf eines neuen, verbesserten Verbindungsannahmeverfahrens. Ferner wird auf die Problematik der Verkehrsparameterfestlegung eingegangen.

Das abschließende Kapitel faßt die in dieser Arbeit gewonnenen Ergebnisse nochmals zusammen und gibt einen Ausblick auf zukünftige Forschungsschwerpunkte.

## Kapitel 2

# Dienste, Architektur und Protokolle des Breitband-ISDN

Im folgenden werden die für die vorliegende Arbeit wichtigen Aspekte des Breitband-ISDN und des Asynchronen Transfer-Modus erörtert. Eine weitergehende Einführung kann z.B. [87] entnommen werden.

## 2.1 Charakterisierung zukünftiger Dienste

Ein zukünftiges universelles Breitband-Kommunikationsnetz muß auf die Integration einer breiten Palette von Kommunikationsdiensten ausgerichtet werden. Dieses Netz soll sowohl die geschäftliche als auch die private Nutzung gestatten, soll Punkt-Punkt, Punkt-Mehrpunkt und "Broadcast"-Verbindungen unterstützen, soll Texte, Graphiken, Daten, Sprache, Festund Bewegtbilder übertragen können und soll sowohl für konstante als auch zeitlich variable Bitraten geeignet sein. Ferner muß die Einbindung von Multimedia-, Mobilkommunikationsund Mehrwertdiensten möglich sein. Durch das weitgespannte Spektrum der hieraus erwachsenden Anforderungen ergeben sich verschiedene Randbedingungen für ein diensteintegrierendes Breitband-Kommunikationsnetz, die im folgenden erörtert werden sollen.

#### 2.1.1 Klassifizierung der Dienste

Kommunikationsdienste unterscheiden sich in vielfacher Hinsicht und können deshalb nach verschiedenen Gesichtspunkten gegliedert werden. Da für die vorliegende Arbeit insbesondere die Netzaspekte von Bedeutung sind, wird im folgenden die in der ITU-T-Empfehlung I.211 [113] vorgeschlagene Klassifizierung der Dienste aus Netzsicht kurz vorgestellt. Darin werden

zwei fundamentale Kategorien von Breitband-Diensten unterschieden — interaktive Dienste und Verteildienste (vgl. Bild 2.1).

Die erste Kategorie bezieht sich auf den Informationsaustausch zwischen zwei oder mehreren Partnern und kann in drei Klassen unterteilt werden:

- Dialogorientierte Dienste sind durch einen bidirektionalen, dialogorientierten Informationsaustausch ohne Zwischenspeicherung von Information gekennzeichnet.
- Nachrichtenorientierte Dienste nehmen zusätzlich eine Zwischenspeicherung bzw. Zwischenbearbeitung der Information vor.
- Dienste für den Zugriff auf gespeicherte Informationen erlauben einen individuellen Zugriff auf Informationen, die in Datenbanken gespeichert sind.

Die zweite Kategorie umfaßt alle Dienste, die Informationen von einer zentralen Quelle an verschiedene Benutzer verteilen. Je nachdem, ob der Benutzer Einfluß auf den Beginn und die Reihenfolge der dargebotenen Informationen hat, werden Dienste mit individueller Präsentationskontrolle und Dienste ohne individuelle Präsentationskontrolle unterschieden.

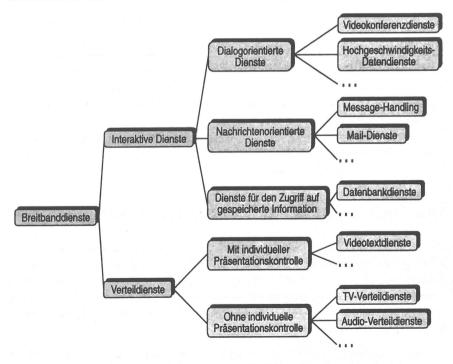


Bild 2.1: Klassifikation von Breitband-Diensten

Für eine genauere Charakterisierung von Dienst- und Netzaspekten diensteintegrierender Digitalnetze wurde die Methodik der Dienst- und Netzattribute entwickelt [111]. Aus den ITU-T-Empfehlungen I.140 [111] und I.211 [113] können die in Tabelle 2.1 aufgeführten Beispiele für Dienstattribute und Attributwerte entnommen werden, die im weiteren Verlauf der Arbeit von Bedeutung sind. Das Hauptaugenmerk liegt auf den für die einzelnen Dienste einzuhaltenden Vorgaben im Hinblick auf die Bitrate (Transferrate) und die Qualität der Übermittlung (Dienstgüte).

Attribut	Attributwerte			
Informationsart	Sprache, Video, Daten,			
Kommunikationsauf- bzwabbau	Permanente Kommunikationsbeziehung, Vorreservierung der Kommunikation, Kommunikation nach Bedarf,			
Transferrate	Maximale Bitrate, Durchsatz,			
Dienstgüte	Mittlere Verzögerung, Bitfehlerrate,			

Tabelle 2.1: Beispiele für Dienstattribute und Attributwerte

Diesen Dienstattributen, die den Dienst aus der Sicht des Benutzers charakterisieren, lassen sich Netzattribute zuordnen, welche die Fähigkeiten des Kommunikationsnetzes beschreiben. Insbesondere wird eine Netzgüte als Gegenstück zur Dienstgüte definiert [104], welche die Leistungsfähigkeit des Kommunikationsnetzes beschreibt. Diese muß im Einklang mit den Anforderungen der abgewickelten Dienste stehen, d.h. das Netz muß eine definierte Netzgüte aufweisen, damit die Dienstgütespezifikationen aller darauf aufbauenden Kommunikationsdienste erfüllt werden können. Die Sicherstellung dieser Netzgüte ist die zentrale Aufgabe der Verkehrssteuerung (vgl. Abschnitt 2.4.2).

#### 2.1.2 Verkehrscharakteristik ausgewählter Dienste

Zur Festlegung und Untersuchung der Verkehrssteuerungsverfahren ist eine genaue Kenntnis der Rahmenbedingungen, d.h. der Verkehrscharakteristik und der Dienstgütespezifikationen der unterstützten Dienste unerläßlich. Die Definition relevanter Verkehrs- und Dienstgüteparameter und die Festlegung geeigneter Parameterwerte ist jedoch ein aktuelles Forschungsgebiet, weshalb die in den folgenden Abschnitten diskutierten Werte nur als Anhaltspunkte zu sehen sind.

Tabelle 2.2 zeigt eine Zusammenstellung der Verkehrscharakteristika verschiedener Dienste. Grundsätzlich lassen sich zwei verschiedene Kategorien unterscheiden: Dienste mit zeitlich

Art der Infor- mation	Codierung bzw. Dienst	Max. Bitrate [Mbit/s]	Min. Bitrate [Mbit/s]	Büschel- faktor	Stan- dardabw. [Mbit/s]	Büschel- dauer [ms]	Quelle
Sprache	PCM [106] <sup>c</sup>	0,064	0,064	1	0	-	[1]
	ADPCM [107] mit Pausen- unterdr.	0,032	0	2,8	0,0153	352	[225]
Bilder	Festbild- übertragung <sup>c</sup>	2	0	23	0,4	500	[38]
Video	CBR-Cod.	2, 34, 140	2, 34, 140	1	0	_	[74]
	Conditional replenishment <sup>c</sup>	10,6	0,6	2,7	1,7	33,3	[164]
	DCT, DPCM, Huffman-Cod.	$5,1 - 27,3^a$	$0,1 - 3,3^a$	1,5 - 5,8°	$0,5 - 4,7^a$	40	[90]
	MPEG	$2,4 - 6,8^{b}$	$0,1 - 0,8^{b}$	$2,0-3,3^{b}$	$0,4 - 1,3^b$	33,3	[190]
Daten	Hochgeschw Datenkomm.	10	0	10	3	13	[39]
	Verbindungs- orientierte Datenkomm.	2, 10, 34, 140	0	1 – 50	keine Angaben	keine Angaben	[6]
	Verbindungs- lose Datenkomm.	2, 10, 34, 140	0	groß	keine Angaben	keine Angaben	[6]

a verschiedene Videoszenen und -quellen

Tabelle 2.2: Verkehrscharakteristik ausgewählter Dienste

**PCM** 

konstanter Bitrate (Constant Bit Rate, CBR) und Dienste mit zeitlich variabler Bitrate (Variable Bit Rate, VBR). Dienste mit variabler Bitrate können weiter unterteilt werden in VBR-Dienste mit diskreten Bitratenwerten und VBR-Dienste mit kontinuierlichen Bitratenwerten. Die erste Gruppe schwankt zwischen einer endlichen Anzahl von Zuständen, wo-

b verschiedene Codierparameter

<sup>&</sup>lt;sup>c</sup> für spätere Untersuchungen verwendet

<sup>(</sup>A)DPCM (Adaptive) Differentielle PCM

Constant Bit Rate CBR

Discrete Cosine Transform DCT Motion Pictures Expert Group **MPEG** Pulscodemodulation

bei jedem Zustand eine feste Bitrate zugeordnet ist. Ein Spezialfall dieser Klasse ist die sporadische Quelle, die zwischen einem Ruhezustand und einem Sendezustand mit fester Spitzenbitrate alterniert. Dieses Verhalten ist charakteristisch für Datendienste. Dagegen führen Kompressionsverfahren zur Redundanz- und Irrelevanzreduktion von Videodaten zu kontinuierlichen Bitratenwerten.

Die Bitratenschwankungen von VBR-Diensten können durch die Extremwerte und die Standardabweichung der Bitrate charakterisiert werden, wogegen CBR-Dienste durch die Spitzenbitrate vollständig beschrieben sind. Ein weiteres wichtiges Maß für die Verkehrsschwankungen repräsentiert der Büschelfaktor, für den eine Vielzahl von Definitionen vorgeschlagen wurde. Im folgenden soll der Büschelfaktor das Verhältnis aus maximaler und mittlerer Bitrate bezeichnen. Das zeitliche Verhalten von VBR-Diensten wird durch die mittlere Büscheldauer beschrieben. Diese Meßgröße entspricht entweder der Dauer einer Sendephase (für sporadischen Quellverkehr) oder der Bilddauer (für Videodienste).

Abhängig von der Codierung der Sprachinformation — Pulscodemodulation (PCM) bzw. adaptive differentielle Pulscodemodulation (ADPCM) — muß für Sprachübermittlung eine maximale Bitrate von 64 bzw. 32 kbit/s zur Verfügung stehen. Die Bitratenanforderungen weiterer Sprachcodierverfahren können aus [1] entnommen werden. Des weiteren treten Sprachpausen zwischen Sätzen, Worten und Silben auf, die bei der Informationsübertragung unterdrückt werden können. Da in diesem Fall der Übertragungskanal nur noch sporadisch belegt wird, führt dies zu einer Reduktion der mittleren Bitrate.

Die Übermittlung von Festbildern führt ebenfalls zu einer sporadischen Belegung des Übertragungskanals für die Dauer eines Bildes. Der anschließende Bearbeitungsvorgang durch den Benutzer dauert vergleichsweise lange, was sich in einer signifikanten Erhöhung des Büschelfaktors äußert. Ferner muß auch die Übertragungsbitrate relativ hoch gewählt werden, um eine zufriedenstellende Antwortzeit zu erhalten.

Für Videodienste ergibt sich ein sehr breites Bitratenspektrum, da die Bitrate sowohl von der Auflösung des Bildes als auch von der darin enthaltenen Redundanz abhängt. Die eingesetzten Codierverfahren reduzieren sowohl die Redundanz innerhalb eines Bildes ("Intra-frame"-Codierverfahren) als auch die Redundanz von aufeinanderfolgenden Bildern ("Inter-frame"-Codierverfahren). Deshalb ist für Bildfernsprechen (niedrige Auflösung, langsame Änderung des Bildinhalts) eine Bitrate von ca. 2 Mbit/s ausreichend, wogegen für hochauflösendes Fernsehen (High Definition Television, HDTV) eine Bitrate von mindestens 140 Mbit/s zur Verfügung stehen sollte.

Die heute gebräuchlichen Codierverfahren zur Redundanz- und Irrelevanzreduktion liefern konstante Bitraten, die sich an der PCM-Kanalhierarchie orientieren (vgl. z.B. ITU-T-Empfehlung H.261 [109]). Da die Redundanz vom Bildinhalt bzw. dessen zeitlicher Änderung abhängt, ergeben sich damit zwangsläufig Schwankungen der Bildqualität. Moderne Codierverfahren zielen demgegenüber auf eine gleichbleibende Bildqualität ab, was zu variablem

Bitratenaufkommen führt. In Tabelle 2.2 sind verschiedene Meßergebnisse für experimentelle VBR-Codierverfahren für Fernsehqualität gegenübergestellt. Diese Resultate und weitere Ergebnisse aus der Literatur bestätigen, daß sowohl das Codierverfahren (vgl. [190]) als auch die jeweilige Videosequenz [90], d.h. die spezielle Anwendung, großen Einfluß auf die Bitratencharakteristik haben. Ferner läßt sich aus der zeitlichen Reihenfolge der Messungen ersehen, daß der technologische Fortschritt zu immer kleineren Bitraten führt.

Für Datendienste ist eine ausgeprägte Abhängigkeit des Bitratenbedarfs von der jeweiligen Anwendung zu beobachten, wie eine Studie heutiger und zukünftiger Datenanwendungen nachweist [158]. Allerdings läßt sich der natürliche Bitratenbedarf von Datendiensten nur sehr schwer abschätzen, da die Charakteristik heutiger Datenanwendungen hauptsächlich durch die Endgeräte, Netze und Protokolle bestimmt ist. Die aus der Literatur bekannten Messungen, die sich auf die Paketgrößen und -abstände in lokalen Netzen beschränken, lassen deshalb nur sehr eingeschränkte Aussagen für die in Tabelle 2.2 aufgeführten Verkehrsparameter zu. Der extrem hohe Büschelfaktor von Datendiensten, d.h. die äußerst sporadische Netzbelastung, stellt einen weiteren Unsicherheitsfaktor dar. In Tabelle 2.2 sind daher nur einige exemplarische Schätzwerte, aufgeschlüsselt nach Hochgeschwindigkeits-, verbindungsorientierter und verbindungsloser Datenkommunikation, wiedergegeben.

Für die Wahl der Spitzenbitrate in zukünftigen Vermittlungssystemen werden wirtschaftliche und qualitative Erwägungen eine Rolle spielen. Eine höhere Spitzenbitrate führt zwar zu kürzeren Übermittlungszeiten, zieht allerdings höhere Übermittlungskosten nach sich. Die typische Büscheldauer wird deshalb durch die maximal zulässige Übermittlungszeit bestimmt werden. Da der technologische Fortschritt zu immer größeren Speicherbausteinen und immer schnelleren Prozessoren führt, wird das Datenvolumen und damit die Spitzenbitrate in Zukunft ansteigen.

### 2.1.3 Dienstgüteanforderungen ausgewählter Dienste

Mit der Deregulierung des Fernmeldewesens wird die Dienstgüte eine zunehmende Bedeutung erlangen, da sie einen wichtigen Wettbewerbsfaktor darstellt. Im Hinblick auf die Dimensionierung und Steuerung des Kommunikationsnetzes sind insbesondere jene Dienstgüteparameter von Interesse, die sich auf die Verzögerung bzw. den Verlust von Informationen beziehen. Die entsprechenden Parametervorgaben können dann, abhängig vom betrachteten Vermittlungskonzept, in konkrete Leistungskenngrößen für die Netzgüte umgesetzt werden.

Die maximal zulässige Verzögerung ist ein äußerst wichtiger Dienstgüteparameter, der insbesondere für Fernsprechen und Bildfernsprechen wichtig ist, da diese Dienste engen zeitlichen Beschränkungen unterliegen und deshalb häufig als Echtzeitdienste bezeichnet werden. Für Fernsprechen wurde innerhalb der ITU-T-Empfehlung G.114 eine maximale Verzögerung

von 50 ms zwischen Sender und Empfänger vorgegeben [105]. Falls eine Echokompensation angewandt wird, erhöht sich dieser Wert auf 400 ms. Alle anderen Dienste haben geringere zeitliche Anforderungen [141, 173].

Ein zweites wichtiges Qualitätsmerkmal stellt die Bitfehlerrate einer Verbindung dar. In Tabelle 2.3 sind einige Beispiele für die zulässigen langzeitigen Bitfehlerraten bei zufällig auftretenden Bitfehlern aufgeführt. Die angegebenen Werte zeigen, daß die Verständlichkeit von Sprache auch bei relativ hohen Bitfehlerraten gewährleistet ist. Ähnliche Aussagen gelten auch bezüglich der Zellverlustwahrscheinlichkeit in ATM-Netzen [173]. Für das diensteintegrierende Digitalnetz (ISDN) wurden vom ITU-TSS weitergehende Vorgaben bezüglich des zeitlichen Verhaltens der Bitfehlerrate gemacht [108]:

- Weniger als 10% der Intervalle von einer Minute Dauer haben eine Bitfehlerrate größer als 10<sup>-6</sup>.
- Weniger als 0,2% der Intervalle von einer Sekunde Dauer haben eine Bitfehlerrate größer als 10<sup>-3</sup>.
- In weniger als 8% der Intervalle von einer Sekunde Dauer treten Bitfehler auf.

D.h. neben der langzeitigen Bitfehlerrate ist auch die zeitliche Korrelation der Bitfehler von Bedeutung. Für die subjektive Beurteilung der Dienstgüte ist insbesondere der zeitliche Abstand aufeinanderfolgender Störungen entscheidend. Die mittlere Dauer und die Stärke dieser Störungen sind weitere wichtige Kenngrößen der Dienstgüte [254].

Breitbanddienste reagieren empfindlicher auf Bitfehler, da der zeitliche Abstand zweier Bit-

Informationsart	Codierung, Protokoll oder Dienst	Bitfehlerrate	Quelle
Spende	Pulscodemodulation (PCM), 64 kbit/s [106]	$3 \cdot 10^{-5}$	[254]
Sprache	Adaptive differentielle PCM, 32 kbit/s [107]	10-4	[254]
Video	Pulscodemodulation	$5 \cdot 10^{-7}$	[254]
	Differentielle Pulscodemodulation, 32 Mbit/s	10-7	[254]
	Inter-frame-Codierung, 1,5 Mbit/s	$10^{-9} - 10^{-10}$	[254]
Daten	Go-back-N	$3 \cdot 10^{-9a}$	[173]
	Selective Reject	$3 \cdot 10^{-6a}$	[173]
	Switched Multi-megabit Data Service	10 <sup>-8b</sup>	[17]

a aus der Zellverlustwahrscheinlichkeit bestimmt

Tabelle 2.3: Zulässige langzeitige Bitfehlerwahrscheinlichkeiten einzelner Dienste

<sup>&</sup>lt;sup>b</sup> aus der Verlustwahrscheinlichkeit für Schicht-3-Protokolldateneinheiten bestimmt

fehler abnimmt, falls die Übertragungsbitrate ansteigt. Videodienste sind daher wesentlich anspruchsvoller als Sprachdienste, was die Fehlerfreiheit der Übermittlung anbelangt. Falls effiziente Codierverfahren zur Redundanz- und Irrelevanzreduktion eingesetzt werden, die auch die Redundanz von aufeinanderfolgenden Bildern einbeziehen ("Inter-frame"-Codierung), wirken sich Bitfehler stärker aus, da der durch ein Übertragungsbit transportierte Informationsgehalt zunimmt. Andererseits können Verfahren, die gezielt Redundanz zur Vorwärts-Fehlerkorrektur einbinden, wesentlich höhere Bitfehlerraten tolerieren [187]. Die Effizienz dieses Konzepts nimmt allerdings ab, falls Fehlerbüschel auftreten.

Für Datendienste ist trotz fehlerbehaftetem Kanal eine fehlerfreie Übermittlung der Information erforderlich. Dies wird durch Übermittlungsprotokolle gewährleistet, die eine Erkennung und Wiederholung gestörter Dateneinheiten vornehmen. Bei Protokollen mit selektiver Wiederholung (Selective Reject, SREJ), wird nur die durch Übermittlungsfehler unbrauchbar gewordene Dateneinheit wiederholt. Die heute gebräuchlichen Protokolle wiederholen jedoch die gesamte Information, die nach der gestörten Dateneinheit bis zur Erkennung des Übermittlungsfehlers gesendet wurde (Go-back-N). Damit der Nutzdatendurchsatz und die Verzögerungszeit innerhalb akzeptabler Grenzen bleiben, ist eine kleinere Bitfehlerrate als bei den SREJ-Protokollen erforderlich.

Für den Hochgeschwindigkeits-Datendienst SMDS (Switched Multi-megabit Data Service) wurden bereits Vorgaben bezüglich der Verlustwahrscheinlichkeit von Schicht-3-Protokolldateneinheiten gemacht [17]. Falls unabhängige Bitfehler vorausgesetzt werden, kann hieraus auf eine Bitfehlerwahrscheinlichkeit von ca.  $10^{-8}$  geschlossen werden. Für Fehlerbüschel erhöht sich dieser Wert, da eine gestörte Dateneinheit unabhängig von der Anzahl der Bitfehler wiederholt werden muß, d.h. für Datendienste wirkt sich eine ungleiche Verteilung der Übermittlungsfehler positiv aus.

## 2.2 Der Asynchrone Transfer-Modus

#### 2.2.1 Vermittlungsprinzip

Der wahlweise Informationstransport zwischen den angeschlossenen Endeinrichtungen stellt die zentrale Aufgabe eines vermittelnden Kommunikationsnetzes dar, wofür zwei grundlegende Vermittlungsverfahren — Durchschalte- und Paketvermittlung — unterschieden werden können. Das heutige Schmalband-ISDN basiert auf der Durchschaltevermittlungstechnik, die einen durchgehenden Kanal fester Bitrate zur ausschließlichen Nutzung durch die zugeordnete Verbindung bereitstellt. Im Hinblick auf die heterogenen und z.T. variablen Bitratenanforderungen der unterstützten Dienste erweist sich diese starre Zuweisung der Netzressourcen als unwirtschaftlich.

In Paketvermittlungsnetzen werden die Nachrichten in einzelne Datenblöcke aufgeteilt, die mit entsprechender Zielinformation zum abschnittsweisen Transport von Vermittlungsknoten zu Vermittlungsknoten versehen werden. Der asynchrone Zugriff der einzelnen Kommunikationsbeziehungen auf die Übertragungseinrichtungen ermöglicht eine flexible und effiziente Betriebsmittelteilung. Die unvermeidlichen Zugriffskonflikte müssen durch die Zwischenspeicherung von Paketen und eine abschnittsweise Fehlerbehebung und Flußkontrolle aufgelöst werden. Die dazu erforderlichen komplexen Protokollmechanismen lassen sich allerdings bei den erwähnten hohen Übertragungsgeschwindigkeiten nicht mehr realisieren.

Dies führte zur Entwicklung von schnellen Paketvermittlungstechniken wie z.B. Fast Packet Switching [152, 237] und Asynchronous Time-Division Switching [73], deren vereinfachte Protokollmechanismen mittels hochintegrierter Halbleiterschaltungen abgewickelt werden können. Die Realisierung erster Prototypen [97, 233, 257] bestätigte die Flexibilität und das Entwicklungspotential dieser Konzepte. Mit der ITU-T-Empfehlung I.121 [110] wurde schließlich der Asynchrone Transfer-Modus als Basis für das zukünftige B-ISDN festgeschrieben. Dieses paketorientierte Vermittlungsverfahren basiert auf Dateneinheiten konstanter Länge, die als Zellen bezeichnet werden. Eine Zelle besteht aus einem Informationsfeld und einem Zellkopf, der eine Kanalkennung und zusätzliche Steuerinformationen enthält (vgl. Abschnitt 2.2.3).

Der Asynchrone Transfer-Modus baut auf der asynchronen Zeitmultiplextechnik auf, bei der die Mehrfachausnutzung eines Übertragungswegs durch eine zeitliche Schachtelung der Informationsströme erreicht wird (vgl. Bild 2.2). Das Übertragungsmedium stellt Zeitschlitze konstanter Dauer zur Verfügung, die genau eine Zelle einer beliebigen Kommunikationsbe-

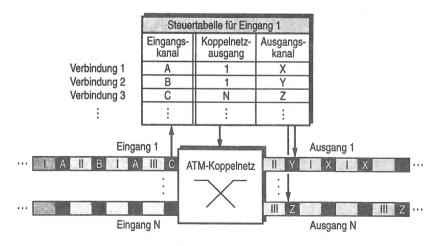


Bild 2.2: Prinzipielle Funktionsweise eines ATM-Koppelnetzes

ziehung aufnehmen können. Da die Zeitschlitze asynchron belegt werden dürfen, kann der Abstand zweier Zellen an den momentanen Bitratenbedarf der Kommunikationsbeziehung angepaßt werden.

Die Vermittlung der Zellen geschieht in einheitlichen Koppelnetzen mittels des Konzepts der virtuellen Verbindung. Dazu werden beim Verbindungsaufbau nur für den jeweiligen Verbindungsabschnitt gültige Kennungen für die logischen Kanäle vereinbart, die einen eindeutigen Weg markieren, dem alle Zellen der betrachteten Verbindung folgen. Diese Kanalkennung wird im Zellkopf mitgeführt und in den Koppelnetzen ausgewertet. Mit Hilfe einer Tabelle, die beim Verbindungsaufbau initialisiert wurde, kann der ankommenden Zelle der entsprechende Koppelnetzausgang zugeordnet und die für den nachfolgenden Verbindungsabschnitt gültige Kanalkennung ins Kopffeld eingetragen werden. Dieses Prinzip ist in Bild 2.2 veranschaulicht.

#### 2.2.2 Konzept der virtuellen Pfade

Der oben beschriebene Ansatz der virtuellen Verbindung wird in ATM-Netzen weiter verfeinert, indem durch sog. virtuelle Pfade ein zusätzliche Hierarchieebene eingeführt wird [114, 188, 217]. Ein virtueller Pfad faßt mehrere virtuelle Kanäle zusammen, die dann als einheitliches Bündel vermittelt, verwaltet und gewartet werden können. Dadurch wird eine flexible Netzarchitektur sowie eine Vereinfachung von Steuerung, Betrieb und Wartung des Netzes erreicht [217].

Dieses Konzept nimmt eine logische Unterteilung des physikalischen Übertragungskanals in einzelne virtuelle Pfade (Virtual Paths, VPs) vor, die wiederum verschiedene virtuelle Kanäle (Virtual Channels, VCs) umfassen können (vgl. Bild 2.3). Die eindeutige Zuordnung einer Zelle zu einem bestimmten virtuellen Pfad bzw. zu einem sich hierin befindlichen virtuellen Kanal wird durch die Aufspaltung der Kanalkennung in zwei Teile, die VP-Kennung (Virtual Path Identifier, VPI) und die VC-Kennung (Virtual Channel Identifier, VCI) erreicht.

Dies legt die in Bild 2.3 gezeigte logische Strukturierung des Vermittlungsnetzes in Bitübertragungsschicht, VP- und VC-Ebene nahe [114, 134]. Eine VC-Verbindung (Virtual Channel Connection, VCC) erstreckt sich zwischen zwei Endpunkten (Endgeräte, Netzkoppeleinheiten, etc.), an denen ein Übergang in die ATM-Anpassungsschicht (vgl. Abschnitt 2.3.4) stattfindet und setzt sich aus einer Kette von VC-Verbindungsabschnitten (Virtual Channel Links, VC-Links) zusammen. Alle Punkte, an denen eine Änderung der VC-Kennung vorgenommen wird, d.h. an denen von der VP- in die VC-Ebene übergegangen wird, markieren den Abschluß eines VC-Verbindungsabschnitts. Dies kann das Einsetzen bzw. Entfernen der VC-Kennung in den Endgeräten oder die Umsetzung der VC-Kennung innerhalb der Vermittlungsknoten sein (in der Regel ist hiermit auch eine Umsetzung der VP-Kennung verknüpft).

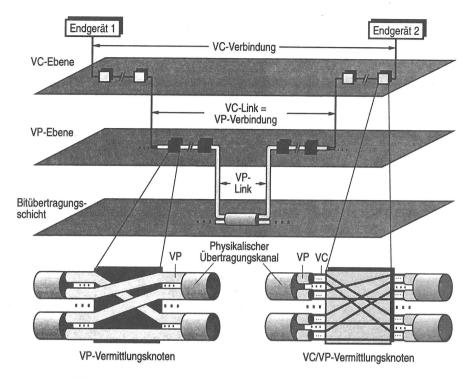


Bild 2.3: Übersicht über virtuelle Pfade und virtuelle Verbindungen in ATM-Netzen

Ein VC-Link nutzt eine VP-Verbindung (Virtual Path Connection, VPC), die wiederum auf einzelnen VP-Verbindungsabschnitten (Virtual Path Links, VP-Links) basiert. Ein VP-Verbindungsabschnitt erstreckt sich zwischen Punkten, an denen die VP-Kennung verändert, d.h. eingesetzt, gelöscht oder umgesetzt wird. Die Funktionalität eines VP-Vermittlungsknotens beschränkt sich auf die Vermittlung der virtuellen Pfade; die virtuellen Kanäle bleiben unverändert. Die eigentliche Informationsübertragung findet in der Bitübertragungsschicht statt. Diese Schicht kann in weitere übertragungstechnische Ebenen untergliedert werden [114], was nicht näher ausgeführt werden soll, da dies keine Bedeutung für die vorliegende Arbeit hat.

VP-Verbindungen können zwischen verschiedenen Benutzern, zwischen Benutzern und VC-bzw. VC/VP-Vermittlungsknoten oder zwischen verschiedenen VC- bzw. VC/VP-Vermittlungsknoten geschaltet werden. Ein permanenter virtueller Pfad wird auf Antrag des Teilnehmers durch den Netzbetreiber eingerichtet. Ferner kann eine VP-Verbindung bei Bedarf mittels entsprechender Signalisier- oder Netzmanagementfunktionen geschaltet werden [112], die allerdings noch nicht spezifiziert sind. Der VP-Verbindungsaufbau kann sowohl vom Teil-

nehmer als auch vom Netzbetreiber initiiert werden. Die an die virtuellen Pfade zugewiesene Übertragungsrate kann konstant oder variabel sein, womit sich deterministische und statistische virtuelle Pfade unterscheiden lassen [217].

Das oben diskutierte Konzept eröffnet eine Reihe von Vorteilen und Anwendungsmöglichkeiten. Insbesondere kann der Aufbau einer VC-Verbindung stark vereinfacht und beschleunigt werden, da nur noch die VC- und VC/VP-Vermittlungsknoten am Aufbau einer VC-Verbindung beteiligt sind, d.h. die hierfür benötigten Signalisier-, Verkehrslenkungs- und Verbindungsannahmeprozeduren und -daten müssen nur in diesen Netzknoten vorhanden sein (vgl. Unterabschnitt 2.4.2.5). Die Zuverlässigkeit und die Auslastung des Netzes läßt sich durch eine flexible Umkonfigurierung von virtuellen Pfaden steigern. Schließlich können virtuelle Pfade bei der Bildung von geschlossenen Benutzergruppen (z.B. in "Virtual Private Networks" oder bei der Kopplung von Lokalen Netzen) bzw. zur gegenseitigen Abgrenzung von Diensten mit stark unterschiedlicher Charakteristik eingesetzt werden.

### 2.2.3 Zellformat

Bei der Wahl eines bestimmten Zellformats müssen verschiedene, sich zum Teil widersprechende Interessen gegeneinander abgewogen werden. Das Informationsfeld sollte möglichst groß gewählt werden, um ein günstiges Verhältnis zwischen Nutzdaten und im Zellkopf mitgeführten Steuerinformationen zu erzielen. Gleichzeitig vereinfacht sich die Realisierung der Koppelelemente, da die zur Bearbeitung des Zellkopfes zur Verfügung stehende Zeit ansteigt. Für ein kurzes Informationsfeld spricht dagegen die geringere Paketier- und Zellübertragungsdauer und die dadurch bedingte Verkleinerung des zur Auflösung von Zellkollisionen bzw. zum Ausgleich von Verzögerungsschwankungen nötigen Pufferspeichers. Die in Bild 2.4 veranschaulichte Festlegung eines 5 Byte großen Kopffeldes und eines 48 Byte großen Informationsfeldes stellt einen Kompromiß zwischen den genannten Gesichtspunkten dar.

Die innerhalb der ITU-T-Empfehlung I.361 [118] definierte Strukturierung des Zellkopfes unterscheidet sich an der Benutzer/Netzschnittstelle (User Network Interface, UNI) und den netzinternen Schnittstellen (Network-Node Interface, NNI) durch das Flußkontrollfeld (Generic Flow Control, GFC). Das entsprechende GFC-Protokoll dient zur Regulierung des Verkehrsflusses vom Teilnehmer zum ATM-Netz und unterstützt die bei gemeinsam genutzten Übertragungsmedien erforderlichen Medienzugriffsprotokolle. Insbesondere sollen Überlastsituationen vermieden werden, damit die vereinbarte Übertragungskapazität und Netzgüte jeder einzelnen Verbindung auch im Teilnehmeranschlußbereich eingehalten werden kann [112]. Allerdings konnte bislang keine Einigung über die genaue Verwendung des GFC-Feldes erzielt werden.

Den größten Teil des Zellkopfes beanspruchen die Felder zur Identifikation des zugehörigen virtuellen Pfades bzw. Kanals. Die Länge der VP-Kennung (VPI) beträgt 8 Bit am UNI und

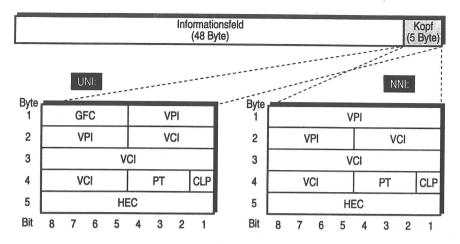


Bild 2.4: ATM-Zellformat an der Benutzer/Netz- und den netzinternen Schnittstellen

wird innerhalb des Netzes auf 12 Bit erweitert. Für die VC-Kennung (VCI) stehen überall 16 Bit zur Verfügung.

Das PT-Feld (Payload Type, PT) dient zur Unterscheidung unterschiedlicher Arten von Zellen. Eine Zusammenstellung der möglichen Codierungen und deren jeweilige Interpretation ist in Tabelle 2.4 zu finden. Dieses Feld dient zur Unterscheidung von

- verschiedenen Arten von Benutzerzellen,
- Zellen für Betrieb und Wartung (Operation and Maintenance, OAM) und
- Zellen für eine schnelle Ressourcenzuteilung (vgl. Unterabschnitt 2.4.2.4).

Die vier Arten von Benutzerzellen ermöglichen sowohl den Austausch einer Benutzerkennung, welche den höheren Protokollschichten zur Verfügung steht, als auch einer Überlastmeldung, die anzeigt, daß mindestens ein Netzknoten zwischen Sender und Empfänger überlastet ist (vgl. auch Unterabschnitt 2.4.3.2).

Ein weiteres Bit innerhalb des Zellkopfes erlaubt die Markierung von Zellen, die bei einer möglichen Überlastsituation bevorzugt verworfen werden sollen. Diese Zellverlustpriorität (Cell Loss Priority, CLP) ermöglicht beispielsweise eine explizite Unterstützung der in Abschnitt 2.1.3 beschriebenen heterogenen Anforderungen der Dienste bezüglich der Zellverlustwahrscheinlichkeit (vgl. Kapitel 5).

Die abschließende, 8 Bit lange Prüfsequenz (Header Error Control, HEC) sichert die im Zellkopf mitgeführte Information gegen Übertragungsbitfehler. Der hierbei verwendete zyklische Abramson-Code ermöglicht die Korrektur von Einzelbitfehlern sowie die Erkennung

PT-Inhalt	Interpretation			
000	Benutzerzelle, keine Überlast, Benutzerkennung = 0			
001	Benutzerzelle, keine Überlast, Benutzerkennung = 1			
010 Benutzerzelle, Überlastmeldung, Benutzerkennung = 0				
011	Benutzerzelle, Überlastmeldung, Benutzerkennung = 1			
100	Zelle für Betrieb und Wartung des zugeordneten VC-Abschnitts			
101	Zelle für Betrieb und Wartung der zugeordneten VC-Verbindung			
110	Zelle zur schnellen Ressourcenverwaltung			
111 Reserviert für zukünftige Anwendungen				

Tabelle 2.4: Codierung des PT-Feldes

bestimmter Mehrfachbitfehler (siehe z.B. [147]). Ferner ist mit Hilfe dieses Feldes eine Erkennung der Zellgrenzen möglich [122].

# 2.3 Protokollarchitektur des Breitband-ISDN

## 2.3.1 Protokoll-Referenzmodell

Moderne Kommunikationssysteme weisen eine hierarchische Protokollstruktur auf, welche die Kommunikationsfunktionen auf mehrere, aufeinander aufbauende, horizontale Schichten verteilt. Beispielsweise umfaßt das von der International Organization for Standardization (ISO) vorgeschlagene Basis-Referenzmodell für offene Kommunikationssysteme [103] sieben verschiedene Protokollschichten. In Anlehnung an dieses Konzept wurde für das

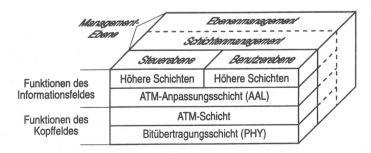


Bild 2.5: Protokoll-Referenzmodell des Breitband-ISDN [115]

Breitband-ISDN das in Bild 2.5 dargestellte geschichtete Protokoll-Referenzmodell definiert [115]. Zusätzliche vertikale Ebenen grenzen die Nutzdatenübertragung gegenüber Signalisierund Managementfunktionen ab.

Die Benutzerebene umfaßt alle Protokollfunktionen, die zum Transfer von Benutzerdaten erforderlich sind und weist eine hierarchische Strukturierung auf. Die Bitübertragungsschicht und die ATM-Schicht beinhalten die für alle Dienste gemeinsamen übertragungs- und vermittlungstechnischen Funktionen der Netzknoten. An den Endpunkten einer virtuellen Verbindung findet ein Übergang in die ATM-Anpassungsschicht (ATM Adaptation Layer, AAL) statt, die für die dienstespezifische Anbindung der ATM-Schicht an die höheren Protokollschichten zuständig ist. Tabelle 2.5 gibt Aufschluß über die Funktionalität der unteren Schichten, die in den folgenden Abschnitten näher erläutert werden.

Die Steuerebene faßt die für Aufbau, Steuerung und Abbau von Rufen bzw. Verbindungen erforderlichen Signalisierprotokolle zusammen. Sie basiert auf den von der ATM-Schicht bereitgestellten Funktionen, d.h. die Signalisierinformationen werden durch die ATM-Schicht übermittelt. Die grundlegenden Aspekte der sich noch in der Standardisierung befindlichen Signalisierprotokolle werden in Abschnitt 2.3.5 diskutiert.

Schicht	Subschicht	Funktionen				
ATM- Anpassungs-	Convergence Sublayer (CS)	Dienstespezifische Funktionen wie z.B. Ausgleich von Verzögerungsschwankungen, Behandlung von Zellverlusten und Bitfehlern				
Schicht	Segmentation and Reassembly (SAR)	Umsetzung der Daten in Zellen und umgekehrt				
ATM-Schicht		Generische Flußkontrolle Erzeugung bzw. Entfernung des Zellkopfes VPI- und VCI-Übersetzung Multiplexen bzw. Demultiplexen von VP- und VC-Verbindungen				
Bitüber- tragungs- schicht	Transmission Convergence (TC)	Entkopplung der Zellraten Erzeugung bzw. Überprüfung der Fehlersicherung Erkennung der Zellgrenzen Rahmenanpassung Rahmenerzeugung bzwrückgewinnung				
	Physical Medium (PM)	Bitsynchronisation Leitungscodierung				

Tabelle 2.5: Strukturierung des Protokoll-Referenzmodells [115]

Die Management-Ebene ist für die Verwaltung und Koordination der Protokollebenen und -schichten zuständig. Das Schichten-Management ist hierarchisch strukturiert und verwaltet die Ressourcen und Daten der jeweiligen Protokollschicht. Ferner sind hier die Protokollfunktionen zum Austausch und zur Bearbeitung von Nachrichten für Betriebs- und Wartungsaufgaben angesiedelt. Die Koordination der verschiedenen Protokollebenen sowie die Verwaltung gemeinsam genutzter Ressourcen und Daten geschieht innerhalb des Ebenen-Managements. Auf eine detaillierte Beschreibung der einzelnen Funktionen wird verzichtet, da sie im vorliegenden Zusammenhang von untergeordneter Bedeutung sind.

# 2.3.2 Bitübertragungsschicht

Nach Tabelle 2.5 werden die Funktionen der Bitübertragungsschicht auf zwei Subschichten, die PM-Subschicht (Physical Medium Sublayer, PM) und die TC-Subschicht (Transmission Convergence Sublayer, TC), aufgeteilt. Die PM-Subschicht stellt alle zur Übertragung eines kontinuierlichen Bitstroms benötigten Funktionen bereit. Dazu zählt die Leitungscodierung, die Bitsynchronisation sowie das Aussenden bzw. der Empfang der entsprechenden elektrischen oder optischen Signale.

Die TC-Subschicht setzt die von der ATM-Schicht angelieferten Zellen in einen kontinuierlichen Bitstrom um, der von der PM-Subschicht übertragen wird. Auf der Empfangsseite müssen die Zellen aus dem Bitstrom wiedergewonnen werden. Die Umsetzung des kontinuierlichen Bitstroms auf einen diskontinuierlichen Strom von Nutzzellen kann in folgende Einzelschritte unterteilt werden:

- Rahmenerzeugung bzw. -rückgewinnung: Das bei der Übertragung verwendete Rahmenformat muß auf der Sendeseite erzeugt und beim Empfang wiedergewonnen werden.
- Rahmenanpassung: Der kontinuierliche Zellstrom muß in die innerhalb des Übertragungssystems verwendete Rahmenstruktur eingefügt (Sendeseite) bzw. aus dieser Rahmenstruktur extrahiert (Empfangsseite) werden. In der ITU-T-Empfehlung I.432 [122] sind für die Benutzer/Netzschnittstelle zwei verschiedene Übertragungsformate vorgesehen, ein Zell-orientiertes Format und eine Rahmen-orientierte Variante basierend auf der synchronen digitalen Hierarchie (Synchronous Digital Hierarchy, SDH). Beide Systeme arbeiten mit einer Übertragungsgeschwindigkeit von 155,520 Mbit/s bzw. 622,080 Mbit/s. Für Betrieb und Wartung des Übertragungssystems wird 1/27 dieser Kapazität verwendet, so daß sich die Nutzbitrate auf 149,760 Mbit/s bzw. 599,040 Mbit/s beläuft.
- Erkennung der Zellgrenzen: Zur Synchronisation der Zellen wird die Fehlersicherung des Zellkopfes verwendet [122]. Dazu werden die Zellgrenzen bitweise verschoben, bis das HEC-Feld einen gültigen Zellkopf anzeigt. Die Synchronisation ist erfolgreich, falls

auch die nachfolgenden sechs Zellen ein fehlerfreies Kopffeld aufweisen. Ferner ist mit dieser Funktion eine Verwürfelung (Scrambling) des Informationsfeldes verknüpft, um eine Fehlsynchronisation auszuschließen.

- Erzeugung bzw. Überprüfung der Fehlersicherung: In Senderichtung wird eine Prüfsequenz zur Fehlersicherung des Zellkopfes berechnet und in das HEC-Feld eingetragen. Die Empfangsseite wertet diese Prüfsequenz aus und unterdrückt Zellen mit nicht korrigierbaren Fehlern.
- Entkopplung der Zellraten: Schließlich müssen Leerzellen eingefügt werden, um den von der ATM-Schicht angelieferten Strom von Nutzzellen in einen kontinuierlichen Zellstrom umzusetzen. Diese Leerzellen werden mit einer speziellen Kennung versehen, damit sie auf der Empfangsseite wieder entfernt werden können.

#### 2.3.3 ATM-Schicht

Die eigentliche Vermittlungsfunktion wird von der ATM-Schicht wahrgenommen, die für den Transport von ATM-Zellen zuständig ist. Die Funktionen dieser Schicht sind unabhängig vom Übertragungsmedium und den unterstützten Diensten und beziehen sich ausschließlich auf die Felder des ATM-Zellkopfes. Das Informationsfeld wird transparent übermittelt und an den Endpunkten einer VC-Verbindung mit der ATM-Anpassungsschicht ausgetauscht.

Die Hauptaufgaben der ATM-Schicht lassen sich aus dem in Abschnitt 2.2.1 angesprochenen Vermittlungsprinzip und der in Abschnitt 2.2.2 erörterten Unterscheidung von virtuellen Pfaden und virtuellen Kanälen ersehen. Sie umfassen die Erzeugung/Entfernung des Zellkopfes an den Endpunkten der VC-Verbindung, das Multiplexen/Demultiplexen von VC-bzw. VP-Verbindungen und die Umsetzung der VCI- und VPI-Felder in den Netzknoten.

Ferner beinhaltet die ATM-Schicht verschiedene Steuerungsfunktionen zur Regulierung des ATM-Zellflusses. Dazu zählen insbesondere die generische Flußkontrolle, die Überwachung bzw. Formung des ATM-Verkehrs von VC- und VP-Verbindungen, die Abwicklung von Protokollen zur schnellen Ressourcenverwaltung, die Verlustprioritätsbehandlung und das Erzeugen von Überlastmeldungen. Auf diese Funktionen wird in Unterkapitel 2.4 ausführlich eingegangen.

## 2.3.4 ATM-Anpassungsschicht

Die ATM-Anpassungsschicht (ATM Adaptation Layer, AAL) schließt eine VC-Verbindung ab und übernimmt die dienstespezifische Anbindung der ATM-Schicht an die höheren Protokollschichten. Die Hauptaufgaben beziehen sich auf die Abbildung der Protokoll-Dateneinheiten auf ATM-Zellen und die individuelle Anpassung der Netzgüte der ATM-Schicht an

die Dienstgüteanforderungen der speziellen Dienste und Anwendungen. Gemäß Tabelle 2.5 wird diese Funktionalität von zwei getrennten Teilschichten erbracht.

- Die SAR-Subschicht (Segmentation and Reassembly, SAR) dient der Umsetzung von Protokoll-Dateneinheiten der höheren Schichten auf das Informationsfeld der Zelle und umgekehrt. Ferner übernimmt diese Subschicht die Erkennung von Übertragungsfehlern wie z.B. Zelleinfügungen, Zellverlusten und Übertragungsbitfehlern.
- Die CS-Subschicht (Convergence Sublayer, CS) enthält die zusätzlichen, dienstbezogenen Anpassungsfunktionen, welche die bei der Übermittlung aufgetretenen Defekte (Verzögerungsschwankungen, Zelleinfügungen, Zellverluste bzw. Übertragungsbitfehler) behandeln oder eine Flußkontrolle durchführen.

Zur optimalen Unterstützung der verschiedenen Dienste und Anwendungen wurden fünf unterschiedliche AAL-Protokolle definiert [120], die jeweils einer von vier AAL-Dienstklassen zugeordnet werden können [119]. Das hierzu verwendete Klassifizierungsschema berücksichtigt die Zeitbeziehung zwischen Sender und Empfänger, den Bitratenbedarf und den Verbindungstyp. Die Charakteristika der hieraus entstehenden Dienstklassen A-D sind in Tabelle 2.6 zusammengefaßt.

	Klasse A	Klasse B	Klasse C	Klasse D		
Zeitbeziehung	vorha	anden	nicht vorhanden			
Bitrate	konstant		variabel			
Verbindungstyp	ve	rbindungsorientiert		verbindungslos		

Tabelle 2.6: Dienstklassen der ATM-Anpassungsschicht [119]

Die AAL-Dienstklasse A stellt einen Kanal konstanter Bitrate (Constant Bit Rate, CBR) zur Emulation von durchschaltevermittelten Kanälen bereit. Der durch die Paketierung entstehende regelmäßige Zellstrom muß auch beim Empfänger zu synchronen Zeittakten ausgespielt werden, um einen kontinuierlichen Bitstrom zurückzugewinnen. Dies erfordert die Erkennung und Behandlung von Zelleinfügungen, Zellverlusten und Bitfehlern, den Ausgleich der im Netz verursachten variablen Zellverzögerungen und eine Rückgewinnung des Sendebittaktes. Die AAL-Dienstklasse B unterstützt Video-, Audio- und Sprachdienste mit zeitlich veränderlicher Bitrate und erfordert ähnliche Funktionen wie die Dienstklasse A.

Die verbleibenden AAL-Dienstklassen C und D sind für verbindungsorientierte und verbindungslose Datenkommunikation vorgesehen. Dies erfordert eine eindeutige Zuordnung der Zellen zu den Protokoll-Dateneinheiten der höheren Schichten, die Erkennung und ggfs. Behebung von Übertragungsfehlern (Bitfehler, Zellverluste, Zelleinfügungen) sowie die Behandlung von teilgefüllten Zellen. Zusätzliche Mechanismen gestatten das Multiplexen/Demultiplexen verschiedener AAL-Verbindungen auf eine ATM-Verbindung.

Basierend auf diesen Dienstklassen wurden bislang fünf verschiedene AAL-Protokollvarianten definiert, die als AAL-Typen bezeichnet werden. Die AAL-Typen 1 bis 4 sind den Dienstklassen A bis D zugeordnet, und der AAL-Typ 5 basiert auf Dienstklasse C. Die detaillierte Spezifikation der umfangreichen AAL-Protokolle und die dazu definierte Strukturierung des Informationsfeldes kann der ITU-T-Empfehlung I.363 [120] entnommen werden.

## 2.3.5 Signalisierungsprinzipien

Die Signalisierung spielt eine Schlüsselrolle bei der Konzeption des Breitband-ISDN, da die Vielfalt der möglichen Verbindungskonfigurationen eine äußerst flexible und mächtige Verbindungssteuerung erfordert. Die Signalisierprotokolle sind Teil der Steuerungsebene des Protokoll-Referenzmodells und somit abgegrenzt von der eigentlichen Nutzdatenübertragung. Das dadurch entstehende, logisch abgesetzte Signalisiernetz übernimmt die Steuerung der Vermittlungsknoten beim Aufbau, der Überwachung und beim Abbau von Verbindungen.

Der Austausch von Signalisiermeldungen zwischen den Instanzen des Signalisiernetzes erfolgt über spezielle Signalisierverbindungen, die auf den von der ATM-Schicht zur Verfügung gestellten virtuellen Verbindungen basieren. Signalisierverbindungen können zwischen Benutzern, zwischen Benutzern und Netzknoten und zwischen Netzknoten bestehen. Für die Anbindung an die höheren Schichten der Signalisierprotokolle wurde ein eigenes ATM-Anpassungsprotokoll (Signalling-AAL, S-AAL) entworfen [123], das auf dem AAL-Typ 5 basiert. Im Gegensatz zur Nutzdatenübertragung muß diese Schicht sowie ausgewählte Teile der höheren Steuerungsprotokolle auch in den Vermittlungsknoten implementiert werden.

Die gegenüber dem Schmalband-ISDN in wesentlichen Punkten erweiterten Verbindungskonfigurationen und -merkmale stellen neue Anforderungen an die Funktionalität der Signalisierung [22, 99, 114, 134]:

- Aufbau, Rekonfiguration und Abbau von VC- und VP-Verbindungen
- Vereinbarung verschiedener Verbindungsattribute (Transferrate, Netzgüte) beim Verbindungsaufbau
- Änderung der Verbindungsattribute während der Verbindung
- Unterstützung von Verteil- und Konferenzdiensten zwischen mehr als zwei Partnern mit Hilfe von Punkt-Mehrpunkt-, Mehrpunkt-Mehrpunkt- und "Broadcast"-Verbindungen
- Unterstützung von Multimedia-Diensten durch mehrere synchronisierte Verbindungen für unterschiedliche Dienstkomponenten
- Dynamische Hinzunahme bzw. Entfernung von Partnern bzw. Dienstkomponenten während eines Rufes

Da Konferenz- und Multimedia-Dienste eine Vielzahl von Einzelverbindungen enthalten, die zudem dynamisch geändert werden können, ist eine grundsätzliche Trennung zwischen Ruf (call) und Verbindung (connection) unumgänglich. Ein Ruf wird hierbei modular aus Einzelverbindungen zusammengesetzt, die z.B. einzelnen Dienstkomponenten von Multimedia-Rufen zugeordnet sein können. Die Steuerung von Rufen und Verbindungen ist hierarchisch angelegt, d.h. der Auf- und Abbau von Verbindungen ist Aufgabe der Verbindungssteuerung, während die übergeordnete Rufsteuerung Koordinationsfunktionen wahrnimmt. Dazu zählt beispielsweise die Synchronisation von Multimedia-Verbindungen, die auf einem Weg durch das Netz geführt werden müssen, um Laufzeitunterschiede auszuschließen.

Die Empfehlungen des ITU-TSS sehen einen evolutionären Ausbau der im Schmalband-ISDN verwendeten Protokolle vor, der in drei Stufen ("Capability Set" 1–3) vollzogen werden soll [22, 99, 124, 134]. Die wesentlichen Eigenschaften dieser drei Versionen des Signalisiersystems sind in Tabelle 2.7 dargelegt.

		Version 1	Version 2	Version 3	
	Verbindungs- konfiguration	Punkt-Punkt- Verbindungen	Punkt-Mehrpunkt- Verbindungen	Broadcast-Verbindungen und Multimedia-Rufe	
	Verkehrsparameter	Spitzenzellrate	weitere Parameter		
	Netzgüteklassen	keine Klassen	Anzeige der Klasse	Vereinbarung der Klasse	

Tabelle 2.7: Charakteristika zukünftiger Signalisierprotokolle (nach [134])

# 2.4 Verkehrsmanagement für ATM-Netze

# 2.4.1 Übersicht und Zusammenwirken der Steuerungsfunktionen

Der Asynchrone Transfer-Modus hat sich aufgrund seiner flexiblen und effektiven Kapazitätszuteilung als universelles Vermittlungsprinzip für zukünftige Breitband-Netze etabliert. Das statistische Multiplexen unterschiedlicher Verkehrsströme erlaubt eine ökonomische Integration verschiedenster Dienste mit heterogenen und z.T. variablen Bitratenanforderungen (vgl. Abschnitt 2.1.2). Die damit verknüpfte hochgradig dynamische Belastung der Netzressourcen und die wechselseitige Beeinflussung der Verbindungen führt jedoch bei Überschreitung einer zulässigen Netzlast zu einer merklichen Beeinträchtigung der Netzgüte. Zur Einhaltung bestimmter Grenzwerte in Bezug auf Zellverluste und -verzögerungen ist ein Verkehrsmanagement notwendig, welches sich an den folgenden Bewertungskriterien ausrichtet [54, 121]:

• Effizienz: Optimierung der Netzauslastung unter der Randbedingung, daß die Netzgüte in iedem Fall eingehalten werden kann.

• Flexibilität: Dienste mit unterschiedlichsten Eigenschaften müssen unterstützt werden.

• Einfachheit: Die Komplexität des Netzes und der Endeinrichtungen sollte aus Kostengründen möglichst gering gehalten werden.

• Robustheit: Die Steuerungsmechanismen sollten möglichst unempfindlich gegenüber den getroffenen idealisierten Annahmen sein.

Die in heutigen paketvermittelten Netzen gebräuchlichen Flußkontroll- und Fehlerbehebungsmechanismen passen den Verkehrsfluß adaptiv an die momentane Lastsituation innerhalb des Netzes an. Die Reaktion auf eine Überlastsituation hängt jedoch von der Signallaufzeit ab, die in Weitverkehrsnetzen relativ große Werte annehmen kann. Bei den betrachteten hohen Übertragungsgeschwindigkeiten müßte daher eine sehr große Informationsmenge zwischengespeichert werden, um Zellverluste vermeiden zu können. Dies erfordert sehr große Puffer und kann zu unzulässig großen Zellverzögerungen führen, die für Echtzeitdienste wie Sprach- oder Videokommunikation nicht tragbar sind. Da die Überlastbehandlung zudem ineffizient sein kann, falls sie auf bereits abgeklungene Überlastsituationen anspricht, sollte sie nur in Ausnahmefällen zum Einsatz kommen. Im Regelfall sollte das Verkehrsmanagement auf die Vermeidung einer Überlastsituation abzielen und den Verkehrsfluß so begrenzen, daß die Mindestnetzgüte eingehalten werden kann.

Das in der ITU-T-Empfehlung I.371 [121] vorgestellte Gesamtkonzept nimmt aus diesem Grund eine Zweiteilung des Verkehrsmanagements vor. Unter dem Oberbegriff Verkehrssteuerung (Traffic Control) wurden verschiedene Mechanismen und Funktionen zusammengefaßt, die auf die Vermeidung einer Überlastsituation abzielen. Die Einzelkomponenten der Verkehrssteuerung müssen zusammenwirken, damit die für die relevanten Netzgüteparameter vorgegebenen statistischen Grenzwerte eingehalten werden können. Die Mechanismen zur Behandlung einer Überlastsituation (Congestion Control) bilden die zweite Komponente des Verkehrsmanagements, welche die Intensität, Ausbreitung und Dauer einer Überlastsituation begrenzt. Eine Überlastsituation bezeichnet einen Zustand, bei dem die vordefinierten Grenzwerte der Netzgüte verletzt werden und kann durch statistisches Multiplexen, unvorhersehbare Fluktuationen der Verkehrsströme oder Fehlfunktionen des Netzes ausgelöst werden. Die Überlastbehandlung muß jedoch aus den oben genannten Gründen diesen Ausnahmefällen vorbehalten bleiben.

Für die Verkehrssteuerung und Überlastbehandlung in ATM-Netzen wurde ein ganzes Bündel von sich ergänzenden Einzelmechanismen spezifiziert [7, 121]:

# • Verkehrssteuerung (Traffic Control)

Zellverlustprioritäten (Cell Loss Priority, CLP):
 Diese Art der Prioritätsbehandlung gestattet eine flexible Anpassung der Zellver-

lustrate an die Bedürfnisse der übergelagerten Dienste und Anwendungen.

- Quellflußkontrolle (Usage Parameter Control, UPC, und Network Parameter Control, NPC):
   Diese Funktion überwacht und regelt die beim Verbindungsaufbau vereinbarten Verkehrsparameter von VC- und VP-Verbindungen an der Benutzer/Netz-Schnittstelle (UPC) und an netzinternen Schnittstellen (NPC).
- Verkehrsformung (Traffic Shaping):
   Dieser Mechanismus führt eine aktive Formung der Verkehrscharakteristik von VC- oder VP-Verbindungen durch.
- Schnelles Ressourcenmanagement (Fast Resource Management, FRM):
   Damit kann eine schnelle Zuweisung von Netzressourcen (Puffer- bzw. Übertragungskapazität) für jedes einzelne Datenbüschel erfolgen.
- Verbindungsannahmesteuerung (Connection Admission Control, CAC):
   Die Verbindungsannahmefunktion entscheidet, unter Berücksichtigung der Verkehrs- und Netzgütevorgaben aller Verbindungen, ob eine neue VC- bzw. VP-Verbindung zugelassen werden kann oder nicht.
- Verwaltung der Netzressourcen (Network Resource Management, NRM):
   Dieser Aufgabenbereich beschränkt sich bisher auf die Verwaltung virtueller Pfade, insbesondere auf deren mittel- und langfristige Kapazitätsanpassung an den aktuellen Bedarf.

# • Überlastbehandlung (Congestion Control)

- Selektives Verwerfen von Zellen (Selective Cell Discarding):
   Bei Überlastsituationen verwirft diese Funktion primär Zellen der niedrigen Verlustpriorität.
- Explizite Überlastanzeige (Explicit Forward Congestion Indication, EFCI): Diese Überlastanzeige signalisiert dem Empfänger, daß mindestens ein Verbindungsabschnitt überlastet ist, worauf entsprechende Protokollmechanismen aktiviert werden können, die eine adaptive Senkung des Verkehrsaufkommens veranlassen.

Für jede dieser Steuerungsfunktionen wurde eine Vielzahl von Mechanismen und Algorithmen vorgeschlagen [10, 130, 213, 240]. Die Auswahl eines speziellen Verfahrens unterliegt nicht der Standardisierung, sondern kann durch den Netzbetreiber frei entschieden werden. Sie sollte sich jedoch an den für das Gesamtkonzept angeführten Kriterien orientieren. Ferner sind nur die Quellflußkontrolle und die Verbindungsannahme bindend vorgeschrieben. Alle anderen Funktionen sind optional und müssen nicht unbedingt implementiert werden.

Der beschriebene Satz von Funktionen erlaubt die Steuerung und Kontrolle des Verkehrsaufkommens in verschiedenen Zeitmaßstäben, entsprechend den dynamischen Schwankungen des Verkehrsflusses. Damit ist eine effiziente und direkte Steuerung der von unterschiedlichen Ursachen herrührenden statistischen Verkehrsschwankungen möglich.

Kurzzeitige Schwankungen des Zellaufkommens können durch die Regelung des Puffer- und Netzzugangs von einzelnen Zellen ausgeglichen werden. Die Büschelförmigkeit des Verkehrs wird am Netzzugang durch die Quellflußkontrolle, schnelle Reservierungs- und geeignete Verkehrsformungsmechanismen kontrolliert bzw. modifiziert. Funktionen, die auf Rückmeldungen des Netzes angewiesen sind (EFCI, FRM), benötigen eine laufzeitabhängige Reaktionszeit zur Änderung der Lastsituation. Eine von der Verbindungsannahme zugelassene Verbindung belastet das ATM-Netz für die gesamte Verbindungsdauer. Mittelfristige statistische Schwankungen der Ruflast können durch eine dynamische Kapazitätsanpassung der virtuellen Pfade bzw. durch adaptive Wegesuchalgorithmen ausgeglichen werden. Eine langfristige Änderung des Verkehrsaufkommens kann demgegenüber nur durch eine verkehrsgerechte Netzplanung und -optimierung aufgefangen werden. Die Zeitkonstanten, in denen die einzelnen Verkehrssteuerungs- und Überlastbehandlungsmaßnahmen die Lastsituation innerhalb des Netzes beeinflussen, sind in Bild 2.6 graphisch wiedergegeben.

Die zeitliche Staffelung der einzelnen Mechanismen führt auf ein hierarchisches Verkehrsmanagement, bei dem die längerfristigen Steuerungsfunktionen auf der kurzzeitigen Regelung des Verkehrs aufbauen. Alle Mechanismen müssen sorgfältig aufeinander abgestimmt werden, um ein möglichst konsistentes Gesamtkonzept zu erhalten. Beispielsweise muß die

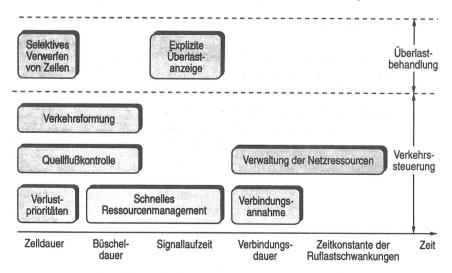


Bild 2.6: Zeitkonstanten der Steuerungsfunktionen

Verbindungsannahme auf den von der Quellflußkontrolle überwachten Verbindungsparametern basieren, um eine zuverlässige Begrenzung des Verkehrsaufkommens zu erzielen. Dabei können eventuelle Unzulänglichkeiten der Quellflußkontrolle kompensiert werden, indem diese Parameter mit einem entsprechenden Sicherheitszuschlag versehen werden.

### 2.4.2 Funktionen zur Verkehrssteuerung

### 2.4.2.1 Dienstgüteklassen und Prioritätsmechanismen

Das in ATM-Netzen angewandte flexible und einheitliche Vermittlungsprinzip unterstützt die Integration einer Vielzahl von Diensten und Anwendungen. Andererseits steht die Gleichbehandlung aller Verbindungen und Zellen in gewissem Widerspruch zu den in Abschnitt 2.1.3 dargelegten heterogenen Dienstgüteanforderungen der unterschiedlichen Dienste. Bei einer einheitlichen Behandlung aller Verbindungen müßte sich die Netzplanung und der Netzbetrieb an den Vorgaben der jeweils anspruchsvollsten Dienste und Anwendungen orientieren, was eine ineffiziente Nutzung der Netzressourcen zur Folge hätte. Dies wäre auch im Hinblick auf die nicht vorhersehbaren Dienstgütespezifikationen zukünftiger Dienste von Nachteil. Die neueren Standardisierungsdokumente [7, 121] sehen daher eine Vereinbarung der gewünschten Netzgüte beim Verbindungsaufbau vor. Dies ermöglicht eine flexible Anpassung des Übermittlungsdienstes an die Erfordernisse der darauf aufsetzenden Kommunikationsdienste und -anwendungen und damit eine Optimierung der Netzauslastung.

Innerhalb des Netzes wird die Differenzierung von Dienstgüteklassen sowohl durch die individuelle Festlegung der Netzgüte virtueller Pfade als auch durch die Verwendung von Zellverlustprioritäten ermöglicht. Virtuelle Pfade mit niedriger Netzgüte können stärker ausgelastet werden als virtuelle Pfade mit höherer Netzgüte. Die Zuweisung von virtuellen Kanälen zu virtuellen Pfaden hat so zu erfolgen, daß die Netzgüte des virtuellen Pfades ausreichend für die Anforderungen der zugeordneten Verbindung ist. Die Verwaltung der Netzgüte virtueller Pfade ist Aufgabe des Netzressourcenmanagements [121].

Der expliziten Unterscheidung von zwei Verlustprioritäten kommt im Hinblick auf die stark streuenden Bitfehlerraten-Vorgaben unterschiedlicher Dienste (vgl. Abschnitt 2.1.3) noch größere Bedeutung zu. Insbesondere ermöglicht die individuelle Kennzeichnung der Priorität jeder einzelnen Zelle zwei verschiedene Netzgüten innerhalb einer Verbindung. Dies ist bei bestimmten Videocodierverfahren, wie z.B. "Layered Coding" [72], nützlich, da diese die Bildinformation in mehrere Komponenten mit unterschiedlicher Bedeutung für die Bildqualität zerlegen. Die wichtigen Bildinhalte werden mit hoher Priorität übermittelt, wogegen für die weniger wichtigen Informationsanteile die niedere Priorität ausreichend ist.

Andererseits bedeutet die Unterscheidung von zwei verschiedenen Prioritäten einen nicht zu unterschätzenden Steuerungsaufwand, da die Verwaltung und Organisation der Netzpuffer, die Quellflußkontrolle sowie die Verbindungsannahme angepaßt bzw. erweitert werden müssen. Daher schließt die in Kapitel 5 durchgeführte Untersuchung von Verlustprioritätsmechanismen neben der Leistungsfähigkeit der Verfahren auch die damit verknüpfte Erhöhung des Implementierungsaufwands und der Netzkomplexität ein.

### 2.4.2.2 Quellflußkontrolle

Die durch das asynchrone Zeitmultiplexprinzip erzielte flexible Kapazitätszuteilung erlaubt einer Verbindung — abweichend von den beim Verbindungsaufbau vereinbarten Grenzwerten — mit einer beliebigen Zellrate zu senden, die nur durch die physikalische Übertragungsgeschwindigkeit begrenzt ist. Um einer Beeinträchtigung der Netzgüte anderer Verbindungen vorzubeugen, wird eine laufende Überwachung aller Verkehrsparameter empfohlen [121] bzw. vorgeschrieben [7].

Die hierfür zuständige Quellflußkontrolle überprüft zunächst, ob die VCI- und VPI-Felder gültige Werte besitzen, d.h. ob eine entsprechende Verbindung überhaupt besteht. Im zweiten Schritt werden dann die für diese Verbindung vereinbarten Verkehrsparameter wie z.B. maximale Zellrate, mittlere Zellrate und maximale Büscheldauer [121] überwacht. Bei einer Überschreitung der Grenzwerte wird die entsprechende Zelle entweder unmittelbar verworfen oder sie wird mit niederer Priorität (CLP-Bit = "1") versehen, damit sie bei Überlastsituationen innerhalb des Netzes selektiv verworfen werden kann (vgl. Unterabschnitt 2.4.3.1). Bei einer extremen Überschreitung der zulässigen Werte kann ein Abbruch der Verbindung bzw. eine Erhöhung des Verbindungstarifs erwogen werden.

Die Überwachung des Verkehrsflusses muß in Echtzeit und getrennt

- für einzelne VC- und VP-Verbindungen,
- am Netzzugang (Usage Parameter Control, UPC) bzw. innerhalb des Netzes (Network Parameter Control, NPC),
- für die hohe und die niedere Zellverlustpriorität sowie
- für jeden einzelnen Verkehrsparameter

erfolgen.

Diese Anforderungen lassen sich nur mit einfachen Mechanismen erfüllen, die sich kostengünstig in Form von speziellen Integrierten Schaltungen realisieren lassen. Ein weiteres Vergleichskriterium potentieller Quellflußkontrollverfahren ist die Korrektheit der Entscheidung über die Zulässigkeit einer Zelle. Insbesondere sollte für Verbindungen, die sich entsprechend den vorgegebenen Verbindungsparametern verhalten, eine außerordentlich niedrige Zellverlustwahrscheinlichkeit im Bereich von  $10^{-10}$  eingehalten werden; andererseits sollte

die Quellflußkontrolle auf jegliche Überschreitung der Grenzwerte angemessen und rasch reagieren.

Bisherige Studien haben gezeigt, daß die Überwachung von statistischen Parametern, wie z.B. der mittleren Zellrate, grundlegende Probleme aufwirft, da die geforderte hochpräzise Erfassung des statistischen Verhaltens auf der Basis von relativ kurzen Stichproben nicht möglich ist [94, 200, 201, 239]. Ein weiteres Problem stellt die im Teilnehmeranschlußbereich und innerhalb des Netzes durch Medienzugriffsprotokolle bzw. asynchrones Multiplexen eingefügte variable Zellverzögerung (Cell Delay Variation, CDV) dar [92, 180, 208], die zu zusätzlichen Schwankungen der gemessenen Zellrate führt. Damit die Zellrate genau erfaßt werden kann, muß die maximale Zellverzögerungsschwankung ebenfalls spezifiziert werden.

Aus diesen Gründen wird eine algorithmische Beschreibung der Verkehrsparameter angestrebt [7, 121], d.h. es wird ein Meßalgorithmus definiert, mit dem festgestellt werden kann, ob die vereinbarten Grenzwerte (Zellabstand  $T_I$ , zulässige Zeitabweichung  $\tau$  zum theoretischen Ankunftszeitpunkt der Zelle) eingehalten sind. Einer von zwei äquivalenten Meßalgorithmen stellt eine Verfeinerung des "Leaky Bucket"-Mechanismus dar [237], welcher die oben genannten Anforderungen am besten erfüllt [199, 200].

Dieser Algorithmus basiert auf einem Zähler mit kontinuierlichem Wertebereich  $\nu \in [0,\tau)$ , der bei jeder Zellankunft, abhängig vom Ankunftsabstand  $T_{Z,A}$  zur vorangegangenen Zelle, auf den Zwischenwert  $\nu' = \max\{\nu - T_{Z,A}, 0\}$  gesetzt wird. Die ankommende Zelle befindet sich in Übereinstimmung mit der Meßvorschrift, falls  $\nu' + T_I < \tau$  ist, und der neue Zählerstand ergibt sich zu  $\nu = \nu' + T_I$ . Im anderen Fall bleibt der Zählerstand auf  $\nu = \nu'$  stehen, und die Zelle muß von der Quellflußkontrolle verworfen oder markiert werden, da sie die Meßvorschrift verletzt.

Dieser generische Meßalgorithmus wurde ursprünglich zur gleichzeitigen Messung von maximaler Zellrate (Kehrwert des minimalen Zellabstands) und maximaler Zellverzögerungsschwankung [121] definiert. Die erweiterten Spezifikationen des ATM-Forums [7] verwenden denselben Algorithmus zur Erfassung einer langzeitigen Zellrate (Sustainable Cell Rate) innerhalb vorgegebener Toleranzgrenzen (Burst Tolerance), welche sich auf die statistischen Verkehrs- und Zellverzögerungsschwankungen beziehen. Die langzeitige Zellrate stellt eine obere Grenze für die mittlere Zellrate der Quelle dar, die sich für ein unendlich großes Meßintervall ergibt. Obwohl kein Quellflußkontrollverfahren standardisiert werden soll, liegt die Verwendung der obigen Meßvorschrift bei der Kontrolle des Quellflusses nahe. Die Definition weiterer Verkehrsparameter ist Gegenstand zukünftiger Standardisierungsbemühungen.

Die Vorhersage der oben beschriebenen Verkehrsparameter beim Verbindungsaufbau erweist sich jedoch auch bei einer algorithmischen Beschreibung des Verkehrs als fundamentales Problem [31, 200, 239]. Beispielsweise zeigen die in Tabelle 2.2 für Videoquellen angegebenen statistischen Verkehrsparameter eine ausgeprägte Abhängigkeit von der speziellen Videosequenz und der Codierung. Quantitative Untersuchungen realer Verkehrsquellen haben

bestätigt [201], daß in der Regel nur die Spitzenzellrate mit der erforderlichen Genauigkeit vorhergesagt und kontrolliert werden kann. Für die weitergehende Einschränkung des Quellverhaltens scheint eine aktive Beeinflussung des Quellverkehrs ein vielversprechender Ansatz zu sein.

#### 2.4.2.3 Formung des Quellverkehrs

Die Formung des Zellstroms von VC- und VP-Verbindungen ermöglicht eine Anpassung an die beim Verbindungsaufbau vereinbarten Verkehrsparameter, um ein unkontrolliertes Verwerfen von Zellen an der Quellflußkontrolle zu vermeiden. Eine Nachbildung der Meßalgorithmen innerhalb des Terminals [216] bzw. eine entsprechende Rückmeldung der Quellflußkontrolle [163], ermöglichen die frühzeitige Erkennung einer Parameterverletzung. In diesem Fall müssen geeignete dienstespezifische Gegenmaßnahmen ergriffen werden. Für zeitkritische Dienste (z.B. Sprach- und Videoübertragung) muß — ähnlich wie bei einer Codierung mit konstanter Bitrate — das Verkehrsaufkommen durch eine Anpassung des Codiermechanismus gesenkt werden. Die hieraus resultierende Schwankung der Netzgüte ist allerdings wesentlich geringer als bei einer Codierung mit konstanter Bitrate bzw. bei einem unkontrollierten Verwerfen von Zellen an der Quellflußkontrolle. Bei zeitunkritischen Diensten, wie z.B. Datenübertragung, kann ein kurzzeitiges Überangebot durch eine Zwischenspeicherung von Zellen überbrückt werden. Die damit verknüpfte Verzögerung der Zellen sollte innerhalb des durch die Netzgütevorgaben gesteckten Rahmens bleiben.

Ferner kann die Verkehrsformung zu einer Glättung des Verkehrs beitragen, welche sich günstig auf die Auslastung der Netzressourcen auswirkt (vgl. dazu auch die Ergebnisse in Kapitel 4). Da die Verkehrsformung eine optionale Funktion für Benutzer und Netzbetreiber ist, können beide von der hierdurch erzielten Kosteneinsparung profitieren. Während der Benutzer innerhalb des Endgeräts weitreichende Möglichkeiten zur Beeinflussung des Verkehrs hat, kann der Netzbetreiber den Verkehrsfluß nur in sehr beschränktem Umfang durch eine Verzögerung der Zellen am Netzzugang modifizieren.

Die bisher vorgeschlagenen Verkehrsformungsmechanismen entkoppeln den von der Verkehrsquelle generierten Verkehr mit Hilfe eines Puffers von dem in das Netz gesendeten Verkehr. Die für die Bedienung des Puffers vorgesehenen Strategien richten sich an folgenden Zielen aus:

- Wiederherstellung der maximalen Zellrate (Cell Spacing), um die durch die variable Zellverzögerung verursachte Büschelung der Zellen auszugleichen [33, 85, 246],
- Reduktion der maximalen Zellrate [64, 205, 256],
- Begrenzung der Büschelgröße [64],
- Begrenzung der mittleren Zellrate innerhalb eines Zeitintervalls [64, 205],

- Begrenzung einer "effektiven" Zellrate [169] oder
- Glättung des Verkehrs durch eine adaptive Anpassung der Zellbedienrate an den momentanen Füllstand des Puffers [19].

Die in [33, 64, 205, 246, 256] beschriebenen Leistungsuntersuchungen bestätigen eine signifikante Verbesserung der Netzauslastung, falls eine Glättung des Verkehrs vorgenommen wird. Die durch die Verzögerung der Zellen bzw. die Änderung der Codierung verursachte Verringerung der Dienstgüte muß innerhalb der für die jeweiligen Dienste gültigen Toleranzgrenzen liegen.

#### 2.4.2.4 Schnelle Ressourcenzuteilung

Datenkommunikation wird eine der ersten und wichtigsten Anwendungen zukünftiger Breitband-Kommunikationsnetze sein. Verkehrsmessungen zeigen, daß sich Datenverkehr durch eine extrem hohe Büschelförmigkeit und relativ große Spitzenbitraten auszeichnet [6, 158]. Unter diesen Randbedingungen wird das statistische Multiplexen verschiedener Verkehrsströme ineffizient, da die Betriebsmittel nur äußerst sporadisch belastet werden (vgl. die Ergebnisse in Kapitel 4). Da Datendienste außerdem relativ unempfindlich gegenüber Verzögerungen sind, wurde eine schnelle Reservierung von Betriebsmitteln für einzelne Datenbüschel (z.B. für die Übertragung einer Datei) angedacht [31, 49, 102, 238].

Die entsprechenden Protokolle reservieren die erforderlichen Netzressourcen entlang des beim Verbindungsaufbau festgelegten Weges nur für die Dauer eines Datenbüschels. Hierfür wird eine Reservierungszelle mit spezieller PT-Kombination (vgl. Tabelle 2.4) verwendet, deren Informationsfeld die angeforderte Puffer- bzw. Übertragungskapazität spezifiziert. Falls diese Kapazität an allen Netzknoten zwischen Quelle und Senke zur Verfügung steht, können die Daten erfolgreich übertragen werden. Eine Blockierung bzw. ein Verlust des Datenbüschels kann durch eine Wiederholung der Reservierung behoben werden. Nachdem die Daten übermittelt wurden, wird die Kapazität wieder freigegeben.

Die verschiedenen Protokollvarianten unterscheiden sich in mehrfacher Hinsicht:

- Reservierung von Puffer- [47, 49, 238] oder Übertragungskapazität [31, 34, 49, 102]:
   Die Reservierung von Pufferkapazität ist insbesondere bei hohen Spitzenzellraten effizienter als die Reservierung von Übertragungskapazität [49]. Allerdings müssen dafür größere Netzpuffer vorgesehen werden, die einen erhöhten Aufwand und zusätzliche Verzögerungen bedeuten.
- Datenübermittlung im Anschluß an die Reservierung [34, 102, 238] bzw. erst nach Erhalt der Reservierungsbestätigung [31, 34, 47, 49, 102]:
   Eine unquittierte Übermittlung des Datenbüschels direkt im Anschluß an die Reservierungszelle wirkt sich günstig auf die Transferzeit aus. Falls jedoch eine Blockierung

auftritt, müssen Mechanismen vorgesehen werden, die alle nachfolgenden Zellen dieser Verbindung selektiv an dem betroffenen Netzknoten verwerfen.

- Wiederholung blockierter Datenbüschel ist Teil des Protokolls [31, 34] oder wird den höheren Protokollschichten überlassen [47, 49, 238]:
   Eine direkte Wiederholung blockierter Dateneinheiten durch das schnelle Reservierungsprotokoll ist zu empfehlen, um die Transferzeit auch im Blockierungsfall in erträglichen Grenzen zu halten. Allerdings erfordert dies eine Zwischenspeicherung der Informationen.
- Freigabe der Netzressourcen zeitgesteuert [31, 34, 238] oder mit Hilfe einer speziellen Zelle [31, 47, 49, 102, 238]:
   Eine zeitgesteuerte Freigabe der belegten Ressourcen ist aufwendig, hat aber den Vorteil, daß das Protokoll weniger fehleranfällig ist.

Die in der Literatur vorgestellten Untersuchungen bestätigen eine Verbesserung der Netzauslastung bei vorgegebener Netzgüte, falls ein schnelles Ressourcenmanagement eingesetzt wird [34]. Die Einhaltung einer bestimmten Netzgüte, d.h. einer maximalen Blockierungswahrscheinlichkeit für Datenbüschel, erfordert weitere Kontrollmechanismen, welche die Quellaktivität überwachen und über die Annahme einer neuen Verbindung entscheiden.

### 2.4.2.5 Verbindungsannahme und Verkehrslenkung

Beim Asynchronen Transfer-Modus ist eine gegenseitige Beeinflussung der Verbindungen unvermeidlich. Daher muß für die Annahme einer neuen Verbindung bzw. für die Neuverhandlung von Verbindungsparametern nicht nur die Netzgüte der betrachteten Verbindung, sondern aller hiervon betroffenen Verbindungen überprüft werden. Die Gesamtentscheidung über die Annahme einer VC- bzw. VP-Verbindung beruht auf Einzelentscheidungen für die jeweiligen Verbindungsabschnitte. Dies ermöglicht einen schrittweisen Verbindungsaufbau wie in herkömmlichen durchschaltevermittelnden Netzen.

Eine Beeinträchtigung der Netzgüte kann ganz ausgeschlossen werden, falls für jede Verbindung die Spitzenbitrate reserviert wird. Zusätzlich muß eine bestimmte Auslastungsgrenze (kleiner 100%) eingehalten werden, um Pufferüberläufe infolge der asynchronen Belegung der Übertragungsleitungen zu begrenzen. Da in der Einführungsphase des Breitband-ISDN die Auslastung des Netzes gering sein wird, liegt diese einfache und sichere Methode der Spitzenbitratenzuteilung nahe.

Mit zunehmender Teilnehmerzahl und der Einführung von Breitband-Diensten wird ein durch statistisches Multiplexen erzielter Auslastungsgewinn interessant. Folglich wurden zahlreiche Verfahren vorgeschlagen, welche das statistische Multiplexen von Verbindungen einbeziehen (siehe z.B. [83, 95, 102, 174, 214]). Für eine detaillierte Beschreibung, Klassifizierung und Untersuchung dieser Ansätze sei auf Kapitel 6 verwiesen. Diese Mechanismen erfordern eine Charakterisierung des statistischen Verhaltens der Verkehrsquellen, was zu den in Unterabschnitt 2.4.2.2 angesprochenen Schwierigkeiten bei der Vereinbarung und Überwachung der entsprechenden Verkehrsparameter führt. Eine Alternative bietet die laufende Messung der Netzauslastung bzw. der Quellaktivität. Die Resultate dieser Messungen sind allerdings mit einer großen Unsicherheit behaftet, da die interessierenden Überlastzustände nur sehr selten auftreten dürfen.

Die Suche und Festlegung eines Weges, den alle Zellen einer VC-Verbindung nehmen, ist eine weitere wichtige Aufgabe der Verbindungsannahme. Die oben beschriebene schrittweise Annahme einer Verbindung ermöglicht ähnliche Verkehrslenkungsverfahren wie in durchschaltevermittelnden Netzen [51]. Allerdings muß die Auswahl eines bestimmten Weges neben den bekannten Lastkriterien [51] weitere Randbedingungen einbeziehen [86]. Beispielsweise sollten alle Verbindungen eines Multimedia-Rufes denselben Weg durch das Netz nehmen, um Laufzeitunterschiede zu vermeiden und den Synchronisationsaufwand beim Empfänger gering zu halten. Ferner darf eine VC-Verbindung nur über virtuelle Pfade mit ausreichender Netzgüte geführt werden [121]. Insbesondere ist das statistische Multiplexen von VC-Verbindungen innerhalb eines virtuellen Pfades nur dann zulässig, wenn alle VC-Verbindugen eine entsprechende Verschlechterung der Netzgüte tolerieren können [121]. Für eine optimale Auslastung sollte die Netzgüte der VC-Verbindung mit der Netzgüte des virtuellen Pfades übereinstimmen, d.h. VC-Verbindungen mit ähnlichen Netzgüteanforderungen sollten innerhalb eines virtuellen Pfades zusammengefaßt werden. Ähnliche Aussagen lassen sich für die Zusammenfassung von VC-Verbindungen mit ähnlicher Verkehrscharakteristik treffen (siehe [186] bzw. Kapitel 6).

Weitere Teilaufgaben der Verbindungsannahme beziehen sich auf die Festlegung der Parameter für die Quellflußkontrolle und den fairen Netzzugang von Verbindungen mit unterschiedlichem Bitratenbedarf. Ohne eine entsprechende Zugangskontrolle steigt die Blockierungswahrscheinlichkeit einer Verbindung mit ihrem Bitratenbedarf an. Die grundlegenden Mechanismen zum Schutz hochbitratiger Verbindungen sind aus der Mehrkanal-Durchschaltevermittlung bekannt [98, 138] und können relativ einfach an die Gegebenheiten in ATM-Netzen angepaßt werden.

## 2.4.2.6 Verwaltung virtueller Pfade

Auch im Hinblick auf die Verkehrssteuerung erweist sich die Einrichtung virtueller Pfade als vorteilhaft. Virtuelle Pfade erlauben eine logische Trennung von VC-Verbindungen mit unterschiedlichen Netzgüte- bzw. Bitratenanforderungen und vereinfachen die Annahme neuer VC-Verbindungen, da die VC-bezogenen Verbindungsannahmealgorithmen und Signalisierfunktionen nur in den VC-Vermittlungsknoten implementiert sein müssen [121]. Andererseits

führt die Aufteilung der Übertragungskapazität auf einzelne virtuelle Pfade zu einer geringeren Auslastung des Netzes, da kein Lastausgleich zwischen virtuellen Pfaden stattfinden kann. Beispielsweise kann ein Pfad blockiert sein, wogegen ein anderer Pfad auf demselben Übertragungsabschnitt noch genügend freie Kapazität besitzt. Bei gemeinsamer Nutzung der Kapazität würde diese Blockierung vermieden werden. Die dadurch erzielte Erhöhung der Netzlast ist aus der Durchschaltevermittlungstechnik unter dem Stichwort Bündelungsgewinn bekannt (vgl. z.B. [149]).

Ein ähnlicher Lastausgleich kann durch eine dynamische Anpassung der Übertragungskapazität der virtuellen Pfade an die stochastischen Schwankungen der Verkehrslast erzielt werden. Dazu muß das VP-Management eine periodische Optimierung der Bitratenzuteilung an virtuelle Pfade vornehmen. Eine detaillierte Studie [37] belegt den Kostenvorteil dieser dynamischen Lösung gegenüber einer statischen Zuteilung bzw. keiner Zuteilung von Übertragungskapazität an virtuelle Pfade (falls den virtuellen Pfaden keine Kapazität zugewiesen wird, muß die VC-Verbindungsannahme auch in VP-Vermittlungsknoten stattfinden).

Die dynamische Bitratenzuteilung an einzelne virtuelle Pfade basiert auf

- der laufenden Messung der Auslastung der virtuellen Pfade (umfaßt sowohl angenommene als auch abgewiesene Verbindungen),
- der Ableitung von Schätzwerten für den zukünftigen Bitratenbedarf und
- der Optimierung der Bitratenzuteilung an einzelne virtuelle Pfade, um die Rufverlustwahrscheinlichkeit zu minimieren bzw. die Auslastung zu maximieren.

In der Literatur wurden verschiedene Optimierungsalgorithmen vorgestellt, die entweder dezentral [37, 222] oder zentral (in den VC-Vermittlungsknoten) [161] ablaufen. Für die zentrale Zuteilung existiert sogar ein exakter Algorithmus [2], falls der zukünftige Bedarf präzise vorhergesagt werden kann. Dies ist jedoch relativ schwierig, da die zentrale Zuteilung infolge des Kommunikationsaufwandes und der komplexen Optimierungsalgorithmen relativ zeitaufwendig ist. Ein dezentraler Algorithmus kann wesentlich schneller auf Laständerungen ansprechen, ermöglicht aber keine globale Optimierung der Ressourcenzuteilung.

# 2.4.3 Funktionen zur Überlastbehandlung

#### 2.4.3.1 Selektives Verwerfen von Zellen

Diese Funktion ist eng mit der in Unterabschnitt 2.4.2.1 beschriebenen Prioritätsbehandlung verknüpft und kann Zellen selektiv verwerfen, falls eine Überlastsituation auftritt. Insbesondere können Zellen selektiv verworfen werden, für die keine Übertragungskapazität reserviert wurde. Beispielsweise wurde in der Literatur [71] eine Unterstützung von verbindungslosen

Diensten mit Hilfe der niedrigen Verlustpriorität angeregt, um die zur Reservierung der Übertragungskapazität erforderliche Zeitspanne überbrücken zu können.

Im Gegensatz zur Prioritätsbehandlung kann bei selektivem Verwerfen von Zellen für die niedere Priorität keine definierte Netzgüte garantiert werden. Die Netzgüte des hochprioritären Zellstroms wird so lange wie möglich gewahrt, indem bei einer Überlastsituation bevorzugt Zellen mit niederer Priorität bzw. Zellen von Verbindungen, die gegen die beim Verbindungsaufbau vereinbarten Verkehrsparameter verstoßen, verworfen werden. Wie dieses Verfahren mit der in Unterabschnitt 2.4.2.1 beschriebenen Prioritätsbehandlung zusammenarbeiten soll, die auch für die niedere Priorität eine definierte Netzgüte vorsieht, geht aus den Standardisierungsdokumenten nicht hervor.

## 2.4.3.2 Explizite Anzeige einer Überlast

Das im ATM-Zellkopf befindliche PT-Feld ermöglicht die explizite Anzeige einer drohenden bzw. bereits bestehenden Überlast. Ein Netzknoten kann diese Anzeige setzen, falls z.B. ein bestimmter Pufferfüllstand überschritten wird. Aufgrund dieser Überlastmeldung können in den Endgeräten implementierte Protokolle eine adaptive Verringerung der Zellrate vornehmen.

In [165] wurde ein Protokoll vorgestellt, welches die Zellrate einer Datenquelle auf einen bestimmten Minimalwert drosselt, falls eine Überlast angezeigt wird. Nach einer gewissen Zeit wird die Zellrate wieder allmählich bis zu einem bestimmten Maximalwert gesteigert. Die in [165] beschriebenen Simulationsergebnisse bestätigen, daß dieser reaktive Mechanismus nur bei kurzen Signallaufzeiten effizient arbeitet. Für längere Signallaufzeiten wird nur noch eine marginale Verbesserung der Zellverlustwahrscheinlichkeit infolge einer allgemeinen Verringerung der Netzauslastung erzielt. Allerdings kann die Dauer der Überlastsituation auf die Signallaufzeit begrenzt werden. Für Sprachquellen wurde ein ähnliches Protokoll beschrieben [258], wobei die Überlastsituation durch eine Anpassung des Codierverfahrens abgebaut wird.

Es soll betont werden, daß sich der Netzbetreiber nicht auf diesen optionalen Mechanismus verlassen sollte, da eine Reaktion nicht zwingend vorgeschrieben ist [121]. Insbesondere zeigen die in "Frame Relay"-Netzen mit ähnlichen Mechanismen gemachten Erfahrungen, daß die Endgeräte in der Regel keinen Gebrauch von dieser Information [128] machen, da Verbindungen, die ihre Bitrate nicht reduzieren, in gleichem Maße vom Abbau der Überlastsituation profitieren wie Verbindungen, welche ihre Bitrate verringern. Auf eine Reduktion der Bitrate wird deswegen verzichtet, um den Gesamtdurchsatz möglichst hoch zu halten. In ATM-Netzen könnte eine Reaktion auf die Überlastanzeige erzwungen werden, indem die Parameter der Quellflußkontrolle entsprechend eingestellt werden.

# Kapitel 3

# Methoden zur Modellierung und Leistungsbewertung von ATM-Netzen

Bedingt durch die funktionelle Komplexität und den hohen Implementierungsaufwand muß die Leistungsfähigkeit von ATM-Systemen schon in der Entwurfsphase mittels geeigneter verkehrstheoretischer Modelle und Analysemethoden abgeschätzt werden. Diese Methoden und Ansätze bilden auch die Grundlage für die spätere Systemdimensionierung und -optimierung bzw. für die Spezifikation adäquater Verkehrssteuerungsalgorithmen.

Ein verkehrstheoretisches Modell beschreibt die Funktionalität des realen Systems unter Verkehrs- und Leistungsgesichtspunkten. Es basiert auf einer abstrakten Beschreibung der Systemstruktur, des Ablaufgeschehens und der regellos eintreffenden Anforderungen an das System bzw. der zufallsabhängigen Reaktionszeiten des Systems. Für die Genauigkeit und die Aussagekraft der Leistungsuntersuchungen spielt die realistische Modellierung des statistischen Verhaltens der Verkehrsquellen durch geeignete stochastische Prozesse eine maßgebliche Rolle.

Das Verkehrsmodell bildet die Grundlage für die nachfolgende simulative bzw. analytische Leistungsbewertung des Systems. Die Verkehrssimulation ist eine sehr mächtige Methode, mit der auch komplexe bzw. detaillierte Verkehrsmodelle auf einem Rechner nachgebildet werden können. Für elementare Verkehrsmodelle existiert ein reichhaltiges Sortiment von analytischen Ansätzen, die durch verschiedene approximative Techniken ergänzt werden. Die nachfolgenden Untersuchungen stützen sich hauptsächlich auf dieser Methode ab, da die analytischen Lösungsalgorithmen direkt in entsprechende Prozeduren zur Verkehrssteuerung eingebunden werden können. Ferner ist eine präzise Erfassung von kleinen Zellverlustwahrscheinlichkeiten möglich. In den folgenden Abschnitten wird eine Einführung in die in dieser Arbeit verwendeten Verkehrsmodelle, Simulationstechniken und Analyseverfahren gegeben.

# 3.1 Modellierung von ATM-Verkehrsströmen

Wie bereits aus den einführenden Bemerkungen des vorigen Kapitels zu entnehmen ist, weisen die in ATM-Netzen auftretenden Verkehre eine Vielzahl von Freiheitsgraden auf und sind daher schwierig zu charakterisieren. Das statistische Verhalten der Verkehrsquellen wird sowohl durch das Teilnehmerverhalten als auch durch die Dienste, Protokolle und Codierverfahren beeinflußt. Zusätzliche Probleme bereiten die tageszeitlichen Schwankungen des Verkehrsangebots und die Abhängigkeiten zwischen verschiedenen Verkehrsströmen.

Das statistische und zeitliche Verhalten des ATM-Verkehrs (Rufe, Zellen) kann durch zufallsabhängige Prozesse charakterisiert werden. Für die Beschreibung und Untersuchung dieser Prozesse werden die nachfolgend erörterten Methoden aus der Stochastik und der Verkehrstheorie angewandt. Ausgehend von der hierarchischen Modellierung des Quellverkehrs wird abschließend ein Überblick über die für die einzelnen Modellierungsebenen verwendeten Prozesse gegeben.

# 3.1.1 Zufallsvariablen, Verteilungen, Momente und Transformationsmethoden

Eine Zufallsgröße wird in der Stochastik dadurch beschrieben, daß jeder möglichen Realisierung der Zufallsgröße eine reelle Zahl, die Zufallsvariable X, zugeordnet wird. Man unterscheidet kontinuierliche Zufallsvariablen (z.B. die Bediendauer) und diskrete Zufallsvariablen wie z.B. die Anzahl belegter Bedieneinheiten. Mehrdimensionale Zufallsvariablen (Zufallsvektoren) eignen sich zur gleichzeitigen Beschreibung verschiedener Zufallsgrößen.

Das statistische Verhalten von Zufallsvariablen wird durch die Verteilungsfunktion

$$F(x) = P\{X \le x\} \tag{3.1}$$

charakterisiert. Die Verteilungsfunktion F(x) gibt die Wahrscheinlichkeit an, daß die Zufallsvariable einen Maximalwert x nicht überschreitet. Die Ableitung der Verteilungsfunktion führt auf die Verteilungsdichtefunktion

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x} \,. \tag{3.2}$$

Diese Wahrscheinlichkeitsdichte geht für diskrete Zufallsvariablen in eine endliche Wahrscheinlichkeit über, was die Angabe einer Verteilung

$$p(x) = P\{X = x\}, x = 0, 1, ...,$$
 (3.3)

ermöglicht. Der Wertebereich von diskreten Zufallsvariablen wird ohne Einschränkung der Allgemeinheit auf den Bereich der natürlichen Zahlen begrenzt.

Eine andere Form der Beschreibung von Zufallsvariablen ermöglichen Transformationsmethoden, wie sie auch in der klassischen Elektrotechnik verwendet werden. Transformationen vereinfachen viele verkehrstheoretische Berechnungen bzw. machen diese überhaupt erst möglich. Die Laplace-Transformation läßt sich sowohl auf kontinuierliche als auch auf diskrete Zufallsvariablen anwenden. Die Laplace-Transformierte (LT) zur Verteilungsdichtefunktion f(x) ergibt sich zu [139]

 $\Phi(s) = \int_{x=-\infty}^{\infty} e^{-sx} \cdot f(x) \, dx . \tag{3.4}$ 

Für positive Zufallsvariablen kann die untere Integrationsgenze nach  $0^-$  verschoben werden. In diesem Fall liefert das obige Integral einen endlichen Wert, falls  $\mathrm{Re}(s) \geq 0$  gilt. Für diskrete Zufallsvariablen wird die Erzeugende Funktion (EF)

$$G(z) = \sum_{x=0}^{\infty} p(x) \cdot z^{x} , |z| \le 1 ,$$
 (3.5)

definiert. Die Erzeugende Funktion ist eng mit der z-Transformation verknüpft und geht aus dieser hervor, indem z durch  $z^{-1}$  ersetzt wird.

Für die grobe Charakterisierung einer Zufallsvariablen reicht deren Mittelwert E[X] und Varianz VAR[X] aus. Für kontinuierliche Zufallsvariablen ergibt sich

$$E[X] = \int_{x=-\infty}^{\infty} x \cdot f(x) dx = -\frac{d\Phi(s)}{ds} \bigg|_{s=0} = -\Phi'(0) \quad \text{bzw.}$$
 (3.6)

$$VAR[X] = \int_{x=-\infty}^{\infty} (x - E[X])^2 \cdot f(x) \, dx = E[X^2] - E[X]^2 = \Phi''(0) - \Phi'(0)^2 \,. \tag{3.7}$$

Im diskreten Fall erhält man

$$E[X] = \sum_{x=0}^{\infty} x \cdot p(x) = \frac{dG(z)}{dz}\Big|_{z=1} = G'(1)$$
 bzw. (3.8)

$$VAR[X] = \sum_{x=0}^{\infty} (x - E[X])^2 \cdot p(x) = E[X^2] - E[X]^2 = G''(1) + G'(1) - G'(1)^2.$$
 (3.9)

Die Varianz ist ein Maß für die Streuung einer Zufallsvariablen. Die Standardabweichung

$$\sigma_X = \sqrt{\text{VAR}[X]} \tag{3.10}$$

und der Variationskoeffizient

$$c_X = \frac{\sigma_X}{E[X]} \tag{3.11}$$

sind dazu äquivalente Beschreibungsformen.

Häufig setzt sich die Zufallsvariable X aus der Summe zweier Zufallsvariablen  $X_1$  und  $X_2$  zusammen. Falls die Zufallsvariablen  $X_1$  und  $X_2$  statistisch unabhängig sind, ergeben sich folgende Beziehungen [150]:

 Die Verteilungsdichtefunktion bzw. die Verteilung der Zufallsvariablen X ergibt sich aus der Faltung der Verteilungsdichtefunktionen bzw. Verteilungen der Zufallsvariablen X<sub>1</sub> und X<sub>2</sub>, d.h.

$$f(x) = f_1(x) * f_2(x)$$
 bzw.  $p(x) = p_1(x) * p_2(x)$ . (3.12)

 In dem durch die Transformationsmethoden definierten Bildbereich geht die relativ komplizierte Faltungsoperation in eine einfache Multiplikation über:

$$\Phi(s) = \Phi_1(s) \cdot \Phi_2(s)$$
 bzw.  $G(z) = G_1(z) \cdot G_2(z)$ . (3.13)

• Die Mittelwerte und Varianzen addieren sich wie die Zufallsvariablen selbst zu

$$E[X] = E[X_1] + E[X_2]$$
 und  $VAR[X] = VAR[X_1] + VAR[X_2]$ . (3.14)

Die im Rahmen dieser Arbeit verwendeten kontinuierlichen und diskreten Verteilungen sowie deren charakteristische Größen sind in Tabelle 3.1 zusammengestellt [78, 150, 200]. Alle Verteilungen können als Sonderfall der allgemeinen und unabhängigen Verteilung (Generally Independent, GI) aufgefaßt werden. Hinsichtlich der Anwendung dieser Verteilungen sind folgende Eigenschaften und Zusammenhänge von Bedeutung:

- Die deterministische Verteilung nimmt insofern eine Sonderstellung ein, als die hierdurch beschriebene Zufallsvariable nur einen festen Wert annehmen kann. Diese Verteilung kann zur Beschreibung der konstanten Zellübertragungsdauer bzw. der konstanten Zellrate von CBR-Verkehrsquellen eingesetzt werden.
- Die negativ-exponentielle Verteilung ist die einzige kontinuierliche Verteilung, welche die Eigenschaft der Gedächtnisfreiheit ("Markoffsche Eigenschaft") besitzt. Ein stochastischer Prozeß, der durch negativ-exponentiell verteilte Ereigniszeitpunkte bestimmt ist, entwickelt sich daher unabhängig von der Vergangenheit, d.h. die Wahrscheinlichkeitsdichte für ein Ereignis ("Ereignisrate") ist unabhängig von der Zeit. Diese Eigenschaft vereinfacht die verkehrstheoretische Beschreibung und Behandlung des Systems erheblich bzw. macht diese überhaupt erst möglich (vgl. Unterkapitel 3.4). Ein Zufallsprozeß mit negativ-exponentiell verteilten Ereignisabständen wird als Poisson-Prozeß bezeichnet, da die Anzahl von Ereignissen innerhalb eines festen Zeitintervalls einer Poisson-Verteilung (siehe unten) gehorcht.
- Die Erlang-k-Verteilung setzt sich aus k negativ-exponentiell verteilten Teilphasen mit den Mittelwerten k zusammen. Dies kann symbolisch durch das in Bild 3.1 dargestellte "Phasenmodell" ausgedrückt werden. Der Variationskoeffizient dieser Verteilung kann diskrete Werte zwischen 0 (deterministische Verteilung) und 1 (negativexponentielle Verteilung) annehmen. Die Erlang-k-Verteilung wird daher häufig eingesetzt, um eine allgemeine Verteilung mit einem Variationskoeffizienten zwischen 0 und 1 zu approximieren.

			γ	Υ	·						
	Variationskoeffizient	0	-	$\frac{1}{\sqrt{k}}$	$\sqrt{1+2\frac{p(1-p)(\overline{x}_1-\overline{x}_2)^2}{[p\overline{x}_1+(1-p)\overline{x}_2]^2}}$	ρ   læ		Variationskoeffizient	$\sqrt{1-rac{1}{x}}$	$\sqrt{\frac{N-\overline{x}}{N\overline{x}}}$	$\frac{1}{\sqrt{x}}$
	Mittelwert	les	les	les	$\frac{p\overline{x}_1 +}{(1-p)\overline{x}_2}$	le		Mittelwert	læ	le:	<del> </del>
Verteilungen	LT $\Phi(s)$	e - xs	$\frac{1}{1+\overline{x}s}$	$\left(\frac{k}{k+\overline{x}s}\right)^k$	$\frac{p}{1+\overline{x}_1s} + \frac{1-p}{1+\overline{x}_2s}$	$e^{-\overline{x}s+(s\sigma)^2/2}$	teilungen	$\operatorname{EF} G(z)$	$\frac{z}{\overline{x}-(\overline{x}-1)z}$	$\left(1 - \frac{\overline{x}}{N} + \frac{\overline{x}z}{N}\right)^N$	$\mathrm{e}^{\overline{x}(z-1)}$
Kontinuierliche Verteilungen	Verteilungsdichtefunktion $f(x)$	$\delta(x-\overline{x})$	$\frac{1}{x}e^{-x/\overline{x}} , \ x \ge 0, \ \overline{x} > 0$	$rac{k}{x} \cdot rac{(kx)^{k-1}}{x^{k-1} \cdot (k-1)!} \mathrm{e}^{-kx/x}, \ x \geq 0, \ \overline{x} > 0, \ k = 1, 2$	$\frac{p}{\bar{x}_1} e^{-x/\bar{x}_1} + \frac{1-p}{\bar{x}_2} e^{-x/\bar{x}_2} ,$ $x \ge 0, \ \bar{x}_1, \ \bar{x}_2 > 0, \ 0 \le p \le 1$	$rac{1}{\sqrt{2\pi}\sigma} e^{-(x-\bar{x})^2/(2\sigma^2)}, \ \sigma > 0$	Diskrete Verteilungen	Verteilung $p(x)$	$\frac{1}{x}\left(1-\frac{1}{x}\right)^{x-1}, \ x=1,2,, \ \overline{x}\geq 1$	$\binom{N}{x} \left(\frac{\overline{x}}{N}\right)^x \left(1 - \frac{\overline{x}}{N}\right)^{N-x},$ $x = 0, 1, \dots, N, \ 0 \le \overline{x} \le N, \ N = 1, 2, \dots$	$\frac{\bar{x}^x}{x!}e^{-\bar{x}}, x = 0, 1,, \bar{x} \ge 0$
	Verteilungstyp	Determinist. Vert. (D)	Negativ-exponentielle Verteilung (M)	Erlang- $k$ -Verteilung $(\mathbf{E}_k)$	Hyperexponentielle Verteilung 2. Ordnung (H <sub>2</sub> )	Normalverteilung		Verteilungstyp	Um 1 Stelle versch. geom. Vert. (GEO <sub>1</sub> )	Binomial-Verteilung (BIN)	Poisson-Verteilung (POIS)

Tabelle 3.1: Charakteristika wichtiger Verteilungen

- Die hyperexponentielle Verteilung 2. Ordnung kann durch das "Phasenmodell" nach Bild 3.2 beschrieben werden, d.h. die zwei Teilphasen mit den Mittelwerten  $\overline{x}_1$  und  $\overline{x}_2$  werden mit den Wahrscheinlichkeiten p bzw. (1-p) ausgewählt. Diese Verteilung eignet sich zur Approximation von allgemeinen Verteilungen mit einem Variationskoeffizienten größer 1.
- Die Normalverteilung nähert die Summenverteilung einer großen Anzahl von unabhängigen Zufallsvariablen mit beliebigen statistischen Eigenschaften an, falls der Beitrag einer einzelnen Zufallsvariablen vernachlässigbar klein bleibt. Der zentrale Grenzwertsatz beweist diese Konvergenz unter sehr allgemeinen Randbedingungen (siehe z.B. [60]).
- Die um 1 Stelle verschobene geometrische Verteilung führt auf einen Bernoulli-Prozeß, bei dem zu jedem diskreten Zeitpunkt ein Ereignis mit der Wahrscheinlichkeit 1/x auftritt. Er weist daher — wie der Poisson-Prozeß im kontinuierlichen Wertebereich die Eigenschaft der Gedächtnisfreiheit auf.
- Die Binomial-Verteilung beschreibt die Anzahl der durch einen Bernoulli-Prozeß zu N Zeitpunkten bzw. von N Bernoulli-Prozessen zu einem Zeitpunkt erzeugten Ereignisse. Jedes Bernoulli-Ereignis tritt mit der Wahrscheinlichkeit  $\overline{x}/N$  ein.
- Die Poisson-Verteilung charakterisiert die Anzahl von Ereignissen innerhalb eines Zeitintervalls der Dauer 1, falls die Ereignisabstände negativ-exponentiell mit Mittelwert 1/x verteilt sind. Sie entsteht aus der Binomial-Verteilung für N→∞.

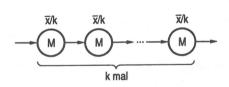


Bild 3.1: Phasenmodell der Erlang-Verteilung k-ter Ordnung

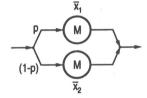


Bild 3.2: Phasenmodell der hyperexponentiellen Verteilung 2. Ordnung

# 3.1.2 Hierarchische Modellierung des Quellverkehrs

In ATM-Netzen findet eine Betriebsmittelteilung sowohl zwischen verschiedenen Verbindungen als auch zwischen den dazugehörigen Datenbüscheln und Zellen statt. Dementsprechend bezieht sich die Steuerung des Verkehrsflusses ebenfalls auf Verbindungen, Datenbüschel und Zellen. Der in Bild 3.3 veranschaulichte hierarchische Modellierungsansatz, der den Verkehrsfluß auf Verbindungs-, Büschel- und Zellebene beschreibt, liegt daher nahe (vgl. auch

[62, 102, 200, 219]). Dieses sehr allgemeine Grundmodell kann nahezu beliebig erweitert werden, indem zusätzliche dienstabhängige Ebenen eingeführt werden. Beispielsweise kann für Multimedia-Dienste eine Rufebene, für Dialogdienste eine Dialogebene [200] und für paketierte Datenübermittlung eine Paketebene [62] definiert werden. Zur Untersuchung der Verkehrssteuerung sind die vorgenannten drei Ebenen ausreichend, da die Rufebene bei der Modellierung der Verbindungsebene und die Dialog- und die Paketebene bei der Modellierung der Büschelebene einbezogen werden können.

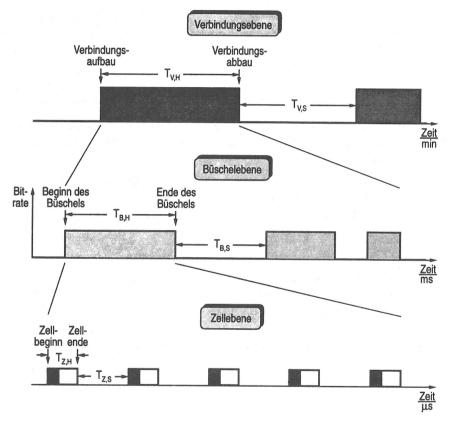


Bild 3.3: Hierarchische Beschreibung einer ATM-Verkehrsquelle

Die statistischen Verkehrsschwankungen auf diesen drei Ebenen werden von verschiedenen Faktoren (Teilnehmerverhalten, Dienst, Protokoll bzw. Codierverfahren) beeinflußt und weisen vielfältige Abhängigkeiten auf. Dies drückt sich in der hierarchischen Strukturierung des Quellmodells aus, d.h. die Verkehrsschwankungen auf einer Ebene beeinflussen alle darunterliegenden Ebenen. Für den in Bild 3.3 veranschaulichten Spezialfall, bei dem der An-

kunftsprozeß auf jeder Ebene zwischen einem aktiven und einem passiven Zustand schwankt, wird dies besonders deutlich: Ein Datenaustausch auf der Büschel- bzw. Zellebene kann nur dann stattfinden, wenn eine Verbindung besteht. Ebenso wird der Zellabstand durch die statistischen Schwankungen des Bitratenbedarfs der Büschelebene moduliert. Allerdings sind für die Büschelebene auch einfachere Modelle mit nur einem Zustand (CBR-Dienste) bzw. komplexere Zustandsmodelle mit mehreren Übertragungsbitraten denkbar (vgl. Unterabschnitt 3.1.2.2).

Die Verweildauern in den aktiven und passiven Zuständen werden durch die Zufallsvariablen  $T_{a,H}$  bzw.  $T_{a,S}$  (a=V,B,Z) beschrieben. Die Indizes V,B und Z bezeichnen die Zugehörigkeit zur Verbindungs-, Büschel- oder Zellebene. Die zeitliche Dynamik der einzelnen Modellebenen weist große Unterschiede auf. Dabei reichen die typischen Zeitkonstanten der Lastschwankungen von Mikrosekunden für die Zellabstände bis hin zu Stunden für die Verbindungsdauern.

Diese extremen Zeitunterschiede bereiten Probleme bei einer gleichzeitigen Untersuchung aller drei Abstraktionsebenen. Andererseits wird dadurch eine getrennte Betrachtung (Dekomposition) der unterschiedlichen Ebenen möglich. Die Verkehrsanalyse kann sich daher auf eine oder zwei Ebenen konzentrieren, wobei der Lastzustand der darüberliegenden Ebenen festgehalten wird und die Lastschwankungen der darunterliegenden Ebenen vernachlässigt werden. Dies führt zu entsprechenden Teilmodellen mit verschiedenen Abstraktionsniveaus, die in den nachstehenden Unterabschnitten vorgestellt werden.

## 3.1.2.1 Modellierung der Verbindungsebene

Wie in Bild 3.3 angedeutet, alterniert die Verbindungsebene zwischen einem aktiven Zustand und einem Ruhezustand, in dem keine Verbindung besteht. Ein Wechsel zwischen diesen Zuständen wird durch einen Verbindungsauf- bzw. -abbau ausgelöst. Häufig werden die Verweildauern in diesen Zuständen durch negativ-exponentielle Verteilungen approximiert, was die Systemanalyse stark vereinfacht. In diesem Fall spricht man auch von Zufallsverkehr zweiter Art. Dieser geht in Zufallsverkehr erster Art über, falls der Überlagerungsverkehr von unendlich vielen Teilnehmern mit jeweils verschwindend kleiner Anrufrate erzeugt wird [150]. In diesem Fall ist der Abstand zwischen zwei aufeinanderfolgenden Verbindungswünschen negativ-exponentiell verteilt.

Obwohl dieses einfache Modell viele Abhängigkeiten wie z.B. die Anrufwiederholung im Blockierungsfall bzw. die zeitlichen Beziehungen verschiedener Verbindungen vernachlässigt und nur den stationären Fall während einer repräsentativen "Hauptverkehrsstunde" [184] beschreibt, hat es sich bei der Untersuchung und Dimensionierung von durchschaltevermittelnden Netzen bewährt.

#### 3.1.2.2 Modellierung der Büschelebene

Die Büschelebene charakterisiert die statistischen Schwankungen des Bitratenbedarfs der Verkehrsquelle. Grundsätzlich lassen sich Dienste mit konstanter, zeitunabhängiger Bitrate und Dienste mit zeitlich variabler Bitrate (VBR) unterscheiden. Eine zeitlich variable Bitrate ergibt sich bei Dialog-Diensten durch den wechselseitigen Dialog, bei Datendiensten infolge der Segmentierung der Daten in den höheren Protokollschichten und bei Video- und Sprachdiensten, falls redundanzmindernde Codierverfahren eingesetzt werden. Diese Vielfalt der Ursachen führt zu einer starken Variation des Quellverhaltens auf der Büschelebene. Stellvertretend sollen folgende Beispiele genannt werden:

- Bei Datendiensten und Sprachübertragung mit Pausenunterdrückung ergeben sich ähnliche Lastwechsel wie auf der Verbindungsebene (vgl. Abschnitt 2.1.2).
- Für VBR-Videodienste hängt die Bitrate direkt vom Informationsgehalt des Bildes ab, weshalb sich die Bitrate zwischen 0 und einer vorgegebenen Maximalbitrate bewegen kann. Da die Codierverfahren das Bild oft block- oder zeilenweise codieren, kann die Bitrate auch während eines Bildes schwanken. In diesem Fall ergeben sich Verkehrsschwankungen und -korrelationen in den unterschiedlichsten Zeitbereichen (Zeilen, Blöcke, Bilder, Szenen ...). Die Nachbildung dieses Verhaltens führt zu sehr aufwendigen Quellmodellen [82]. In zunehmendem Maße wird jedoch eine Zwischenspeicherung der Bildinformation vorgenommen, um die Bitrate während eines Bildes konstant halten zu können. Dies führt zu einer Verminderung der Bitratenstreuung und somit zu einer Erhöhung des zulässigen Verkehrsangebots [207, 224].

Alle diese Beispiele haben die stufenweise Änderung der Bitrate gemeinsam. Daher wird die Büschelebene zweckmäßigerweise durch einen endlichen Zustandsautomaten mit L+1 Zuständen beschrieben, wobei jedem Zustand l, l=0,1,...,L, eine bestimmte Zellrate r(l) zugeordnet ist. Der Zustandsautomat verbleibt für eine allgemein und unabhängig verteilte Zeitdauer  $T_B(l)$  in einem bestimmten Quellzustand l. Die mittlere Verweildauer im Zustand l wird mit  $t_B(l)$  bezeichnet. Die Übergangswahrscheinlichkeiten zwischen den einzelnen Quellzuständen werden in einer Übergangsmatrix Q=[q(m|l)] (l,m=0,1,...,L) zusammengefaßt. Dabei bezeichnet q(m|l) die Wahrscheinlichkeit für einen Übergang vom Zustand l in den Zustand m. Ein entsprechender Zustandsautomat mit drei Zuständen ist in Bild 3.4(a) veranschaulicht.

### 3.1.2.3 Modellierung der Zellebene

Die zeitliche Folge der Zellen auf der Übertragungsleitung kann durch deren Zwischenankunftsabstände  $T_{Z,A} = T_{Z,H} + T_{Z,S}$  beschrieben werden. Für unabhängig und identisch verteilte Ankunftsabstände gehört der Zellankunftsprozeß zur Klasse der Erneuerungsprozesse. Die Modellierung der Zellebene beruht in der Regel auf zeitdiskreten Erneuerungsprozessen, da die Zellen innerhalb des durch die Zellübertragungsdauer  $T_{Z,H}=\Delta t$  vorgegebenen Zeitrasters  $i\cdot \Delta t$ , i=0,1,..., gesendet werden.

Die von einer CBR-Verkehrsquelle produzierten Zellankünfte können im Idealfall durch einen deterministischen Ankunftsprozeß modelliert werden. Der Jitter der dadurch entsteht, daß der Zellabstand kein ganzzahliges Vielfaches der Zellübertragungsdauer beträgt, wird hierbei vernachlässigt. Für einfache Betrachtungen auf Zellebene kann dieser deterministische Ankunftsprozeß durch einen Bernoulli-Prozeß bzw. sein zeitkontinuierliches Analogon, den Poisson-Prozeß, approximiert werden.

Eine Modellierung von Verkehrsquellen mit stufenförmiger Variation der Bitrate erlaubt der allgemein modulierte deterministische Prozeß (Generally Modulated Deterministic Process, GMDP) [146]. Dieser Prozeß modelliert sowohl das Verhalten auf Büschel- als auch auf Zellebene. Die Beschreibung der Büschelebene basiert auf dem im vorigen Unterabschnitt vorgestellten Zustandsautomaten mit L+1 Zuständen. Im Zustand l ergeben sich konstante Zellabstände  $t_Z(l)$ , entsprechend dem Kehrwert der Zellrate r(l). Die Zustandsdauern  $T_B(l)$  werden in dieser Arbeit durch zeitdiskrete Verteilungen beschrieben. Die Diskretisierungseinheit entspricht dem für den jeweiligen Zustand gültigen Zellabstand  $t_Z(l)$ , so daß während der Verweildauer im Zustand l eine ganzzahlige Anzahl von Zellen erzeugt wird. Die Verweildauer im Ruhezustand kann beliebig verteilt sein. Das Verhalten eines dreistufigen allgemein modulierten deterministischen Prozesses ist in Bild 3.4 wiedergegeben.

Im allgemeinen Fall gehört der GMDP nicht zur Klasse der Erneuerungsprozesse, da die Modulation auf Büschelebene zu hochgradig korrelierten Ankunftsabständen führt. Falls

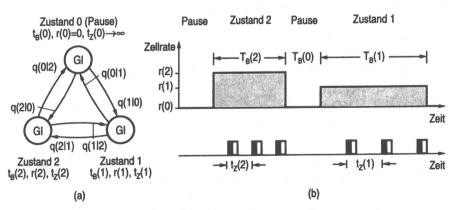


Bild 3.4: Allgemein modulierter deterministischer Prozeß mit drei Zuständen

- (a) Zustands-Übergangs-Modell
- (b) Zeitliches Verhalten auf Büschel- und Zellebene

die Anzahl der innerhalb eines Zustands gesendeten Zellen einer geometrischen Verteilung (GEO<sub>1</sub>) gehorcht, kann das Korrelationsverhalten einfach berechnet werden [25]. Die Erzeugende Funktion und die Momente der Ankunftsabstände sind in [231] abgeleitet.

Ein wichtiger Sonderfall des GMDP ist die sporadische Quelle (auch "Burst-Silence-", "Talkspurt-Silence-" oder "On/Off-Quelle" genannt), bei der sich Aktivitäts- und Ruhephasen abwechseln. Die Anzahl der während einer aktiven Phase gesendeten Zellen ist geometrisch (GEO<sub>1</sub>) verteilt, d.h. nach jeder gesendeten Zelle findet mit gleicher Wahrscheinlichkeit ein Übergang in den Ruhezustand statt. Der resultierende Ankunftsprozeß besitzt daher die Erneuerungseigenschaft. Die Ruhephase kann negativ-exponentiell oder geometrisch verteilt sein. Dieses Modell wurde ursprünglich für paketierte Sprachübertragung entwickelt [91], läßt sich allerdings auch auf Datendienste übertragen.

# 3.2 Modellierung des Kommunikationsnetzes

Die oben beschriebenen Modellierungsebenen spiegeln sich auch in der Beschreibung des Systemverhaltens wider, d.h. auch das ATM-Netz kann auf Verbindungs-, Büschel- und Zellebene modelliert werden. Je nach Anwendungsfall wird ein einzelner Verbindungsabschnitt oder eine ganze Referenzverbindung betrachtet.

# 3.2.1 Modellierung eines Verbindungsabschnitts

Das Grundmodell für die nachfolgenden Leistungsuntersuchungen ist der asynchrone Multiplexer, der die Zellen verschiedener Verbindungen auf eine nach dem asynchronen Multiplexprinzip betriebene Ausgangsleitung umsetzt. Ein typisches Beispiel für einen asynchronen Multiplexer ist der Ausgangspuffer eines Koppelelements mit der daran angeschlossenen Übertragungsleitung. Andere Pufferungskonzepte wie Eingangspuffer mit Mehrfachzugriff oder kombinierte Ein-/Ausgangspuffer weisen eine ähnliche Leistungsfähigkeit auf, lassen sich aber einfacher realisieren [232]. Auch diese Pufferungsstrategien können daher näherungsweise durch das generische Modell des asynchronen bzw. statistischen Multiplexers nach Bild 3.5 beschrieben werden.

Der Multiplexer ist durch die endliche Puffergröße S und die konstante Zellübertragungsdauer  $\Delta t$  festgelegt. Da die Einhaltung der Zellreihenfolge vorgeschrieben ist, werden die Zellen in ihrer Ankunftsreihenfolge aus dem Pufferspeicher entnommen (First In First Out, FIFO). Mit dem Multiplexer sind N unabhängige Verkehrsquellen direkt verbunden. Dies führt zu einer leichten Überschätzung der Zellverluste, da sich die auf einer gemeinsamen Eingangsleitung befindlichen Verbindungen nicht gegenseitig blockieren können [207]. Bei dem vorliegenden zeitdiskreten Modell können Zellankünfte und Bedienenden gleichzeitig

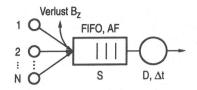


Bild 3.5: Verkehrstheoretisches Modell eines statistischen Multiplexers

auftreten. Deshalb muß festgelegt werden, ob zuerst das Ankunftsereignis (Arrival First, AF) oder das Bedienendeereignis (Departure First, DF) bearbeitet wird [76].

Eine einfache Klassifikation von einstufigen Bediensystemen ist mit Hilfe der erweiterten Kendallschen Notation (siehe z.B. [150]) möglich. Diese erlaubt es, die Ankunfts- und Bedienprozesse, die Anzahl der Bedieneinheiten und die Größe der Warteschlange zu spezifizieren. Für das vorliegende Warte-Verlustsystem mit N allgemeinen (General, G) Ankunftsprozessen, deterministischem Bedienprozeß, einer Bedieneinheit und S Warteplätzen lautet die Kurznotation  $N\cdot G/D/1/(S+1)$ . Für Verlustsysteme wird die Größe des Wartespeichers explizit mit S=0 angegeben, wogegen bei reinen Wartesystemen  $(S\to\infty)$  keine Angaben gemacht werden. Eine genauere Beschreibung des Ankunftsprozesses ist mit den in Tabelle 3.1 eingeführten Abkürzungen möglich. Eine von der FIFO-Strategie abweichende Abfertigungsdisziplin muß ausdrücklich angegeben werden. In der vorliegenden Arbeit wird ferner von der AF-Strategie für die Behandlung gleichzeitiger Ankunfts- und Bedienendeereignisse ausgegangen, falls keine gegenteiligen Angaben gemacht werden.

Auf der Büschelebene ist der diskrete Charakter der Zellankünfte und -bedienungen nicht mehr sichtbar und wird durch einen kontinuierlichen Zu- bzw. Abfluß von Information ersetzt. Diese makroskopische Betrachtungsweise führt auf das "Fluid Flow"- oder "UAS"- Modell (Uniform Arrival and Service, UAS) [4, 236]. Die Zuflußrate ist durch die Summe der Zellraten aller N Verbindungen bestimmt, und die Abflußrate entspricht dem Zelltakt des Übertragungsabschnitts. Falls die Zuflußrate größer als die Abflußrate ist, steigt der Füllstand des Puffers an, während er im umgekehrten Fall abnimmt. Beim Erreichen des maximalen Pufferfüllstands S geht die Differenz aus Zu- und Abflußrate verloren.

Da die typische Zeitkonstante der Laständerungen auf Büschelebene die Übertragungsdauer einer Zelle um Größenordnungen übersteigt, wird die Zwischenpufferung der Zellen häufig ganz vernachlässigt (S=0). Das hierdurch entstehende Verlustsystem kann sehr einfach analysiert werden, da die momentane Zuflußrate den Systemzustand vollständig beschreibt. Sobald die Zuflußrate die Abflußrate übersteigt, geht der überschüssige Verkehr verloren. Im umgekehrten Fall treten keine Verluste auf. Auftretende Verluste verteilen sich proportional zur momentanen Ankunftsrate auf die einzelnen Verbindungen.

Auf Verbindungsebene verhält sich ein ATM-Netz ähnlich wie ein durchschaltevermittelndes

Netz, d.h. es wird als reines Verlustsystem betrieben. Beim Eintreffen eines neuen Verbindungswunsches prüft die Verbindungsannahmefunktion, ob die Verbindung auf dem betrachteten Verbindungsabschnitt zulässig ist oder abgewiesen werden muß und damit verloren geht. Diese Annahmeentscheidung hängt von den Verkehrs- und Netzgütespezifikationen aller Verbindungen — einschließlich der neuen Verbindung — ab.

Für homogenen Verkehr, d.h. alle Verbindungen haben identische Verkehrs- und Netzgütevorgaben, kann eine eindeutige Obergrenze  $\widehat{N}$  für die Anzahl zulässiger Verbindungen angegeben werden. Eine neue Verbindung wird nur dann angenommen, falls  $N<\widehat{N}$  Verbindungen bestehen. Mit der zusätzlichen Annahme eines Poisson-Ankunftsprozesses und allgemein und unabhängig verteilten Bedienzeiten (General Independent, GI) kann der Verbindungsabschnitt durch ein Verlustsystem vom Typ  $M/GI/\widehat{N}/\widehat{N}$  modelliert werden. Die Analyse dieses Systems führt auf die bekannte Erlang-Zustandsverteilung [150].

Ein heterogener Verkehrsmix, der sich aus K verschiedenen Verkehrsklassen zusammensetzt, führt auf eine K-dimensionale Beschreibung des Verbindungszustands durch den Verbindungsvektor  $(N_1, N_2, ..., N_K)$ . Die Komponente  $N_k$ , k = 1, 2, ..., K, bezeichnet die Anzahl der Verbindungen aus Verkehrsklasse k, wobei  $N_1 + N_2 + ... + N_K = N$  gilt. Die zulässigen Verbindungszustände sind durch die Verbindungsannahmefunktion definiert und werden durch eine (K-1)-dimensionale Grenzkurve, die Verbindungsannahmekurve, abgegrenzt. Die analytische Lösung dieses Verlustsystems gelingt für Poisson-Ankünfte und führt auf eine mehrdimensionale Erlang-Zustandsverteilung [245].

# 3.2.2 Netzmodellierung

Aufbauend auf den im vorigen Abschnitt beschriebenen Verkehrsmodellen einzelner Verbindungsabschnitte können Gesamtmodelle für eine ganze Verbindung bzw. ein ganzes ATM-Netz entworfen werden. Hierzu müssen die Teilmodelle der Verbindungsabschnitte entsprechend der Topologie des ATM-Netzes miteinander verknüpft werden. Das hierdurch entstehende Verkehrsmodell beschreibt die physikalische Netzstruktur. Anschließend wird für jede Verkehrsquelle ein bestimmter Weg durch das Netz festgelegt, dem alle von dieser Quelle ausgehenden Anforderungen (Zellen, Büschel oder Verbindungswünsche) folgen.

Dieses allgemeine Vorgehen soll exemplarisch für die Zellebene erläutert werden. In diesem Fall kann das ATM-Netz durch ein Warteschlangennetz nachgebildet werden, bei dem jede Bedienstation einen Verbindungsabschnitt des ATM-Netzes repräsentiert. Die Bedienstationen sind entsprechend der Struktur des ATM-Netzes miteinander verbunden, und die virtuellen Verbindungen markieren den Weg der von den Verkehrsquellen ausgehenden Zellströme. Für ein reales ATM-Netz ist ein solches Warteschlangenmodell nicht mehr handhabbar, weshalb im folgenden Kapitel nur eine einzelne Referenzverbindung betrachtet wird.

Das Warteschlangenmodell vereinfacht sich damit auf die in Bild 3.6 skizzierte Kette aus J Bedienstationen. Die Referenzverbindung wird durch den von anderen Verbindungen eingekoppelten Querverkehr beeinflußt.

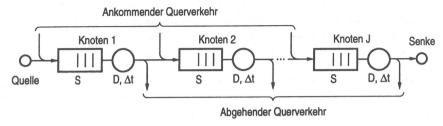


Bild 3.6: Verkehrstheoretisches Modell einer Referenzverbindung auf Zellebene

## 3.3 Simulationstechnik

Die ereignisgesteuerte Simulation ist eine universell einsetzbare Methode zur Leistungsuntersuchung von Rechner- und Kommunikationssystemen. Hierzu wird das Verkehrsmodell einschließlich des zufallsabhängigen Ablaufgeschehens auf einem Rechner nachgebildet. Die während der Simulation durchgeführten Verkehrs- und Leistungsmessungen lassen auch für komplexe Systemmodelle, die keiner mathematischen Analyse zugänglich sind, quantitative Aussagen zu. Ferner ist eine Validierung von analytischen Ansätzen, die auf vereinfachenden Modellannahmen bzw. approximativen Lösungsverfahren beruhen, möglich. Im folgenden wird die in dieser Arbeit verwendete Methodik der stationären Systemsimulation erörtert; eine ausführliche Behandlung dieser Thematik ist in [151, 192] zu finden.

Ein Simulationsprogramm bildet die strukturellen Komponenten der im vorigen Unterkapitel beschriebenen Verkehrsmodelle (z.B. Verkehrsquellen, Warteschlangen, Bedieneinheiten oder Zellen) auf geeignete Datenstrukturen ab. Diese speichern die Parameter, den momentanen Zustand und die Meßzähler der zugeordneten Modellkomponente. Aus den Meßzählern werden nach dem Ablauf der Simulation statistische Leistungskenngrößen wie die Zellverlustwahrscheinlichkeit oder die mittlere Wartezeit ermittelt. Die funktionalen Abläufe und die erforderlichen Meßfunktionen werden durch modellspezifische Programmprozeduren nachgebildet, welche die Datenstrukturen entsprechend manipulieren.

Der zeitliche Ablauf der Simulation wird durch Ereignisse wie "Ankunft einer Zelle" oder "Ende einer Zellübertragung" und den sich daraus ergebenden Folgeereignissen gesteuert. Die Ereignisse werden mit Hilfe eines Ereignis-Kalenders verwaltet, in den alle zukünftigen Ereignisse sequentiell einsortiert werden. Das Simulationsprogramm bearbeitet diese Ereignisse chronologisch und überspringt die zwischen den Ereignissen liegende Zeit. Dazu nimmt

die Ablaufsteuerung das jeweils nächste Ereignis aus dem Kalender und setzt die Systemzeit auf den zugehörigen Wert. Anschließend werden die durch dieses Ereignis ausgelösten Zustandsänderungen vorgenommen, die Meßzähler entsprechend aktualisiert und die daraus abgeleiteten Folgeereignisse im Ereignis-Kalender vorgeplant. Die Zeitspanne zwischen dem momentanen Ereignis und einem daraus hervorgehenden Folgeereignis (Bedienzeit, Ankunftsabstand) ist meist zufallsabhängig und muß entsprechend einer vorgegebenen Verteilung mit Hilfe eines Pseudo-Zufallszahlengenerators "ausgewürfelt" werden.

Diese stochastische Komponente führt zu Unsicherheiten bei der meßtechnischen Erfassung der Leistungskenngrößen. Mit Hilfe von Methoden aus der beurteilenden Statistik kann die Abweichung des Meßwerts vom tatsächlichen Wert quantifiziert werden. Dazu wird der Simulationslauf in mehrere hinreichend lange Teiltests unterteilt, die getrennt voneinander ausgewertet werden. Mit Hilfe der Stichprobentheorie kann aus diesen Einzelergebnissen (Stichproben) das Gesamtergebnis und das dazugehörige Vertrauensintervall bestimmt werden [151, 192]. Das Vertrauensintervall gibt den "Toleranzbereich" an, in dem sich der tatsächliche Wert mit einer bestimmten Wahrscheinlichkeit, der statistischen Aussagesicherheit, befindet.

Die Größe des Vertrauensintervalls nimmt zu, je seltener ein Ereignis auftritt. Für eine direkte Erfassung der angestrebten Zellverlustwahrscheinlichkeiten im Bereich von  $10^{-10}$  müßten ca.  $10^{12}$  Zellen simuliert werden, falls die Zellverluste unkorreliert auftreten [81]. Die durch die Büschelebene verursachte hochgradige Korrelation der Zellabstände führt allerdings zu büschelförmigen Zellverlusten, weshalb der obige Wert mit der mittleren Anzahl der innerhalb eines solchen "Verlustbüschels" verlorenen Zellen multipliziert werden muß. Dieser Faktor ist in erster Näherung proportional zur mittleren Büscheldauer (vgl. die Ergebnisse im folgenden Kapitel). Mit den heutigen skalaren Rechnern können bei vertretbarem Rechenzeitaufwand (weniger als 1 Tag) nur Zellverlustwahrscheinlichkeiten größer  $10^{-6}$  mit zufriedenstellender Genauigkeit erfaßt werden. Ähnliche Einschränkungen gelten für die gleichzeitige Simulation von Verbindungs-, Büschel- und Zellebene.

In letzter Zeit wurde deshalb eine Reihe von Verfahren zur beschleunigten Simulation seltener Ereignisse vorgeschlagen, die folgendermaßen klassifiziert werden können [207]:

- Hybride Verfahren, die Simulation und Analyse kombinieren [143].
- Statistische Verfahren, welche die Streuung des Ergebnisses reduzieren [182, 241].
- Verfahren, die eine Extrapolation der Meßergebnisse nach bestimmten Gesetzmäßigkeiten vornehmen.

Eine ausgezeichnete Übersicht über die grundlegenden Methoden und die verschiedenen Ansätze kann [207] entnommen werden. Einige dieser Ansätze wurden auch für die Simulation von einfachen ATM-Verkehrsmodellen eingesetzt [143, 182, 241]. Die Anwendung dieser

Verfahren ist jedoch auf Spezialfälle begrenzt und kann zu einer signifikanten Verfälschung der Ergebnisse führen, falls die für diese Verfahren geltenden Randbedingungen und Einschränkungen nicht beachtet werden.

Daher wurde in dieser Arbeit die oben eingeführte "Standard-Simulationstechnik" verwendet, für die eine umfangreiche Simulationsbibliothek mit vorgefertigten Programm-Modulen zur Verfügung stand. Diese Module übernehmen die Verkehrs- und Zufallszahlenerzeugung, die Kalenderverwaltung, die Messung statistischer Kenngrößen und die statistische Auswertung des Simulationslaufs, so daß nur noch die modellspezifischen Datenstrukturen und Programmprozeduren implementiert werden müssen. Ferner ist die oben angegebene Meßgrenze von  $10^{-6}$  vollkommen ausreichend für die angestrebte Validierung analytischer Ansätze.

## 3.4 Grundlagen der verkehrstheoretischen Analyse

Der analytischen Behandlung der oben diskutierten Verkehrsmodelle kommt im Hinblick auf die vorliegende Problemstellung eine Schlüsselrolle zu. Analytische Zusammenhänge zwischen der Verkehrsbelastung und der daraus resultierenden Netzgüte sind überaus wichtig für das grundlegende Verständnis des statistischen Multiplexens, für die Definition relevanter Verkehrs- und Netzgüteparameter und für den Entwurf adäquater Verkehrssteuerungsmechanismen. Insbesondere können einfache analytische Ansätze die Basis für die bei der Verbindungsannahme bzw. beim Bitratenmanagement von virtuellen Pfaden eingesetzten Algorithmen bilden. Im folgenden wird eine kurze Einführung in die in dieser Arbeit verwendeten analytischen Methoden gegeben. Eine systematische und ausführliche Zusammenstellung verkehrstheoretischer Methoden kann der einschlägigen Fachliteratur entnommen werden [45, 78, 139, 150, 206].

Für die verkehrstheoretische Systemanalyse muß im ersten Schritt eine geeignete Beschreibung des zufallsabhängigen zeitlichen Ablaufgeschehens gefunden werden. Hierzu wird ein stochastischer Zustandsprozeß  $\{X(t),\,t\geq 0\}$  definiert, der die zeitliche Änderung des Systemzustands X beschreibt. Der Zustand X(t)=x charakterisiert beispielsweise die Anzahl der sich zum Zeitpunkt t im System befindlichen Anforderungen. Abhängig vom Verkehrsmodell kann sowohl der Systemzustand X als auch die Zeit t diskrete oder kontinuierliche Werte annehmen. Den Kern der Verkehrsanalyse bildet die Bestimmung der Verteilungsfunktion  $P\{X(t)\leq x\}$  oder — für diskrete Zustandswerte — der Verteilung  $P\{X(t)=x\}$ . Aus den Zustandswahrscheinlichkeiten können dann einfache statistische Leistungskenngrößen wie die Verlustwahrscheinlichkeit oder die mittlere Wartezeit ermittelt werden.

Die prinzipielle Vorgehensweise soll anhand des für diese Arbeit wichtigen Spezialfalls eines zeitdiskreten stochastischen Prozesses mit diskreten Zustandswerten x, x=0,1,..., erläutert werden. Der stochastische Prozeß wird zu aufeinanderfolgenden Zeitpunkten  $t_i=i\cdot \Delta t$ ,

i=0,1,..., beobachtet. Die zwischen diesen Zeitpunkten stattfindenden Zustandsübergänge werden durch Übergangswahrscheinlichkeiten beschrieben. Die Übergangswahrscheinlichkeit ist die bedingte Wahrscheinlichkeit  $P\{X(t_{i+1})=x_{i+1}|X(t_i)=x_i,...,X(t_0)=x_0\}$ , daß das System zum folgenden Zeitpunkt  $t_{i+1}$  in den Zustand  $x_{i+1}$  übergeht, unter der Bedingung, daß es sich zu den vorangegangenen Zeitpunkten  $t_j, j=0,1,...,i$ , in den Zuständen  $x_j$  befunden hat. Nach dem Gesetz der bedingten Wahrscheinlichkeit ergibt sich damit folgender Zusammenhang, aus dem die Zustandswahrscheinlichkeiten rekursiv ausgerechnet werden könnten:

$$P\{X(t_{i+1}) = x_{i+1}, ..., X(t_0) = x_0\} = P\{X(t_i) = x_i, ..., X(t_0) = x_0\} \cdot P\{X(t_{i+1}) = x_{i+1} | X(t_i) = x_i, ..., X(t_0) = x_0\} ,$$

$$i = 0, 1, ...$$
(3.15)

Die oben beschriebene prinzipielle Vorgehensweise ist allerdings nur dann praktikabel, wenn die zukünftige Entwicklung des stochastischen Prozesses durch eine begrenzte Anzahl von Vorgängerzuständen bestimmt ist. Der einfachste Fall ist der Markoff- $Proze\beta$ , der die Eigenschaft der Gedächtnisfreiheit aufweist, d.h. seine zukünftige Entwicklung ist nur vom gegenwärtigen Zustand und nicht von seiner Vergangenheit abhängig. Mit anderen Worten: Der Zustand zum folgenden Zeitpunkt  $t_{i+1}$  hängt nur vom momentanen Zustand  $X(t_i)$  und nicht von den Zuständen zu früheren Zeitpunkten  $t_0, t_1, ..., t_{i-1}$  ab. Bei diskreten Zustandswerten spricht man daher von einer Markoff-Kette. Für die Übergangswahrscheinlichkeiten läßt sich dieses Verhalten in die folgende mathematische Schreibweise fassen:

$$P\{X(t_{i+1}) = x_{i+1} | X(t_i) = x_i, ..., X(t_0) = x_0\} = P\{X(t_{i+1}) = x_{i+1} | X(t_i) = x_i\}$$

$$= q(x_{i+1} | x_i).$$
(3.16)

An dieser Stelle wurde die zusätzliche Annahme gemacht, daß es sich um einen homogenen Markoff-Prozeß handelt, bei dem die Übergangswahrscheinlichkeiten zeitunabhängig sind. Damit kann Gleichung 3.15 in die folgende vereinfachte Form gebracht werden:

$$P\{X(t_{i+1}) = x_{i+1}, X(t_i) = x_i\} = P\{X(t_i) = x_i\} \cdot q(x_{i+1}|x_i).$$
(3.17)

Nach dem Gesetz der totalen Wahrscheinlichkeit ergibt sich schließlich der folgende Zusammenhang zwischen den Zustandsverteilungen zu den Zeitpunkten  $t_i$  und  $t_{i+1}$ :

$$P\{X(t_{i+1}) = x_{i+1}\} = \sum_{x_i} P\{X(t_i) = x_i\} \cdot q(x_{i+1}|x_i) .$$
 (3.18)

Diese Gleichung ermöglicht eine rekursive Bestimmung der Zustandswahrscheinlichkeiten ausgehend von den zum Zeitpunkt  $t_0$  vorgegebenen Zustandswahrscheinlichkeiten. Für den stationären Fall  $(t_i \to \infty)$  werden die Zustandswahrscheinlichkeiten zeitunabhängig, d.h. es gilt  $P\{X(t_{i+1}) = x\} = P\{X(t_i) = x\} = P\{X = x\} = p(x)$ . Die obige Gleichung führt dann

auf das folgende lineare Gleichungssystem zur Bestimmung der Zustandswahrscheinlichkeiten:

$$p(y) = \sum_{x} p(x) \cdot q(y|x), \quad x, y = 0, 1, \dots$$
 (3.19)

Häufig werden die Zustandswahrscheinlichkeiten durch den Zeilenvektor P = (p(0), p(1), ...) und die Übergangswahrscheinlichkeiten durch die Matrix Q = [q(y|x)], x, y = 0, 1, ..., ausgedrückt, womit Gleichung 3.19 in die folgende Darstellungsform überführt werden kann:

$$P = P \cdot Q . \tag{3.20}$$

Die Gedächtnisfreiheit der zeitdiskreten Markoff-Kette drückt sich auch dadurch aus, daß die Zustandsdauern geometrisch verteilt sein müssen, was für viele Anwendungsfälle eine unzulässige Einschränkung darstellt. Demgegenüber sind für die Klasse der Semi-Markoff-Prozesse beliebige zeitdiskrete und zeitkontinuierliche Verteilungen der Zustandsdauern zulässig. Unmittelbar nach einem Zustandswechsel verliert auch ein Semi-Markoff-Prozeß sein Gedächtnis, da die gedächtnisbehaftete Zustandsphase abgeschlossen ist. Diese Zeitpunkte werden häufig als Regenerationszeitpunkte bezeichnet, da die zukünftige Prozeßentwicklung nur von dem zu diesem Zeitpunkt bestehenden Zustand abhängig ist. Da sich der Prozeß zu den Regenerationszeitpunkten wie die obige Markoff-Kette verhält, spricht man von einer eingebetteten Markoff-Kette. Die für die zeitdiskrete Markoff-Kette abgeleiteten Gleichungen können unverändert für die eingebettete Markoff-Kette übernommen werden, falls die Beobachtungszeitpunkte  $t_i$  so gewählt werden, daß sie mit den Regenerationszeitpunkten zusammenfallen.

Abschließend soll die zeitkontinuierliche Markoff-Kette betrachtet werden, deren Zustandsdauern negativ-exponentiell verteilt sind. Die negativ-exponentielle Verteilung ergibt sich als Grenzfall der geometrischen Verteilung, indem der Grenzübergang  $\Delta t \to 0$  durchgeführt wird. Ebenso entsteht die zeitkontinuierliche Markoff-Kette aus der zeitdiskreten Markoff-Kette, indem die Zeitdauer  $\Delta t$  zwischen zwei aufeinanderfolgenden Zeitpunkten  $t_i$  und  $t_{i+1}$  infinitesimal klein gewählt wird. Damit kann aus Gleichung 3.18 die folgende Beziehung, die unter der Bezeichnung Kolmogoroff-Vorwärts-Gleichung bekannt ist, abgeleitet werden:

$$\lim_{\Delta t \to 0} P\{X(t + \Delta t) = y\} = \lim_{\Delta t \to 0} \sum_{x} P\{X(t) = x\} \cdot q(y|x)$$

$$\lim_{\Delta t \to 0} \frac{P\{X(t + \Delta t) = y\} - P\{X(t) = y\}}{\Delta t} = \lim_{\Delta t \to 0} \sum_{x \neq y} P\{X(t) = x\} \cdot \frac{q(y|x)}{\Delta t} - P\{X(t) = y\} \cdot \frac{1 - q(y|y)}{\Delta t}$$

$$\frac{dP\{X(t) = y\}}{dt} = \sum_{x \neq y} P\{X(t) = x\} \cdot \lambda(y|x) - P\{X(t) = y\} \cdot \sum_{x \neq y} \lambda(x|y), \qquad (3.21)$$

wobei die für die geometrische Verteilung gültigen Übergangswahrscheinlichkeiten in die für die negativ-exponentielle Verteilung gültigen Übergangsraten

$$\lim_{\Delta t \to 0} \frac{q(y|x)}{\Delta t} = \lambda(y|x) \quad \text{und} \quad \lim_{\Delta t \to 0} \frac{1 - q(y|y)}{\Delta t} = \lim_{\Delta t \to 0} \sum_{x \neq y} \frac{q(x|y)}{\Delta t} = \sum_{x \neq y} \lambda(x|y) \quad (3.22)$$

übergehen. Die Übergangsrate  $\lambda(y|x)$  bezieht sich auf die Übergänge vom Zustand x in den Zustand y.

Für den stationären Fall  $(t \to \infty)$  werden die Zustandswahrscheinlichkeiten wiederum zeitunabhängig, d.h. es gilt  $P\{X(t) = x\} = p(x)$ . Die Kolmogoroff-Vorwärts-Gleichungen gehen damit in die Gleichungen des statistischen Gleichgewichts über:

$$p(y) \sum_{x \neq y} \lambda(x|y) = \sum_{x \neq y} p(x) \cdot \lambda(y|x) . \tag{3.23}$$

Diese Gleichungen erlauben eine einfache Interpretation: Die Übergänge aus dem Zustand x heraus und die Übergänge in den Zustand x hinein stehen im Gleichgewicht.

# Kapitel 4

# Leistungsuntersuchung generischer ATM-Systemmodelle

In diesem Kapitel werden — basierend auf der im vorigen Kapitel vorgestellten Methodik der Verkehrsmodellierung und -analyse — die charakteristischen Leistungskenngrößen eines statistischen Multiplexers bzw. einer ganzen Referenzverbindung ermittelt. Dabei konzentrieren sich die Untersuchungen hauptsächlich auf die Fragestellung des statistischen Multiplexens von Verbindungen mit variabler Bitrate. Die daraus gewonnenen Erkenntnisse und Schlußfolgerungen sind bei der Konzeption der Verkehrssteuerung von Nutzen.

## 4.1 Leistungsanalyse eines statistischen Multiplexers

Die große Bedeutung des statistischen Multiplexers für die Modellierung und Leistungsbewertung von ATM-Netzen wird aus den zahlreichen Veröffentlichungen, die sich mit dieser Thematik befassen, ersichtlich. Eine Übersicht über die Modellierungsaspekte der wichtigsten analytischen Lösungsansätze wurde in Tabelle 4.1 zusammengestellt.

	Homogener Verkehrsmix CBR-Quellen   Sporad. Quellen		Heterogener Verkehrsmix CBR-Quellen   Sporad. Quellen	
Endliche Puffergröße	[43, 132, 250]	[12, 100, 156, 175] [198, 225, 236, 260]		[14, 226]
Unendliche Puffergröße	[21, 50, 53] [243]	[4, 18, 46, 61] [91, 157, 181]	[209, 242]	[140, 220]

Tabelle 4.1: Modellierungsaspekte analytischer Lösungsansätze für statistische Multiplexer

Die Überlagerung von verschiedenen CBR-Verkehrsströmen mit gleichen Zellabständen (homogener Verkehrsmix) führt zu einem periodischen Ankunftsmuster der Zellen, das durch die Phasenlage der einzelnen Verkehrsquellen bestimmt ist. Die Belegung des Multiplexers weist dieselbe Periodizität auf und ist abhängig vom betrachteten Ankunftsmuster der Zellen. Der Puffer ist beispielsweise permanent leer, falls pro Zeitschlitz maximal eine Quelle sendet. Andererseits kann ein maximaler Pufferfüllstand von N-1 Zellen nur dann erreicht werden, wenn alle Verbindungen im gleichen Zeitschlitz eine Zelle senden.

Die starre Phasenlage führt außerdem dazu, daß in jeder Periode dieselben Verbindungen Zellverluste erleiden, d.h. einige Verbindungen verlieren alle Zellen, während die restlichen Verbindungen ungestört übertragen können. Analytische Ansätze berechnen daher mittlere Leistungskenngrößen, die alle möglichen Phasenlagen und deren Auftrittswahrscheinlichkeiten berücksichtigen. Die Auftrittswahrscheinlichkeit eines bestimmten Ankunftsmusters ergibt sich aus der Tatsache, daß eine Quelle einen der zur Verfügung stehenden Zeitschlitze zufällig, d.h. gleichverteilt, auswählt.

Für den Spezialfall des  $N\cdot D/D/1$ -Systems wurden sehr effiziente exakte Lösungsalgorithmen zur Bestimmung der Verteilung des Pufferfüllstands vorgeschlagen [21, 50, 53, 243]. Dagegen sind für endliche Puffergrößen nur relativ aufwendige exakte Lösungsalgorithmen [43, 132], bzw. eine auf dem System mit unendlicher Puffergröße basierende Näherungslösung [250], bekannt. Falls CBR-Verkehrsströme mit unterschiedlicher Periodendauer überlagert werden  $(\sum_k N_k \cdot D_k/D/1$ -System), sind — mit endlichem Aufwand — nur approximative Lösungen möglich [209, 242].

Das statistische Multiplexen von VBR-Verkehrsquellen wurde, von wenigen Ausnahmen abgesehen, ausschließlich für die Überlagerung von sporadischen Verkehrsquellen (vgl. Unterabschnitt 3.1.2.3) analytisch behandelt. Die Komplexität dieses relativ einfachen Modells ist immer noch so groß, daß nur approximative Lösungsmethoden bekannt sind, die folgendermaßen klassifiziert werden können:

• Lösung mit Hilfe einer zweidimensionalen eingebetteten Markoff-Kette: Dieser Ansatz beschreibt das System zu den eingebetteten Zeitpunkten durch die Anzahl aktiver Quellen und den Füllstand des Puffers. Da diese Systembeschreibung den Ankunftsprozeß nur unvollständig charakterisiert, müssen zusätzliche Näherungsannahmen getroffen werden. In [260] werden die Zellankünfte der aktiven Quellen innerhalb des durch den minimalen Zellabstand vorgegebenen Zeitintervalls zufällig verteilt. Mit dieser Annahme kann zu den Intervallgrenzen eine Markoff-Kette eingebettet werden. Das resultierende lineare Gleichungssystem wird mit Hilfe numerischer Verfahren aufgelöst.

In [46] wird dagegen ein gleichförmiger Ankunftsstrom angenommen, d.h. der Pufferfüllstand ändert sich nicht, falls genau so viele Quellen im aktiven Zustand sind, daß die Ankunftsrate der Bedienrate entspricht. Andererseits wird der Pufferfüllstand

entsprechend der überschüssigen Zellrate (Differenz aus Ankunfts- und Bedienrate) inkrementiert, falls die momentane Ankunftsrate über der Bedienrate liegt. Im umgekehrten Fall gelten analoge Überlegungen für die Verringerung des Pufferfüllstands. Mit diesen Annahmen kann wiederum eine Markoff-Kette eingebettet werden, wobei sich zu den Regenerationszeitpunkten entweder der Pufferfüllstand oder die Anzahl aktiver Quellen ändert. Die mathematische Behandlung des Systems basiert auf den Methoden der Matrixgeometrie [178]. Eine direkte numerische Auflösung der Gleichungen ist nicht möglich, da der Puffer als unendlich groß angenommen wird. Dieser Ansatz überschätzt — ähnlich wie das "Fluid Flow"-Modell — die Wahrscheinlichkeit, daß das System leer ist, da die kurzzeitigen Fluktuationen des Ankunftsprozesses vernachlässigt werden.

• Anpassung eines Ersatzankunftsprozesses: Dieser Lösungsansatz versucht, die statistischen Eigenschaften des Überlagerungsverkehrs durch einen analytisch handhabbaren Ersatzankunftsprozeß nachzubilden. Die einfachsten Ansätze ersetzen den von hinreichend vielen Verkehrsquellen erzeugten Überlagerungsverkehr durch einen Poissonbzw. einen Bernoulli-Ankunftsprozeß. Hierbei wird die Tatsache ausgenutzt, daß der Überlagerungsverkehr von vielen Verkehrsquellen innerhalb eines kleinen Zeitintervalls gegen einen Poisson-Prozeß strebt (local poissonification), falls jede Verkehrsquelle maximal eine Zelle innerhalb dieses Zeitintervalls sendet und einen verschwindend kleinen Beitrag zum Gesamtverkehr leistet. Das resultierende M/D/1/(S+1)-bzw. GEO<sub>1</sub>/D/1/(S+1)-System kann relativ einfach mit Hilfe einer zu den Bedienendezeitpunkten eingebetteten Markoff-Kette gelöst werden (siehe z.B. [200]). Diese einfachen Ersatzmodelle liefern nur für kleine Puffergrößen zufriedenstellende Ergebnisse (vgl. [225] und die in Abschnitt 4.1.3 diskutierten Ergebnisse).

Andere Ersatzprozesse steuern die Ankunftsrate des Poisson-Prozesses durch eine Markoff-Kette mit mehreren Zuständen. Der resultierende Markoff-Modulierte Poisson-Prozeß (MMPP) weist zusätzliche Freiheitsgrade auf und kann auch das Korrelationsverhalten des Überlagerungsverkehrs nachbilden. Der häufig verwendete zweistufige MMPP ist durch vier Parameter — die Ankunftsraten und die mittleren Verweildauern in den beiden Quellzuständen charakterisiert. Diese Parameter werden dadurch festgelegt, daß bestimmte statistische Kenngrößen des MMPPs an den Überlagerungsprozeß angepaßt werden [12, 91, 175]. Die Analyse des resultierenden MMPP/D/1(/(S+1))-Systems basiert auf der Methode der eingebetteten Markoff-Kette. In der Regel werden matrixanalytische Methoden zur numerisch effizienten Auflösung des Gleichungssystems eingesetzt [13, 91, 179]. Die für die Anpassung des Ersatzankunftsprozesses angewandte Methode erweist sich als entscheidend für die Qualität der Ergebnisse. Die in [12] vorgeschlagene Parameteranpassung liefert die genauesten Resultate.

• "Fluid Flow"-Modell: Dieses Modell vernachlässigt den diskreten Charakter der Zell-

ankünfte und -bedienungen und beschreibt das System auf der Büschelebene, indem ein kontinuierlicher Informationszu- und -abfluß angenommen wird. Die Analyse führt auf ein homogenes System linearer Differentialgleichungen erster Ordnung, dessen allgemeine Lösung durch die Eigenwerte und Eigenvektoren der daraus abgeleiteten charakteristischen Matrix gegeben ist. Unter den in Tabelle 4.1 angegebenen Randbedingungen wurden verschiedene exakte [4, 140, 220, 236], approximative [14, 18, 226] und asymptotische Lösungsverfahren [4, 58] abgeleitet. Eine Verallgemeinerung dieser Ansätze ist in Anhang A zu finden. Die Vernachlässigung der durch die Paketierung der Information ausgelösten "mikroskopischen" Verkehrsschwankungen führt zu einer Unterschätzung der charakteristischen Leistungskenngrößen für kleine Puffergrößen [236].

#### 4.1.1 Exakte Analyse

In diesem Abschnitt wird eine exakte analytische Lösung für das Multiplexen von N sporadischen Verkehrsströmen vorgestellt. Prinzipiell könnte dieser Ansatz auf den allgemeinen Fall mit heterogener Verkehrszusammensetzung übertragen werden, jedoch führt die exponentielle Zunahme des Zustandsraums dazu, daß selbst bei homogenen Verkehren Parametereinschränkungen hingenommen werden müssen. Mit vertretbarem Aufwand lassen sich nur Systeme mit relativ großer Spitzenbitrate lösen. Andererseits erlaubt diese exakte Lösung wichtige Rückschlüsse für den im nachfolgenden Abschnitt besprochenen Dekompositionsansatz.

## 4.1.1.1 Modellierungsaspekte

Die analytische Behandlung des Verkehrsmodells nach Bild 3.5 wird hauptsächlich durch die Komplexität des Ankunftsprozesses erschwert. Der Ankunftsverkehr wird durch N sporadische Quellen erzeugt, die auf der Büschelebene zwischen einem Sende- und einem Ruhezustand alternieren. Die Ruhedauer kann nur ganzzahlige Vielfache der elementaren Zeiteinheit  $\Delta t$  betragen und weist eine um 1 Stelle verschobene geometrische Verteilung (GEO<sub>1</sub>) mit dem Mittelwert  $t_{B,S}$  auf. Im Sendezustand wird eine GEO<sub>1</sub>-verteilte Anzahl von Zellen mit einem äquidistanten Zellabstand von d Zeitschlitzen generiert. Die mittlere Anzahl der innerhalb einer Sendephase generierten Zellen wird mit b abgekürzt.

Die sporadische Quelle ist zwar gedächtnisbehaftet, weist jedoch die Erneuerungseigenschaft auf, da innerhalb der Sendephase eine geometrisch verteilte Anzahl von Zellen erzeugt wird. Nachdem eine Zelle gesendet wurde, verbleibt die Quelle mit Wahrscheinlichkeit 1-1/b im Sendezustand bzw. verläßt diesen mit der komplementären Wahrscheinlichkeit 1/b (vgl. Tabelle 3.1). Ebenso wird im Ruhezustand zu jedem Zeitschlitz entschieden, ob die Quelle im Ruhezustand bleibt (Wahrscheinlichkeit  $1-\Delta t/t_{B,S}$ ) oder in den aktiven Zustand übergeht

(Wahrscheinlichkeit  $\Delta t/t_{B,S}$ ). Der Ankunftsabstand zweier aufeinanderfolgender Zellen kann daher mit Hilfe des in Bild 4.1 dargestellten Phasenmodells charakterisiert werden.

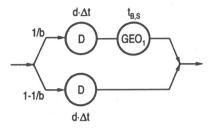


Bild 4.1: Phasenmodell für die Ankunftsabstände der sporadischen Verkehrsquelle

#### 4.1.1.2 Analytischer Ansatz

Der Zellankunftsprozeß wird nun zu aufeinanderfolgenden Zeitpunkten  $t_i = i \cdot \Delta t, i = 0, 1, ...,$  beobachtet, wobei die Anzahl der zum Zeitpunkt  $t_i$  ankommenden Zellen durch die Zufallsvariable  $A(t_i)$  beschrieben wird. Nach Gleichung 3.15 gilt

$$\begin{split} \mathbb{P}\{A(t_{i+1}) = a_{i+1},...,A(t_0) = a_0\} &= \mathbb{P}\{A(t_i) = a_i,...,A(t_0) = a_0\} \cdot \\ & \mathbb{P}\{A(t_{i+1}) = a_{i+1} | A(t_i) = a_i,...,A(t_0) = a_0\} \;, \quad (4.1) \\ & a_i = 0,1,...,N \;, \quad i = 0,1,... \end{split}$$

Die Anzahl der zum folgenden Zeitpunkt  $t_{i+1}$  eintreffenden Zellen hängt im vorliegenden Fall nur von den Zellankünften zu den Zeitpunkten  $t_{i-d+1}, t_{i-d+2}, ..., t_i$  ab. Dieses zeitlich begrenzte "Gedächtnis" des Ankunftsprozesses geht aus dem oben beschriebenen Phasenmodell hervor: Jede Quelle, die zum Zeitpunkt  $t_{i-d+1}$  eine Zelle gesendet hat, sendet mit Wahrscheinlichkeit 1-1/b zum Zeitpunkt  $t_{i+1}$  eine Zelle, da sie in der Sendephase bleibt; jede Quelle, die keine Zelle innerhalb des Zeitintervalls  $[t_{i-d+1}, t_i]$  gesendet hat, befindet sich in der Ruhephase und sendet mit Wahrscheinlichkeit  $\Delta t/t_{B,S}$  eine Zelle zum Zeitpunkt  $t_{i+1}$ . Die Übergangswahrscheinlichkeiten können daher — analog zur zeitdiskreten Markoff-Kette — folgendermaßen vereinfacht werden:

$$P\{A(t_{i+1}) = a_{i+1} | A(t_i) = a_i, ..., A(t_0) = a_0\}$$

$$= P\{A(t_{i+1}) = a_{i+1} | A(t_i) = a_i, ..., A(t_{i-d+1}) = a_{i-d+1}\}$$

$$= q_A \left(a_{i+1} | a_{i-d+1}, N - \sum_{j=i-d+1}^{i} a_j\right). \tag{4.2}$$

Die  $a_{i+1}$  zum Zeitpunkt  $t_{i+1}$  ankommenden Zellen setzen sich demnach aus zwei Anteilen zusammen:

• Zum einen bleiben j der  $a_{i-d+1}$  zum Zeitpunkt  $t_{i-d+1}$  aktiven Quellen im Sendezustand und senden auch zum Zeitpunkt  $t_{i+1}$  eine Zelle. Die Wahrscheinlichkeit für diese Ankünfte kann durch eine Binomial-Verteilung charakterisiert werden. Mit Hilfe von Tabelle 3.1 erhält man

$$q_{A,1}(j|a_{i-d+1}) = \binom{a_{i-d+1}}{j} \left(1 - \frac{1}{b}\right)^j \cdot \left(\frac{1}{b}\right)^{a_{i-d+1}-j}, \quad j = 0, 1, ..., a_{i-d+1}. \tag{4.3}$$

Ferner gehen k der insgesamt N - \(\sum\_{j=i-d+1}^i a\_j = N - A\_i\) passiven Quellen zum Zeitpunkt t<sub>i+1</sub> in den aktiven Zustand \(\text{über}\). Analog zu den obigen \(\text{Überlegungen k\text{\text{\text{onnen}}}}\) die von diesen Quellen stammenden Zellen aus der folgenden Binomial-Verteilung bestimmt werden:

$$q_{A,2}(k|N-A_i) = \binom{N-A_i}{k} \left(\frac{\Delta t}{t_{B,S}}\right)^k \cdot \left(1 - \frac{\Delta t}{t_{B,S}}\right)^{N-A_i-k} \ , \quad k = 0,1,...,N-A_i \ . \ (4.4)$$

Da die Quellen als statistisch unabhängig angenommen werden, ergeben sich die gesuchten Übergangswahrscheinlichkeiten aus einer Faltung dieser Binomial-Verteilungen, d.h. für die Übergangswahrscheinlichkeiten müssen alle möglichen Kombinationen  $j+k=a_{i+1}$  berücksichtigt werden.

$$q_{A}(a_{i+1}|a_{i-d+1}, N - A_{i}) = \sum_{j=\max(0, a_{i+1} - N + A_{i})}^{\min(a_{i+1}, a_{i-d+1})} \binom{a_{i-d+1}}{j} \left(1 - \frac{1}{b}\right)^{j} \cdot \left(\frac{1}{b}\right)^{a_{i-d+1} - j} \cdot \binom{N - A_{i}}{a_{i+1} - j} \binom{N - A_{i}}{t_{B,S}} \cdot \left(1 - \frac{\Delta t}{t_{B,S}}\right)^{N - A_{i} - a_{i+1} + j},$$

$$a_{i+1} = 0, 1, \dots, N - A_{i} + a_{i-d+1}. \tag{4.5}$$

Mit der abkürzenden Schreibweise  $p_A(a_{i+1},...,a_{i-d+1}) = P\{A(t_{i+1}) = a_{i+1},...,A(t_{i-d+1}) = a_{i-d+1}\}$  kann Gleichung 4.1 in

$$p_A(a_{i+1},...,a_{i-d+1}) = p_A(a_i,...,a_{i-d+1}) \cdot q_A\left(a_{i+1}|a_{i-d+1},N - \sum_{j=i-d+1}^i a_j\right)$$
(4.6)

umgeformt werden. Falls alle möglichen Ausgangszustände  $a_{i-d+1}$  berücksichtigt werden, ergibt sich nach dem Gesetz der totalen Wahrscheinlichkeit der folgende Zusammenhang, aus dem der Zellankunftsprozeß rekursiv bestimmt werden kann:

$$p_A(a_{i+1}, ..., a_{i-d+2}) = \sum_{a_{i-d+1}=0}^{N} p_A(a_i, ..., a_{i-d+1}) \cdot q_A\left(a_{i+1} | a_{i-d+1}, N - \sum_{j=i-d+1}^{i} a_j\right) . \quad (4.7)$$

Für die Untersuchung des statistischen Multiplexers muß der Belegungszustand, d.h. die Anzahl der sich in der Bedieneinheit und der Warteschlange befindlichen Zellen, in die Beschreibung des Systemzustands einbezogen werden. Der Belegungszustand zum folgenden Zeitpunkt  $t_{i+1}$  ergibt sich eindeutig aus dem aktuellen Belegungszustand  $X(t_i) = x_i$  und den  $a_{i+1}$  neu eintreffenden Zellen. Der Systemzustand kann daher schrittweise aus dem zum vorigen Zeitpunkt gültigen Systemzustand mit Hilfe der folgenden Rekursionsbeziehung bestimmt werden:

$$p_{G}(x_{i+1}, a_{i+1}, ..., a_{i-d+2}) = \sum_{a_{i-d+1}=0}^{N} p_{G}(x_{i}, a_{i}, ..., a_{i-d+1}) \cdot q_{A} \left( a_{i+1} | a_{i-d+1}, N - \sum_{j=i-d+1}^{i} a_{j} \right),$$

$$x_{i} = 0, 1, ..., S + 1$$

$$(4.8)$$

und

$$x_{i+1} = \begin{cases} \min(S+1, x_i + a_{i+1}) - \min(1, x_i) & \text{für die Bearbeitungsreihenfolge AF} \\ \min(S+1, x_i + a_{i+1} - \min(1, x_i)) & \text{für die Bearbeitungsreihenfolge DF} \end{cases}$$
(4.9)

Der Belegungszustand des Multiplexers erhöht sich um die  $a_{i+1}$  neu eintreffenden Zellen, wobei die endliche Größe des Pufferspeichers den maximalen Füllstand auf S+1 Zellen begrenzt (1 Zelle in der Bedieneinheit, S Zellen im Wartespeicher). Gleichzeitig tragen Zellen, die das System verlassen, zu einer Verringerung des Belegungszustands um  $\min(1, x_i)$  Zellen bei. Die Bildung des Minimums drückt die Tatsache aus, daß der Belegungszustand nur dann dekrementiert wird, falls sich zum vorangegangenen Taktzeitpunkt  $t_i$  eine Zelle im System befunden hat  $(x_i > 0)$ . Ferner ist der Belegungszustand abhängig von der Reihenfolge, in welcher gleichzeitige Zellankünfte und -abgänge bearbeitet werden.

Die Strategie DF (Departure First) erweist sich im Hinblick auf die Zellverlustwahrscheinlichkeit günstiger als die Strategie AF (Arrival First), da eine Zelle, die das System verläßt, ihren Platz für eine im gleichen Zeitschlitz ankommende Zelle freigibt. Aus Gleichung 4.9 geht hervor, daß die Bearbeitungsreihenfolge DF bei einer um einen Speicherplatz reduzierten Puffergröße auf dasselbe Systemverhalten wie die Bearbeitungsreihenfolge AF führt, falls die Anzahl der in einem Zeitschlitz eintreffenden Zellen auf den Wert S+1 begrenzt ist. Für alle anderen Fälle gilt diese Beziehung zumindest näherungsweise. Eine ausführliche Diskussion dieser Strategien ist in [76, 200, 232] zu finden.

Im stationären Fall,  $i \to \infty$ , werden die Zustandswahrscheinlichkeiten zeitunabhängig und es könnte — analog zur zeitdiskreten Markoff-Kette — ein lineares Gleichungssystem zur Bestimmung der Zustandswahrscheinlichkeiten abgeleitet werden. Da der Zustandsraum allerdings extrem groß werden kann und die Übergangsmatrix nur äußerst dünn besetzt ist, bietet sich eine iterative Bestimmung der Zustandswahrscheinlichkeiten an. Dazu wird Gleichung 4.8, ausgehend vom Leerzustand ( $p_G(0,0,...,0)=1$ ), schrittweise ausgewertet. Da diese Iteration den Einschwingvorgang des Systems nachvollzieht, konvergieren die Zustandswahrscheinlichkeiten auf ihre stationären Werte. Die Iteration wird abgebrochen, wenn die Summe der absoluten Wahrscheinlichkeitsänderungen zwischen zwei aufeinanderfolgenden

Iterationsschritten unter eine Schranke von  $10^{-10}$  fällt. Die Verteilung des Belegungszustands des Multiplexers errechnet sich aus:

$$p_{Z}(x) = \sum_{a_{i+1}=0}^{N} \sum_{a_{i}=0}^{N} \cdots \sum_{a_{i-d+2}=0}^{N} p_{G}(x, a_{i+1}, a_{i}, ..., a_{i-d+2}).$$
 (4.10)

#### 4.1.1.3 Charakteristische Kenngrößen

Abschließend sollen die wichtigsten Parameter und Leistungskenngrößen des untersuchten Systems abgeleitet werden:

#### • Verkehrsparameter:

Da die sporadische Quelle ein ausgeprägtes büschelförmiges Verhalten aufweist, wird sie häufig mit Hilfe des Büschelfaktors Z beschrieben, der das Verhältnis aus maximaler und mittlerer Bitrate angibt. Für die sporadische Quelle errechnet sich die mittlere Bitrate aus der maximalen Bitrate und der Wahrscheinlichkeit, daß sich die Quelle im Sendezustand befindet. Mit Hilfe der mittleren Sendedauer  $b \cdot d \cdot \Delta t$  und der mittleren Ruhedauer  $t_{B,S}$  ergibt sich schließlich

$$Z = \frac{b \cdot d \cdot \Delta t + t_{B,S}}{b \cdot d \cdot \Delta t} \,. \tag{4.11}$$

Das Verkehrsangebot  $A_Z$ , d.h. die mittlere Anzahl von Zellen, die während eines beliebigen Zeitschlitzes ankommen, stellt einen weiteren wichtigen Verkehrsparameter dar. Da eine Quelle im Mittel alle  $d \cdot Z$  Zeitschlitze eine Zelle sendet, erhält man

$$A_Z = \frac{N}{d \cdot Z} \ . \tag{4.12}$$

#### • Leistungskenngrößen des Multiplexers:

Die Auslastung des Multiplexers ergibt sich aus der Verteilung des Belegungszustands. Von den  $x \ge 1$  innerhalb des Systems befindlichen Zellen wird eine Zelle bedient, während die restlichen x-1 Zellen warten. Die mittlere Auslastung der Bedieneinheit bzw. der Warteschlange wird durch den Verkehrswert

$$Y_Z = \sum_{x=1}^{S+1} p_Z(x) = 1 - p_Z(0)$$
 (4.13)

bzw. die mittlere Warteschlangenlänge

$$\Omega_Z = \sum_{x=1}^{S+1} p_Z(x) \cdot (x-1) = \sum_{x=2}^{S+1} p_Z(x) \cdot (x-1)$$
 (4.14)

charakterisiert.

#### • Netzgüteparameter:

Von besonderer Bedeutung sind die Netzgüteparameter der ATM-Schicht, die sich auf den Verlust bzw. die Verzögerung der Zellen beziehen. Da alle Verkehrsquellen identische statistische Eigenschaften aufweisen, erfahren alle ATM-Verbindungen dieselbe Netzgüte. Die Differenz aus dem angebotenen und dem bedienten Verkehr entspricht dem Verkehrsanteil, der aufgrund der endlichen Größe des Puffers verloren geht; bezogen auf den angebotenen Verkehr ergibt sich die Verlustwahrscheinlichkeit

$$B_Z = \frac{A_Z - Y_Z}{A_Z} = 1 - \frac{Y_Z}{A_Z} \,. \tag{4.15}$$

Die mittlere Wartezeit aller angenommenen Zellen kann nach dem Gesetz von Little [150] bestimmt werden. Dieses Gesetz stellt einen allgemeingültigen Zusammenhang her zwischen der mittleren Anzahl der im System befindlichen Anforderungen (mittlere Warteschlangenlänge  $\Omega_Z$ ), der mittleren Verweilzeit einer Anforderung im System (mittlere Wartezeit  $w_Z$ ) und der mittleren Ankunftsrate der Anforderungen, die ins System gelangen ( $Y_Z$  Zellen pro Zeiteinheit  $\Delta t$ ):

$$w_Z = \frac{\Omega_Z}{Y_Z} \cdot \Delta t \ . \tag{4.16}$$

#### 4.1.1.4 Ergebnisse

In diesem Unterabschnitt wird — ausgehend von einem von N=8 Quellen erzeugten Referenzverkehr — der Einfluß der verschiedenen Quellparameter auf die Zellverlustwahrscheinlichkeit systematisch untersucht. Die Verkehrsquellen senden in der aktiven Phase alle d=4 Zeitschlitze eine Zelle. Dies entspricht einer Spitzenbitrate von 34 Mbit/s bei der zur Verfügung stehenden Nutzbitrate von ca. 136 Mbit/s. Diese relativ große Spitzenbitrate mußte gewählt werden, da der numerische Aufwand der exakten Lösung mit steigenden Werten für d exponentiell zunimmt. Der Büschelfaktor und die mittlere Büschelgröße wurden auf die für Video- und Datendienste repräsentativen Werte von Z=5 bzw. b=100 eingestellt. Gleichzeitige Zellankünfte und -abgänge werden in der Bearbeitungsreihenfolge AF behandelt.

Unter diesen Randbedingungen sind zwei verschiedene Blockierungssituationen möglich. Einerseits treten die für die asynchrone Zeitmultiplextechnik charakteristischen Zellkollisionen auf, falls zu einem Zelltakt mehrere Verbindungen auf die Ausgangsleitung des Multiplexers zugreifen wollen. Zusätzlich bewirkt das Prinzip des statistischen Multiplexens langanhaltende Überlastsituationen, falls die momentane Bitrate aller Verbindungen die Übertragungskapazität überschreitet. Eine anschauliche Interpretation der beiden Blockierungseffekte ist in Bild 4.2 gegeben, das die Überlagerung verschiedener Verkehre wiedergibt. Im linken Teil des Bildes treten nur die Blockierungen auf Zellebene auf, da die Gesamtbitrate unter der Bitrate

der Übertragungsleitung liegt. Im rechten Teil des Bildes erkennt man eine langanhaltende Überlastsituation, bei der über einen relativ großen Zeitraum mehr Zellen ankommen als bedient werden können.

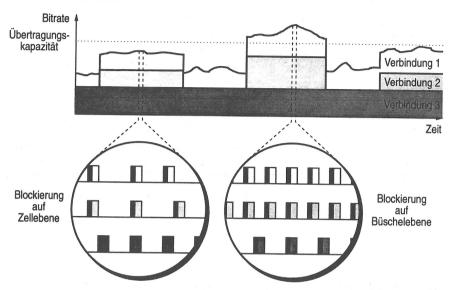


Bild 4.2: Veranschaulichung der auf Büschel- und Zellebene auftretenden Blockierungseffekte

Diese grundlegend unterschiedlichen Blockierungseffekte führen zu dem in Bild 4.3 gezeigten charakteristischen Verlauf der Zellverlustwahrscheinlichkeit in Abhängigkeit von der Puffergröße. Für kleine Puffergrößen resultieren die Zellverluste überwiegend aus den durch die Asynchronität der Zellankünfte verursachten kurzzeitigen Verkehrsschwankungen. Diese Komponente der Zellverlustwahrscheinlichkeit klingt bei einer Vergrößerung des Puffers schnell ab, da der Puffer die kurzzeitigen Verkehrsschwankungen kompensieren kann. Die von der Büschelebene herrührenden, langanhaltenden Verkehrsschwankungen können dagegen nur durch relativ große Puffer ausgeglichen werden. Deshalb werden die durch eine Überlastsituation auf Büschelebene ausgelösten Zellverluste ab einer gewissen Puffergröße dominant. Dies führt zu zwei deutlich unterscheidbaren Kurvenbereichen mit stark unterschiedlicher Kurvensteigung.

Die Ergebnisse in Bild 4.3 zeigen zudem, daß die Steigung im hinteren Teil der Kurve von der mittleren Büschelgröße abhängt. Die Verlustwahrscheinlichkeit nimmt umso stärker ab, je weniger Zellen pro Datenbüschel gesendet werden, d.h. je kürzer die auf der Büschelebene auftretenden Überlastsituationen sind. Diese Resultate bestätigen, daß sich die von

einigen Verkehrsformungsverfahren angestrebte Begrenzung der Büschelgröße günstig auf die Leistungskenngrößen des Multiplexers auswirkt. Das skizzierte Verhalten wird an den folgenden Extremfällen besonders deutlich:

- Falls die Büschelgröße auf eine Zelle begrenzt wird, treten keine langanhaltenden Überlastsituationen auf, d.h. die Kurve fällt gleichmäßig ab und weist keine Änderung der Steigung für große Puffergrößen auf.
- Falls die mittlere Dauer der Büschel- und Ruhezustände gegen unendlich strebt, führt dies zu einem horizontalen Kurvenverlauf, wenn die Puffergröße den zum Ausgleich der kurzzeitigen Verkehrsschwankungen erforderlichen Minimalwert überschreitet. Dieser Sonderfall wurde in [100, 101] unter der Bezeichnung quasistationäre Näherung behandelt.

Der Einfluß des Büschelfaktors auf die Zellverlustwahrscheinlichkeit des statistischen Multiplexers ist in Bild 4.4 wiedergegeben. Die Anzahl der Verkehrsquellen, der minimale Zellabstand und die mittlere Büschelgröße wurden konstant gehalten, womit sich bei steigendem Büschelfaktor eine Verringerung des Verkehrsangebots einstellt. Die Änderung des Verkehrsangebots macht sich sowohl bei den Anfangswerten als auch bei den Steigungen in beiden Kurvenabschnitten bemerkbar. Mit anderen Worten: Sowohl die Blockierungen auf Zellebene als auch die Blockierungen auf Büschelebene sind lastabhängig.

Ferner zeigen diese Ergebnisse, daß für büschelförmigen Verkehr ein Gewinn durch statistisches Multiplexen erzielt werden kann, d.h. bei vorgegebener Zellverlustwahrscheinlichkeit können mehr Verbindungen zugelassen werden, falls der Büschelfaktor der Quellen zunimmt.

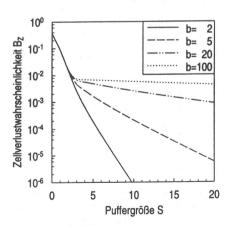


Bild 4.3: Einfluß der Büschelgröße auf die Zellverlustwahrscheinlichkeit

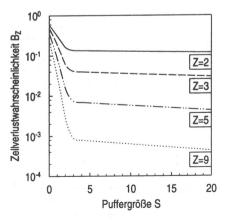


Bild 4.4: Einfluß des Büschelfaktors auf die Zellverlustwahrscheinlichkeit

Allerdings kann diese Zunahme der Verbindungszahl die Abnahme des Verkehrsangebots einer einzelnen Verbindung nicht vollständig kompensieren, so daß sich eine Erhöhung des Büschelfaktors letztendlich negativ auf das zulässige Verkehrsangebot auswirkt.

Ein weiterer wichtiger Verkehrsparameter ist die Spitzenbitrate der Verkehrsquellen, die in Bild 4.5 variiert wurde. Gleichzeitig wurde die Anzahl der Verkehrsquellen so angepaßt, daß das Verkehrsangebot den konstanten Wert  $A_Z=0,4$  Erlang annimmt. Da das Verkehrsangebot festgehalten wird, bleiben die Zellverluste infolge der kurzzeitigen Verkehrsschwankungen auf Zellebene nahezu unbeeinflußt. Der hintere Kurvenast weist hingegen eine Parallelverschiebung auf, d.h. die Anfangswerte sind von der Spitzenbitrate der Quellen abhängig, wogegen die Steigung nahezu unverändert beibehalten wird. Dieses Verhalten bleibt auch für größere Werte von d erhalten, wie die in Bild 4.14 dargestellten Ergebnisse belegen. Die Ergebnisse bestätigen ferner, daß sich die von manchen Verfahren zur Verkehrsformung angestrebte Reduktion der Spitzenbitrate der Verkehrsquellen günstig auf die Netzauslastung auswirkt, da ein höherer Multiplexgewinn erzielt werden kann. Dies läßt sich darauf zurückführen, daß sich die Bitratenschwankungen der einzelnen Verkehrsquellen umso besser ausgleichen, je mehr Verkehrsquellen überlagert werden.

Während kurzzeitige Überlastsituationen durch die Netzpuffer überbrückt werden können, müssen die längerfristigen Verkehrsschwankungen der Büschelebene durch geeignete Verbindungsannahmealgorithmen kontrolliert werden, die für eine Begrenzung der Verbindungszahl sorgen. In Bild 4.6 ist die Abhängigkeit der Verlustwahrscheinlichkeit von der Verbindungszahl wiedergegeben. Die Begrenzung des Verkehrsangebots hat einen Einfluß auf die Anfangswerte und die Steigungen in beiden Kurvenbereichen (vgl. auch die Ergebnisse in Bild 4.4).

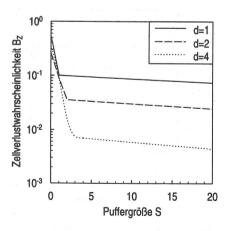


Bild 4.5: Einfluß der Spitzenbitrate auf die Zellverlustwahrscheinlichkeit

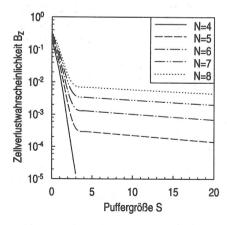


Bild 4.6: Einfluß der Verbindungszahl auf die Zellverlustwahrscheinlichkeit

Falls die Verbindungszahl auf bzw. unter den minimalen Zellabstand d=4 sinkt, steht für jede Verbindung die Spitzenbitrate zur Verfügung und langanhaltende Überlastsituationen können ganz vermieden werden.

## 4.1.2 Leistungsuntersuchung auf Zellebene

Für eine weitergehende Untersuchung der im vorigen Abschnitt angesprochenen Blockierungseffekte ist es vorteilhaft, das Verkehrsmodell in zwei verschiedene Teilmodelle für die Zell- und die Büschelebene zu zerlegen. In der Literatur wurde hierzu eine ganze Reihe verschiedener Ansätze vorgeschlagen [14, 62, 102, 144, 157, 181, 198, 219]. Die auf Zellebene auftretenden Zugriffskonflikte auf die Ausgangsleitung des Multiplexers können mit Hilfe eines M/D/1(/(S+1))- [14, 62, 102, 144, 198, 219], eines GI/D/1- [157] oder eines  $N\cdot D/D/1$ -Systems [181] erfaßt werden. Für die Büschelebene wird der Puffer entweder ganz vernachlässigt [102, 198, 219] oder es wird das in Abschnitt 3.2.1 eingeführte "Fluid Flow"-Modell verwendet [14, 62, 144, 157, 181].

Dieser Abschnitt beschränkt sich auf die durch die asynchronen Zellankünfte ausgelösten Zugriffskonflikte auf die Ausgangsleitung des Multiplexers. Dazu wird jeder Verbindung die Spitzenbitrate zur Verfügung gestellt, um Überlastsituationen auf Büschelebene ausschließen zu können. Die von der Büschelebene ausgelösten längerfristigen Blockierungssituationen sind Gegenstand von Abschnitt 4.1.3.

## 4.1.2.1 Überlagerung homogener sporadischer Quellen

Die Zuweisung der Spitzenbitrate bedeutet, daß die Verbindungszahl N kleiner oder gleich dem minimalen Zellabstand d ist. Diese Eigenschaft führt zu einer wesentlichen Vereinfachung der im vorigen Abschnitt beschriebenen exakten Analyse. Der nachfolgend vorgestellte Ansatz für endliche Puffergröße verallgemeinert die in [43] behandelte Überlagerung von CBR-Verkehrsströmen  $(b \to \infty)$  auf die Überlagerung von sporadischen Quellen. Für unendliche Puffergröße ist ein einfacherer Lösungsansatz bekannt [181].

Die oben eingeführte Randbedingung  $N \leq d$  führt dazu, daß sich innerhalb einer Periode von N aufeinanderfolgenden Zeitschlitzen mindestens ein Zeitpunkt finden läßt, zu dem nur die neu angekommenen Zellen im System sind (vgl. Bild 4.7). Zu diesem Zeitpunkt erneuert sich der Systemzustand x, d.h. er wird unabhängig von den zu früheren Zeitpunkten eingetroffenen Zellen. Dies führt zu einer Entkopplung der einzelnen Perioden, womit sich die Analyse auf einen beliebig gewählten Zeitausschnitt  $[t_1, t_N]$  beschränken kann. Im stationären Fall können die während dieser Periode eintreffenden Zellen schrittweise mit Hilfe der folgenden

Übergangswahrscheinlichkeiten bestimmt werden:

$$q_{A}(a_{i+1}|N-A_{i}) = \binom{N-A_{i}}{a_{i+1}} \left(\frac{1}{d \cdot Z-i}\right)^{a_{i+1}} \cdot \left(1 - \frac{1}{d \cdot Z-i}\right)^{N-A_{i}-a_{i+1}}, \quad i = 0, 1, ..., N-1,$$

$$(4.17)$$

mit

$$A_{i} = \begin{cases} 0 & \text{für } i = 0\\ \sum_{j=1}^{i} a_{j} & \text{für } i = 1, 2, ..., N - 1 \end{cases}$$
 (4.18)

Diese Übergangswahrscheinlichkeit entspricht der bedingten Wahrscheinlichkeit, daß zum Zeitpunkt  $t_{i+1}$  genau  $a_{i+1}$  Zellen am Multiplexer eintreffen, falls zu den i vorhergehenden Zeitschlitzen insgesamt  $A_i$  Zellen angekommen sind. Die angegebene Binomial-Verteilung, die auch für den Fall N>d gilt, kann mit Hilfe von Gleichung 4.7 nachgewiesen werden und läßt sich folgendermaßen interpretieren: Jede der  $N-A_i$  Quellen, die noch keine Zellen gesendet haben, sendet mit einer bestimmten Wahrscheinlichkeit eine Zelle im nächsten Zeitschlitz. Die Sendewahrscheinlichkeit  $1/[d \cdot Z - i]$  errechnet sich aus dem mittleren Zellabstand  $d \cdot Z$  und der Bedingung, daß diese Quellen keine Zelle zu den i vorangegangenen Zeitschlitzen gesendet haben. An dieser Stelle soll betont werden, daß diese Beziehung nur vom minimalen Zellabstand d Gebrauch macht, d.h. die analytische Lösung gilt für beliebige Verteilungen der Sende- und Ruhedauern der Verkehrsquellen (vgl. auch [181]). Außerdem können die Sende- und Ruhephasen beliebige Korrelationen aufweisen.

Da der Belegungszustand zu Beginn des betrachteten Zeitintervalls  $[t_1, t_d]$  nicht bekannt ist, wird ein Ersatzsystem definiert, das zum Zeitpunkt  $t_0$  leer sei. Der Belegungszustand dieses

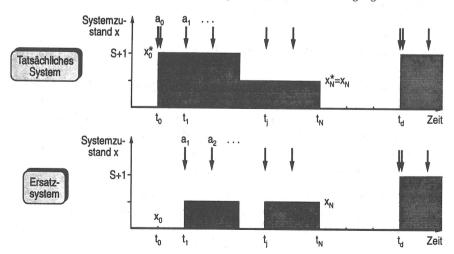


Bild 4.7: Stichproben der Zustandsprozesse des tatsächlichen Systems und des Ersatzsystems ( $N=6,\,d=9,\,S=1,\,$ Bearbeitungsreihenfolge DF)

Ersatzsystems kann — ausgehend vom Anfangszustand  $x_0 = 0$  — mit Hilfe der folgenden aus Gleichung 4.8 abgeleiteten Rekursionsbeziehung bestimmt werden:

$$p_G(x_{i+1}, A_{i+1}) = \sum_{A_i=0}^{A_{i+1}} p_G(x_i, A_i) \cdot q_A(A_{i+1} - A_i | N - A_i) , \quad i = 0, 1, ..., N - 1$$
 (4.19)

und

$$x_{i+1} = \begin{cases} \min(S+1, x_i + A_{i+1} - A_i) - \min(1, x_i) & \text{für die Bearbeitungsreihenfolge AF} \\ \min(S+1, x_i + A_{i+1} - A_i - \min(1, x_i)) & \text{für die Bearbeitungsreihenfolge DF} \end{cases}. \tag{4.20}$$

Das tatsächliche System kann sich allerdings in einem anderen Ausgangszustand  $x_0^* \geq 0$  befinden und weist in diesem Fall ein vom obigen Ersatzsystem abweichendes Systemverhalten auf. Es läßt sich jedoch ein Zeitpunkt  $t_j$   $(1 \leq j \leq N)$  finden, zu dem nur die neu eingetroffenen Zellen im tatsächlichen System sind. Da bei identischem Ankunftsmuster der Zellen der Belegungszustand des Ersatzsystems nur kleiner oder gleich dem Belegungszustand des tatsächlichen Systems sein kann  $(x_i \leq x_i^*)$ , befinden sich zum Zeitpunkt  $t_j$  auch innerhalb des Ersatzsystems nur die neu eingetroffenen Zellen. Ab diesem Zeitpunkt sind daher beide Belegungsprozesse synchronisiert, d.h. es gilt  $x_i = x_i^*$  für  $i \geq j$ . Die beschriebenen Zusammenhänge sind in Bild 4.7 veranschaulicht. Nach dem N-ten Zeitschlitz ist diese Synchronisation der Belegungsprozesse spätestens eingetreten, und das Ersatzsystem befindet sich im eingeschwungenen Zustand. Die Verteilung des Belegungszustands x zu einem beliebig gewählten Zeitpunkt entspricht damit der Verteilung des Belegungszustands  $x_N$ .

$$p_Z(x) = \sum_{A_N=0}^{N} p_G(x, A_N) . (4.21)$$

Die charakteristischen Leistungskenngrößen können aus den im vorigen Abschnitt abgeleiteten Gleichungen 4.11 - 4.16 bestimmt werden.

Innerhalb der ersten  $N \leq d$  Zeitschlitze ist der durch die Überlagerung von sporadischen Verkehrsquellen entstehende Zellankunftsprozeß nicht unterscheidbar von einem durch CBR-Verkehrsquellen mit einem entsprechend vergrößerten Zellabstand von  $d \cdot Z$  Zeitschlitzen generierten Ankunftsverkehr (vgl. Gleichung 4.17). Da die analytische Lösung nur die ersten N Zeitschlitze betrachtet, ergeben sich somit identische Leistungskenngrößen für beide Ankunftsprozesse. Dieses Verhalten wurde bereits in [181] für den Fall unendlicher Puffergröße mit Hilfe der Verteilung der Pufferlänge bewiesen.

#### 4.1.2.2 Heterogene Verkehrszusammensetzung

Für heterogene Ankunftsverkehre werden die für Verbindung n, n = 1, 2, ..., N, gültigen Quellparameter mit dem Index n versehen. Die Spitzenbitratenzuteilung führt in diesem

Fall auf die Bedingung  $\sum_{n=1}^{N} 1/d_n \leq 1$ . Die analytische Behandlung dieses allgemeinen Falls ist nur dann möglich, wenn der minimale Zellabstand aller Verkehrsquellen denselben Wert aufweist, d.h. es gilt  $d_n = d, \forall n = 1, 2, ..., N$ . Das prinzipielle analytische Vorgehen, das auf dem in [207] beschriebenen Ansatz basiert, soll kurz skizziert werden.

Unter den genannten Voraussetzungen kann wiederum ein beliebiges Zeitintervall  $[t_1, t_d]$  betrachtet werden, in dem a, a = 0, 1, ..., N, Zellen eintreffen können. Unter der Bedingung, daß genau a Zellen eintreffen, erhält man die Zustandsverteilung  $p_Z(x|a)$  aus der oben beschriebenen Lösung des  $a \cdot D/D/1/(S+1)$ -Systems. Nach dem Gesetz der totalen Wahrscheinlichkeit müssen diese Teilergebnisse mit der Wahrscheinlichkeit  $p_B(a)$ , daß sich während des betrachteten Zeitintervalls a Quellen im aktiven Zustand befinden, gewichtet werden, um die Zustandsverteilung zu einem beliebigen Zeitpunkt zu erhalten:

$$p_Z(x) = \sum_{a=0}^{N} p_Z(x|a) \cdot p_B(a) . \tag{4.22}$$

Verbindung n sendet während des Zeitintervalls  $[t_1, t_d]$  eine Zelle mit der Wahrscheinlichkeit  $1/Z_n$ . Nach Gleichung 3.12 liefert die Faltung der zugeordneten Bernoulli-Verteilungen die gesuchte Verteilung  $p_B(a)$ , welche die Summe der Zellankünfte innerhalb des betrachteten Zeitintervalls beschreibt.

Für die Überlagerung von Ankunftsverkehren mit unterschiedlicher Spitzenbitrate ist keine exakte analytische Lösung bekannt. Die in [181] für heterogene CBR-Verkehrsströme erhaltenen Ergebnisse deuten darauf hin, daß der homogene Fall eine obere Grenze für die Zellverlustwahrscheinlichkeit liefert, falls das Verkehrsangebot und die Anzahl der Verkehrsquellen konstant gehalten wird. Daher liefert der im vorigen Unterabschnitt diskutierte homogene Verkehrsmix auch für heterogene Ankunftsverkehre eine sichere Abschätzung der Zellverlustwahrscheinlichkeit.

## 4.1.2.3 Einfache Näherungsmodelle

Die numerische Auswertung der für homogene Ankunftsverkehre abgeleiteten Rekursionsbeziehung nach Gleichung 4.19 ist insbesondere für große Zellabstände d aufwendig. Andererseits kann die Abhängigkeit der zu aufeinanderfolgenden Zeitschlitzen eintreffenden Zellen gerade in diesem Fall vernachlässigt werden, wodurch sich die Systemanalyse signifikant vereinfacht, da nur noch der Belegungszustand des Multiplexers in die Systembeschreibung eingeht. Der Ankunftsprozeß, d.h. die Anzahl der zu einem beliebigen Zeitpunkt eintreffenden Zellen, kann durch eine unabhängig und identisch verteilte diskrete Zufallsvariable A beschrieben werden. Diese Zufallsvariable gehorcht einer Binomial-Verteilung, da jede der N Quellen mit der Wahrscheinlichkeit  $1/(d \cdot Z)$  eine Zelle zu dem betreffenden Zeitpunkt

sendet. Aus diesen Überlegungen ergibt sich

$$P\{A = a\} = \binom{N}{a} \left(\frac{1}{d \cdot Z}\right)^{a} \cdot \left(1 - \frac{1}{d \cdot Z}\right)^{N-a}, \quad a = 0, 1, ..., N. \tag{4.23}$$

Die analytische Behandlung des hieraus entstehenden  $D^{[A]}/D/1/(S+1)$ -Systems kann mit Hilfe einer einfachen zeitdiskreten Markoff-Kette erfolgen [23]. Die aus dieser Analyse erhaltene Zellverlustwahrscheinlichkeit überschätzt den tatsächlichen Wert, da die Abhängigkeiten des Ankunftsprozesses zu einer geringfügigen negativen Korrelation der Zellankünfte in aufeinanderfolgenden Zeitschlitzen führen: Falls in den vorangegangenen d-1 Zeitschlitzen viele Zellen angekommen sind, können im folgenden Zeitschlitz nur wenige Quellen eine Zelle generieren und umgekehrt. Damit ist der Belegungszustand des Multiplexers und die Anzahl der neu ankommenden Zellen negativ korreliert, was sich günstig auf die Zellverlustwahrscheinlichkeit auswirkt.

Das  $D^{[A]}/D/1/(S+1)$ -Modell liefert nur dann exakte Resultate, wenn die Anzahl der Verkehrsquellen gegen unendlich strebt, da für diesen Grenzfall keine Korrelation der Zellankünfte auftritt. Einen formalen Beweis für dieses Verhalten liefert der folgende Grenzübergang für den Ankunftsprozeß nach Gleichung 4.17:

$$\lim_{\substack{N \to \infty \\ dZ \to \infty \\ \frac{dZ}{dZ} = A_Z}} \binom{N - A_i}{a_{i+1}} \left(\frac{1}{dZ - i}\right)^{a_{i+1}} \cdot \left(1 - \frac{1}{dZ - i}\right)^{N - A_i - a_{i+1}} = \lim_{\substack{N \to \infty \\ dZ = A_Z}} \binom{N}{a_{i+1}} \left(\frac{A_Z}{N}\right)^{a_{i+1}} \cdot \left(1 - \frac{A_Z}{N}\right)^{N}$$

$$= \frac{A_Z^{a_{i+1}}}{a_{i+1}!} \cdot e^{-A_Z} . \tag{4.24}$$

Die zum Zeitpunkt  $t_{i+1}$  eintreffenden Zellen gehorchen somit einer Poisson-Verteilung, unabhängig von den Zellankünften zu früheren Zeitpunkten. Die nachfolgenden Ergebnisse zeigen, daß das dadurch entstehende  $\mathrm{D}^{[A]}/\mathrm{D}/1/(S+1)$ -Modell mit Poisson verteilten Gruppenankünften eine Obergrenze für die Zellverluste des ursprünglichen Verkehrsmodells liefert. Dieser Grenzwert ist nur abhängig vom Verkehrsangebot  $A_Z$ .

Dieses Modell kann durch sein zeitkontinuierliches Analogon, das M/D/1/(S+1)-Modell, approximiert werden. Bei diesem System sind Ankünfte und Bedienungen zu beliebigen Zeitpunkten möglich. Eine neu eintreffende Zelle, die eine freie Bedieneinheit antrifft, wird unmittelbar bedient und muß nicht bis zum Beginn des nächsten Zeitschlitzes warten. Daher liefert diese Näherung etwas kleinere Zellverlustwahrscheinlichkeiten als das  $D^{[A]}/D/1/(S+1)$ -Modell (für die Bearbeitungsreihenfolge AF).

#### 4.1.2.4 Ergebnisse

In Bild 4.8 wird das asynchrone Multiplexen von deterministischen Verkehrsströmen mit unterschiedlichen Zellabständen d betrachtet. Die Verbindungszahl N wurde so gewählt, daß

das Verkehrsangebot den konstanten Wert  $A_Z=N/d=0,85$  annimmt. Aus den Ergebnissen ist zu entnehmen, daß die Zellverlustwahrscheinlichkeit mit steigendem Zellabstand, d.h. sinkender Bitrate der Verkehrsquellen zunimmt. Der Sonderfall des  $\mathrm{D}^{[A]}/\mathrm{D}/1/(S+1)$ -Systems, d.h.  $N,d\to\infty$ , stellt eine obere Grenze für die Zellverlustwahrscheinlichkeit dieses Systems dar. Die Ergebnisse liegen umso dichter an dieser Schranke, je größer der Zellabstand und je geringer die Auslastung des Systems ist.

Bild 4.9 zeigt ähnliche Resultate für die Überlagerung von sporadischen Verkehrsströmen, falls keine Überlast auf Büschelebene auftreten kann. In diesem Beispiel beträgt der minimale Zellabstand 40 Zeitschlitze. Die Anzahl der Quellen wurde so angepaßt, daß ein Verkehrsangebot  $A_Z=0,25$  beibehalten wird. Das Verhältnis zwischen mittlerer Sende- und mittlerer Ruhedauer wird durch den Büschelfaktor Z festgelegt. Die Absolutwerte dieser Größen haben keinen Einfluß auf die Zellverlustwahrscheinlichkeiten. Es ist zu erkennen, daß die Ergebnisse mit zunehmendem Büschelfaktor näher an die Grenzkurve des  $\mathrm{D}^{[A]}/\mathrm{D}/\mathrm{1}/(S+1)$ -Systems rücken. Dies ist darauf zurückzuführen, daß sich der Multiplexer so verhält, als ob N CBR-Verkehrsquellen mit einem Zellabstand von  $d\cdot Z$  Zeitschlitzen überlagert würden (vgl. Unterabschnitt 4.1.2.1).

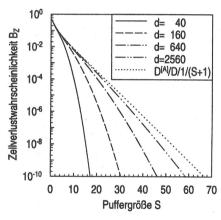


Bild 4.8: Einfluß des Zellabstands von CBR-Verkehrsströmen auf die Zellverlustwahrscheinlichkeit

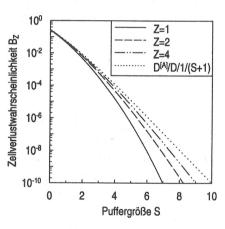
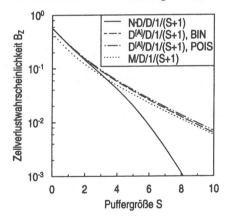


Bild 4.9: Einfluß des Büschelfaktors auf die Zellverlustwahrscheinlichkeit

Ein Vergleich der im vorigen Abschnitt behandelten elementaren Verkehrsmodelle wurde in Bild 4.10 durchgeführt. Als Referenz dienen die Ergebnisse des  $N \cdot D/D/1/(S+1)$ -Systems, die das exakte Systemverhalten wiedergeben. Da die bisherigen Ergebnisse gezeigt haben, daß die Abweichungen für wenige Verkehrsquellen und hohes Verkehrsangebot stärker ausgeprägt sind, wurde  $A_Z=0,85$  und d=40 gewählt.

Falls die negative Korrelation der in aufeinanderfolgenden Zeitschlitzen ankommenden Zellen vernachlässigt wird, führt das obige Modell auf ein  $D^{[A]}/D/1/(S+1)$ -System mit Binomial verteilten Gruppenankünften. Dies führt zu einer Überschätzung der Zellverlustwahrscheinlichkeit. Für unendlich viele Verkehrsquellen geht die Binomial-Verteilung (BIN) in die Poisson-Verteilung (POIS) über, wodurch sich eine weitere geringfügige Erhöhung der Zellverlustwahrscheinlichkeit ergibt. Das zeitkontinuierliche Analogon dieses Modells, das M/D/1/(S+1)-System, liefert etwas kleinere Werte für die Zellverlustwahrscheinlichkeit als das  $D^{[A]}/D/1/(S+1)$ -System. Diese Ergebnisse können mit den im vorhergehenden Unterabschnitt gemachten Bemerkungen erklärt werden.

Für die Systemdimensionierung und -steuerung müssen die oben beschriebenen Zusammenhänge umgekehrt werden, d.h. es ist das maximal zulässige Verkehrsangebot bei vorgegebener Zellverlustwahrscheinlichkeit gesucht. In Bild 4.11 ist das zulässige Verkehrsangebot für die Überlagerung homogener CBR-Verkehre bei einer maximalen Zellverlustwahrscheinlichkeit von  $10^{-10}$  aufgetragen. Für  $N \leq S+1$  Verkehrsquellen treten keine Zellverluste auf und das System kann mit  $\hat{A}_Z=1$  ausgelastet werden. Mit steigender Quellenzahl nimmt das zulässige Verkehrsangebot ab, bis schließlich der Endwert des  $\mathrm{D}^{[A]}/\mathrm{D}/1/(S+1)$ -Systems erreicht wird. Dieser Endwert hängt von der Größe des Puffers ab. Für eine Puffergröße von 64 Zellen kann ein Verkehrsangebot von ca. 0,85 Erlang zugelassen werden.



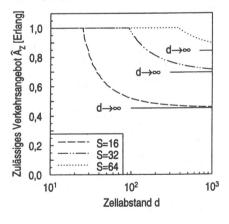


Bild 4.10: Vergleich verschiedener Zellankunftsprozesse

Bild 4.11: Zulässiges Verkehrsangebot für CBR-Verkehr

## 4.1.3 Leistungsuntersuchung auf Büschelebene

Dieser Abschnitt beschäftigt sich mit den durch eine Überlastsituation auf Büschelebene verursachten längerfristigen Blockierungen, die dann auftreten, wenn das Prinzip des stati-

stischen Multiplexens angewandt wird. Auf der Büschelebene kann der statistische Multiplexer durch das in Abschnitt 3.2.1 eingeführte "Fluid Flow"-Modell beschrieben werden, das den diskreten Charakter der Zellankünfte und -bedienungen durch einen kontinuierlichen Zu- bzw. Abfluß von Information ersetzt. Die infolge des asynchronen Netzbetriebs unvermeidlichen kurzzeitigen Zellkollisionen werden von diesem Modell vernachlässigt. Eine ausführliche Beschreibung dieses Modells und die analytische Behandlung für allgemeine Markoff-modulierte Ankunftsprozesse wird aus Gründen der Übersichtlichkeit in Anhang A ausgelagert.

Im folgenden werden die für sporadischen Quellverkehr erzielten Ergebnisse verallgemeinert und vertieft. Die den Untersuchungen zugrundeliegende Referenzquelle wurde bereits in Unterabschnitt 4.1.1.4 eingeführt und ist in Bild 4.12 in Form des Zustands-Übergangs-Diagramms veranschaulicht. Die sporadische Quelle schwankt zwischen einem passiven und einem aktiven Zustand. Die Verweildauern in diesen Zuständen sind negativ-exponentiell verteilt und weisen die Mittelwerte  $1/\alpha$  bzw.  $1/\beta$  auf. Im aktiven Zustand produziert diese Quelle einen kontinuierlichen Zellstrom mit der Ankunftsrate  $\hat{r}$ . Die Quellparameter können aus Bild 4.12 entnommen werden. Die Bedienrate des Multiplexers wurde zu  $C=1,0001/\Delta t$  gewählt, um die Gleichheit zwischen Ankunfts- und Bedienrate ausschließen zu können. Dies ist eine Grundvoraussetzung der in Anhang A beschriebenen Analyse.

Zustands-Übergangs-Diagramm:

# $0 \qquad \hat{r}$

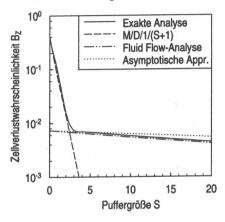
#### Parameterwerte:

 $\alpha = 1 / [(Z - 1) \cdot b \cdot d \cdot \Delta t] = 0,000625 / \Delta t$   $\beta = 1 / [b \cdot d \cdot \Delta t] = 0,0025 / \Delta t$  $\hat{r} = 1 / [d \cdot \Delta t] = 0,25 / \Delta t$ 

Bild 4.12: "Fluid Flow"-Modell der Referenzquelle

Zunächst soll die ebenfalls in Unterabschnitt 4.1.1.4 motivierte Zerlegung der Leistungsmodellierung und -analyse in die Zell- und die Büschelebene validiert werden. Die in Bild 4.13 dargestellten Ergebnisse bestätigen, daß sich die durch das M/D/1/(S+1)- und das "Fluid Flow"-Modell errechneten Verlustwahrscheinlichkeiten gegenseitig ergänzen. Die kurzzeitigen Zellkollisionen können durch das M/D/1/(S+1)-Modell beschrieben werden und sind nur für kleine Puffergrößen entscheidend. Die langzeitigen, von der Büschelebene herrührenden Lastschwankungen werden von diesem Modell vernachlässigt. Bei einer Vergrößerung des Puffers klingt die durch das M/D/1/(S+1)-Modell beschriebene Zellverlustkomponente relativ schnell ab und die durch das "Fluid Flow"-Modell definierte Verlustwahrscheinlichkeit wird dominant. Für große Puffergrößen ist der schon erwähnte exponentielle Abfall der Zellverlustwahrscheinlichkeit zu beobachten. Die durch Gleichung A.51 definierte asymptotische Approximation liefert eine obere Grenze für die Verlustwahrscheinlichkeit auf Büschelebene.

Das folgende Beispiel verdeutlicht den schon in Bild 4.3 untersuchten Einfluß der Büschelgröße auf die Verlustwahrscheinlichkeit. Für das "Fluid Flow"-Modell läßt sich zudem beweisen [156], daß sich die Verlustwahrscheinlichkeit nicht ändert, falls sowohl die Puffergöße als auch die Büschel- und die Ruhedauer mit einem beliebigen Faktor multipliziert werden. Mit anderen Worten: Die Verlustwahrscheinlichkeit hängt nur vom Verhältnis aus Büschelund Puffergröße und nicht von den zugehörigen Absolutwerten ab. Dieser interessante Sachverhalt ist in Bild 4.14 durch entsprechende Hilfslinien veranschaulicht. Die nachfolgend beschriebenen Ergebnisse können daher auch für andere Büschelgrößen interpretiert werden. Beispielsweise müssen die an der Abszisse stehenden Puffergrößen nur halbiert werden, falls die mittlere Büschelgröße der Referenzquelle auf 50 Zellen reduziert wird.



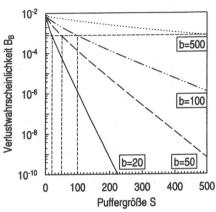


Bild 4.13: Validierung des Dekompositionsansatzes

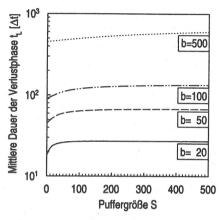
Bild 4.14: Einfluß der Büschelgröße auf die Verlustwahrscheinlichkeit

Neben der Verlustwahrscheinlichkeit hat auch die Korrelation der Verluste Einfluß auf die Dienstgüte der übergelagerten Dienste und Anwendungen (vgl. Abschnitt 2.1.3). Beim "Fluid Flow"-Modell treten nur dann Verluste auf, wenn der Puffer voll belegt ist und die Ankunftsrate die Bedienrate übersteigt. Die Korrelation der Zellverluste drückt sich dadurch aus, daß das System zwischen Phasen, in denen Verluste auftreten, und Phasen, in denen keine Verluste auftreten, alterniert (vgl. Bild A.5). Dieses Verhalten wird mit den folgenden, in Anhang A definierten Meßgrößen erfaßt:

- Mittlere Dauer der Verlustphase
- Mittlerer zeitlicher Abstand zweier Verlustphasen
- Verlustwahrscheinlichkeit innerhalb einer Verlustphase

Die in den Bildern 4.15–4.17 gezeigten Ergebnisse belegen, daß die durch die Zwischenpufferung erzielte Verringerung der Verlustwahrscheinlichkeit ausschließlich auf eine Vergrößerung

des zeitlichen Abstands zweier Verlustphasen zurückzuführen ist. Andererseits äußert sich eine Vergrößerung des Puffers sogar in einer Verlängerung und einer Verstärkung der Verlustphasen (vgl. Bilder 4.15 bzw. 4.17). Dieses auf den ersten Blick paradoxe Systemverhalten läßt sich damit erklären, daß nur extreme Überlastsituationen zu einem Pufferüberlauf führen, wogegen kurzzeitige Blockierungen durch den Puffer ausgeglichen werden können. Die in Bild A.5 skizzierte Stichprobe des Zustandsprozesses veranschaulicht diesen Sachverhalt. Über ähnliche Ergebnisse wird auch in der Literatur berichtet [20, 24, 28, 156].



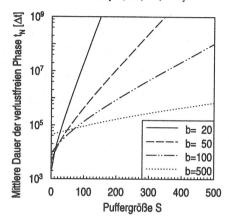


Bild 4.15: Mittlere Dauer der Verlustphase

Bild 4.16: Mittlere Dauer der verlustfreien Phase

Für die weitergehende Interpretation dieser Zusammenhänge sind die im folgenden abgeleiteten Näherungsbeziehungen nützlich. Bei einem minimalen Zellabstand von d Zeitschlitzen tritt eine Überlastsituation auf, sobald sich d+1 Quellen im Sendezustand befinden. Falls der Puffer voll ist, geht somit mindestens jede (d+1)-te Zelle verloren. Die Überlastsituation wird dadurch beendet, daß eine dieser d+1 Quellen in die Ruhephase übergeht. Die entsprechende Übergangsrate ist umgekehrt proportional zur mittleren Sendedauer  $d \cdot b \cdot \Delta t$ . Für die mittlere Verlustdauer und die Verlustwahrscheinlichkeit innerhalb der Verlustphase gelten demnach folgende Grenzwerte:

$$t_L \ge \frac{d \cdot b \cdot \Delta t}{d+1} = 0, 8 \cdot b \cdot \Delta t \quad \text{und} \quad B_L \ge \frac{1}{d+1} = 0, 2.$$
 (4.25)

Die tatsächlichen Ergebnisse liegen nur knapp über den aus diesen Gleichungen ermittelten Minimalwerten. Ferner muß betont werden, daß diese Grenzwerte nur von den Quellparametern und nicht von der Netzlast abhängen, d.h. die Verbindungsannahmefunktion kann diese Leistungskenngrößen nicht kontrollieren. Dagegen weist der mittlere zeitliche Abstand zweier Verlustphasen eine starke Lastabhängigkeit auf, da er hauptsächlich durch die Verlustwahrscheinlichkeit bestimmt ist (vgl. Gleichung A.53). Diese Abhängigkeit führt dazu,

daß mit der Begrenzung der Verlustwahrscheinlichkeit auch eine indirekte Kontrolle des zeitlichen Abstands zweier Verlustphasen möglich ist. Dies erleichtert die Verkehrssteuerung, da nur die Verlustwahrscheinlichkeit erfaßt werden muß. Die untersuchten Leistungskenngrößen fließen jedoch bei der Festlegung der maximal zulässigen Verlustwahrscheinlichkeit ein.

Alle bisherigen Untersuchungen bezogen sich auf das sporadische Quellmodell mit negativexponentiell verteilten Büschel- und Ruhedauern. Da diese Annahmen für reale Quellen nicht haltbar sind, soll die Sensitivität der Ergebnisse im Hinblick auf eine Änderung des Quellmodells untersucht werden. Bild 4.18 zeigt den Einfluß der Variationskoeffizienten der Büschel- und der Ruhedauer (durch  $c_{B,H}$  bzw.  $c_{B,S}$  bezeichnet). Die Büschel- und die Ruhephasen wurden dabei mit Hilfe einer  $E_2$  bzw. einer  $H_2$ -Verteilung modelliert, deren statistische Kenngrößen aus Tabelle 3.1 entnommen werden können. Ferner wurde für die  $H_2$ -Verteilung die in [151] beschriebene Symmetriebedingung zugrundegelegt. Dabei werden die Verweigungswahrscheinlichkeiten umgekehrt proportional zu den mittleren Phasendauern gewählt.

Die in Bild 4.18 skizzierten Ergebnisse verdeutlichen den Einfluß des Variationskoeffizienten der Büschel- und der Ruhedauer, der sich mit zunehmender Puffergröße verstärkt. Erwartungsgemäß führt eine Verkleinerung des Variationskoeffizienten zu einer Verringerung der Verlustwahrscheinlichkeit. Dabei ist zu erkennen, daß der Streuung der Büscheldauer im Hinblick auf die Verlustwahrscheinlichkeit größeres Gewicht zukommt als der Streuung der Ruhedauer. Die Verlustwahrscheinlichkeit des pufferlosen Systems ist unabhängig von der Verteilung der Büschel- und der Ruhedauer. Dies ist darauf zurückzuführen, daß der

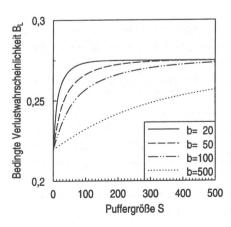


Bild 4.17: Verlustwahrscheinlichkeit innerhalb der Verlustphase

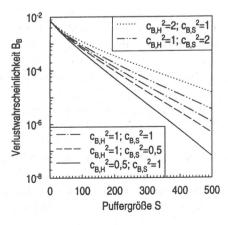


Bild 4.18: Einfluß der Variationskoeffizienten der Büschel- und der Ruhephase

Systemzustand nur vom momentanen Quellzustand und nicht von der Dauer seines Bestehens abhängt (vgl. auch Gleichung A.45). Diese grundsätzlichen Aussagen werden durch die in [18, 221] beschriebenen, ähnlich gelagerten Untersuchungen bestätigt. Darüberhinaus konnte in [221] anhand des MMPP/M/1/(S+1)-Systems nachgewiesen werden, daß auch die Korrelation aufeinanderfolgender Büschel- und Ruhephasen einen vergleichbaren, allerdings schwächer ausgeprägten Einfluß auf die Verlustwahrscheinlichkeit hat.

Eine weitere Einschränkung der sporadischen Quelle resultiert aus der Tatsache, daß sich die Quelle nur im Sende- bzw. im Ruhezustand befinden kann. Demgegenüber sind für die wirklichkeitsnahe Modellierung von Videoquellen mehr als zwei verschiedene Quellzustände mit unterschiedlicher Sendebitrate erforderlich. In [164] wurde das in Bild 4.19 veranschaulichte Videoquellmodell vorgeschlagen, das aus der Überlagerung von L sporadischen Quellen hervorgeht. Jede dieser sogenannten "Miniquellen" ist durch die Senderate  $\hat{r}$ , die mittlere Büscheldauer  $1/\beta$  und die mittlere Ruhedauer  $1/\alpha$  charakterisiert. Die von der Videoquelle generierte Bitrate kann daher durch eine Binomial-Verteilung beschrieben werden.

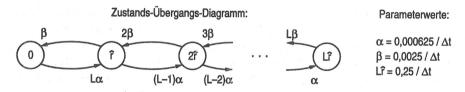
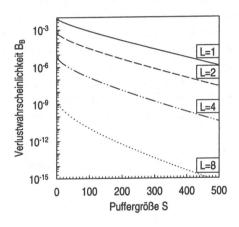


Bild 4.19: Überlagerung von L sporadischen Quellen

In Bild 4.20 wurde die Anzahl der Miniquellen bei gleichbleibender Spitzenbitrate  $L\hat{r}$  variert. Die mittlere Büschelgröße und der Büschelfaktor der Miniquellen entsprechen den für die bisherigen Untersuchungen verwendeten Werten. Ferner wurden, wie in den obigen Beispielen, N=8 Videoquellen überlagert. Unter den genannten Voraussetzungen bleibt die mittlere Bitrate einer Videoquelle konstant, wogegen der quadrierte Variationskoeffizient der Bitrate  $(c_B^2)$  umgekehrt proportional zur Anzahl der Miniquellen L ist (für L=1, 2, 4, 8 ergibt sich  $c_B^2=4, 2, 1, 0,5$ ). Erwartungsgemäß sinkt die Verlustwahrscheinlichkeit bei einer Verkleinerung des Variationskoeffizienten, d.h. die sporadische Quelle weist das im Hinblick auf die Verlustwahrscheinlichkeit schlechteste Quellverhalten auf (vgl. auch Abschnitt 6.5.2). Eine andere Interpretation ergibt sich dadurch, daß die Miniquellen als einzelne Verkehrsquellen aufgefaßt werden. Dies bestätigt die zu Bild 4.5 gemachten Aussagen, wonach mit sinkender Spitzenbitrate eine Abnahme der Verlustwahrscheinlichkeit eintritt.

Bislang wurden ausschließlich homogene Verkehrsmischungen betrachtet, bei denen alle Quellen identische Verkehrsparameter aufweisen. Im folgenden wird das generelle Systemverhalten bei heterogener Verkehrsbelastung anhand ausgewählter Beispiele aufgezeigt. Es werden jeweils zwei Verkehrsklassen $^1$  betrachtet, die sich entweder hinsichtlich der Büschelgröße b, dem Büschelfaktor Z oder dem minimalen Zellabstand d unterscheiden. Die beiden anderen Kenngrößen sind für beide Verkehrsklassen gleich und stimmen mit den entsprechenden Parametern der Referenzquelle überein.

In Bild 4.21 wurden zwei Verkehrsklassen mit unterschiedlicher Büschelgröße überlagert. Die erste Klasse hat eine mittlere Büschelgröße von  $b_1=500$  Zellen, während die zweite Klasse eine zehnmal kleinere Büschelgröße aufweist. Die Gesamtzahl der Verbindungen wurde auf dem konstanten Wert von  $N_1+N_2=8$  belassen. Die Ergebnisse bekräftigen die aus dem homogenen Fall bekannte Erhöhung der Verlustwahrscheinlichkeit bei steigender Büschelgröße, d.h. mit zunehmendem Verkehrsanteil von Klasse 1 steigt die Verlustwahrscheinlichkeit an. Dies kann darauf zurückgeführt werden, daß langanhaltende Überlastsituationen die Wahrscheinlichkeit für einen Pufferüberlauf vergrößern. Aus demselben Grund ist die Verlustwahrscheinlichkeit der Verkehrsklasse 1 höher als die Verlustwahrscheinlichkeit der zweiten Verkehrsklasse. Diese Unterschiede sind allerdings nur schwach ausgeprägt.



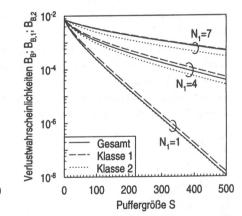


Bild 4.20: Verlustwahrscheinlichkeit für mehrstufige Verkehrsquellen

Bild 4.21: Verkehrsmischung aus zwei Klassen mit unterschiedlicher Büschelgröße

Die in Bild 4.22 betrachtete Verkehrsmischung setzt sich aus zwei Verkehrsklassen mit den Büschelfaktoren  $Z_1=2$  und  $Z_2=9$  zusammen. Die Gesamtzahl der Verbindungen wurde wiederum zu  $N_1+N_2=8$  gewählt. Die Verlustwahrscheinlichkeit erhöht sich mit zunehmendem Verkehrsanteil von Klasse 1, da die Verbindungen dieser Verkehrsklasse eine höhere

 $<sup>^{1}</sup>$ Quellen mit gleichen statistischen Eigenschaften werden innerhalb einer Verkehrsklasse zusammengefaßt. Die Kenngrößen von Verkehrsklasse k (k = 1, 2) werden mit dem Index k versehen.

mittlere Bitrate aufweisen. Dies deckt sich mit den in Bild 4.4 für homogenen Ankunftsverkehr erhaltenen Ergebnissen. Dagegen ist die Verlustwahrscheinlichkeit für die Verkehrsklasse mit dem größeren Büschelfaktor höher. Dies läßt sich anhand des folgenden Sonderfalls nachvollziehen [244]: Betrachtet man die Überlagerung einer CBR-Quelle und einer sporadischen Quelle, deren Maximalbitraten jeweils der Übertragungsbitrate entsprechen, so tritt eine Überlastsituation auf, sobald sich die sporadische Quelle im Sendezustand befindet. Dagegen kann die CBR-Quelle in den Ruhephasen der sporadischen Quelle ungestört senden und erleidet deshalb weniger Zellverluste. Beim pufferlosen System stehen die Verlustwahrscheinlichkeiten im selben Verhältnis wie die Büschelfaktoren. Mit steigender Anzahl von Verkehrsquellen verwischen diese Unterschiede zunehmend, da der Verkehrsanteil einer einzelnen Verbindung abnimmt.

Ein weiterer Verkehrsparameter, der einen signifikanten Einfluß auf die Verlustwahrscheinlichkeit ausübt, ist die Spitzenbitrate der Verkehrsquellen. Abschließend werden daher zwei verschiedene Verkehrsklassen mit einem minimalen Zellabstand von  $d_1=4$  bzw.  $d_2=20$  Zeitschlitzen betrachtet. Die Anzahl der von beiden Verkehrsklassen bestehenden Verbindungen wurde so gewählt, daß das Verkehrsangebot auf einem konstanten Wert von 0,4 Erlang gehalten wird, d.h. es gilt  $N_1+N_2/5=8$ . Der Büschelfaktor und die Büschelgröße wurden wiederum zu  $Z_1=Z_2=5$  bzw.  $b_1=b_2=100$  gewählt. Die mit diesen Vorgaben erhaltenen Ergebnisse sind in Bild 4.23 dargestellt. Das prinzipielle Verhalten läßt sich wiederum mit den aus den homogenen Verkehrsbetrachtungen erhaltenen Ergebnissen deuten: Die Verlustwahrscheinlichkeit ist umso geringer, je größer der Verkehrsanteil von Verkehrsklasse 2 ist, da sich die Bitratenschwankungen mit zunehmender Verbindungszahl besser ausmitteln. Ferner

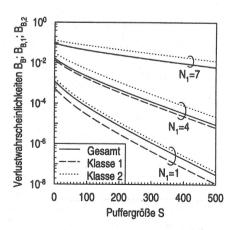


Bild 4.22: Verkehrsmischung aus zwei Klassen mit unterschiedlichem Büschelfaktor

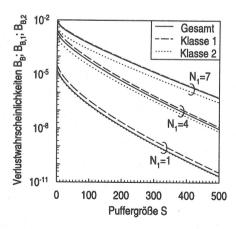


Bild 4.23: Verkehrsmischung aus zwei Klassen mit unterschiedlicher Spitzenbitrate

wirkt sich die höhere Spitzenbitrate von Verkehrsklasse 1 dahingehend aus, daß diese eine überdurchschnittlich hohe Verlustwahrscheinlichkeit aufweist.

Insgesamt läßt sich sagen, daß die für heterogene Verkehrszusammensetzung erhaltenen Ergebnisse mit dem aus den homogenen Beispielen bekannten Systemverhalten übereinstimmen, d.h. die aus den vorherigen Untersuchungen gezogenen Schlußfolgerungen besitzen auch für den allgemeinen Fall Gültigkeit. Ferner zeigten sich nur geringe Unterschiede in Bezug auf die Verlustwahrscheinlichkeit verschiedener Verbindungen. Dies bestätigt die in [183] für das pufferlose "Fluid Flow"-Modell gemachten Aussagen, wonach sich die Verlustwahrscheinlichkeiten verschiedener Verbindungen für realistische Fälle um maximal zwei Zehnerpotenzen unterscheiden.

## 4.2 Leistungsanalyse einer Referenzverbindung

Um abschließende Aussagen über die Netzgüte der ATM-Schicht treffen zu können, muß eine ganze VC-Verbindung betrachtet werden, die sich aus mehreren VC-Verbindungsabschnitten zusammensetzt. Da jeder Verbindungsabschnitt zu den lastabhängigen Zellverlusten und -verzögerungen beiträgt, müssen die im vorigen Unterkapitel diskutierten Teilmodelle der einzelnen Verbindungsabschnitte zu der in Bild 3.6 dargestellten Warteschlangenkette aus J Bedienstationen verknüpft werden. Die Reihenfolge der Bedienstationen spiegelt den Weg der Referenzverbindung durch das Kommunikationsnetz wider. An jeder Bedienstation können zusätzliche Verbindungen ein- bzw. ausgekoppelt werden. Vereinfachend soll angenommen werden, daß die Verkehre aller Verbindungen von sporadischen Quellen mit identischen statistischen Eigenschaften erzeugt werden.

Dieses Modell fällt in die Klasse der offenen, rückkopplungsfreien Warteschlangennetze, für die eine Vielzahl analytischer Methoden existiert (siehe z.B. [16, 145]). Allerdings sind diese Ansätze für die vorliegende Problemstellung ungeeignet, da sie von abweichenden Modellannahmen für die Ankunfts- und Bedienprozesse bzw. für die netzinterne Verzweigung des Verkehrs ausgehen. Darüberhinaus sind diese Ansätze nur für die mittlere Transferzeit anwendbar, während die Netzgüte der ATM-Schicht durch das Zellverlustverhältnis und die Zellverzugsschwankungen charakterisiert wird [117]. Da die Bestimmung dieser Leistungskenngrößen ungleich schwieriger ist, sind hier nur approximative Lösungsansätze bekannt, die entweder auf sehr einfachen Modellannahmen beruhen [32, 79, 80, 194] oder nur für Spezialfälle gültig sind [30, 185]. Nachfolgend wird deshalb ein eigener Lösungsansatz entwickelt, der eine obere Grenze für die Zellverzugsschwankungen einer Referenzverbindung bestimmt.

In der ITU-T-Empfehlung I.356 [117] werden die zwischen zwei getrennten Meßpunkten eingefügten Zellverzugsschwankungen durch die sogenannte "2-point Cell Delay Variation"

beschrieben. Diese Meßgröße erfaßt die zeitliche Differenz aus der Transferzeit der betrachteten Zelle und einer festgelegten Referenztransferzeit. Bei den nachfolgenden Betrachtungen entspricht diese Referenztransferzeit der Summe der konstanten Signallaufzeiten. Die daraus resultierende "2-point Cell Delay Variation" wird vereinfachend als Transferzeit bezeichnet.

## 4.2.1 Approximative Analyse

Für die analytische Behandlung des in Bild 3.6 gezeigten Warteschlangenmodells wird die Referenzverbindung zunächst in Teilmodelle für die einzelnen Verbindungsabschnitte zerlegt. Aus den daraus ermittelten Teilergebnissen können die charakteristischen Leistungskenngrößen der Referenzverbindung abgeleitet werden. Bei dieser allgemein gebräuchlichen Vorgehensweise müssen allerdings zwei entscheidende Punkte beachtet werden:

- Die Zerlegung des Warteschlangenmodells in die entsprechenden Bedienstationen muß die Verformung des Ankunftsverkehrs an den vorangegangenen Bedienstationen einbeziehen.
- Die Zusammenfassung der für die einzelnen Bedienstationen erhaltenen Teilresultate muß deren Korrelation an aufeinanderfolgenden Bedienstationen berücksichtigen.

Eine umfassende Simulationsstudie [154] zeigt, daß unter den gegebenen Randbedingungen die netzinterne Verformung des Verkehrs bei der Bestimmung der charakteristischen Leistungskenngrößen vernachlässigt werden kann. Dies ist darauf zurückzuführen, daß die durch die Puffergröße definierte maximale Zellverzögerung wesentlich geringer als die typische Büscheldauer der Verkehrsquellen ist. Deshalb bleiben die längerfristigen Verkehrsschwankungen innerhalb des Netzes nahezu erhalten. Auf der Zellebene verursachen die unvermeidlichen Zellverluste und Zellverzugsschwankungen geringfügige Modifikationen des Verkehrs. Insbesondere können die Zellen eine Bedienstation nur seriell verlassen, weshalb an der folgenden Bedienstation nur die auf getrennten Eingangsleitungen eintreffenden Zellen kollidieren können. Allerdings kann diese Verkehrsglättung, ebenso wie die sehr selten auftretenden Zellverluste, für realistische Fälle vernachlässigt werden [154, 251].

Demgegenüber führt die langzeitige, von der Büschelebene herrührende Korrelation des Verkehrsflusses zu einer positiven Korrelation der Zellverzögerungen für aufeinanderfolgende Verbindungsabschnitte. Dies vergrößert die Streuung der Gesamttransferzeit, weshalb dieses Korrelationsverhalten bei der in Unterabschnitt 4.2.1.2 vorgestellten Zusammenfassung der für die einzelnen Verbindungsabschnitte ermittelten Teilergebnisse einbezogen wird.

## 4.2.1.1 Analyse einer einzelnen Bedienstation

In diesem Unterabschnitt wird eine einzelne Bedienstation j (j = 1, 2, ..., J) betrachtet, wobei auf eine zusätzliche Indizierung der dieser Bedienstation zugeordneten Größen aus Gründen

der Übersichtlichkeit verzichtet wird. Die wesentlichen Eigenschaften der an dieser Station eintreffenden Verkehrsströme stimmen mit der ursprünglichen, am Eingang des Netzes auftretenden Quellcharakteristik überein. Damit kann die oben diskutierte Dekomposition des Gesamtmodells in ein M/D/1/(S+1)-Modell für die Zellebene und ein "Fluid-Flow"-Modell für die Büschelebene angewandt werden.

Die analytische Behandlung des M/D/1/(S+1)-Systems mit Hilfe einer eingebetteten Markoff-Kette ist in [78] ausführlich erörtert. Diese Analyse liefert die Wahrscheinlichkeit  $\pi(x)$ , daß eine abgehende Zelle genau x wartende Zellen an der betrachteten Bedienstation zurückläßt. Da keine Gruppenankünfte bzw. -bedienungen möglich sind, entspricht dies auch der Wahrscheinlichkeit, daß eine ankommende Zelle, die nicht verloren geht, x Anforderungen innerhalb des Systems vorfindet (ein Beweis dieser für allgemeine Ankunfts- und Bedienprozesse geltenden Identität findet sich in [45]). Die Durchlaufzeit der betrachteten Zelle liegt innerhalb des Zeitintervalls  $(x \cdot \Delta t, (x+1) \cdot \Delta t]$ . Alle Durchlaufzeiten, die innerhalb dieses Zeitintervalls liegen, werden auf den nächsthöheren Wert  $(x+1) \cdot \Delta t$  aufgerundet, da die Durchlaufzeit aufgrund der getakteten Betriebsweise des ATM-Netzes nur ganzzahlige Vielfache der Zellübertragungsdauer betragen kann. Mit diesen Vorbemerkungen ergibt sich folgende Verteilung für die auf die Zellübertragungsdauer normierte Durchlaufzeit des M/D/1/(S+1)-Systems:

$$p_{D,M/D/1/(S+1)}(x+1) = \pi(x), \quad x = 0, 1, ..., S.$$
 (4.26)

Die Verlustwahrscheinlichkeit des M/D/1/(S+1)-Systems ergibt sich zu [78]

$$B_{\text{M/D/1/(S+1)}} = 1 - \frac{1}{\pi(0) + A_Z}$$
 (4.27)

Die von der Büschelebene herrührenden, längerfristigen Verkehrsschwankungen überlagern sich den kurzzeitigen Verkehrsschwankungen der Zellebene und führen zu zusätzlichen Zellverzögerungen und -verlusten. Diese Anteile werden mit Hilfe des "Fluid Flow"-Modells erfaßt, das in Anhang A ausführlich beschrieben ist. Der Überlagerungsverkehr der N sporadischen Verkehrsquellen kann durch N+1 globale Quellzustände beschrieben werden (vgl. Bild 4.19). Im globalen Quellzustand l (l=0,1,...,N) befinden sich l Einzelquellen in der Sendephase und N-l Einzelquellen in der Ruhephase. Da sich eine Quelle mit der Wahrscheinlichkeit 1/Z in der Sendephase und mit der komplementären Wahrscheinlichkeit 1-1/Z in der Ruhephase befindet, können die Auftrittswahrscheinlichkeiten der Quellzustände l durch die folgende Binomial-Verteilung beschrieben werden:

$$p_B(l) = \binom{N}{l} \left(\frac{1}{Z}\right)^l \cdot \left(1 - \frac{1}{Z}\right)^{N-l}, \quad l = 0, 1, ..., N.$$
 (4.28)

Das gesamte Systemverhalten kann durch die in Anhang A abgeleitete Verbundwahrscheinlichkeit F(l,x) charakterisiert werden. Diese beschreibt die Wahrscheinlichkeit, daß sich l

Quellen im aktiven Zustand befinden und der Pufferfüllstand kleiner gleich x ist. Um die Korrelation der auf den verschiedenen Verbindungsabschnitten auftretenden Verluste und Verzögerungen erfassen zu können, müssen aus dieser Verbundverteilung auf den globalen Quellzustand l bezogene Leistungskenngrößen ermittelt werden. Falls die dem Quellzustand l zugeordnete Ankunftsrate  $l \cdot \hat{r}$  die Bedienrate C des Multiplexers übersteigt und der Puffer voll belegt ist, so geht die überschüssige Rate  $l \cdot \hat{r} - C$  verloren (vgl. auch Bild A.5). Die bedingte Wahrscheinlichkeit, daß der Puffer voll belegt ist, falls sich l Quellen im aktiven Zustand befinden, ergibt sich zu  $[p_B(l) - F(l, S^-)]/p_B(l)$ . Aus dieser Wahrscheinlichkeit und dem Verhältnis der für den Quellzustand l gültigen Verlust- und Ankunftsraten erhält man die auf den Quellzustand l bezogene Verlustwahrscheinlichkeit:

$$B_B(l) = \begin{cases} \frac{[l \cdot \hat{r} - C] \cdot [p_B(l) - F(l, S^-)]}{l \cdot \hat{r} \cdot p_B(l)} & \text{für } l \cdot \hat{r} > C \\ 0 & \text{für } l \cdot \hat{r} \le C \end{cases}$$
(4.29)

Die in Gleichung A.35 angegebene Zellverlustwahrscheinlichkeit vereinfacht sich im vorliegenden Sonderfall zu

$$B_{B} = \frac{\sum_{l=\lceil C/\hat{\tau} \rceil}^{N} [l \cdot \hat{\tau} - C] \cdot [p_{B}(l) - F(l, S^{-})]}{\sum_{l=1}^{N} l \cdot \hat{\tau} \cdot p_{B}(l)} = \frac{\sum_{l=\lceil C/\hat{\tau} \rceil}^{N} [l \cdot \hat{\tau} - C] \cdot [p_{B}(l) - F(l, S^{-})]}{\hat{\tau} \cdot N \cdot \frac{1}{Z}} . \quad (4.30)$$

Dabei bezeichnet  $\lceil x \rceil$  die kleinste ganze Zahl größer oder gleich x. Ferner ist die auf den Quellzustand l bedingte Verteilung der Durchlaufzeit von Interesse. Die Durchlaufzeiten werden wiederum auf ganzzahlige Vielfache der Zellübertragungsdauer aufgerundet, d.h. der Verteilungswert an der Stelle  $x \cdot \Delta t$  entspricht der Wahrscheinlichkeit, daß die Durchlaufzeit innerhalb des Zeitintervalls  $((x-1)\cdot \Delta t,x\cdot \Delta t]$  liegt. Abgesehen von den Randwerten bei x=0 und x=S treten die dazugehörigen Pufferfüllstände mit der Wahrscheinlichkeit  $[F(l,x)-F(l,x-1)]/p_B(l)$  auf, falls sich l Quellen im aktiven Zustand befinden. Diese Wahrscheinlichkeit muß noch mit dem Verhältnis der in den Zuständen l und (x-1,x] angenommenen Ankunftsrate  $l\cdot \hat{r}$  und der gesamten im Quellzustand l angenommenen Ankunftsrate  $[1-B_B(l)]\cdot l\cdot \hat{r}$  gewichtet werden, um die auf den Quellzustand l bedingte Durchlaufzeitverteilung  $p_{D,B}(x|l)$  zu erhalten. Aus analogen Überlegungen für die Verteilungswerte bei x=0 und x=S erhält man

$$p_{D,B}(x|l) = \begin{cases} \frac{F(l,0)}{[1 - B_B(l)] \cdot p_B(l)} & \text{für } x = 0 \\ \frac{F(l,x) - F(l,x-1)}{[1 - B_B(l)] \cdot p_B(l)} & \text{für } x = 1,2, \\ \dots, S-1 \\ \frac{[F(l,S^-) - F(l,S-1)] \cdot l \cdot \hat{r} + [p_B(l) - F(l,S^-)] \cdot \min(l \cdot \hat{r},C)}{[1 - B_B(l)] \cdot p_B(l) \cdot l \cdot \hat{r}} & \text{für } x = S. \end{cases}$$

$$(4.31)$$

Abschließend müssen die für das M/D/1/(S+1)- und das "Fluid Flow"-Modell erzielten Teilresultate miteinander verknüpft werden. Da Blockierungen auf Zell- und Büschelebene auftreten können, ergibt die Addition der aus den zugeordneten Teilmodellen errechneten Verlustwahrscheinlichkeiten eine gute Näherung für die auf den Quellzustand l bezogene Zellverlustwahrscheinlichkeit  $B_Z(l)$  und für die gesamte Zellverlustwahrscheinlichkeit  $B_Z$  (vgl. auch Bild 4.13):

$$B_Z(l) \approx B_{\rm M/D/1/(S+1)} + B_B(l)$$
 bzw.  $B_Z \approx B_{\rm M/D/1/(S+1)} + B_B$ . (4.32)

Die durch kurz- und langfristige Verkehrsfluktuationen verursachten Zellverzögerungen summieren sich ebenfalls auf. Die längerfristigen Verkehrsschwankungen führen zu einer gewissen "Grundverzögerung" von y Zeitschlitzen, die sich aus dem "Fluid Flow"-Modell ableiten läßt. Die Anzahl der für die Zwischenspeicherung kurzzeitiger Verkehrsfluktuationen zur Verfügung stehenden Pufferplätze reduziert sich somit auf den Wert S-y. Die daraus resultierenden additiven Verzögerungsanteile werden daher mit Hilfe eines M/D/1/(S-y+1)-Modells erfaßt. Falls beide Anteile als statistisch unabhängig betrachtet werden, ergibt sich die auf den Quellzustand l bedingte Transferzeitverteilung

$$p_{D,Z}(x|l) \approx \sum_{y=0}^{x-1} p_{D,B}(y|l) \cdot p_{D,M/D/1/(S-y+1)}(x-y)$$
, für  $x = 1, 2, ..., S+1$ . (4.33)

#### 4.2.1.2 Analyse der Referenzverbindung

Die Leistungskenngrößen der gesamten Referenzverbindung sollen aus den für die einzelnen Verbindungsabschnitte ermittelten Teilergebnissen bestimmt werden. Häufig wird in diesem Fall von der Unabhängigkeit der für die verschiedenen Bedienstationen gültigen Leistungskenngrößen ausgegangen (siehe z.B. [32, 185, 194]). Allerdings führt die Büschelförmigkeit und die verbindungsorientierte Verzweigung des Verkehrs zu einer positiven Korrelation der Leistungskenngrößen an zwei aufeinanderfolgenden Bedienstationen: Wenn sich überdurchschnittlich viele der durch beide Bedienstationen verlaufenden Verbindungen im aktiven Zustand befinden, ergeben sich an beiden Bedienstationen große Verlustwahrscheinlichkeiten bzw. Verzögerungen und umgekehrt.

Falls diese positive Korrelation vernachlässigt wird, ergibt sich dadurch eine leichte Überschätzung der Gesamtzellverlustwahrscheinlichkeit, wie die in Abschnitt 4.2.2 beschriebenen Ergebnisse bestätigen. Mit der Annahme, daß die an den einzelnen Bedienstationen auftretenden Zellverluste unabhängig sind, wird eine Zelle der Referenzverbindung mit der Wahrscheinlichkeit  $\prod_{j=1}^{J} (1 - B_{Z,j})$  erfolgreich übermittelt (die Zellverlustwahrscheinlichkeit an Bedienstation j wurde mit  $B_{Z,j}$  bezeichnet). Die durch die Unabhängigkeitsannahme definierte obere Schranke für die Zellverlustwahrscheinlichkeit der Referenzverbindung lautet

somit

$$B_Z = 1 - \prod_{j=1}^{J} (1 - B_{Z,j}) \approx \sum_{j=1}^{J} B_{Z,j}$$
 (4.34)

Für die Transferzeitverteilung liefert die Unabhängigkeitsannahme jedoch unbefriedigende Ergebnisse, da die Wahrscheinlichkeiten für große Transferzeiten unterschätzt werden. Deshalb bezieht der nachfolgend vorgestellte analytische Ansatz die positive Korrelation der Durchlaufzeiten an aufeinanderfolgenden Netzknoten ein. Zunächst soll das in Bild 4.24 dargestellte Grundmodell mit zwei Bedienstationen betrachtet werden. Die Anzahl der durch die beiden letzten Bedienstationen J-1 und J verlaufenden Verbindungen werde mit  $N_{J-1}$  bzw.  $N_J$  bezeichnet. Nach der vorletzten Bedienstation zweigen  $N_D$  Verbindungen vom Referenzpfad ab, während  $N_A=N_J+N_D-N_{J-1}$  Verbindungen an der letzten Bedienstation hinzukommen. Das beschriebene Modell kann in eine äquivalente Bedienstation überführt werden, an die im folgenden Schritt die vorangehende Bedienstation J-2 angefügt werden kann. Dieses schrittweise Vorgehen ermöglicht die analytische Behandlung der gesamten Warteschlangenkette.

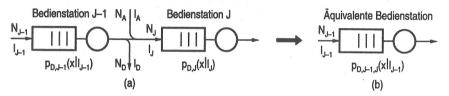


Bild 4.24: (a) Grundmodell aus zwei Bedienstationen (b) Äquivalente Bedienstation

Ferner bezeichnen die Größen  $l_{J-1}$ ,  $l_D$ ,  $l_A$  und  $l_J$  die Anzahl aktiver Verbindungen, die eine markierte Zelle der Referenzverbindung auf ihrem Weg durch die Warteschlangenkette an den jeweiligen Verzweigungspunkten vorfindet. Eine Verbindung befindet sich genau dann im aktiven Zustand, wenn die letzte von dieser Verbindung an dem betrachteten Punkt angekommene Zelle nicht die letzte Zelle eines Büschels war. Da sich die Zellen verschiedener Verbindungen nicht überholen können, trifft die Referenzzelle diejenigen Verbindungen, die beide Bedienstationen durchlaufen, an jeder Bedienstation im selben Zustand an, d.h. es gilt

$$l_J = l_{J-1} - l_D + l_A . (4.35)$$

Damit ist die Anzahl der an den Bedienstationen J-1 und J aktiven Verbindungen positiv korreliert. Diese Abhängigkeit wird durch die bedingte Wahrscheinlichkeit  $p_L(l_J|l_{J-1})$  ausgedrückt, die im folgenden bestimmt werden soll. Gemäß Gleichung 4.35 setzen sich die an Bedienstation J aktiven Verbindungen aus zwei voneinander unabhängigen Anteilen zusammen:

- $l_{J-1} l_D$  aktive Verbindungen stammen von Bedienstation J-1, und
- l<sub>A</sub> aktive Verbindungen werden an Bedienstation J neu eingekoppelt.

Die  $l_{J-1}-l_D$  von Bedienstation J-1 stammenden aktiven Verbindungen schließen die Referenzverbindung ein, die sich im aktiven Zustand befindet und deren Weg zu Bedienstation J führt. Die Auswahl der  $l_D$  abzweigenden aktiven Verbindungen aus den restlichen  $l_{J-1}-1$  an Bedienstation J-1 aktiven Verbindungen stellt ein einfaches kombinatorisches Problem dar, dessen Lösung auf eine hypergeometrische Wahrscheinlichkeitsverteilung (vgl. [60]) für die durchgehenden aktiven Verbindungen führt. Somit ergibt sich die erste Teilkomponente der gesuchten Verteilung  $p_L(l_J|l_{J-1})$  zu

$$p_{L,1}(l_{J-1} - l_D|l_{J-1}) = \frac{\binom{l_{J-1} - 1}{l_D} \cdot \binom{N_{J-1} - l_{J-1}}{N_D - l_D}}{\binom{N_{J-1} - 1}{N_D}},$$

$$l_{J-1} - l_D = 1, 2, ..., \min(N_{J-1} - N_D, l_{J-1}).$$

$$(4.36)$$

Der Zähler von Gleichung 4.36 beschreibt die Anzahl der Möglichkeiten,  $l_D$  aktive und  $N_D-l_D$  passive abzweigende Verbindungen aus insgesamt  $l_{J-1}-1$  aktiven bzw.  $N_{J-1}-l_{J-1}$  passiven Verbindungen auszuwählen. Dieser Wert muß auf die Anzahl der insgesamt möglichen Kombinationen zur Auswahl von  $N_D$  abzweigenden Verbindungen aus  $N_{J-1}-1$  Verbindungen bezogen werden, um die angegebene Wahrscheinlichkeit zu erhalten.

Die Anzahl der an Bedienstation J neu eingekoppelten aktiven Verbindungen gehorcht einer Binomial-Verteilung, da sich jede Verbindung mit der Wahrscheinlichkeit 1/Z im aktiven Zustand befindet:

$$p_{L,2}(l_A) = \binom{N_A}{l_A} \left(\frac{1}{Z}\right)^{l_A} \cdot \left(1 - \frac{1}{Z}\right)^{N_A - l_A}, \quad l_A = 0, 1, ..., N_A.$$
 (4.37)

Schließlich ergibt sich die gesuchte Verteilung aus einer Faltung der durch die Gleichungen 4.36 und 4.37 definierten Wahrscheinlichkeitsverteilungen, da alle Verbindungen statistisch unabhängig sind.

$$p_{L}(l_{J}|l_{J-1}) = \sum_{l_{A}=\max(0,l_{J}-l_{J-1},l_{J}-N_{J-1}+N_{D})}^{\min(N_{A},l_{J}-1)} p_{L,1}(l_{J}-l_{A}|l_{J-1}) \cdot p_{L,2}(l_{A}) , \qquad (4.38)$$

$$l_{J} = 1, 2, ..., \min(N_{A}+l_{J-1}, N_{J}) .$$

Mit Hilfe dieser bedingten Wahrscheinlichkeit kann die Abhängigkeit der an den Bedienstationen J-1 und J auftretenden Durchlaufzeiten erfaßt werden. Hierzu wird zunächst die Verteilung  $p_{D,J-1,J}(x|l_{J-1},l_J)$  definiert, welche die auf die Quellzustände  $l_{J-1}$  und  $l_J$  bezogene Gesamtdurchlaufzeit durch die Bedienstationen J-1 und J beschreibt. Diese Verteilung

ergibt sich aus der Faltung der bedingten Durchlaufzeitverteilungen durch die Bedienstationen J-1 und J, da diese näherungsweise unabhängig voneinander sind.

$$p_{D,J-1,J}(x|l_{J-1},l_J) = p_{D,J-1}(x|l_{J-1}) * p_{D,J}(x|l_J) = \sum_{y=1}^{x-1} p_{D,J-1}(y|l_{J-1}) \cdot p_{D,J}(x-y|l_J) . \tag{4.39}$$

Die bedingten Durchlaufzeiten der Bedienstationen J-1 und J wurden mit den Indizes J-1 bzw. J versehen. Daraus ergibt sich nach dem Gesetz der totalen Wahrscheinlichkeit die bedingte Durchlaufzeitverteilung der Ersatzbedienstation nach Bild 4.24(b) zu

$$p_{D,J-1,J}(x|l_{J-1}) = \sum_{l_{J}=1}^{\min(N_A + l_{J-1}, N_J)} p_L(l_J|l_{J-1}) \cdot p_{D,J-1,J}(x|l_{J-1}, l_J) . \tag{4.40}$$

Die beschriebene Abbildung zweier aufeinanderfolgender Bedienstationen auf eine äquivalente Bedienstation ist der Schlüssel zu einer schrittweisen Analyse der gesamten Warteschlangenkette. Nach dem besprochenen Schritt wird an die Ersatzbedienstation die unmittelbar vorangehende Bedienstation J-2 angefügt. Das so entstehende System aus zwei Bedienstationen kann wiederum auf ein Ersatzsystem mit nur einer Bedienstation abgebildet werden. Dieses schrittweise Vorgehen wird fortgesetzt, bis der Anfang der Warteschlangenkette erreicht ist. Man erhält die auf die Anzahl der am ersten Knoten aktiven Quellen bedingte Verteilung der Transferzeit  $p_{D,1,J}(x|l_1)$  der Referenzverbindung.

Nach dem Gesetz der totalen Wahrscheinlichkeit muß diese bedingte Transferzeitverteilung mit der Wahrscheinlichkeit für  $l_1$  aktive Verbindungen gewichtet werden, um die Transferzeitverteilung der Referenzverbindung zu erhalten. Diese Wahrscheinlichkeit gehorcht einer Binomial-Verteilung, da sich jede der  $N_1-1$  quer zur Richtung der Referenzverbindung verlaufenden Verbindungen ("Querverbindungen") mit der Wahrscheinlichkeit 1/Z im aktiven Zustand befindet. Zusätzlich können die am ersten Knoten auftretenden Zellverluste dadurch berücksichtigt werden, daß diese Teilergebnisse mit der Wahrscheinlichkeit für die Annahme einer im Quellzustand  $l_1$  eintreffenden Zelle gewichtet werden. Die Verteilung der gesamten Transferzeit ergibt sich, indem die daraus entstehende Summe auf die Wahrscheinlichkeit, daß eine ankommende Zelle nicht verloren geht, bezogen wird:

$$p_{D,1,J}(x) = \frac{\sum_{l_1=1}^{N_1} {N_1-1 \choose l_1-1} \left(\frac{1}{Z}\right)^{l_1-1} \cdot \left(1-\frac{1}{Z}\right)^{N_1-l_1} \cdot \left[1-B_{Z,1}(l_1)\right] \cdot p_{D,1,J}(x|l_1)}{1-B_{Z,1}} . \tag{4.41}$$

Abschließend soll erwähnt werden, daß die vorgestellte Methodik auch auf allgemein modulierte deterministische Ankunftsprozesse anwendbar ist. Die analytische Behandlung einer einzelnen Bedienstation kann dabei auf beliebigen Ansätzen basieren, welche auf den Quellzustand bedingte Leistungskenngrößen ermitteln.

#### 4.2.2 Ergebnisse

Ziel dieses Abschnitts ist die Bestimmung der charakteristischen Leistungskenngrößen einer Referenzverbindung. Die Untersuchungen beziehen sich auf den zu Beginn dieses Kapitels definierten Referenzverkehr². Die Kapazität der Netzpuffer wurde auf 64 Zellen begrenzt. Die Transferzeiten werden in Form der komplementären Verteilungsfunktion aufgetragen, welche die Wahrscheinlichkeit angibt, daß eine bestimmte Transferzeit überschritten wird. Dies liefert eine übersichtliche und aussagekräftige Darstellung der Ergebnisse.

Einführend soll der prinzipielle Lösungsansatz für einen einzelnen Verbindungsabschnitt validiert werden. Die in Bild 4.26 aufgetragene komplementäre Verteilungsfunktion der Transferzeit gibt die charakteristischen Blockierungseffekte auf der Zell- und der Büschelebene wieder. Diese werden durch unterschiedliche analytische Modelle erfaßt, wobei die Analyse die Transferzeiten überschätzt. Dies ist hauptsächlich darauf zurückzuführen, daß die aus den beiden Teilmodellen errechneten kontinuierlichen Transferzeiten jeweils auf ganzzahlige Vielfache der Zellübertragungsdauer  $\Delta t$  aufgerundet werden. Diese Rundungsfehler führen zu einer durchschnittlichen Erhöhung der Transferzeit um  $\Delta t$ , wie die in Bild 4.26 gezeigten Ergebnisse belegen.

Im folgenden werden die Untersuchungen auf die in Bild 4.24(a) dargestellte Warteschlangenkette ausgedehnt, die aus zwei Bedienstationen besteht. An diesem einfachen Warteschlangenmodell soll die Gültigkeit des den analytischen Untersuchungen zugrundeliegenden Modellierungsansatzes überprüft werden, der das Warteschlangennetz auf eine Kette von Bedienstationen reduziert. Dazu wird die Transferzeit dieser Anordnung mit der Transferzeit des in Bild 4.25 gezeigten Netzbeispiels verglichen, bei dem die am zweiten Knoten eingekoppelten Querverbindungen über einen zusätzlichen Multiplexer geführt werden.

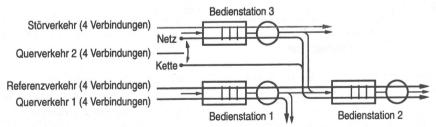


Bild 4.25: Beispiel eines einfachen Warteschlangennetzes

Der bei dem betrachteten Warteschlangennetz an Bedienstation 2 eingekoppelte Querverkehr unterscheidet sich von dem am Netzzugang auftretenden Verkehr aufgrund folgender Einflußfaktoren:

 $<sup>^{2}</sup>$ Es werden N=8 Verkehrsströme mit einem Büschelfaktor Z=5, einer mittleren Büschelgröße von b=100 Zellen und einem minimalen Zellabstand von d=4 Zeitschlitzen überlagert.

- Der Querverkehr wird an Bedienstation 3 geglättet, d.h. die zugehörigen Zellen treffen seriell an Bedienstation 2 ein.
- An Bedienstation 3 können Zellverluste auftreten, womit sich das Verkehrsangebot an Bedienstation 2 verringert.

Insgesamt führt dies zu einer Verringerung der Transferzeiten der Referenzverbindung. Allerdings zeigen die in Bild 4.27 dargestellten Ergebnisse, daß die Glättung des Querverkehrs bei dem betrachteten Beispiel vernachlässigt werden kann, da sich die Zellverzugsschwankungen des Warteschlangennetzes und der Warteschlangenkette nur geringfügig unterscheiden. Dies ist auch der Grund dafür, daß die beschriebenen Einflüsse bei der vorgestellten Analyse nicht berücksichtigt wurden. Die Analyse überschätzt die Transferzeit hauptsächlich aufgrund der oben beschriebenen Rundungsfehler, weshalb die analytischen Resultate gegenüber den Simulationsergebnissen um ca. zwei Zeiteinheiten  $\Delta t$  verschoben sind. Es soll jedoch betont werden, daß das grundsätzliche Systemverhalten erfaßt wird und die Analyse eine obere Grenze für die Transferzeit liefert.

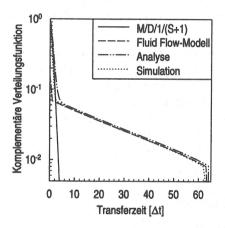


Bild 4.26: Transferzeit über einen Verbindungsabschnitt

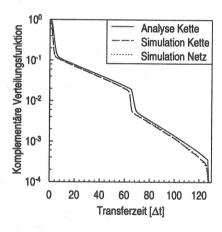


Bild 4.27: Transferzeiten für ein Warteschlangennetz sowie eine Warteschlangenkette

Bemerkenswert ist der treppenförmige Abfall der komplementären Verteilungsfunktion, der für ein realistisches Verkehrsangebot sogar noch deutlicher ausgeprägt ist (vgl. Bild 4.30). Die unterschiedlichen Kurvenbereiche lassen sich wie folgt interpretieren. Im Normalfall tritt an keiner Bedienstation eine Überlastsituation auf, so daß die komplementäre Verteilungsfunktion im Bereich kleiner Transferzeiten rasch abfällt. Die im anschließenden flachen Bereich auftretenden größeren Transferzeiten ergeben sich dadurch, daß an einer der beiden

Bedienstationen eine von der Büschelebene herrührende Überlastsituation auftritt. Nach einem Übergangsbereich folgt ein weiteres Flachstück, in dem sich an beiden Bedienstationen große Transferzeiten ergeben, d.h. beide Bedienstationen sind überlastet. Der abschließende Steilabfall wird durch Zellverluste verursacht, die auf die endliche Größe des Pufferspeichers zurückzuführen sind.

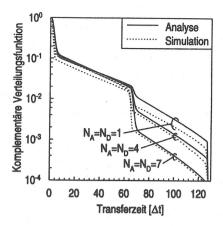
Für die eben behandelte, aus zwei seriell angeordneten Bedienstationen bestehende Warteschlangenkette soll der Einfluß des Querverkehrs genauer untersucht werden. Dazu wird die Anzahl der an den Bedienstationen 1 und 2 eingekoppelten Querverbindungen variiert. Gleichzeitig wird die Anzahl der dem Referenzpfad folgenden Verbindungen so verändert, daß die Gesamtzahl der Verbindungen auf dem konstanten Wert von  $N_1 = N_2 = 8$  bleibt.

Die in Bild 4.28 dargestellten Ergebnisse zeigen, daß sich die beschriebene Änderung der Verbindungstopologie hauptsächlich auf den "Schwanz" der komplementären Verteilungsfunktion auswirkt. Falls die Anzahl der Querverbindungen  $N_A$  und  $N_D$  abnimmt, steigt die Wahrscheinlichkeit für große Transferzeiten an. Dies läßt sich durch die zunehmende Anzahl der auf dem Referenzpfad verlaufenden Verbindungen erklären, die zu der oben erwähnten positiven Korrelation der Durchlaufzeiten an aufeinanderfolgenden Netzknoten führt. Unabhängige Durchlaufzeiten ergeben sich, wenn nur die Referenzverbindung beide Netzknoten durchläuft ( $N_A = N_D = 7$ ). Dieses Szenario führt auf die geringsten Transferzeiten. Gleichzeitig zeigt dies, daß die in der Literatur verwendete Unabhängigkeitsannahme zu einer deutlichen Unterschätzung der Transferzeiten führen kann.

Dagegen bestätigen die in Tabelle 4.2 zusammengefaßten Ergebnisse, daß die Unabhängigkeitsannahme bei der Bestimmung der Zellverlustwahrscheinlichkeit durchaus gerechtfertigt ist. Die positive Korrelation der Zellverluste an beiden Bedienstationen und die Glättung des Verkehrs führen zu der erwarteten Senkung der Zellverlustwahrscheinlichkeit, falls der Querverkehr abnimmt. Mit Ausnahme des Falles  $N_A=N_D=7$  überschätzt die Analyse die tatsächlichen Zellverlustwahrscheinlichkeiten.

Im folgenden werden die Untersuchungen auf Verbindungen mit mehr als zwei Verbindungsabschnitten ausgedehnt. Dem Referenzpfad folgen vier Verbindungen, und an jeder Bedienstation werden vier zusätzliche Querverbindungen eingekoppelt, die unmittelbar nach der betrachteten Bedienstation wieder abzweigen. Die in Bild 4.29 bzw. Tabelle 4.3 wiedergegebenen Transferzeiten und Zellverluste bestätigen die für die vorigen Beispiele gemachten Aussagen.

Zum Abschluß dieses Abschnitts soll die Transferzeit für eine realistische Verkehrsbelastung ermittelt werden, bei der die Zellverlustwahrscheinlichkeit an einem Knoten bei ca.  $10^{-10}$  liegt. Mit den bisher verwendeten Verkehrsparametern ist dies nur dann möglich, wenn jeder Verbindung die Spitzenbitrate zugewiesen wird. Deshalb werden die Quellparameter für die folgende Untersuchung entsprechend dem in Tabelle 2.2 eingeführten Hochgeschwindigkeits-Datenkommunikationsdienst gewählt. Die Puffergröße des Multiplexers wurde bei 64 Zellen



Querverkehr $N_A = N_D$	Analyse	Simulation		
1	$3,059 \cdot 10^{-3}$	$2,004 \cdot 10^{-3} \\ \pm 1,36 \cdot 10^{-4}$		
4		$2,753 \cdot 10^{-3}$ $\pm 2,62 \cdot 10^{-4}$		
7		$3,393 \cdot 10^{-3}$ $\pm 3,71 \cdot 10^{-4}$		

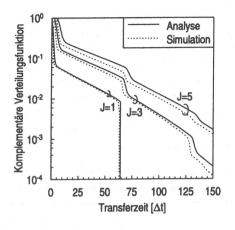
Bild 4.28: Transferzeit über zwei Verbindungsabschnitte

Tabelle 4.2: Zellverlustwahrscheinlichkeit über zwei Verbindungsabschnitte

belassen, und die Nettoübertragungsbitrate beträgt 135,64 Mbit/s (dabei wurde angenommen, daß das gesamte Informationsfeld für Nutzinformation zur Verfügung steht). Unter den genannten Randbedingungen muß die Anzahl der auf den verschiedenen Wegen eingerichteten Verbindungen gegenüber dem zuvor behandelten Beispiel verdreifacht werden.

Die in Bild 4.30 skizzierten Ergebnisse zeigen einen ausgeprägten stufenförmigen Abfall der komplementären Verteilungsfunktion der Transferzeit. Dieser stufenförmige Abfall führt dazu, daß die Transferzeit, die mit einer Wahrscheinlichkeit von  $10^{-10}$  überschritten wird, nur geringfügig über der Durchlaufzeit eines Netzknotens liegt. Dies ist in Bild 4.30 durch entsprechende Hilfslinien verdeutlicht. Die variablen Zellverzögerungen können somit ausgeglichen werden, indem die erste Zelle beim Empfänger innerhalb der ATM-Anpassungsschicht um die maximale Durchlaufzeit eines Netzknotens verzögert wird. Diese Verzögerung ist insbesondere dann erforderlich, wenn ein kontinuierlicher Bitstrom wiedergewonnen werden muß (AAL-Dienstklassen A und B). Alle Zellen, die eine größere Transferzeit aufweisen, gehen verloren, da sie nicht mehr in den abgehenden Bitstrom eingefügt werden können. Die durch verspätet eintreffende Zellen hervorgerufenen Zellverluste liegen damit in derselben Größenordnung wie die durch die endliche Puffergröße verursachten Zellverluste.

Da die Durchlaufzeit eines Netzknotens für realistische Puffergrößen weit unterhalb den für die empfindlichsten Dienste spezifizierten Grenzwerten liegt (vgl. Abschnitt 2.1.3), muß die innerhalb des Netzes eingefügte variable Zellverzögerung von der Verkehrssteuerung nicht in Betracht gezogen werden. Mit anderen Worten: Falls die geforderte Zellverlustwahrscheinlichkeit eingehalten werden kann, bleiben auch die Zellverzögerungen innerhalb des durch die übergelagerten Dienste und Anwendungen gesteckten Rahmens.



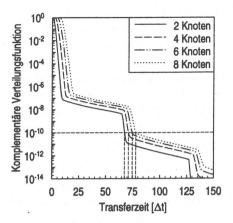


Bild 4.29: Transferzeit über mehrere Verbindungsabschnitte (Referenzverkehr)

Bild 4.30: Transferzeit über mehrere Verbindungsabschnitte für realistische Verkehrsbelastung

Anzahl der Netzknoten $J$	1	2	3	4	5
Analyse	$1,531 \cdot 10^{-3}$	$3,059 \cdot 10^{-3}$	$4,586 \cdot 10^{-3}$	$6,109 \cdot 10^{-3}$	$7,631 \cdot 10^{-3}$
Simulation	$1,546 \cdot 10^{-3} \\ \pm 9,61 \cdot 10^{-5}$	$2,753 \cdot 10^{-3}  \pm 2,62 \cdot 10^{-4}$	$3,623 \cdot 10^{-3}$ $\pm 3,32 \cdot 10^{-4}$	$4,412 \cdot 10^{-3}$ $\pm 5,42 \cdot 10^{-4}$	$5,324 \cdot 10^{-3} \\ \pm 7,24 \cdot 10^{-4}$

Tabelle 4.3: Zellverlustwahrscheinlichkeit über mehrere Verbindungsabschnitte

Etwas größere Zellverzögerungen ergeben sich dann, wenn ausschließlich Verbindungen mit konstanter Bitrate überlagert werden, bei denen ein kontinuierlicher Abfall der komplementären Transferzeitverteilung auftritt. Die in [194] für die Überlagerung geometrischer Ankunftsprozesse erhaltenen Ergebnisse bestätigen allerdings die generelle Aussage, wonach die variable Zellverzögerung gegenüber der Gesamtverzögerung vernachlässigt werden kann. Für weitere Studien sei auf [52, 177] verwiesen.

## 4.3 Schlußfolgerungen für die Verkehrssteuerung in ATM-Netzen

In diesem Unterkapitel sollen die wichtigsten Erkenntnisse der vorangegangenen Untersuchungen nochmals stichwortartig zusammengefaßt werden. Innerhalb des Netzes können

zwei grundlegende Blockierungssituationen unterschieden werden, die sich jeweils den auf der Zell- und der Büschelebene auftretenden Verkehrsfluktuationen zuordnen lassen. Kurzzeitige Zellkollisionen werden dadurch ausgelöst, daß Zellen von verschiedenen Verbindungen gleichzeitig am Multiplexer eintreffen. Die von diesen Blockierungen ausgehenden Zellverluste hängen hauptsächlich von der Anzahl der Verbindungen und dem Verkehrsangebot ab und können durch entsprechende Netzpuffer fast vollständig ausgeglichen werden. Eine obere Grenze für die verbleibenden Zellverluste läßt sich aus dem M/D/1/(S+1)-System ableiten. Umgekehrt kann mit Hilfe dieses Modells ein maximal zulässiges Verkehrsangebot für eine vorgegebene Zellverlustwahrscheinlichkeit definiert werden.

Der schwankende Bitratenbedarf von VBR-Verbindungen führt zu zusätzlichen, langanhaltenden Blockierungssituationen, falls die Übertragungsbitrate überschritten wird. Die primäre Aufgabe der Verkehrssteuerung ist die Begrenzung dieser Blockierungen, so daß eine vordefinierte Netzgüte eingehalten werden kann. Gleichzeitig sollte ein Multiplexgewinn erzielt werden, der sich auf folgende Ausgleichseffekte zurückführen läßt:

- Die statistischen Verkehrsschwankungen verschiedener Verbindungen können sich gegenseitig ausmitteln und führen zu einer Glättung des Überlagerungsverkehrs auf dem betrachteten Übertragungsabschnitt. Der dadurch erzielte Multiplexgewinn kann mit Hilfe des pufferlosen "Fluid Flow"-Modells abgeschätzt werden und hängt nur von der Verteilung des Bitratenbedarfs ab. Das zeitliche Verhalten der Bitratenschwankungen ist ohne Bedeutung. Die Verkehrsschwankungen gleichen sich umso besser aus, je geringer der Verkehrsanteil einer einzelnen Verbindung ist.
- Mit Hilfe der Netzpuffer können kurzzeitige Blockierungen überbrückt werden, was zu einer weiteren Erhöhung des Multiplexgewinns beiträgt. Die verbleibenden Zellverluste und -verzögerungen können aus dem "Fluid Flow"-Modell bestimmt werden. Die durch die Zwischenspeicherung von Zellen erzielte Verringerung der Zellverlustwahrscheinlichkeit ist jedoch äußerst sensitiv in Bezug auf das zeitliche Verhalten der Verkehrsquellen. Insbesondere kann nur dann eine signifikante Erhöhung des Multiplexgewinns erzielt werden, wenn die Größe des Puffers die mittlere Büschelgröße übersteigt.

Weitere Untersuchungen bestätigten die dominierende Rolle der Zellverlustwahrscheinlichkeit für die Netzgüte der ATM-Schicht. Andere Leistungskenngrößen, die das zeitliche Verhalten der Zellverluste oder die Zellverzögerungen betreffen, können indirekt mit Hilfe der Zellverlustwahrscheinlichkeit gesteuert werden. Ferner zeigte sich, daß die Zellverlustwahrscheinlichkeiten von Verbindungen mit unterschiedlichen Verkehrsparametern nur geringfügig voneinander abweichen. Schließlich konnten die in der Literatur getroffenen Annahmen, wonach sich die auf den einzelnen Verbindungsabschnitten eingefügten Zellverluste näherungsweise addieren und die innerhalb des Netzes auftretende Verformung des Verkehrs vernachlässigbar ist, bestätigt werden. Dies ermöglicht die abschnittsweise Annahme einer neuen Verbindung.

### Kapitel 5

### Untersuchung von Verlustprioritätsmechanismen

Die Ergebnisse des vorigen Kapitels bestätigen, daß die Zellverlustwahrscheinlichkeit der entscheidende Netzgüteparameter ist, der das zulässige Verkehrsangebot begrenzt. Da die Empfindlichkeit gegenüber Zellverlusten von Dienst zu Dienst variiert (vgl. Tabelle 2.3), liegt die in Unterabschnitt 2.4.2.1 diskutierte Bereitstellung von zwei Übermittlungsdiensten mit unterschiedlicher Zellverlustwahrscheinlichkeit auf der Hand. Dies ermöglicht eine flexible Anpassung der durch die ATM-Schicht zur Verfügung gestellten Netzgüte an die Dienstgüteanforderungen der darauf aufsetzenden Dienste und Anwendungen. Dem hierdurch erzielten Auslastungsgewinn steht allerdings eine nicht zu unterschätzende Erhöhung des Steuerungsaufwands gegenüber. Der nachfolgend durchgeführte Vergleich unterschiedlicher Realisierungskonzepte bezieht daher neben dem erzielten Auslastungsgewinn auch die Erhöhung der Netzkomplexität mit ein.

#### 5.1 Anwendungsbeispiele und Anforderungen

Verlustprioritäten ermöglichen das selektive Verwerfen unwichtiger Zellen, falls innerhalb des Netzes eine Überlastsituation auftritt. Mit Hilfe einer entsprechenden Puffersteuerung können Zellen mit hoher Verlustpriorität vor einer Blockierung durch niederprioritäre Zellen zuverlässig geschützt werden. Diese Eigenschaft ist für eine Reihe von Anwendungen interessant.

Primär ist hierbei die flexible Auswahl eines im Hinblick auf die Dienstgüteanforderungen des Kommunikationsdienstes geeigneten Übermittlungsdienstes zu nennen (siehe z.B. [227]). Eine hochqualitative Übermittlung ist nur für relativ wenige Anwendungen, wie beispielsweise Videoübertragung, erforderlich. Sprachdienste, die relativ robust gegenüber Zellverlusten

sind, können dagegen mit Hilfe der niederen Prioritätsklasse übermittelt werden. Darüberhinaus zerlegen moderne Codierverfahren für Sprach- [195, 259], Bild- [40] und Videosignale [70, 72] die zu übertragende Information in verschiedene Komponenten mit unterschiedlicher Bedeutung für die subjektiv empfundene Ton- und Bildqualität. Die Zuordnung dieser Informationsströme auf die beiden Prioritätsklassen muß innerhalb des Endgeräts erfolgen. Im Gegensatz zum vorigen Anwendungsfall können die Zellen einer virtuellen Verbindung unterschiedliche Verlustprioritäten besitzen.

In Verbindung mit der Quellflußkontrolle wurde angedacht, hochprioritäre Zellen, welche die überwachten Verkehrsparameter überschreiten, in niederprioritäre Zellen umzuwandeln [7, 15, 54, 55, 65, 121]. Der Vorteil dieses Konzepts liegt darin, daß die so markierten Zellen nur dann abgewiesen werden, wenn dies aufgrund einer innerhalb des Netzes auftretenden Überlast auch tatsächlich nötig ist. Allerdings können bei einem unkontrollierten Markieren von Zellen keinerlei Angaben in Bezug auf die Zellverlustwahrscheinlichkeit innerhalb der niederprioritären Klasse gemacht werden. Um die bisher genannten Anwendungsbeispiele kombinieren zu können, sehen die neueren ATM-Spezifikationen [7, 121] eine strikte Begrenzung des gesamten innerhalb der niederen Prioritätsklasse gesendeten Verkehrs vor, d.h. eine markierte Zelle wird verworfen, falls das für die niedere Prioritätsklasse vereinbarte Verkehrskontingent überschritten würde. Die Verkehrsaufteilung innerhalb des Endgeräts muß sich an den für die hohe und die niedere Priorität definierten Vorgaben orientieren. Ein erster Ansatz zur gleichzeitigen Formung des hoch- und niederprioritären Verkehrs ist in [203] zu finden.

Ein weiteres Anwendungsszenario ergibt sich aus der Vernetzung von Lokalen Netzen über ein ATM-Netz. Die Kopplung dieser unterschiedlichen Netztypen wird dadurch erschwert, daß Lokale Netze auf den unteren Protokollschichten verbindungslos arbeiten, wogegen das ATM-Netz verbindungsorientiert betrieben wird. Falls verbindungsloser Verkehr an der Netzkoppeleinheit eintrifft, muß zunächst die innerhalb des ATM-Netzes benötigte Übertragungskapazität abgeschätzt und reserviert werden, bevor dieser Verkehr ins ATM-Netz gelangen darf. Anstatt die während dieser Zeit eintreffende Information in der Netzkoppeleinheit zwischenzuspeichern, wird erwogen, diesen Verkehr mit niedriger Verlustpriorität zu übermitteln, bis die angeforderte Übertragungskapazität zugeteilt ist. Allerdings muß auch bei diesem Anwendungsbeispiel eine Beschränkung des von der Netzkoppeleinheit markierten Verkehrs stattfinden. In einem konkreten Realisierungsvorschlag [71] wird dies dadurch erreicht, daß der niederprioritäre Verkehr nur die auf dem zugeordneten virtuellen Pfad frei verfügbare Übertragungskapazität ausschöpfen darf. Die Netzkoppeleinheit wird durch periodische Meldungen über die auf allen abgehenden virtuellen Pfaden verfügbare Übertragungskapazität laufend unterrichtet.

Aus den diskutierten Anwendungsbeispielen ergeben sich wichtige Anhaltspunkte für die Konzeption und Realisierung geeigneter Verlustprioritätsmechanismen. Eine entscheidende Randbedingung ergibt sich aus der Tatsache, daß die meisten Anwendungsbeispiele von beiden Prioritätsklassen innerhalb einer virtuellen Verbindung Gebrauch machen. Da die Zellreihenfolge innerhalb von virtuellen Verbindungen zwingend vorgeschrieben ist [112], darf die zeitliche Reihenfolge der hoch- und niederprioritären Zellen nicht verändert werden. Dies kann entweder dadurch erreicht werden, daß alle Netzpuffer nach dem FIFO-Prinzip arbeiten oder die Zellreihenfolge beim Empfänger wiederhergestellt wird. Da für die Wiederherstellung der korrekten Zellreihenfolge eine Reihenfolgenummer erforderlich wäre, die z.B. für den AAL-Typ 5 nicht vorgesehen ist [120], wird diese Lösungsvariante bei den nachfolgenden Betrachtungen ausgeklammert. Deshalb darf die zeitliche Reihenfolge der Zellen innerhalb des Netzes nicht verändert werden. Insbesondere verbietet dies eine zeitliche Priorisierung von Zellen.

In Bezug auf die Leistungsfähigkeit der einzelnen Mechanismen ist zu erwähnen, daß für beide Prioritätsklassen bestimmte Vorgaben bezüglich der Zellverlustwahrscheinlichkeit eingehalten werden müssen. Unter diesen Voraussetzungen sollten die Verfahren — unabhängig vom Ankunftsprozeß der Zellen oder der momentanen Prioritätszusammensetzung — eine möglichst große Netzauslastung erzielen, um einen wirtschaftlichen Betrieb des Netzes gewährleisten zu können. In Unterkapitel 5.3 wird auf dieser Basis ein Leistungsvergleich der im folgenden Unterkapitel diskutierten Mechanismen durchgeführt.

Ein weiteres Vergleichskriterium ist der für die Prioritätsverwaltung der Netzpuffer hinzukommende Steuerungs- und Implementierungsaufwand. Die Unterstützung verschiedener Verlustprioritäten ist nur dann attraktiv, wenn dadurch ein Kostenvorteil erzielt wird, d.h. die durch die Erhöhung der Netzauslastung erzielte Kostenersparnis sollte gegenüber den zusätzlichen Steuerungskosten überwiegen. Dies führt zu einfachen Prioritätsmechanismen, die nur auf elementare Operationen zurückgreifen und damit eine schnelle und kostengünstige "Hardware-Realisierung" zulassen. Ähnliche Forderungen leiten sich auch daraus ab, daß der Pufferzugriff in Echtzeit erfolgen muß.

# 5.2 Beschreibung, Modellierung und Implementierung der Mechanismen

In der Literatur wurden verschiedene Verlustprioritätsmechanismen mit einer Vielzahl unterschiedlicher Varianten vorgeschlagen. Bild 5.1 gibt einen Überblick über die Konzepte, die diesen Ansätzen zugrundeliegen. Die einfachste Möglichkeit, verschiedene Zellverlustwahrscheinlichkeiten zu erzielen, basiert auf der Verwendung getrennter physikalischer Wege für die unterschiedlichen Prioritätsklassen [142, 159]. Die für niederprioritäre Zellen reservierten Wege können stärker ausgelastet werden, da eine höhere Zellverlustwahrscheinlichkeit toleriert werden kann.

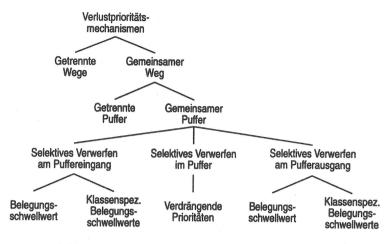


Bild 5.1: Klassifikation von Verlustprioritätsmechanismen

Bei einem gemeinsam genutzten Weg für beide Prioritätsklassen sind Eingriffe in die Pufferverwaltung unumgänglich. Der Puffer kann beispielsweise in zwei logisch getrennte Teilbereiche unterteilt werden, die jeweils einer Prioritätsklasse fest zugeordnet sind [70, 159]. Bei einer gemeinsamen Nutzung des Pufferspeichers müssen Zellen selektiv, entsprechend ihrer Priorität, verworfen werden. Dies kann sich auf ankommende Zellen, bereits im Puffer befindliche Zellen oder zur Bedienung anstehende Zellen beziehen, d.h. die Priorisierung kann am Eingang, innerhalb oder am Ausgang des Puffers stattfinden.

Eingangsseitig können Zellen selektiv verworfen werden, indem eine zusätzliche Zugangsbeschränkung eingeführt wird. Folgende Zugangssteuerungen können unterschieden werden:

- Klassenspezifische Belegungsschwellwerte [15, 41]: Eine ankommende Zelle wird abgewiesen, falls die Anzahl der von dieser Prioritätsklasse wartenden Zellen einen bestimmten Maximalwert erreicht hat.
- Globaler Belegungsschwellwert [15, 29, 142]: Eine ankommende niederprioritäre Zelle wird abgewiesen, falls der Pufferfüllstand einen vorgegebenen Schwellwert überschreitet.

Eine weitergehende Prioritätskontrolle ergibt sich, wenn auch bereits im Puffer befindliche Zellen selektiv verworfen werden können. Da dies nur dann Sinn macht, wenn Platz für eine neu ankommende Zelle geschaffen werden muß, spricht man auch von verdrängenden Prioritäten [227].

Am Ausgang des Puffers wird die Priorisierung der Zellströme dadurch erreicht, daß eine zur Bedienung anstehende niederprioritäre Zelle zugunsten einer nachfolgenden hochprioritären

Zelle selektiv verworfen werden kann. Die Entscheidung, ob eine niederprioritäre Zelle verworfen bzw. bedient wird, orientiert sich wiederum am Pufferfüllstand [162] oder an der Anzahl der sich im Puffer befindlichen hochprioritären Zellen [230].

Daneben sind Prioritätsmechanismen denkbar, welche mehrere der beschriebenen Grundkonzepte anwenden. Konkrete Vorschläge kombinieren verdrängende Prioritäten mit dem selektiven Verwerfen von niederprioritären Zellen am Pufferein- [135] bzw. -ausgang [230]. Darüberhinaus wurde in [162] ein Mechanismus untersucht, der die Zellen sowohl am Eingang als auch am Ausgang des Puffers selektiv verwerfen kann. Noch häufiger werden Verlustprioritäten mit zeitlichen Prioritäten verknüpft, um sowohl auf die Zellverluste als auch auf die Zellverzögerungen Einfluß nehmen zu können (siehe z.B. [9, 75, 170]). Da dies jedoch zwangsläufig zu einer Störung der Zellreihenfolge führt, werden diese Mechanismen nicht weiter betrachtet. Zeitliche Prioritäten erübrigen sich auch aus dem Gesichtspunkt, daß die innerhalb der ATM-Schicht eingefügten Zellverzugsschwankungen die Dienstgüte der darauf aufbauenden Dienste und Anwendungen nicht einschränken.

#### 5.2.1 Getrennte Wegeführung

Dieses Verfahren trennt die hoch- und niederprioritären Zellen bereits beim Netzzugang in unterschiedliche Verkehrsströme auf, die das Netz auf verschiedenen physikalischen Wegen durchlaufen. Jede netzinterne Leitung wird in der Regel nur von einer Prioritätsklasse genutzt, d.h. es findet keine Betriebsmittelteilung zwischen den unterschiedlichen Verkehrsströmen statt. Bei ungünstiger Prioritätszusammensetzung kann allerdings eine Mischung der Prioritäten auf einzelnen Übertragungsabschnitten nötig werden. Die dafür verwendeten Leitungen dürfen nur so stark ausgelastet werden, daß die für die hohe Priorität angestrebte Zellverlustwahrscheinlichkeit eingehalten werden kann.

Aufgrund der sehr einfachen Funktionsweise soll dieses Verfahren bei der in Unterkapitel 5.3 durchgeführten Leistungsuntersuchung einbezogen werden. Da die Verwaltung der Netzpuffer nicht geändert werden muß, können alle Verbindungsabschnitte durch das in Bild 3.5 gezeigte Modell eines statistischen Multiplexers beschrieben werden. Die Leistungsuntersuchung auf Zellebene basiert daher auf dem in Bild 5.2 skizzierten Verkehrsmodell. Die hochund niederprioritären Verkehre werden innerhalb der Prioritätsklassen H ("High") und L ("Low") zusammengefaßt.

Diese Lösungsvariante hat den Vorteil, daß das Puffermanagement und die Quellflußkontrolle beibehalten werden können, d.h. dieses Verfahren könnte ohne "Hardware"-Änderungen in bereits bestehende Systeme eingebunden werden. Die gesamte Prioritätsbehandlung wird von der Verbindungssteuerung übernommen, wobei insbesondere die Verkehrslenkung beeinflußt wird, da bei der Auswahl eines Weges die Priorität der Verbindung berücksichtigt werden

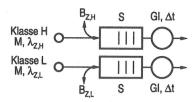


Bild 5.2: Verkehrsmodell für getrennte Wegeführung

muß. Auf die Verbindungsannahme wirkt sich dieses Verfahren nur durch veränderte Parameterwerte aus; der eigentliche Annahmealgorithmus kann beibehalten werden. Im Vergleich zu den übrigen Prioritätsmechanismen ergeben sich die geringsten Kostensteigerungen.

Andererseits ist die Anwendbarkeit und die Flexibilität dieses Konzepts stark eingeschränkt, da jede Verbindung nur Zellen einer einzigen Prioritätsklasse enthalten darf. Viele der vorgenannten Applikationen scheiden damit aus. Ferner kann dieses Verfahren nur dann angewendet werden, wenn — zumindest auf einzelnen Verbindungsabschnitten — getrennte physikalische Wege zur Auswahl stehen. Für eine zufriedenstellende Leistungsfähigkeit sollten sogar wesentlich mehr als zwei alternative Wege zur Verfügung stehen (vgl. Abschnitt 5.3.1.3).

#### 5.2.2 Selektives Verwerfen am Puffereingang

Die zweite Gruppe von Verfahren regelt den Pufferzugang für jede Prioritätsklasse individuell, abhängig vom momentanen Belegungszustand des Puffers. Die einfachste Variante reserviert  $S_H$  Pufferplätze für die hohe und  $S_L = S - S_H$  Pufferplätze für die niedrige Verlustpriorität, d.h. der Puffer wird in zwei logisch getrennte Teilbereiche aufgespalten [70, 159]. Die Implementierung dieses Verfahrens ist sehr einfach, da die Pufferzugangssteuerung nur die Anzahl der im Puffer befindlichen hoch- und niederprioritären Zellen auf die Werte  $S_L$  bzw.  $S_H$  beschränken muß.

Andererseits könnte durch eine gemeinsame Nutzung des Puffers eine höhere Netzauslastung erzielt werden. Die strikte (logische) Aufteilung des Puffers kann dadurch aufgehoben werden, daß die Summe der Belegungsschwellwerte  $S_H$  und  $S_L$  die Puffergröße S übersteigen darf. In diesem Fall wird die Anzahl der für die hohe bzw. niedere Priorität reservierten Pufferplätze auf  $S - S_L$  bzw.  $S - S_H$  begrenzt [41], womit  $S_L + S_H - S$  Warteplätze für beide Prioritätsklassen zur Verfügung stehen. Dieses unter der Bezeichnung "Movable Boundary" bekannte Prinzip wird auch für die Kanalaufteilung in hybriden Vermittlungssystemen angewandt. In [15] wurde ein sinnvoller Sonderfall dieses Verfahrens behandelt, bei dem nur die Anzahl der im Puffer befindlichen niederprioritären Zellen begrenzt wird, während die hohe Priorität den Puffer vollständig belegen kann.

Alle bislang vorgestellten Zugangsmechanismen haben den Nachteil, daß die Entscheidung über die Annahme einer ankommenden Zelle unabhängig vom Belegungszustand der jeweils anderen Prioritätsklasse getroffen wird, falls der Puffer nicht vollständig belegt ist. Für einen effizienten Schutz des hochprioritären Verkehrs ist es allerdings günstig, niederprioritäre Zellen auch dann abzuweisen, wenn der Puffer hauptsächlich durch hochprioritäre Zellen belegt ist. Aus diesem Grund verwirft das in [15, 29, 142] beschriebene Schwellwertverfahren ankommende Zellen in Abhängigkeit vom globalen Pufferfüllstand. Zellen mit hoher Priorität werden nur bei vollständig belegtem Puffer abgewiesen, während Zellen die der niederen Prioritätsklasse angehören, bereits bei Überschreitung eines Pufferfüllstands  $S_L < S$  verworfen werden.

Unter den genannten Strategien weist das Schwellwertverfahren die höchste Leistungsfähigkeit auf, wie die in [15, 159] beschriebenen Studien bestätigen. Ferner konnte in [230] nachgewiesen werden, daß das Schwellwertverfahren auch allen anderen Konzepten, welche die Zellen ausschließlich beim Pufferzugang abweisen, überlegen ist. Der in [230] geführte mathematische Beweis stützt sich auf die Minimierung der durch Zellverluste verursachten Gesamtkosten, wobei die Kosten eines Zellverlusts in Anlehnung an die für die betreffende Klasse einzuhaltende Zellverlustwahrscheinlichkeit festgelegt werden. Da das Schwellwertverfahren außerdem Implementierungsvorteile gegenüber den übrigen Varianten aufweist, soll nur dieser Mechanismus in den nachfolgenden Leistungsvergleich einbezogen werden. Das hierzu verwendete Verkehrsmodell ist in Bild 5.3 wiedergegeben.

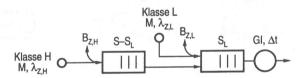


Bild 5.3: Verkehrsmodell des Schwellwertsystems

Für die besprochenen Verfahren kann die Priorität jeder einzelnen Zelle individuell festgelegt werden. Dies führt in mehrfacher Hinsicht zu einer Erhöhung des Steuerungs- und Implementierungsaufwands. Beispielsweise verdoppelt sich der Aufwand für die Quellflußkontrolle, da der Gesamtverkehr und der Verkehr mit hoher Verlustpriorität getrennt überwacht werden muß [7, 121]<sup>1</sup>. Die Verbindungsannahmefunktion muß ebenfalls überprüfen, ob die für beide Prioritätsklassen geforderte Netzgüte eingehalten werden kann, falls eine zusätzliche Verbindung angenommen wird. Dies führt auch für diese Steuerungsfunktion in erster Näherung zu einer Verdopplung des Speicherplatzbedarfs und des Rechenaufwands. Schließlich muß

<sup>&</sup>lt;sup>1</sup>Die in den Spezifikationen des ATM Forums [7] vorgesehene Option, den hoch- und niederprioritären Verkehr getrennt zu überwachen, führt auf denselben Implementierungsaufwand.

die Prioritätsbehandlung von den Signalisierprotokollen und den Endgeräten unterstützt werden.

Neben diesen für alle Verlustprioritätsmechanismen erforderlichen Änderungen muß die Eingangssteuerung der Netzpuffer eine Prioritätsbehandlung durchführen. Bild 5.4 zeigt die hierzu nötigen Funktionsblöcke und deren Zusammenwirken. Das Prioritätsbit der am Puffereingang anliegenden Zelle wird zunächst an eine Umsetzungstabelle angelegt, aus der sich der für die betreffende Prioritätsklasse gültige Schwellwert ergibt. Der anschließende Vergleich mit dem aktuellen Pufferfüllstand liefert die Entscheidung über die Annahme (log. "1") bzw. Ablehnung der Zelle. Falls die Zelle angenommen wird, muß der für die Speicherung des Pufferfüllstands zuständige Zähler inkrementiert werden. Ebenso muß dieser Zähler für jede aus dem Puffer ausgelesene Zelle dekrementiert werden. Im Rahmen einer Studienarbeit [8] wurden die einzelnen Funktionsblöcke dieser Prioritätssteuerung bis auf Gatterebene verfeinert. Der Implementierungsaufwand für das Schwellwertverfahren liegt bei ca. 100 Gatteräquivalenten, was gegenüber dem gesamten Schaltungsaufwand vernachlässigt werden kann. Die zusätzliche Verarbeitungszeit ist ebenfalls unkritisch und liegt im Bereich von 50 ns für die zugrundegelegte CMOS-Technologie.

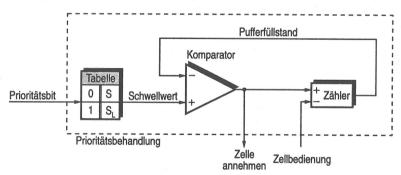


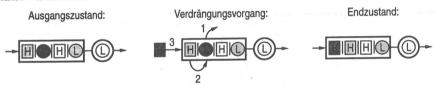
Bild 5.4: Blockschaltbild des Schwellwertverfahrens

#### 5.2.3 Verdrängende Prioritäten

Eine weitergehende Bevorzugung der hohen Priorität ergibt sich dann, wenn eine Zelle mit niedriger Verlustpriorität auch nachträglich aus dem Puffer entfernt werden kann, um für eine eintreffende hochprioritäre Zelle Platz zu schaffen [88, 227]. Dieses Verdrängungsprinzip gewährleistet, daß eine Zelle mit hoher Priorität nur dann verloren geht, wenn der Puffer ausschließlich mit Zellen der Klasse H belegt ist. Demgegenüber wird eine Zelle mit niedriger Priorität zurückgewiesen, falls keine freien Pufferplätze vorhanden sind.

In der Literatur wurden verschiedene Verfeinerungen dieses Grundprinzips vorgeschlagen. In [230] wurde bewiesen, daß sich die innerhalb der Klasse H auftretenden Zellverluste weiter reduzieren, falls sich niederprioritäre Zellen gegenseitig aus dem Puffer drängen können. Diese globale Verdrängung wird in Abschnitt 5.3.1.3 ausführlich diskutiert. Dagegen wird bei dem in [227] gemachten Vorschlag eine Verdrängung nur dann zugelassen, wenn die Anzahl der im Puffer gespeicherten niederprioritären Zellen einen bestimmten Mindestwert überschreitet. Durch eine Anpassung dieses Schwellwerts läßt sich das Verhältnis der für die Klasse H und die Klasse L erhaltenen Zellverlustwahrscheinlichkeiten ändern, während die gesamte Zellverlustwahrscheinlichkeit konstant bleibt. Die in Abschnitt 5.3.1.3 beschriebenen Ergebnisse zeigen, daß der durch diesen zusätzlichen Freiheitsgrad erzielte Gewinn vernachlässigbar ist.

Der Verdrängungsvorgang erfordert die in Bild 5.5 veranschaulichten Teilschritte zur Sicherung der Zellreihenfolge. Im ersten Schritt muß die zu verdrängende niederprioritäre Zelle ausgewählt werden. Die Reihenfolge, in der die niederprioritären Zellen aus dem Puffer verdrängt werden, ist durch die Verdrängungsstrategie geregelt. Aus Implementierungsgesichtspunkten sollte entweder die zuerst ("First In First Drop", FIFD) oder die zuletzt ("Last In First Drop", LIFD) eingeschriebene niederprioritäre Zelle verdrängt werden. Im nächsten Schritt müssen alle dahinterliegenden Zellen um einen Pufferplatz verschoben werden, damit die neu ankommende Zelle im abschließenden Schritt auf den letzten Pufferplatz eingespeichert werden kann.



Schritt 1: Auswahl und Entfernen einer Zelle der Klasse L (Verdrängungsstrategie LIFD)

Schritt 2: Nachrücken aller folgenden Zellen

Schritt 3: Speichern der ankommenden Zelle der Klasse H

Bild 5.5: Verdrängung einer niederprioritären Zelle durch eine hochprioritäre Zelle

Die Implementierung der angesprochenen Reihenfolgesicherung ist relativ aufwendig und erfordert eine Neustrukturierung des Puffermanagements, das nun in drei Hierarchieebenen gegliedert werden kann. Auf der untersten Ebene erfolgt die physikalische Speicherung des Zellinhalts innerhalb eines "Random Access Memory" (RAM). Die Anfangsadressen, unter denen die verschiedenen Zellen gespeichert sind, werden von der darüberliegenden Ebene verwaltet.

Diese mittlere Ebene sorgt für die Einhaltung der Bedienreihenfolge und die Zuteilung freier Pufferplätze an eintreffende Zellen. Falls nur ankommende bzw. zur Bedienung anstehende Zellen verworfen werden dürfen, reicht es aus, die Speicheradressen der zuerst und der zuletzt eingespeicherten Zellen in entsprechenden Registern zu halten, da die Pufferplätze des physikalischen Speichers zyklisch belegt werden. Beim Verdrängungsverfahren hingegen muß diese Ebene die Anfangsadressen aller im Puffer befindlichen Zellen verwalten, da diese zyklische Belegungsreihenfolge nicht mehr gegeben ist. Alle Adreßeinträge müssen entsprechend ihrer Bedienreihenfolge in eine Warteschlange eingereiht werden. Diese Warteschlange muß entweder als verkettete Liste [8, 42, 44] oder als Schieberegister [8, 42, 223] realisiert werden, um den in Bild 5.5 veranschaulichten Verdrängungsvorgang zu ermöglichen.

Die oberste Hierarchieebene übernimmt die Steuerung des Verdrängungsvorgangs, der in verschiedenen Teilschritten abläuft. Zunächst muß überprüft werden, ob die für eine Verdrängung notwendigen Vorbedingungen gegeben sind. Im folgenden Schritt wird die Warteposition der zur Verdrängung vorgesehenen Zelle bestimmt. Dazu könnte beispielsweise eine weitere Warteschlange eingerichtet werden, welche die Wartepositionen aller niederprioritären Zellen zwischenspeichert. Die in Bild 5.6 gezeigte Lösungsvariante vermeidet dies, indem die in der darunterliegenden Ebene angesiedelte Warteschlange neben der physikalischen Speicheradresse auch das Prioritätsbit jeder Zelle zwischenspeichert. Abschließend wird der in der mittleren Ebene ablaufende Verdrängungsvorgang angestoßen.

Eine prinzipielle Möglichkeit zur Realisierung des Verdrängungsverfahrens ist in Bild 5.6 gezeigt. Die Anfangsadressen der innerhalb des physikalischen Speichers befindlichen Zellen

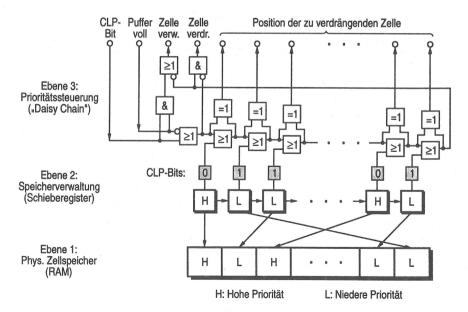


Bild 5.6: Eine mögliche Pufferarchitektur für verdrängende Prioritäten

werden bei dieser Lösung mit Hilfe eines Schieberegisters verwaltet (eine detaillierte Beschreibung dieser Schaltungskomponente ist in [223] zu finden). Aus den an die Speicheradressen angehängten Prioritätsbits wird über das "Daisy Chain"-Verfahren [148] die Position der zu verdrängenden Zelle ermittelt. Die für das Verwerfen bzw. Verdrängen einer Zelle benötigten Steuersignale werden mit Hilfe einer einfachen kombinatorischen Logik erzeugt. Für weitere Lösungsvorschläge, welche die Zugriffszeit durch eine hierarchische Suche verkürzen, sei auf [8] verwiesen.

Die dort durchgeführte Abschätzung des Schaltungsaufwands und der Verarbeitungszeit bestätigt, daß das Verdrängungsverfahren keine grundsätzlichen Realisierungsprobleme aufwirft. Der zusätzliche Implementierungsaufwand liegt allerdings deutlich höher als beim Schwellwertverfahren. Die genauen Werte hängen von der gewählten Pufferarchitektur ab. Das Verdrängungsverfahren ist insbesondere dann attraktiv, wenn für die Warteschlangenverwaltung ohnehin eine verkettete Liste bzw. ein Schieberegister verwendet werden muß, da die Zellen nicht in FIFO-Reihenfolge ausgelesen werden. Dies ist beispielsweise dann der Fall, wenn ein Eingangspuffer mit wahlfreiem Zugriff [223] oder ein zentraler Zellspeicher [44] verwendet wird. In diesen Fällen ist der für die Prioritätssteuerung nötige Zusatzaufwand wesentlich geringer als der gesamte für die Verwaltung des Puffers nötige Schaltungsaufwand.

#### 5.2.4 Selektives Verwerfen am Pufferausgang

Wie zuvor angedeutet, können zur Bedienung anstehende niederprioritäre Zellen verworfen werden, falls der gesamte Pufferfüllstand oder die Anzahl der im Puffer befindlichen hochprioritären Zellen einen bestimmten Schwellwert überschreitet. Damit können zusätzliche Pufferplätze für eventuell eintreffende hochprioritäre Zellen bereitgehalten werden.

Diese Verfahren sind den Prioritätsmechanismen, die ankommende Zellen selektiv verwerfen, sowohl aus Leistungs- als auch aus Implementierungsgesichtspunkten unterlegen. Die Leistungsfähigkeit der ausgangsseitigen Prioritätssteuerung wird hauptsächlich dadurch eingeschränkt, daß die Priorisierung nicht direkt durch den Ankunftsprozeß — die eigentliche Ursache möglicher Überlastsituationen — gesteuert wird. Ferner tragen die unmittelbar vor der Bedienung verworfenen Zellen zur Belastung des Pufferspeichers bei. Diese "Blindlast" wirkt sich nachteilig auf das zulässige Gesamtverkehrsangebot aus. Die in [162] beschriebenen analytischen Ergebnisse bekräftigen diese qualitativen Überlegungen. Die Implementierung dieser Mechanismen ist ebenfalls aufwendiger als die Implementierung einer eingangsseitigen Prioritätssteuerung (vgl. Bild 5.4), da innerhalb eines Zelltaktes unter Umständen mehrere Zellen aus dem Zellpuffer entfernt werden müssen. Aufgrund dieser offensichtlichen Nachteile werden diese Verfahren keiner detaillierten Untersuchung unterzogen.

#### 5.3 Vergleichende Leistungsbewertung

#### 5.3.1 Leistungsuntersuchung auf Zellebene

Dieser Abschnitt dient einem Leistungsvergleich der oben beschriebenen Prioritätsmechanismen. Da die Netzpuffer hauptsächlich für den Ausgleich kurzzeitiger Überlastsituationen zuständig sind, können die Zellankünfte in beiden Prioritätsklassen durch Poisson-Verkehre mit den Ankunftsraten  $\lambda_{Z,H}$  und  $\lambda_{Z,L}$  modelliert werden (vgl. Abschnitt 4.1.2). Im Rahmen einer Studienarbeit wurden die nachfolgenden Ansätze auch auf geometrische Ankunftsprozesse übertragen [93]. Die beschriebenen Mechanismen sind auch für andere Anwendungsbereiche interessant, weshalb der Analyse eine allgemein verteilte Bedienzeit zugrundegelegt wird. Es wird weiter angenommen, daß beide Prioritätsklassen dieselbe Bedienzeitverteilung aufweisen. Falls keine Prioritätssteuerung innerhalb der Netzknoten vorgesehen ist, führt dies auf das bekannte Warte-Verlustsystem vom Typ M/GI/1/(S+1), welches z.B. in [78] behandelt wird. Im folgenden werden ähnliche Lösungsansätze für das Schwellwert- und das Verdrängungsverfahren vorgestellt.

#### 5.3.1.1 Analyse des Schwellwertverfahrens

Das in Bild 5.3 gezeigte Verkehrsmodell des Schwellwertverfahrens kann mit Hilfe einer eingebetteten Markoff-Kette (vgl. Unterkapitel 3.4) analysiert werden. Die Regenerationszeitpunkte dieser eingebetteten Markoff-Kette entsprechen den Zeitpunkten, zu denen die gedächtnisbehaftete Bedienphase abgeschlossen ist und eine Zelle das System verläßt. Der zu den Regenerationszeitpunkten auftretende Systemzustand kann durch einen Wahrscheinlichkeitsvektor  $\pi_Z = (\pi_Z(0), \pi_Z(1), ..., \pi_Z(S))$  beschrieben werden, dessen x-te Komponente  $\pi_Z(x)$  (x=0,1,...,S) die Wahrscheinlichkeit angibt, daß eine abgehende Zelle x Zellen im Puffer zurückläßt. Die Übergangswahrscheinlichkeiten zwischen den Systemzuständen x und y zu zwei aufeinanderfolgenden Regenerationszeitpunkten können, wie in Unterkapitel 3.4 erläutert, innerhalb der Übergangsmatrix Q = [q(y|x)] zusammengefaßt werden. Nach Gleichung 3.20 gilt

$$\boldsymbol{\pi}_Z = \boldsymbol{\pi}_Z \cdot \boldsymbol{Q} \ . \tag{5.1}$$

Für die Zustandsübergänge zwischen zwei eingebetteten Zeitpunkten ist die Anzahl der während dieser Zeitspanne in den Puffer eingeschriebenen Zellen entscheidend. Da der Puffer in verschiedene Teilbereiche aufgeteilt ist, müssen folgende Fälle unterschieden werden:

• Für Übergänge zwischen den Zuständen  $x \leq S_L$  und  $y \leq S_L - 1$  können alle am Multiplexer eintreffenden Zellen in den Puffer gelangen. Die entsprechende Zellankunftsrate

beträgt  $\lambda_Z = \lambda_{Z,L} + \lambda_{Z,H}$ . Während einer Bedienzeit, die der Verteilungsdichtefunktion  $f_{Z,H}(t)$  gehorcht, treffen a Zellen mit folgender Wahrscheinlichkeit ein [150]:

$$q_1(a) = \int_{t=0}^{\infty} \frac{(\lambda_Z t)^a}{a!} \exp\left(-\lambda_Z t\right) \cdot f_{Z,H}(t) dt.$$
 (5.2)

• Die Zustandsübergänge für  $x \geq S_L + 1$  und  $y \geq S_L$  hängen dagegen nur von den Zellankünften innerhalb der Klasse H ab, da in diesen Systemzuständen alle niederprioritären Zellen abgewiesen werden. Analog zum obigen Fall ergibt sich folgende Wahrscheinlichkeit für a Zellankünfte mit hoher Priorität während einer Bedienung:

$$q_2(a) = \int_{t=0}^{\infty} \frac{(\lambda_{Z,H} t)^a}{a!} \exp\left(-\lambda_{Z,H} t\right) \cdot f_{Z,H}(t) dt.$$
 (5.3)

• Bei den restlichen möglichen Zustandsübergängen  $(x \leq S_L, y \geq S_L)$  wird die Pufferbelegungsschwelle  $S_L$  überschritten. Für die ersten  $a_1 = S_L + 1 - x$  Zellankünfte gilt noch die gesamte Ankunftsrate  $\lambda_Z$ , da der gemeinsame Pufferbereich noch nicht voll belegt ist. Anschließend treffen noch weitere  $a - a_1$  Zellen mit der reduzierten Ankunftsrate  $\lambda_{Z,H}$  ein, da alle niederprioritären Zellen nach Erreichen der Belegungsschwelle  $S_L$  abgewiesen werden. Die entsprechenden Ankunftswahrscheinlichkeiten werden mit  $q_3(a_1,a)$  bezeichnet und könnten im Prinzip wie die Ankunftsverteilungen  $q_1(a)$  und  $q_2(a)$  berechnet werden. Für deterministische Bedienzeiten ist diese Lösung allerdings numerisch instabil, da Terme mit alternierendem Vorzeichen aufsummiert werden müssen.

Eine numerisch stabile Lösung ergibt sich aus folgender Überlegung: Da es sich um Poisson-Ankünfte handelt, gehört jede ankommende Zelle mit der Wahrscheinlichkeit  $\lambda_{Z,L}/\lambda_Z$  zur niederen Priorität und mit der komplementären Wahrscheinlichkeit  $\lambda_{Z,H}/\lambda_Z$  zur hohen Priorität. Falls innerhalb einer Bedienphase insgesamt b Zellen eintreffen und der gemeinsam genutzte Teil des Puffers nach  $a_1$  Zellankünften voll belegt ist, so wird jede der verbleibenden  $b-a_1$  Zellen mit der Wahrscheinlichkeit  $\lambda_{Z,L}/\lambda_Z$  verworfen, da es sich um eine niederprioritäre Zelle handelt. Die Summation über alle möglichen Werte  $b \geq a$  liefert die Ankunftswahrscheinlichkeit

$$q_3(a_1, a) = \sum_{b=a}^{\infty} q_1(b) \binom{b-a_1}{a-a_1} \left(\frac{\lambda_{Z,H}}{\lambda_Z}\right)^{a-a_1} \cdot \left(\frac{\lambda_{Z,L}}{\lambda_Z}\right)^{b-a} . \tag{5.4}$$

Diese Reihe konvergiert sehr schnell, weshalb die numerische Auswertung schon nach wenigen Teilschritten abgebrochen werden kann.

Aus diesen Ankunftswahrscheinlichkeiten können die Übergangswahrscheinlichkeiten wie beim M/GI/1/(S+1)-System bestimmt werden. Die Folgezustände y < S werden vom Ausgangszustand x > 0 erreicht, indem a = y - x + 1 Zellen in der dazwischenliegenden Bedienphase in den Puffer eingeschrieben werden. Falls sich das System nach einem Regenerationszeitpunkt im leeren Zustand befindet (x = 0), muß zunächst diejenige Zelle eintreffen,

deren Bedienende den nächsten Regenerationszeitpunkt kennzeichnet. Während diese Zelle bedient wird, gelangen weitere a Zellen in den Puffer, was zum Folgezustand y=a führt. Die Übergangswahrscheinlichkeiten in den noch verbleibenden Folgezustand y=S ergeben sich aus der Überlegung, daß die Summe der Übergangswahrscheinlichkeiten der vom Zustand x ausgehenden Übergänge den Wert 1 ergeben muß. In Bild 5.7 sind die Übergangswahrscheinlichkeiten aus den Zuständen x=1 und  $x=S_L+1$  veranschaulicht.

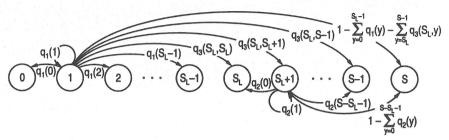


Bild 5.7: Zustandsübergänge aus den Zuständen 1 und  $S_L+1$ 

Mit diesen Vorüberlegungen kann die Übergangsmatrix Q = [q(y|x)] angegeben werden:

Es läßt sich einfach zeigen, daß die durch die Übergangsmatrix definierten Übergangsgleichungen linear abhängig sind. Eine eindeutige Lösung ergibt sich, indem die Zustandswahr-

scheinlichkeit  $\pi_Z(0)$  zunächst auf einen beliebigen Wert gesetzt wird. Mit dieser Annahme können die Zustandswahrscheinlichkeiten  $\pi_Z(1), \, \pi_Z(2), \, ..., \, \pi_Z(S)$  nacheinander aus den Übergangsgleichungen in die Zustände 0, 1, ..., S-1 bestimmt werden. Abschließend müssen alle berechneten Werte entsprechend der Normierungsbedingung

$$\sum_{x=0}^{S} \pi_Z(x) = 1 \tag{5.6}$$

neu skaliert werden, um die durch den falschen Anfangswert für  $\pi_Z(0)$  verursachten Abweichungen zu korrigieren (siehe z.B. [200]).

Im nächsten Lösungsschritt muß die von einer ankommenden Zelle angetroffene Zustandsverteilung  $p_Z(x)$  berechnet werden, aus der sich die charakteristischen Leistungskenngrößen ergeben. In [45] wird dazu der folgende Satz bewiesen, der für Bediensysteme vom Typ G/G/1/(S+1) gilt: Eine ankommende Anforderung, die in das Bediensystem gelangen kann, "sieht" mit derselben Wahrscheinlichkeit den Systemzustand x wie eine Anforderung, die das System verläßt. Dies kann damit begründet werden, daß nach einem durch eine einzelne Ankunft hervorgerufenen Zustandsübergang von x nach x+1 genau ein durch ein Bedienende ausgelöster Übergang in Gegenrichtung stattfinden muß, bevor ein erneuter Zustandsübergang von x nach x+1 möglich ist. Für das vorliegende System führt diese Überlegung auf den folgenden Zusammenhang:

$$\pi_{Z}(x) = \begin{cases} \frac{\lambda_{Z} \cdot p_{Z}(x)}{\lambda_{Z} - \lambda_{Z,H} \cdot p_{Z}(S+1) - \lambda_{Z,L} \sum_{y=S_{L}+1}^{S+1} p_{Z}(y)} & \text{für} & x = 0, 1, ..., S_{L} \\ \frac{\lambda_{Z,H} \cdot p_{Z}(x)}{\lambda_{Z} - \lambda_{Z,H} \cdot p_{Z}(S+1) - \lambda_{Z,L} \sum_{y=S_{L}+1}^{S+1} p_{Z}(y)} & \text{für} & x = S_{L} + 1, S_{L} + 2, ..., S \end{cases}$$

$$(5.7)$$

Da eine ankommende Zelle auch ein voll belegtes System (x=S+1) antreffen kann, ist zur Berechnung der Zustandswahrscheinlichkeiten  $p_Z(x)$  eine weitere Gleichung nötig. Diese erhält man aus der mittleren Zuflußrate, die sich aus dem Produkt zwischen der mittleren Ankunftsrate und der Summe der Zustandswahrscheinlichkeiten aller Systemzustände, bei denen eine ankommende Zelle ins System gelangen kann, ergibt. Eine andere Beziehung für die mittlere Zuflußrate liefert das Gesetz von Little, indem die Auslastung der Bedieneinheit auf die mittlere Bedienzeit  $\Delta t$  bezogen wird. Die Bedieneinheit ist mit der Wahrscheinlichkeit  $1-p_Z(0)$  belegt, da für Poisson-Ankunftsprozesse die von ankommenden Zellen angetroffenen Zustandswahrscheinlichkeiten mit den im zeitlichen Mittel geltenden Zustandswahrscheinlichkeiten  $p_Z(x)$  übereinstimmen (vgl. [249]: Poisson Arrivals See Time Averages, PASTA). Falls die oben beschriebenen Zusammenhänge für die mittlere Zuflußrate gleichge-

setzt werden, ergibt sich

$$\lambda_{Z,H} \cdot [1 - p_Z(S+1)] + \lambda_{Z,L} \cdot \left[ 1 - \sum_{y=S_L+1}^{S+1} p_Z(y) \right] = \frac{1 - p_Z(0)}{\Delta t} . \tag{5.8}$$

Zusammen mit Gleichung 5.7 ergeben sich die zu Ankunftszeitpunkten gültigen Zustandswahrscheinlichkeiten

$$p_{Z}(x) = \begin{cases} \frac{\pi_{Z}(x)}{\pi_{Z}(0) + \lambda_{Z} \cdot \Delta t} & \text{für } x = 0, 1, ..., S_{L} \\ \frac{\lambda_{Z}}{\lambda_{Z,H}} \frac{\pi_{Z}(x)}{\pi_{Z}(0) + \lambda_{Z} \cdot \Delta t} & \text{für } x = S_{L} + 1, S_{L} + 2, ..., S \\ 1 - \frac{1}{\pi_{Z}(0) + \lambda_{Z} \cdot \Delta t} \left[ 1 + \frac{\lambda_{Z,L}}{\lambda_{Z,H}} \sum_{y = S_{L} + 1}^{S} \pi_{Z}(y) \right] & \text{für } x = S + 1 \end{cases}$$
(5.9)

Schließlich erhält man die Zellverlustwahrscheinlichkeiten  $B_{Z,H}$  und  $B_{Z,L}$ , indem die Zustandswahrscheinlichkeiten der zugeordneten Blockierungszustände aufsummiert werden:

$$B_{Z,H} = p_Z(S+1)$$
 und  $B_{Z,L} = \sum_{x=S_L+1}^{S+1} p_Z(x)$ . (5.10)

#### 5.3.1.2 Analyse des Verdrängungsverfahrens

Beim Verdrängungssystem haben beide Prioritätsklassen Zugriff auf den gesamten Zellpuffer. Im Unterschied zum M/GI/1/(S+1)-System kann eine Zelle mit niedriger Verlustpriorität aus dem Puffer verdrängt werden, falls eine hochprioritäre Zelle ankommt und der Puffer vollständig belegt ist (vgl. Bild 5.5). Für die analytischen Betrachtungen wird angenommen, daß immer die zuletzt angekommene Zelle verdrängt wird (Verdrängungsstrategie LIFD). Der Verdrängungsvorgang hat allerdings keinen Einfluß auf den Systemzustand und die Gesamtzahl der verlorenen Zellen, da für jede Zelle, die einen voll belegten Puffer antrifft, genau eine Zelle verloren geht. Da beide Prioritätsklassen zudem dieselbe Bedienzeitverteilung aufweisen, verhält sich das Verdrängungsverfahren in Bezug auf die Zustandsverteilung und die gesamte Zellverlustwahrscheinlichkeit wie das entsprechende M/GI/1/(S+1)-System. Insbesondere gilt zwischen den Zellverlustwahrscheinlichkeiten  $B_{Z,H}$  und  $B_{Z,L}$  sowie der Zellverlustwahrscheinlichkeit  $B_Z$  des M/GI/1/(S+1)-Systems der folgende Zusammenhang [227]:

$$\lambda_{Z,H} \cdot B_{Z,H} + \lambda_{Z,L} \cdot B_{Z,L} = (\lambda_{Z,H} + \lambda_{Z,L}) \cdot B_Z = \lambda_Z \cdot B_Z. \tag{5.11}$$

Da die gesamte Zellverlustwahrscheinlichkeit  $B_Z$  einfach bestimmt werden kann (vgl. [78]), konzentrieren sich die nachfolgenden Ausführungen auf die Zellverlustwahrscheinlichkeit von niederprioritären Zellen (die Zellverlustwahrscheinlichkeit  $B_{Z,H}$  kann dann aus Gleichung

5.11 bestimmt werden). Die Zellverlustwahrscheinlichkeit  $B_{Z,L}$  ergibt sich aus der Verbund-wahrscheinlichkeit  $P\{x, \text{Bedienung}\}$ , daß eine beliebige niederprioritäre Zelle den Systemzustand x antrifft und nicht verloren geht bzw. verdrängt wird.

$$B_{Z,L} = 1 - \sum_{x=0}^{S+1} P\{x, \text{Bedienung}\}$$
 (5.12)

Zur Bestimmung der Verbundwahrscheinlichkeit  $P\{x, Bedienung\}$  wird eine niederprioritäre Zelle von der Ankunft bis zur Bedienung bzw. Verdrängung beobachtet [88]. Diese Zelle findet mit der Wahrscheinlichkeit  $p_Z(x)$  (x=0,1,...,S+1) den Systemzustand x vor. Diese Wahrscheinlichkeiten entsprechen den Zustandswahrscheinlichkeiten des M/GI/1/(S+1)-Systems, da der Systemzustand durch eine Verdrängung nicht verändert wird und die niederprioritären Zellen entsprechend einem Poisson-Ankunftsprozeß eintreffen. Falls die Zelle auf ein leeres System trifft, wird sie sofort bedient, während sie bei vollem Puffer verloren geht. Damit gilt

$$P{0, Bedienung} = p_Z(0)$$
 und  $P{S+1, Bedienung} = 0$ . (5.13)

Für die verbleibenden Wahrscheinlichkeiten muß der Warteprozeß einer nach ihrer Ankunft auf dem Warteplatz x befindlichen Zelle nachvollzogen werden. Diese Zelle rückt zum folgenden Bedienende auf den Warteplatz x-1 vor, falls sie bis dahin nicht aus dem System verdrängt wurde, d.h. falls zwischenzeitlich nicht mehr als a=S-x hochprioritäre Zellen eintreffen. Unter der Bedingung, daß die Zelle einen beliebigen Warteplatz y  $(1 \le y \le x)$  erreicht hat, rückt sie zum nächsten Bedienende auf den Warteplatz y-1 vor, falls während ihrer bisherigen Wartezeit nicht mehr als a=S-y Zellen mit hoher Priorität angekommen sind. Dieser Sachverhalt wird in Bild 5.8 anhand eines Beispiels erläutert.

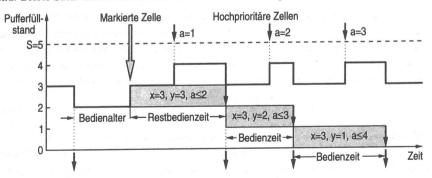


Bild 5.8: Warteprozeß einer niederprioritären Zelle

Der Warteprozeß der markierten Zelle kann durch die Verbundverteilung d(x, y, a) charakterisiert werden, welche die Wahrscheinlichkeit angibt, daß eine niederprioritäre Zelle auf der

Warteposition x eintrifft und bis zum Warteplatz y vorrücken kann, während zwischenzeitlich a Zellen mit hoher Priorität ankommen. Die Wahrscheinlichkeit, daß die Zelle nach ihrer Ankunft die Warteposition x einnimmt und a hochprioritäre Zellen innerhalb der Restbedienzeit der momentan bedienten Zelle ankommen, werde mit  $q_2(x,a)$  bezeichnet und wird weiter unten abgeleitet. Auf den nachfolgenden Warteplätzen 1,2,...,x-1 verbringt die Zelle eine komplette Bedienzeit, so daß die Anzahl der in diesen Zuständen eintreffenden hochprioritären Zellen durch die in Gleichung 5.3 definierte Verteilung  $q_2(a)$  bestimmt ist. Die Zellankünfte innerhalb verschiedener Zeitintervalle sind statistisch unabhängig, so daß sich die Gesamtverteilung der Zellankünfte aus der Faltung der für die einzelnen Bedienintervalle gültigen Verteilungen ergibt (vgl. Gleichung 3.12). Mit den obigen Überlegungen kann der folgende rekursive Algorithmus zur Bestimmung der Verbundwahrscheinlichkeiten d(x,y,a) angegeben werden [88]:

Schritt 0:

$$d(x, x - 1, a) = \begin{cases} q_2(x, a) & \text{für} & a = 0, 1, ..., S - x \\ 0 & \text{für} & a > S - x \end{cases}$$
 (5.14)

Schritt z (z = 1, 2, ..., x - 1):

$$d(x, x - 1 - z, a) = \begin{cases} d(x, x - z, a) * q_2(a) & \text{für} & a = 0, 1, ..., S - x + z \\ 0 & \text{für} & a > S - x + z \end{cases}$$
 (5.15)

Der Warteplatz 0 entspricht der Bedieneinheit, aus der eine Zelle nicht mehr verdrängt werden kann, d.h.

$$P\{x, \text{Bedienung}\} = \sum_{a=0}^{S-1} d(x, 0, a) .$$
 (5.16)

Die Anfangsbedingung der obigen Rekursionsbeziehung erhält man aus der Verbundwahrscheinlichkeit  $q_2(x,a)$ , daß die markierte Zelle auf der Warteposition x eintrifft und bis zum folgenden Bedienende a Zellen mit hoher Priorität ankommen. Mit der Verbundwahrscheinlichkeit  $f_R(x,t)$  d $t=P\{\text{Zelle trifft auf den Warteplatz } x, \text{Restbedienzeit der in Bedienung befindlichen Zelle liegt im Zeitintervall } [t, t+dt)\}$  ergibt sich (vgl. Gleichung 5.3)

$$q_2(x,a) = \int_{t=0}^{\infty} \frac{(\lambda_{Z,H} t)^a}{a!} \exp(-\lambda_{Z,H} t) \cdot f_R(x,t) dt.$$
 (5.17)

Mit Hilfe der Methode der Supplementaritätsvariablen kann die Verbundwahrscheinlichkeit  $f_A(x,u)$  du bestimmt werden, daß zu einem beliebigen Zeitpunkt (bzw. zu einem Ankunftszeitpunkt) der Zustand x besteht und das Bedienalter der momentan bedienten Zelle im Zeitintervall  $[u,u+\mathrm{d}u)$  liegt. Aus [206] kann die zugehörige Wahrscheinlichkeitsdichte  $f_A(x,u)$  entnommen werden:

$$f_A(x,u) = \lambda_Z \exp\left(-\lambda_Z u\right) \cdot \left[1 - F_{Z,H}(u)\right] \cdot \left[p_Z(0) \frac{(\lambda_Z u)^{x-1}}{(x-1)!} + \sum_{y=0}^{x-1} p_Z(x-y) \frac{(\lambda_Z u)^y}{y!}\right] . (5.18)$$

Darin bezeichnet  $F_{Z,H}(u)$  die Verteilungsfunktion der Bedienzeit. Unter dieser Voraussetzung erhält man eine Restbedienzeit der Dauer t, falls die gesamte Bedienzeit u+t Zeiteinheiten beträgt. Die auf das "Bedienalter" u bedingte Wahrscheinlichkeit, daß die Bedienzeit der momentan bedienten Zelle innerhalb des Zeitintervalls [u+t,u+t+dt) liegt, ergibt sich aus [150] zu

 $\frac{f_{Z,H}(u+t)}{1 - F_{Z,H}(u)} \, \mathrm{d}t \,. \tag{5.19}$ 

Die gesuchte Verbundwahrscheinlichkeit  $f_R(x,t)dt$  erhält man, indem die Integration über alle möglichen Werte u durchgeführt wird. Aus den Gleichungen 5.18 und 5.19 ergibt sich

$$f_{R}(x,t) dt = \int_{u=0}^{\infty} \frac{f_{Z,H}(u+t)}{1 - F_{Z,H}(u)} dt \cdot f_{A}(x,u) du$$

$$f_{R}(x,t) = \int_{u=0}^{\infty} \lambda_{Z} f_{Z,H}(u+t) \cdot \exp\left(-\lambda_{Z} u\right) \cdot \left[ p_{Z}(0) \frac{(\lambda_{Z} u)^{x-1}}{(x-1)!} + \sum_{y=0}^{x-1} p_{Z}(x-y) \frac{(\lambda_{Z} u)^{y}}{y!} \right] du.$$
(5.20)

Abschließend sollen die zur numerischen Auswertung der Lösung nötigen Einzelschritte nochmals zusammengefaßt werden. Mit den Gleichungen 5.3, 5.20 und 5.17 werden zunächst die Ankunftsverteilungen  $q_2(a)$  und  $q_2(x,a)$  bestimmt. Anschließend müssen die Gleichungen 5.13-5.16 ausgewertet werden. Die Verlustwahrscheinlichkeiten für nieder- und hochprioritäre Zellen ergeben sich schließlich aus den Gleichungen 5.12 und 5.11.

#### 5.3.1.3 Ergebnisse für Poisson-Ankunftsverkehre

Dieser Abschnitt befaßt sich mit einem Leistungsvergleich der besprochenen Prioritätsmechanismen. Einführend soll das grundlegende Systemverhalten des Schwellwertverfahrens und des Verdrängungssystems aufgezeigt werden. Beim Schwellwertverfahren kann der Grad der Priorisierung durch eine Änderung des Schwellwerts  $S_L$  flexibel eingestellt werden, wie die in Bild 5.9 dargestellten Ergebnisse belegen. Für  $S_L=S=64$  entspricht das Schwellwertverfahren dem M/D/1/(S+1)-System, weshalb beide Prioritätsklassen dieselbe Zellverlustwahrscheinlichkeit aufweisen. Eine Verkleinerung des Schwellwerts bewirkt eine nahezu exponentielle Verringerung der Zellverlustwahrscheinlichkeit  $B_{Z,H}$  auf Kosten einer nahezu exponentiell verlaufenden Zunahme der Zellverlustwahrscheinlichkeit  $B_{Z,L}$ . Eine Änderung des Verkehrsangebots äußert sich hauptsächlich in einer Parallelverschiebung der Kurven. Da die Steigungen der Kurven auch bei einer Änderung der Puffergröße nahezu erhalten bleiben, bestimmt die Größe des für die hohe Priorität reservierten Pufferbereichs das Verhältnis zwischen den Zellverlustwahrscheinlichkeiten  $B_{Z,L}$  und  $B_{Z,H}$ . Beispielsweise müssen für die hohe Priorität  $S-S_L=6$  Pufferplätze reserviert werden, um für die gegebene Prioritätszusammensetzung ein Verhältnis  $B_{Z,L}/B_{Z,H}$  von ca.  $10^4$  zu erreichen. Diese Faustregel ist

bei der Systemdimensionierung nützlich. Ein Suchalgorithmus zur Schwellwertoptimierung ist in [196] beschrieben.

Die Leistungsfähigkeit des Verdrängungsverfahrens kann durch die Verdrängungsreihenfolge beeinflußt werden. In Bild 5.10 sind die in Abschnitt 5.2.3 eingeführten Verdrängungsstrategien LIFD, RANDOM, FIFD und FIFD mit globaler Verdrängung von niederprioritären Zellen gegenübergestellt. Da nur für die Strategie LIFD eine analytische Lösung bekannt ist, wurden die anderen Verdrängungsstrategien mit Hilfe einer Systemsimulation untersucht. Alle Simulationsergebnisse wurden mit 95%-Vertrauensintervallen versehen. Die Ergebnisse zeigen, daß die Zellverlustwahrscheinlichkeit  $B_{Z,H}$  abnimmt, wenn die früher eingetroffenen Zellen zuerst verdrängt werden. Später eingetroffene Zellen bleiben länger im Puffer und können daher mit größerer Wahrscheinlichkeit verdrängt werden, womit sie zu einer Verringerung der hochprioritären Zellverluste beitragen können. Dies veranschaulicht auch den in [230] geführten Beweis, daß die Strategie FIFD mit globaler Verdrängung die kleinste Zellverlustwahrscheinlichkeit  $B_{Z,H}$  liefert. Die Leistungsunterschiede zwischen den einzelnen Verdrängungsstrategien sind allerdings nur schwach ausgeprägt, so daß die weiteren Untersuchungen auf die Verdrängungsstrategie LIFD beschränkt werden, für die eine analytische Lösung bekannt ist.

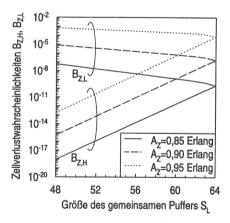


Bild 5.9: Zellverlustwahrscheinlichkeit des Schwellwertverfahrens für S=64 und  $A_{Z,H}/A_Z=0,5$ 

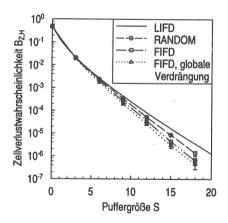


Bild 5.10: Vergleich unterschiedlicher Verdrängungsstrategien für S = 64 und  $A_{Z,H} = A_{Z,L} = 0,45$ 

In Bild 5.11 wurden die Zellverlustwahrscheinlichkeiten  $B_{Z,H}$  und  $B_{Z,L}$  gegeneinander aufgetragen. Falls kein Prioritätsmechanismus implementiert ist, sind beide Zellverlustwahrscheinlichkeiten gleich groß. Die Zellverlustwahrscheinlichkeit  $B_{Z,H}$  kann reduziert werden, indem bei Überlastsituationen bevorzugt Zellen mit niederer Priorität verworfen werden. Im

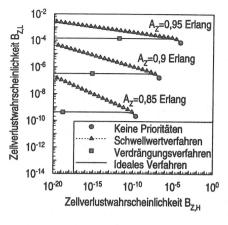
Idealfall werden die Zellverluste von der hohen in die niedere Priorität verlagert, während die gesamte Zellverlustwahrscheinlichkeit auf dem durch das M/D/1/(S+1)-System definierten Minimalwert bleibt. Der in Gleichung 5.11 definierte Erhaltungssatz für die Zellverlustwahrscheinlichkeiten liefert daher eine untere Grenze für  $B_{Z,L}$  in Abhängigkeit von  $B_{Z,H}$ . Im vorliegenden Spezialfall  $\lambda_Z=2\cdot\lambda_{Z,H}=2\cdot\lambda_{Z,L}$  gilt die folgende Beziehung:

$$B_{Z,L} = 2 \cdot B_Z - B_{Z,H}$$
  
 $B_{Z,L} \approx 2 \cdot B_Z, \text{ für } B_{Z,H} << B_{Z,L}.$  (5.21)

Dieser Zusammenhang zeigt, daß die Zellverlustwahrscheinlichkeit innerhalb der niederen Priorität nahezu konstant ist, unabhängig von der Zellverlustwahrscheinlichkeit  $B_{Z,H}$ . Die für das Verdrängungsverfahren erhaltenen Zellverlustwahrscheinlichkeiten liegen auf dieser Idealkurve. Ein beliebiger, auf dem Kurvenabschnitt zwischen dem Verdrängungssystem und dem System ohne Prioritäten liegender Zwischenwert wäre dadurch erreichbar, daß eine Verdrängung nur mit einer bestimmten Wahrscheinlichkeit stattfindet. Demgegenüber können die außerhalb dieses Bereichs liegenden Idealwerte nur mit Hilfe von Bedienprioritäten (auch zeitliche Prioritäten genannt) erreicht werden. Ähnlich kleine Zellverlustwahrscheinlichkeiten  $B_{Z,H}$  — bei allerdings deutlich höheren Zellverlusten innerhalb der niederen Priorität — lassen sich mit dem Schwellwertverfahren erzielen. Die für dieses Verfahren markierten Punkte kennzeichnen unterschiedliche Schwellwerte  $S_L = 0, 1, ..., S$ .

In Bild 5.12 werden die Prioritätsmechanismen in Bezug auf das maximal zulässige Verkehrsangebot  $\hat{A}_Z$  bewertet, das sich aus folgenden Überlegungen ableitet:

- Keine Prioritäten: Das Verkehrsangebot des M/D/1/(S+1)-Systems muß so begrenzt werden, daß die für die hohe Priorität vorgegebene maximale Zellverlustwahrscheinlichkeit  $\hat{B}_{Z,H}$  eingehalten werden kann.
- Wegetrennung: Die auf den verschiedenen Wegen zulässigen Auslastungen müssen sich an dem für die jeweilige Prioritätsklasse gültigen Grenzwert  $\hat{B}_{Z,H}$  bzw.  $\hat{B}_{Z,L}$  orientieren. Das im günstigsten Fall zulässige Gesamtangebot ergibt sich aus der mit dem jeweiligen Verkehrsanteil gewichteten Addition dieser Teilauslastungen.
- Schwellwertverfahren: Bei diesem Verfahren muß neben dem Verkehrsangebot auch der Schwellwert S<sub>L</sub> so eingestellt werden, daß die in Bezug auf die Zellverlustwahrscheinlichkeit gemachten Vorgaben in beiden Prioritätsklassen eingehalten werden können.
   Für die nachfolgenden Untersuchungen wurde der Schwellwert so eingestellt, daß das maximale Verkehrsangebot erzielt wird.
- Verdrängungsverfahren: Das Verkehrsangebot wird durch diejenige Prioritätsklasse begrenzt, deren zulässige Zellverlustwahrscheinlichkeit zuerst überschritten wird. In der Regel wird die in der anderen Prioritätsklasse zulässige Zellverlustwahrscheinlichkeit deutlich unterschritten.



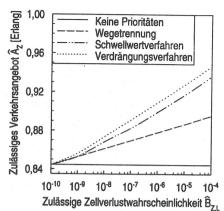


Bild 5.11: Zellverlustwahrscheinlichkeiten in beiden Prioritätsklassen für S = 64 und  $A_{Z,H}/A_Z = 0.5$ 

Bild 5.12: Zulässiges Verkehrsangebot in Abhängigkeit von der Zellverlustwahrscheinlichkeit für Klasse L

- Ideales Verfahren: Als Vergleichsmaßstab wird ein ideales Prioritätsverfahren definiert, welches die Netzauslastung unter den gegebenen Randbedingungen optimiert. Für das zulässige Verkehrsangebot gelten die folgenden prinzipiellen Grenzen:
  - Das Verkehrsangebot der hohen Priorität darf die für das System ohne Prioritäten geltende Lastgrenze nicht überschreiten, da dies in jedem Fall zu einer Verletzung des zugehörigen Grenzwerts  $\hat{B}_{Z,H}$  führen würde.
  - Ein ideales Verfahren sollte Zellen nur dann verwerfen, wenn der Puffer voll belegt ist, weshalb die gesamte Zellverlustwahrscheinlichkeit aus dem äquivalenten M/D/1/(S+1)-System berechnet werden kann. In diesem Fall gilt der Erhaltungssatz für die Zellverlustwahrscheinlichkeiten nach Gleichung 5.11. Im Idealfall verteilen sich die Zellverluste so auf die beiden Prioritätsklassen, daß beim Erreichen des zulässigen Verkehrsangebots die Zellverlustwahrscheinlichkeiten  $B_{Z,H}$  und  $B_{Z,L}$  den zugeordneten Grenzwerten  $\hat{B}_{Z,H}$  und  $\hat{B}_{Z,L}$  entsprechen. Dies führt zu einer Maximierung der gesamten Zellverlustwahrscheinlichkeit und damit zu einer Maximierung des zulässigen Verkehrsangebots, das aus dem M/D/1/(S+1)-Modell bestimmt werden kann.

Das Minimum der oben definierten Lastschranken entspricht dem im günstigsten Fall möglichen Verkehrsangebot. Für eine reale Implementierung dieses idealen Verfahrens muß eine priorisierte Bedienung (zeitliche Priorisierung) mit verdrängenden Prioritäten kombiniert werden. Damit die Zellverlustwahrscheinlichkeiten  $B_{Z,H}$  und  $B_{Z,L}$  optimal an die entsprechenden Vorgaben angepaßt werden können, darf eine Verdrängung nur

mit einer bestimmten Wahrscheinlichkeit stattfinden.

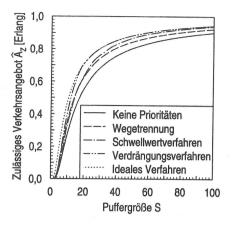
Der auf dieser Grundlage durchgeführte Leistungsvergleich basiert auf folgenden Systemvoraussetzungen: Puffergröße S=64 Zellen, Prioritätszusammensetzung  $A_{Z,H}/A_Z=0,5$ , maximale Zellverlustwahrscheinlichkeiten  $\hat{B}_{Z,L}=10^{-6}$  und  $\hat{B}_{Z,H}=10^{-10}$ .

Bild 5.12 bestätigt, daß Verlustprioritäten zu einer Erhöhung des zulässigen Verkehrsangebots führen. Den größten Gewinn verbucht das Verdrängungsverfahren, das unter den gegebenen Randbedingungen eine dem idealen System vergleichbare Leistungsfähigkeit aufweist (aus Gründen der Übersichtlichkeit wurde deshalb auf die Darstellung der idealen Ergebnisse verzichtet). Das Schwellwertverfahren kommt diesen Ergebnissen ebenfalls nahe, während bei getrennten Wegen deutliche Leistungseinbußen in Kauf genommen werden müssen.

In Bild 5.13 wird der Zusammenhang zwischen dem zulässigen Verkehrsangebot und der Puffergröße näher beleuchtet. Diese Ergebnisse stützen die oben gemachten Aussagen in Bezug auf die relative Leistungsfähigkeit der verschiedenen Mechanismen, wobei die Angebotsunterschiede mit wachsender Puffergröße abnehmen.

Die durch Verlustprioritäten erzielten Leistungsvorteile werden noch deutlicher, falls die obige Fragestellung umgekehrt wird. Tabelle 5.1 gibt die für ein vorgegebenes Verkehrsangebot erforderliche Puffergröße wieder. Das Verdrängungsverfahren spart innerhalb des betrachteten Lastbereichs zwischen 37,5% und 42,6% des benötigten Pufferplatzes ein. Die Differenz zwischen den Puffergrößen des Schwellwertverfahrens und des Verdrängungsverfahrens bleibt in etwa konstant bei 6 Pufferplätzen. Dies entspricht dem vorher ermittelten Wert für die Größe des für die hohe Priorität reservierten Pufferbereichs.

Abschließend soll der Einfluß der Prioritätszusammensetzung  $A_{Z,H}/A_Z$  untersucht werden (vgl. Bild 5.14). Das Verdrängungsverfahren erzielt wiederum die höchste Auslastung, wobei zwei Bereiche unterschieden werden können. Falls der Verkehrsanteil der niederen Priorität überwiegt, bleibt das zulässige Verkehrsangebot nahezu konstant. Dies ist darauf zurückzuführen, daß das Verkehrsangebot in diesem Bereich durch die niedrige Priorität begrenzt wird. Demgegenüber ist für große Werte von  $A_{Z,H}/A_Z$  die Zellverlustwahrscheinlichkeit  $B_{Z,H}$ entscheidend, was sich in einer rapiden Abnahme des zulässigen Verkehrsangebots äußert. In ungünstigen Fällen kann sogar die für das Schwellwertverfahren gültige Lastgrenze unterschritten werden. Auch das Schwellwertverfahren kann in einem breiten Bereich eine signifikante Erhöhung des zulässigen Verkehrsangebots erzielen, wogegen bei getrennten Wegen eine lineare Abnahme des Verkehrsangebots auftritt. Für das Schwellwertverfahren wurden zwei verschiedene Varianten untersucht. Die erste Variante geht von einem statischen Schwellwert aus, der im gezeigten Beispiel bei  $S_L=57$  liegt. Die zweite Variante, die eine adaptive Anpassung des Belegungsschwellwerts an die aktuelle Prioritätszusammensetzung vornimmt, bringt nur eine geringfügige weitere Erhöhung der Netzlast, so daß der hierfür erforderliche Zusatzaufwand nicht gerechtfertigt erscheint.



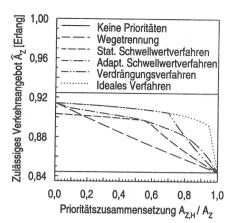


Bild 5.13: Zulässiges Verkehrsangebot in Abhängigkeit von der Puffergröße

Bild 5.14: Zulässiges Verkehrsangebot in Abhängigkeit von der Prioritätszusammensetzung

$\begin{array}{c c} & \text{Angebot } A_Z \\ & \text{[Erl.]} \\ \text{Puffergr\"{o}Be } S \end{array}$	0,7	0,75	0,8	0,85	0,9	0,95
Ohne Prioritäten	32	39	50	67	100	197
Schwellwertverfahren	24	29	36	46	65	120
Verdrängungsverfahren	20	24	30	40	59	113

Tabelle 5.1: Pufferdimensionierung für verschiedene Prioritätsmechanismen

#### 5.3.2 Leistungsuntersuchung auf Büschelebene

In diesem Abschnitt soll untersucht werden, wie sich die Prioritätsmechanismen bei büschelförmiger Belastung verhalten. Da diese Untersuchungen auf relativ komplexe Quell- und Systemmodelle führen, wird hauptsächlich auf Systemsimulationen zurückgegriffen. Zusätzliche analytische Ergebnisse wurden aus dem in Anhang A.8 für das Schwellwertverfahren vorgestellten "Fluid Flow"-Modell gewonnen. Dieses Modell verallgemeinert die in [57, 68] für die Überlagerung homogener, zweistufiger Verkehrsquellen beschriebenen Ansätze. Das in [68] behandelte "Fluid Flow"-Modell des Schwellwertverfahrens bezieht sich auf die Überlagerung von sporadischen Quellen mit identischen statistischen Eigenschaften, die sowohl hoch- als auch niederprioritären Verkehr erzeugen. Dieser Ansatz wurde in [57] auf eine beliebige Anzahl von Prioritätsklassen ausgedehnt. Ferner können die dort betrachteten Quellen in beiden Quellzuständen Verkehr erzeugen.

Weitere analytische Modelle des Schwellwertverfahrens mit korrelierten Ankunftsverkeh-

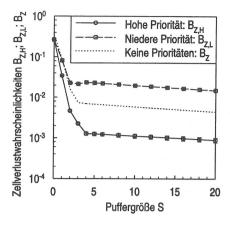
ren wurden in [11, 29, 69, 155] behandelt. Die Zellankünfte werden durch einen Markoff-Modulierten Poisson-Prozeß (MMPP) [11], einen allgemeinen Markoffschen Ankunftsprozeß (Markovian Arrival Process, MAP) [69] oder einen Markoff-Modulierten Bernoulli-Prozeß (MMBP) [29, 155] modelliert. Das Verdrängungsverfahren wurde für MMPP- [211] und MAP-Verkehre [69] untersucht. Allerdings muß einschränkend hinzugefügt werden, daß die mit den obigen Ankunftsprozessen erzielten Ergebnisse zum Teil erheblich von den für reale Überlagerungsverkehre geltenden Zellverlustwahrscheinlichkeiten abweichen.

Die in Bild 5.15 für die Verdrängungsstrategie LIFO gezeigten Simulationsergebnisse beziehen sich auf den in Unterabschnitt 4.1.1.4 eingeführten Referenzverkehr². Dabei senden 6 Verkehrsquellen Zellen mit hoher Priorität, während die verbleibenden Verkehrsquellen niederprioritären Verkehr erzeugen. Der Anteil des hochprioritären Verkehrs mußte relativ groß gewählt werden, damit innerhalb dieser Prioritätsklasse überhaupt Zellverluste auftraten. Dennoch wird durch die Prioritätsbehandlung eine signifikante Verringerung der hochprioritären Zellverluste erreicht. Da das Verdrängungsverfahren die Zellverluste von der hohen in die niedere Priorität verlagert, nimmt die Zellverlustwahrscheinlichkeit des niederprioritären Verkehrs zu. Ferner sind die im vorigen Kapitel ausführlich erörterten Blockierungseffekte auf Zell- und Büschelebene deutlich zu erkennen.

Das Schwellwertverfahren wurde unter denselben Voraussetzungen untersucht. Dabei wurden für die hohe Priorität zwei Pufferplätze reserviert, damit die zwischen hochprioritären Zellen auftretenden kurzzeitigen Zellkollisionen innerhalb des reservierten Pufferbereichs aufgelöst werden können. Die in Bild 5.16 skizzierten Ergebnisse bestätigen, daß das M/D/1/(S+1)-Modell in Verbindung mit dem "Fluid Flow"-Modell eine genaue und anschauliche Beschreibung des Systemverhaltens liefert. Das M/D/1/(S+1)-Modell spiegelt die für kleine Puffergrößen entscheidenden, kurzzeitigen Zellkollisionen wider. Für große Puffer sind hingegen die durch das "Fluid Flow"-Modell beschriebenen Bitratenschwankungen ausschlaggebend. Aus dem Vergleich der für das Verdrängungs- und das Schwellwertverfahren erhaltenen Zellverlustwahrscheinlichkeiten kann ferner gefolgert werden, daß beide Mechanismen für realistische Puffergrößen eine vergleichbare Leistungsfähigkeit aufweisen. Die weiteren Untersuchungen beschränken sich deshalb auf das Schwellwertverfahren.

Anhand eines realistischen Verkehrsszenarios soll die durch diesen Mechanismus erzielte Erhöhung der Netzauslastung quantifiziert werden. Eine Anwendung, die auf dem niederprioritären Übermittlungsdienst aufsetzen könnte, ist die in Tabelle 2.2 charakterisierte Hochgeschwindigkeits-Datenkommunikation. Die entsprechenden Datenverkehre werden durch sporadische Quellen modelliert. Der hochprioritäre Verkehr wird von der in [164] beschriebenen Videoquelle erzeugt, deren Verkehrsparameter in Tabelle 2.2 aufgeführt sind. Das Quellverhalten wird mit Hilfe des in Bild 4.19 dargestellten Quellmodells nachgebildet. Eine Video-

 $<sup>^2</sup>$ Es werden N=8 Verkehrsquellen mit einem Büschelfaktor Z=5, einer mittleren Büschelgröße von b=100 Zellen und einem minimalen Zellabstand von d=4 Zeitschlitzen überlagert.



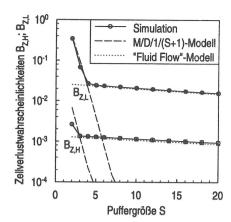


Bild 5.15: Einfluß des Verdrängungsverfahrens auf die Zellverluste bei sporadischem Quellverkehr

Bild 5.16: Einfluß des Schwellwertverfahrens auf die Zellverluste bei sporadischem Quellverkehr

quelle setzt sich dabei aus 20 Einzelquellen zusammen; jede Einzelquelle sendet sporadischen Verkehr mit einem Büschelfaktor Z=5, einer mittleren Büschelgröße von b=809 Zellen und einem minimalen Zellabstand von d=140 Zeitschlitzen. Die Puffergröße entspricht S=64 Zellen, wovon  $S-S_L=7$  Pufferplätze für die hohe Priorität reserviert werden.

Bild 5.17 zeigt die unter diesen Randbedingungen mögliche Netzauslastung in Form einer Grenzkurve, welche zulässige von unzulässigen Verbindungszuständen trennt. Für einen zulässigen Verbindungszustand werden die spezifizierten Verlustwahrscheinlichkeiten in beiden Prioritätsklassen eingehalten; demgegenüber werden die entsprechenden Werte bei einem verbotenen Verbindungszustand in mindestens einer Prioritätsklasse überschritten. Für die hohe und die niedrige Priorität wurden maximale Verlustwahrscheinlichkeiten von  $\hat{B}_{B,L}=10^{-2}$  bzw.  $\hat{B}_{B,H}=10^{-4}$  zugrundegelegt, damit die Güte der Approximation durch Simulation validiert werden kann.

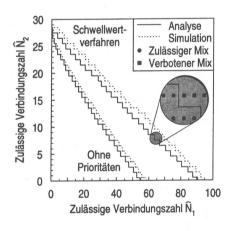
Die analytischen Ergebnisse basieren auf den in den Anhängen A.7 und A.8 vorgestellten asymptotischen Näherungen, da die exakte Lösung aufgrund der großen Anzahl von Quellzuständen nicht ausgewertet werden kann. Diese Approximation überschätzt die Zellverlustwahrscheinlichkeiten (vgl. Bild 4.13), so daß die mit diesem Ansatz ermittelte Grenzkurve unter der exakten Auslastungsgrenze liegt. Die asymptotische Näherung liefert somit eine vorsichtige Abschätzung des möglichen Multiplexgewinns. Da sich Analyse- und Simulationsergebnisse nur geringfügig unterscheiden, dienen die aus der asymptotischen Näherung gewonnenen Resultate als Referenzwerte für die im folgenden Kapitel untersuchten Verbindungsannahmestrategien. Dies ermöglicht die Erfassung der in ATM-Netzen angestrebten

Zellverlustwahrscheinlichkeiten.

In Bild 5.18 werden verschiedene Strategien zur Verkehrssteuerung verglichen, wobei die Verkehrs- und Systemparameter wie im vorigen Beispiel gewählt wurden. Die Grenzkurven für die zulässigen Verbindungszustände wurden wie folgt ermittelt:

- Spitzenbitratenzuweisung: Die Summe der Spitzenbitraten aller Verbindungen liegt unter der Übertragungsgeschwindigkeit.
- Statistisches Multiplexen: Die mit Hilfe von Gleichung A.51 abgeschätzte Verlustwahrscheinlichkeit liegt unter dem Grenzwert  $\hat{B}_{B,H} = 10^{-10}$ .
- Statistisches Multiplexen und Verlustprioritäten: Die aus den Gleichungen A.70 und A.71 errechneten Verlustwahrscheinlichkeiten liegen unter den Grenzwerten  $\hat{B}_{B,H} = 10^{-10}$  bzw.  $\hat{B}_{B,L} = 10^{-6}$ .

25



Spitzenbitratenzuweisung
Statistisches Multiplexen
Statistisches Multiplexen
mit Verlustprioritäten

15

10

10

10

10

10

10

20

30

40

Zulässige Verbindungszahl Ñ<sub>1</sub>

Bild 5.17: Erhöhung der Netzauslastung durch Verlustprioritäten

Bild 5.18: Vergleich verschiedener Grundkonzepte zur Verkehrssteuerung

Die Ergebnisse zeigen, daß der durch das statistische Multiplexen erzielte Auslastungsgewinn im vorliegenden Fall signifikant erhöht werden kann, indem zwischen verschiedenen Verlust-prioritäten unterschieden wird. Allerdings muß angemerkt werden, daß diese Steigerung der Netzauslastung — ähnlich wie der gewöhnliche Multiplexgewinn — von den Verkehrsparametern der verschiedenen Verbindungen abhängt.

Wie in Unterkapitel 5.1 diskutiert, kann eine Verkehrsquelle auch von beiden Prioritäten Gebrauch machen. Den in Bild 5.19 wiedergegebenen Ergebnissen liegt das zu Beginn des Abschnitts definierte Referenzsystem zugrunde. Im Unterschied zu dem dort betrachteten Szenario sendet nun jede Quelle hoch- und niederprioritären Verkehr im Verhältnis 3:1, d.h. nach drei hochprioritären Zellen wird eine Zelle mit niederer Priorität gesendet. Die Ergebnisse unterscheiden sich nur geringfügig von den in Bild 5.16 gezeigten und lassen sich mit Hife der dort gemachten Bemerkungen interpretieren. Für die hohe Priorität ergeben sich Abweichungen zwischen den Simulations- und Analyseergebnissen. Diese können darauf zurückgeführt werden, daß innerhalb eines Büschels nur eine ganzzahlige Anzahl von hoch- und niederprioritären Zellen gesendet werden kann. Dadurch kommt es zu zusätzlichen Schwankungen der Ankunftsraten in beiden Verkehrsklassen, die sich in einer höheren Zellverlustwahrscheinlichkeit für hochprioritäre Zellen niederschlagen.

Eine typische Anwendung, die auf beide Prioritäten zurückgreift, ist die Videoübertragung. Dabei zerlegen die Codierverfahren die Bildinformation in verschiedene Anteile mit unterschiedlicher Bedeutung für die Bildqualität. Da aus der Literatur kein Quellmodell für diese Codierverfahren bekannt ist, wird den folgenden Untersuchungen wiederum die in [164] beschriebene Videoquelle zugrundegelegt, d.h. das gesamte Verkehrsaufkommen wird durch dieses Quellmodell beschrieben (die in [191] angegebenen Meßwerte stützen die Annahme, daß das Gesamtverkehrsaufkommen unabhängig von der Prioritätszusammensetzung ist). Die Schwellwerte, bei denen hoch- und niederprioritäre Zellen abgewiesen werden, wurden wiederum auf S=64 und  $S_L=57$  Zellen festgelegt. Bild 5.20 zeigt die zulässige Anzahl von Verbindungen, falls der Anteil des hochprioritären Verkehrs variiert wird. Es soll nochmals betont werden, daß sich diese Ergebnisse nicht verallgemeinern lassen, da sie von den speziellen Eigenschaften der Verkehrsquellen abhängen.

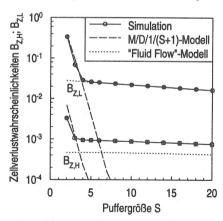


Bild 5.19: Schwellwertverfahren bei sporadischem Quellverkehr und gemischten Prioritäten

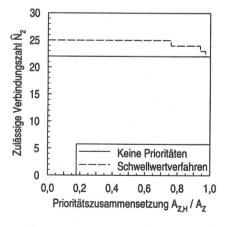


Bild 5.20: Netzauslastung in Abhängigkeit von der Prioritätszusammensetzung

# Kapitel 6

# Untersuchung von Verbindungsannahmestrategien

Kernpunkt dieses Kapitels ist der Vergleich grundlegender Ansätze zur Verbindungsannahme in ATM-Netzen. Die in den vorhergehenden Kapiteln erörterten Modelle, Methoden und Ergebnisse bilden die Grundlage für die Konzeption und Leistungsuntersuchung der verschiedenen Verbindungsannahmealgorithmen.

# 6.1 Anforderungen und Vergleichskriterien

Die Verbindungsannahmefunktion, die in Unterabschnitt 2.4.2.5 einführend behandelt wurde, spielt eine Schlüsselrolle bei der Sicherung der Netzgüte, da sie die Häufigkeit von Überlastsituationen einschränken kann und eine langfristige Steuerung des Verkehrsflusses ermöglicht. Damit erhalten die in Abschnitt 2.4.1 genannten Anforderungen an das Verkehrsmanagement eine besondere Bedeutung für die Verbindungsannahme. Im folgenden werden diese Anforderungen so konkretisiert, daß sie als Vergleichsgrundlage für die nachfolgenden Untersuchungen dienen können.

Effizienz: Die Hauptaufgabe der Verbindungsannahmefunktion ist die Sicherung einer vordefinierten Netzgüte [121], d.h. es dürfen nur soviele Verbindungen angenommen werden, daß die Netzgütevorgaben eingehalten werden können. Unter dieser Randbedingung sollte die Verbindungsannahmefunktion einen möglichst hohen Multiplexgewinn erzielen, um einen kostengünstigen Netzbetrieb zu gewährleisten. Die einzelnen Verfahren können dabei auf der Basis der Verbindungsannahme-Grenzkurven verglichen werden, die "zulässige" von "unzulässigen" Verbindungszuständen trennen (vgl. Bild 5.17).

Flexibilität: Die Verbindungsannahmefunktion sollte das gesamte Dienstespektrum unterstützen, d.h. sie muß unabhängig von den speziellen Dienstattributen der Verbindungen

sein. Insbesondere sollte sich der Annahmealgorithmus nicht auf fest vorgegebene Quellmodelle, wie beispielsweise das sporadische Quellmodell (vgl. Unterabschnitt 3.1.2.3), beschränken. Ferner muß eine Kooperation mit anderen Verkehrssteuerungsverfahren stattfinden. Beispielsweise muß die Verbindungsannahmefunktion verschiedene Verlustprioritäten unterstützen und die vom Netzressourcenmanagement vorgenommenen Kapazitätsänderungen von virtuellen Pfaden in Betracht ziehen.

Einfachheit: Der Rechenzeit- und Speicherplatzbedarf der verschiedenen Algorithmen ist ein weiteres wichtiges Vergleichskriterium, das über die zusätzlichen Steuerungskosten Aufschluß gibt. Darüberhinaus müssen für den Aufbau einer neuen Verbindung bestimmte Zeitvorgaben eingehalten werden. So betragen die in der Empfehlung I.352 [116] für den Mittelwert und das 95%-Perzentil der Verbindungsaufbauzeit vorgeschriebenen Grenzwerte 4,5 bzw. 8,35 s für die in der Empfehlung G.114 [105] definierte Interkontinentalverbindung. Für die Verbindungsannahme steht allerdings nur ein Bruchteil dieser Zeit zur Verfügung, d.h. die Entscheidung über die Annahme bzw. Ablehnung eines Verbindungswunschs sollte in wenigen Millisekunden getroffen werden. Dieser Punkt umfaßt auch den zusätzlichen Aufwand, der durch die Messung von Verbindungsparametern entsteht.

Robustheit: Die Verbindungsannahme basiert auf idealisierten Verkehrs- und Systemannahmen und sollte möglichst unempfindlich gegenüber den in der Realität auftretenden Abweichungen sein. Beispielsweise darf sich die Verbindungsannahme nicht an den vom Benutzer angegebenen Quellparametern orientieren. Vielmehr sollten zusätzliche Verkehrsmessungen einbezogen werden, oder es sollte der für das ATM-Netz ungünstigste Verkehr als Entscheidungsgrundlage dienen.

Die Vielfalt der oben diskutierten Anforderungen und Vergleichskriterien zeigt, daß eine objektive und umfassende Bewertung der unterschiedlichen Ansätze zur Verbindungsannahme außerordentlich schwierig ist. Darüberhinaus widersprechen sich manche der oben genannten Anforderungen. So wird ein komplexer Algorithmus, der auf einer genauen Verkehrsbeschreibung basiert, einen höheren Multiplexgewinn erzielen als ein einfacher Algorithmus, der nur wenige Verkehrsparameter berücksichtigt. Weitere Probleme ergeben sich daraus, daß das Potential der verschiedenen Ansätze noch nicht vollkommen ausgeschöpft ist. Beispielsweise beziehen die meisten in der Literatur beschriebenen Ansätze keine Verlustprioritäten ein und gehen von einer festen Übertragungskapazität aus. In vielen Fällen wäre eine entsprechende Erweiterung der Annahmealgorithmen allerdings durchaus denkbar.

Aus den genannten Gründen wird nur die Leistungsfähigkeit der verschiedenen Algorithmen einer detaillierten quantitativen Untersuchung unterzogen. Zu den restlichen Vergleichskriterien werden bei der Beschreibung der Verfahren einige qualitative Bemerkungen gemacht, falls sich Einschränkungen in Bezug auf den praktischen Einsatz ergeben. Nur für den abschließend vorgeschlagenen zweistufigen Algorithmus wird eine detaillierte Untersuchung aller beschriebenen Kriterien durchgeführt.

# 6.2 Ansätze zur Verbindungsannahme

In diesem Unterkapitel werden die wichtigsten aus der Literatur bekannten Verfahren zur Verbindungsannahme besprochen und qualitativ bewertet. Um den Rahmen dieser Arbeit nicht zu sprengen, mußte eine Vorauswahl getroffen werden, die sich auf die in verschiedenen Studien- und Diplomarbeiten durchgeführten Untersuchungen abstützt [56, 136, 176, 202, 204, 235, 247]. Ferner muß angemerkt werden, daß die Algorithmen vielfach nicht in ihrer ursprünglichen Form beschrieben werden. Dies hat im wesentlichen zwei Gründe: Einerseits führt dies auf eine übersichtlichere Darstellung der Grundkonzepte dieser Algorithmen, andererseits konnten im Rahmen der vorliegenden Arbeit entstandene Verallgemeinerungen bzw. Verbesserungen einbezogen werden.

## 6.2.1 Klassifizierung der Verbindungsannahmealgorithmen

Entsprechend der Bedeutung, die der Verbindungsannahmefunktion im Rahmen des in Unterkapitel 2.4 vorgestellten Verkehrsmanagements zukommt, wurden in der Literatur zahlreiche Algorithmen zur Verbindungsannahme vorgeschlagen. Diese Ansätze unterscheiden sich in mehrfacher Hinsicht und können deshalb nach verschiedenen Gesichtspunkten klassifiziert werden. Insbesondere differieren die Verfahren in Bezug auf die Bestimmung der Quell- und Netzgüteparameter. Die Quellparameter können entweder gemessen werden oder sie gehen aus den beim Verbindungsaufbau ausgehandelten Verkehrsparametern hervor. Diese Konzepte werden in Unterkapitel 6.5 vergleichend gegenübergestellt.

In diesem Abschnitt liegt das Hauptaugenmerk auf den eigentlichen Verbindungsannahmealgorithmen, welche die relevanten Netzgüteparameter aus den Quell- und Systemparametern abschätzen. Die unterschiedlichen Ansätze lassen sich in zwei grundlegende Kategorien einteilen:

- Direkte Verfahren: Diese Verfahren stellen einen direkten Zusammenhang her zwischen den Verkehrs- und Systemparametern einerseits und den Netzgüteparametern andererseits. Dieser Zusammenhang wird in der Regel aus geeigneten analytischen Ansätzen abgeleitet.
- Indirekte Verfahren: Diese Verfahren definieren eine effektive, äquivalente oder virtuelle Bedienrate, die eine indirekte Verknüpfung zwischen den Verkehrs- und den Netzgüteparametern schafft. Diese Rate hängt von den Verkehrsparametern der betrachteten Verbindung(en) ab und muß innerhalb des Netzes reserviert werden, damit die vorgeschriebene Netzgüte eingehalten werden kann.

#### 6.2.2 Direkte Verfahren

Obwohl eine Fülle von Ansätzen für die analytische Behandlung eines statistischen Multiplexers bekannt ist (vgl. Unterkapitel 4.1), lassen sich die direkten Verfahren zur Verbindungsannahme auf zwei grundlegende Ansätze zurückführen. Der erste Ansatz stützt sich auf die
in Kapitel 4 erläuterte Dekompositionsmethode, welche die auf der Zell- und der Büschelebene auftretenden Blockierungseffekte durch getrennte Teilmodelle erfaßt. Der zweite Ansatz
schätzt die relevanten Netzgüteparameter mit Hilfe des Zählprozesses der Zellankünfte ab.
Dieser beschreibt die Anzahl der innerhalb eines bestimmten Zeitintervalls eintreffenden Zellen [200].

#### 6.2.2.1 Faltungsalgorithmus

Dieser Algorithmus basiert auf der in Unterkapitel 4.1 ausführlich erläuterten Zerlegung der Verkehrs- und Leistungsmodellierung in die Zell- und Büschelebene. Wie in Abschnitt 4.1.2 aufgezeigt, können die kurzzeitigen Zellkollisionen mit Hilfe eines M/D/1/(S+1)-Modells beschrieben werden. Damit läßt sich ein maximales Verkehrsangebot  $\hat{A}_Z$  definieren, das einer maximal zulässigen Zellverlustwahrscheinlichkeit  $\hat{B}_Z$  entspricht (vgl. Bild 4.11). Das M/D/1/(S+1)-System darf beispielsweise nur zu ungefähr 85% ausgelastet werden, falls eine maximale Zellverlustwahrscheinlichkeit  $\hat{B}_Z = 10^{-10}$  und eine Puffergröße von 64 Zellen zugrundegelegt wird. Diese Überlegungen führen zu folgendem Annahmekriterium:

$$A_Z = \frac{1}{C} \sum_{n=1}^{N} \overline{r}_n \le \widehat{A}_Z . \tag{6.1}$$

Darin bezeichnet N die Anzahl bestehender Verbindungen, einschließlich des neuen Verbindungswunschs;  $\overline{r}_n$  entspricht der mittleren Zellrate von Verbindung n (n = 1, 2, ..., N) und C beschreibt die Bedienrate.

Die auf der Büschelebene auftretenden, längerfristigen Schwankungen der Ankunftsrate können mit Hilfe eines "Fluid Flow"-Modells erfaßt werden. Allerdings ist dieses Modell aufgrund seines hohen Rechenzeit- und Speicherplatzbedarfs nicht geeignet für die Verbindungsannahme in ATM-Netzen. In der Literatur vorgeschlagene Verbindungsannahmealgorithmen basieren daher häufig auf dem in Anhang A.7 behandelten pufferlosen "Fluid Flow"-Modell [102, 125, 127, 137, 160, 174, 219]. Dieses Modell liefert eine obere Grenze für die Verlustwahrscheinlichkeit auf Büschelebene, da die Pufferung von Überlastsituationen vernachlässigt wird. Die mit dieser Approximation errechneten Leistungskenngrößen stimmen bei unendlicher Büschelgröße mit denen des "Fluid Flow"-Modells überein. Mit sinkender Büschelgröße nimmt die Genauigkeit ab.

Unter den obigen Modellannahmen ist das System ausschließlich durch den momentanen Zustand der Verkehrsquellen bestimmt, da der gedächtnisbehaftete Puffer entfällt (Pufferfüllstand  $x \equiv 0$ ). Aus diesem Grund spielt das zeitliche Verhalten der Schwankungen der

Ankunftsrate keine Rolle bei der analytischen Behandlung dieses Modells; wichtig ist nur die Wahrscheinlichkeitsverteilung der Ankunftsrate (vgl. Bild 6.1). Dies führt zu dem in Anhang A.7 erläuterten Ansatz, der die charakteristischen Leistungskenngrößen aus den Zustandswahrscheinlichkeiten der Verkehrsquellen bestimmt. Die Auswertung dieser prinzipielen Lösung ist jedoch sehr rechenintensiv, da die Quellzustände jeder Verbindung individuell berücksichtigt werden müssen.

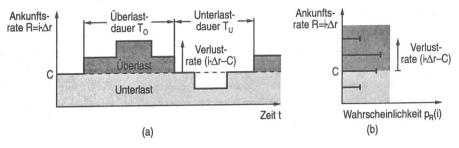


Bild 6.1: Zustandsprozeß des pufferlosen Systems

- (a) Zeitliches Verhalten
- (b) Zustandsverteilung

Eine signifikante Verringerung des numerischen Aufwands ergibt sich, wenn nicht einzelne Quellzustände, sondern — wie in Bild 6.1 angedeutet — die von allen Verbindungen generierte Gesamtankunftsrate R betrachtet wird. Diese Zufallsvariable setzt sich aus den Ankunftsraten der einzelnen Verbindungen zusammen und kann durch eine diskrete Verteilung  $P\{R=i\cdot\Delta r\}=p_R(i),\ i=0,1,...,\hat{r}/\Delta r,$  beschrieben werden, da nur eine endliche Anzahl von Quellzuständen existiert. Die dieser Verteilung zugrundeliegende Diskretisierungseinheit  $\Delta r$  entspricht dem größten gemeinsamen Teiler aller möglichen Quellraten. Falls die Verbindungen statistisch unabhängig sind, errechnet sich die Verteilung  $p_R(i)$  aus der Faltung der für die einzelnen Verbindungen gültigen Verteilungen  $p_{R,n}(i)$  (vgl. Gleichung 3.12):

$$p_R(i) = p_{R,1}(i) * p_{R,2}(i) * \dots * p_{R,N}(i) .$$
(6.2)

Die Ankunftsratenverteilung einer Verbindung wird entweder gemessen oder beim Verbindungsaufbau festgelegt (vgl. Unterkapitel 6.5).

Für die nachfolgenden Betrachtungen wird ferner die Verteilung  $p_{R,-n}$  definiert, welche die Summe der Ankunftsraten aller Verbindungen mit Ausnahme von Verbindung n beschreibt. Es gilt

$$p_{R,-n}(i) = p_{R,1}(i) * p_{R,2}(i) * \dots * p_{R,n-1}(i) * p_{R,n+1}(i) * \dots * p_{R,N}(i) \; . \tag{6.3}$$

Aus den oben eingeführten Wahrscheinlichkeitsverteilungen kann eine Reihe verschiedener Leistungskenngrößen ermittelt werden. Die Blockierungswahrscheinlichkeit  $p_C$  ergibt sich,

indem alle Verteilungswerte, für welche die Ankunftsrate  $i \cdot \Delta r$  über der Bedienrate C liegt, aufsummiert werden [137]:

$$p_C = \sum_{i=\lfloor C/\Delta_T \rfloor + 1}^{\widehat{r}/\Delta_T} p_R(i) . \tag{6.4}$$

Dabei bezeichnet  $\lfloor x \rfloor$  die größte ganze Zahl kleiner oder gleich x. Eine weitere wichtige Leistungskenngröße ist die gesamte Verlustwahrscheinlichkeit  $B_B$ , die das Verhältnis aus verlorener und gesendeter Information beschreibt. Falls die Ankunftsrate  $i \cdot \Delta r$  die Bedienrate C übersteigt, geht die überschüssige Information mit der Rate  $i \cdot \Delta r - C$  verloren (vgl. Bild 6.1). Die mittlere Verlustrate ergibt sich, indem die Verlustraten aller Überlastzustände mit den entsprechenden Auftrittswahrscheinlichkeiten  $p_R(i)$  gewichtet und anschließend aufsummiert werden. Bezogen auf die mittlere Ankunftsrate  $\overline{r}$  erhält man die Gesamtverlustwahrscheinlichkeit [137]

$$B_B = \frac{1}{r} \sum_{i=|C/\Delta r|+1}^{\widehat{r}/\Delta r} (i \cdot \Delta r - C) \cdot p_R(i) . \tag{6.5}$$

Wie in Abschnitt 4.1.3 ausgeführt, unterscheiden sich die Verlustwahrscheinlichkeiten verschiedener Verbindungen. Diese Unterschiede können mit Hilfe von verbindungsindividuellen Verlustwahrscheinlichkeiten  $B_{B,n}$  erfaßt werden. Dazu wird die Verbindung n gesondert betrachtet. Diese Verbindung sendet mit der Ankunftsrate  $j \cdot \Delta r$   $(j=0,1,...,\hat{r}_n/\Delta r)$ , während die restlichen Verbindungen die Summenankunftsrate  $i \cdot \Delta r$  erzeugen. In einem Überlastzustand  $(j+i) \cdot \Delta r > C$  ist die Verlustrate von Verbindung n proportional zur gesamten Verlustrate  $(j+i) \cdot \Delta r - C$  und proportional zu ihrem momentanen Verkehrsanteil j/(j+i). Mit den im vorigen Fall angestellten Überlegungen ergibt sich [137]:

$$B_{B,n} = \frac{1}{\bar{r}_n} \sum_{j=1}^{\hat{r}_n/\Delta r} \sum_{i=|C/\Delta r|-j+1}^{(\hat{r}-\hat{r}_n)/\Delta r} [(j+i) \cdot \Delta r - C] \cdot \frac{j}{j+i} \cdot p_{R,n}(j) \cdot p_{R,-n}(i) . \tag{6.6}$$

Das zeitliche Verhalten der Zellverluste hängt hauptsächlich mit der mittleren Unterlastdauer zusammen, die in Bild 6.1 eingezeichnet ist. Zunächst wird die Anzahl der innerhalb eines beliebig gewählten Zeitintervalls  $T, T \to \infty$ , von Quelle n ausgelösten Übergänge von einem Überlast- in einen Unterlastzustand bestimmt. Die mittlere Anzahl der innerhalb dieses Zeitintervalls zwischen den Quellraten j und  $j^*$  stattfindenden Übergänge beträgt  $p_{R,n}(j) \cdot \lambda_n(j^*|j) \cdot T$ . Dieser Zusammenhang gilt unabhängig von der Verteilung der Verweildauer im Zustand j und ist nur von der Übergangsrate  $\lambda_n(j^*|j)$  vom Zustand j in den Zustand  $j^*$  abhängig. Falls alle Zustandswechsel von j nach  $j^*$  berücksichtigt werden, für die  $(j+i) \cdot \Delta r > C$  und  $(j^*+i) \cdot \Delta r \le C$  gilt, so ergibt sich die mittlere Anzahl der von Verkehrsquelle n innerhalb des Zeitintervalls T verursachten Übergänge von Über- in Unterlastzustände zu

$$\sum_{j=1}^{\widehat{r}_n/\Delta r} \sum_{i=\lfloor C/\Delta r \rfloor - j+1}^{\lfloor C/\Delta r \rfloor} p_{R,n}(j) \cdot p_{R,-n}(i) \cdot \left( \sum_{j^*=0}^{\lfloor C/\Delta r \rfloor - i} \lambda_n(j^*|j) \cdot T \right) . \tag{6.7}$$

Die mittlere Anzahl aller innerhalb des betrachteten Zeitintervalls T auftretenden Unterlastphasen ergibt sich, indem die von allen Verkehrsquellen ausgelösten Unterlastphasen aufsummiert werden.

Im Mittel befindet sich der Zustandsprozeß die Zeitdauer  $(1-p_C) \cdot T$  im Unterlastzustand während des betrachteten Zeitintervalls T. Wird dieser Wert durch die mittlere Anzahl der Unterlastphasen dividiert, ergibt sich die mittlere Unterlastdauer

$$t_U = \mathrm{E}[T_U] = \frac{1 - p_C}{\sum_{n=1}^{N} \sum_{j=1}^{\widehat{r}_n/\Delta r} \sum_{i=\lfloor C/\Delta r \rfloor - j+1}^{\lfloor C/\Delta r \rfloor} p_{R,n}(j) \cdot p_{R,-n}(i) \cdot \left(\sum_{j^*=0}^{\lfloor C/\Delta r \rfloor - i} \lambda_n(j^*|j)\right)}$$
(6.8)

Bemerkenswerterweise ist diese Beziehung unabhängig von der Verteilung der Zustandsdauern der Verkehrsquellen, da sie nur von den Übergangsraten  $\lambda_n(j^*|j)$  abhängt. In der Literatur wurden ähnliche Beziehungen für den Sonderfall negativ-exponentiell verteilter Zustandsdauern abgeleitet [189, 228]. In diesem Fall können auch höhere Momente der Verlustdauer bestimmt werden [189].

Der beschriebenen Liste ließen sich noch weitere Leistungskenngrößen anfügen, was die Vielseitigkeit dieses Ansatzes unterstreicht. Andererseits bedeutet die Faltung der Verteilungen  $p_{R,n}$  einen großen Rechen- und Speicherungsaufwand. Deshalb werden im folgenden einige Hinweise zur effizienten Implementierung des Faltungsalgorithmus gegeben:

- Da in der Regel nur ein kleiner Teil der möglichen Ankunftsraten i Δr auftreten kann, sollten nur die Verteilungswerte größer Null mit den dazugehörigen Indizes i zwischengespeichert und bei der Faltungsoperation berücksichtigt werden. Ein entsprechender Algorithmus ist in [204] beschrieben.
- Da die Faltungsoperation kommutativ ist, sollten zunächst die für die einzelnen Verbindungsklassen gültigen Verteilungen berechnet werden<sup>1</sup>. Anschließend werden diese Verteilungen miteinander gefaltet, woraus sich die Gesamtverteilung ergibt.

Die Ankunftsratenverteilung einer Klasse mit zwei oder drei Verbindungszuständen kann direkt bestimmt werden, indem die entsprechende Binomial- bzw. Trinomialverteilung ausgewertet wird. Falls die Verbindungen mehr als drei Quellzustände besitzen, kann die Faltung durch eine geeignete Faktorisierung der Quellanzahl beschleunigt werden. Beispielsweise ergibt sich die Verteilung der Ankunftsrate von 36 identischen Verkehrsquellen, indem die Ankunftsratenverteilung von 18 Quellen mit sich selbst gefaltet wird. Diese Verteilung ergibt sich wiederum aus der Ankunftsratenverteilung von 9 Quellen usw.

<sup>&</sup>lt;sup>1</sup>Eine Verbindungsklasse faßt Verbindungen mit identischen statistischen Eigenschaften zusammen.

Die Faltungssumme kann zwischengespeichert werden, so daß für den nächsten Verbindungswunsch lediglich eine Faltungsoperation nötig ist. Allerdings müßte bei einem Verbindungsabbau eine inverse Faltung durchgeführt werden, um die nach dem Verbindungsabbau gültige Verteilung der Ankunftsrate zu bestimmen. Dies ist zwar prinzipiell möglich, führt allerdings zu numerischer Instabilität, da Differenzen auftreten [56, 125].

Trotz dieser Vereinfachungen ist der hohe numerische Aufwand der Faltungen ein großer Nachteil dieser Annahmestrategie.

#### 6.2.2.2 Normalverteilungsapproximation

Eine einfache Näherung für die Verteilung der Ankunftsrate ergibt sich aus dem zentralen Grenzwertsatz [60], der sich für die vorliegende Problemstellung folgendermaßen formulieren läßt: Die Verteilung der Ankunftsrate strebt gegen eine Gaußsche Normalverteilung, falls die Anzahl der Verbindungen gegen unendlich geht  $(N \to \infty)$ , jede Verbindung nur einen verschwindend kleinen Verkehrsanteil beisteuert  $(\bar{r}_n \to 0, \forall n = 1, 2, ..., N)$  und alle Verbindungen statistisch unabhängig sind. Die Gesamtankunftsrate kann für diesen Grenzfall mit Hilfe der folgenden Verteilungsdichtefunktion beschrieben werden (vgl. Tabelle 3.1):

$$f_R(r) = \frac{1}{\sqrt{2\pi\sigma_R}} \exp\left[-\frac{1}{2} \cdot \left(\frac{r-\overline{r}}{\sigma_R}\right)^2\right]$$
 (6.9)

Darin bezeichnet  $\overline{r}$  den Mittelwert und  $\sigma_R$  die Standardabweichung der Gesamtankunftsrate. Falls die Verbindungen statistisch unabhängig sind, können diese Kenngrößen aus den für die Einzelverbindungen geltenden Parameterwerten bestimmt werden:

$$\overline{r} = \sum_{n=1}^{N} \overline{r}_n$$
 und  $\sigma_R^2 = \sum_{n=1}^{N} \sigma_{R,n}^2$ . (6.10)

Aus der Normalverteilung können nun die interessierenden Leistungskenngrößen ermittelt werden. Dies soll exemplarisch anhand der Gesamtverlustwahrscheinlichkeit erläutert werden, die in Gleichung 6.5 definiert ist. Da es sich bei der Normalverteilung um eine kontinuierliche Verteilung handelt, geht die dort beschriebene Summe in folgendes Integral über:

$$B_B \approx \frac{1}{\bar{r}} \int_{r=C}^{\infty} f_R(r) \cdot (r - C) dr$$

$$B_B \approx \frac{1}{\bar{r}} \int_{r=C}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_R} \exp\left[-\frac{1}{2} \cdot \left(\frac{r - \bar{r}}{\sigma_R}\right)^2\right] (r - C) dr$$
(6.11)

$$B_B \approx \frac{\sigma_R}{\sqrt{2\pi r}} \exp\left[-\frac{1}{2} \cdot \left(\frac{C - \overline{r}}{\sigma_R}\right)^2\right] + \frac{\overline{r} - C}{2\overline{r}} \operatorname{erfc}\left(\frac{C - \overline{r}}{\sqrt{2}\sigma_R}\right).$$
 (6.12)

Dabei wurde die komplementäre Fehlerfunktion verwendet, für die folgende Definition gilt:

$$\frac{1}{2}\operatorname{erfc}\left(\frac{x}{\sqrt{2}}\right) = \frac{1}{\sqrt{2\pi}} \int_{y=x}^{\infty} \exp\left(-\frac{y^2}{2}\right) \, \mathrm{d}y \,. \tag{6.13}$$

Für die komplementäre Fehlerfunktion sind eine Vielzahl von Approximationen bekannt (siehe z.B. [83, 219]). Andere Leistungskenngrößen wie die Überlastwahrscheinlichkeit [102, 129, 219, 244] oder die verbindungsindividuelle Verlustwahrscheinlichkeit [202] können ebenfalls einfach bestimmt werden.

Die Auswertung von Gleichung 6.12 erfordert nur einen sehr geringen numerischen Aufwand, d.h. dieser Algorithmus wird den für die Verbindungsannahme einzuhaltenden zeitlichen Randbedingungen gerecht. Deshalb bildet er den Kern von vielen in der Literatur vorgeschlagenen Verbindungsannahmefunktionen [102, 129, 219, 244].

Andererseits weicht die tatsächliche Verteilung der Ankunftsrate umso stärker von der Normalverteilung ab, je größer der Verkehrsanteil einer einzelnen Verbindung ist. Daher können unter Umständen zuviele Verbindungen angenommen werden, wie die in Unterkapitel 6.3 beschriebenen Ergebnisse bestätigen. Dies kann größtenteils vermieden werden, falls für Verbindungen, deren maximale Ankunftsrate einen bestimmten Wert überschreitet, die Spitzenbitrate zugewiesen wird [244]. Dies führt nur zu einer geringfügigen Verringerung des Multiplexgewinns, da dieser mit zunehmender Spitzenbitrate ohnehin abnimmt (vgl. Bild 4.5). Ein anderer Ansatz wird in [129] vorgestellt. Dieser versucht für einzelne Verbindungen eine Normalverteilung für die Zellankunftsrate mit Hilfe einer entsprechenden Quellflußkontrollfunktion zu erzwingen. Unter diesen Bedingungen läßt sich auch der Überlagerungsverkehr durch eine Gaußsche Normalverteilung beschreiben.

#### 6.2.2.3 "Large Deviation"-Approximation

Die Normalverteilung weist in der Umgebung des Mittelwerts den kleinsten relativen Fehler auf, während für die Randwerte der Verteilung signifikante Abweichungen auftreten. Dies wird beispielsweise daran deutlich, daß die Normalverteilung auch für negative Ankunftsraten eine von Null verschiedene Verteilungsdichte liefert. Für die Bestimmung der sehr kleinen Zellverlustwahrscheinlichkeiten sind jedoch gerade die Randwerte der Verteilung ausschlaggebend. In [102] wurde aus diesem Grund die "Large Deviation Theory" angewandt, um verbesserte Schätzwerte für die Verlustwahrscheinlichkeit zu erhalten. Eine formale Herleitung dieser Methodik und eine Fülle von Anwendungsbeispielen kann aus [36] entnommen werden. Die nachfolgende Ableitung, die sich auf die in [243] gegebene Interpretation stützt, liefert einen wesentlich einfacheren und anschaulicheren Zugang zu dieser Theorie.

Zunächst wird die Erzeugende Funktion  $G_{R,n}(z)$  definiert, welche die Verteilung der von Verbindung n generierten Ankunftsrate beschreibt. Der Gültigkeitsbereich der Variablen z wurde gegenüber Gleichung 3.5 erweitert, da nur endliche Ankunftsraten auftreten können.

$$G_{R,n}(z) = \sum_{i=0}^{\hat{r}_n/\Delta r} p_{R,n}(i) \cdot z^i , |z| < \infty .$$
 (6.14)

Falls alle Verbindungen statistisch unabhängig sind, errechnet sich die Erzeugende Funktion für die Verteilung der Gesamtankunftsrate mit Hilfe von Gleichung 3.13 zu

$$G_R(z) = \sum_{i=0}^{\hat{r}/\Delta r} p_R(i) \cdot z^i = \prod_{n=1}^{N} G_{R,n}(z) , \quad |z| < \infty .$$
 (6.15)

Nun wird die folgende Transformation definiert, welche die Wahrscheinlichkeitsverteilung  $p_R(i)$  auf eine neue Wahrscheinlichkeitsverteilung  $p_R^*(i)$  abbildet:

$$p_R^*(i) = \frac{z^i}{G_R(z)} \cdot p_R(i) , \ |z| < \infty .$$
 (6.16)

Falls die Verteilung der Gesamtankunftsrate durch die in Gleichung 6.2 gegebene Beziehung ersetzt wird, läßt sich zeigen, daß sich die Verteilung  $p_R^*(i)$  aus einer Faltung der entsprechenden Einzelverteilungen  $p_{R,n}^*(i)$  errechnet:

$$\begin{array}{lcl} p_R^*(i) & = & \frac{z^i}{G_R(z)} \cdot [p_{R,1}(i) * p_{R,2}(i) * \ldots * p_{R,N}(i)] \\ \\ p_R^*(i) & = & p_{R,1}^*(i) * p_{R,2}^*(i) * \ldots * p_{R,N}^*(i) \; , \end{array} \tag{6.17}$$

mit

$$p_{R,n}^{*}(i) = \frac{z^{i}}{G_{R,n}(z)} \cdot p_{R,n}(i) , \quad n = 1, 2, ..., N .$$
 (6.18)

Dies bedeutet, daß die Verteilung  $p_R^*(i)$  die Summe von unabhängigen Zufallsvariablen beschreibt und deshalb ebenso wie die ursprüngliche Verteilung durch eine Normalverteilung approximiert werden kann. Diese Normalverteilung weist die höchste Genauigkeit um den verschobenen Mittelwert  $\bar{r}^*$  auf. Mit der in Gleichung 3.8 gegebenen Definition des Mittelwerts ergibt sich

$$\overline{r}^* = \Delta r \cdot \sum_{i=0}^{\widehat{r}/\Delta r} i \cdot \frac{z^i}{G_R(z)} \cdot p_R(i) = \Delta r \cdot \frac{z}{G_R(z)} \sum_{i=1}^{\widehat{r}/\Delta r} i \cdot z^{i-1} \cdot p_R(i) = \Delta r \cdot \frac{z}{G_R(z)} \cdot \frac{\mathrm{d}G_R(z)}{\mathrm{d}z} . \tag{6.19}$$

Um die Zellverlustwahrscheinlichkeit genau abschätzen zu können, wird der Wert z so gewählt, daß der verschobene Mittelwert  $\overline{r}^*$  der Zellbedienrate C entspricht:

$$\Delta r \cdot \frac{z}{G_R(z)} \cdot \frac{\mathrm{d}G_R(z)}{\mathrm{d}z} \bigg|_{z=z^*} = C. \tag{6.20}$$

Es existiert nur dann eine Lösung für  $z^*$ , falls der triviale Fall  $\hat{r} < C$  ausgenommen wird, bei dem keine Zellverluste auftreten können. Für einfache Verteilungen wie die Binomial- oder die Poissonverteilung erhält man eine geschlossene Lösung für  $z^*$ . Im allgemeinen Fall muß auf eine numerische Routine zur Nullstellenbestimmung zurückgegriffen werden.

Die Varianz der Verteilung  $p_R^*(i)$  kann aus der in Gleichung 3.9 gegebenen Definition abgeleitet werden. Mit den in Gleichung 6.19 ausgeführten Zwischenschritten erhält man:

$$\sigma_R^{*2} = \Delta r^2 \cdot \frac{z^2}{G_R(z)} \cdot \frac{\mathrm{d}^2 G_R(z)}{\mathrm{d}z^2} \bigg|_{z=z^*} + C \cdot \Delta r - C^2 .$$
 (6.21)

Somit ergibt sich die an die Verteilung  $p_R^*(i)$  angepaßte Normalverteilung zu

$$f_R^*(r) = \frac{1}{\sqrt{2\pi}\sigma_R^*} \exp\left[-\frac{1}{2} \cdot \left(\frac{r-C}{\sigma_R^*}\right)^2\right] . \tag{6.22}$$

Aus der Umkehrung der in Gleichung 6.16 definierten Abbildung ergibt sich schließlich die folgende Näherungslösung für die ursprüngliche Verteilung:

$$f_R(r) = \frac{G_R(z^*)}{z^{*^{r/\Delta r}}} \frac{1}{\sqrt{2\pi}\sigma_R^*} \exp\left[-\frac{1}{2} \cdot \left(\frac{r-C}{\sigma_R^*}\right)^2\right]$$
 (6.23)

Aus dieser Verteilungsdichtefunktion lassen sich die relevanten Leistungskenngrößen bestimmen. Die folgenden Ausführungen beschränken sich auf die den späteren Ergebnissen zugrundeliegende Gesamtverlustwahrscheinlichkeit. In der Literatur wird dagegen die einfacher zu bestimmende Blockierungswahrscheinlichkeit verwendet [102]. Durch Einsetzen von Gleichung 6.23 in Gleichung 6.11 erhält man folgende Gesamtverlustwahrscheinlichkeit:

$$B_{B} \approx \frac{1}{\overline{r}} \int_{r=C}^{\infty} \frac{G_{R}(z^{*})}{z^{*r/\Delta r}} \frac{1}{\sqrt{2\pi}\sigma_{R}^{*}} \exp\left[-\frac{1}{2} \cdot \left(\frac{r-C}{\sigma_{R}^{*}}\right)^{2}\right] \cdot (r-C) dr$$

$$B_{B} \approx \frac{G_{R}(z^{*})}{z^{*C/\Delta r}} \cdot \frac{\sigma_{R}^{*}}{\overline{r}} \cdot \left\{\frac{1}{\sqrt{2\pi}} - \frac{\sigma_{R}^{*} \cdot \ln(z^{*})}{2\Delta r} \cdot \exp\left[\frac{1}{2} \left(\frac{\sigma_{R}^{*} \cdot \ln(z^{*})}{\Delta r}\right)^{2}\right] \cdot \operatorname{erfc}\left[\frac{\sigma_{R}^{*} \cdot \ln(z^{*})}{\sqrt{2}\Delta r}\right]\right\}.$$

$$(6.24)$$

Die Auswertung des besprochenen Algorithmus führt zu einem beträchtlichen numerischen Aufwand, da die Nullstelle des in Gleichung 6.20 definierten Polynoms bestimmt werden muß. Allerdings zeigte sich bei den in Unterkapitel 6.3 durchgeführten Untersuchungen ein deutlicher Rechenzeitvorteil gegenüber dem Faltungsalgorithmus.

#### 6.2.2.4 Obere Schranke der Zellverlustwahrscheinlichkeit

Eine andere Gruppe von direkten Annahmealgorithmen schätzt die Netzgüteparameter aus dem Zählprozeß der Zellankünfte innerhalb eines beliebigen Zeitintervalls der Dauer t ab. In der Literatur wurden obere Schranken für die Blockierungs- [198] und die Gesamtzellverlustwahrscheinlichkeit [59, 210, 212] angegeben. Da aus der Sicht des Benutzers die Zellverlustwahrscheinlichkeit die wichtigere Leistungskenngröße darstellt, beschränken sich die nachfolgenden Untersuchungen auf die in [59, 210, 212] beschriebenen Ansätze.

Der Zählprozeß kennzeichnet die Anzahl aller Zellankünfte innerhalb eines Zeitintervalls der Dauer t und wird durch die Zufallsvariable  $A_t$  beschrieben. Diese Zufallsvariable entspricht der Summe der Zufallsvariablen  $A_{t,n}$ , n=1,2,...,N, welche die Zählprozesse der einzelnen Verbindungen charakterisieren. Damit wird eine einfache Beschreibung des Überlagerungsverkehrs verschiedener statistisch unabhängiger Verkehrsströme möglich. Nach Gleichung 3.12 liefert die Faltung der Verteilungen  $p_{A,n}(a_t) = P(A_{t,n} = a_t)$  die gesuchte Verteilung  $p_A(a_t) = P(A_t = a_t)$ , die den Zählprozeß des Überlagerungsverkehrs beschreibt:

$$p_A(a_t) = p_{A,1}(a_t) * p_{A,2}(a_t) * \dots * p_{A,N}(a_t) .$$
 (6.25)

Einzelheiten zur Bestimmung des Zählprozesses einer einzelnen Verkehrsquelle können aus [200] entnommen werden. Insbesondere wird dort der Zählprozeß einer sporadischen Quelle abgeleitet.

In [210] wurden verschiedene Approximationen angegeben, mit denen eine obere Grenze für die Zellverlustwahrscheinlichkeit bestimmt werden kann. Die folgende Näherung liefert die genauesten Ergebnisse und wurde für den Leistungsvergleich verwendet:

$$B_Z \le \frac{\sum_{a_t = Ct+1}^{\infty} (a_t - Ct) \cdot p_A(a_t)}{\mathbb{E}[A_t]} = \frac{\sum_{a_t = Ct+1}^{\infty} (a_t - Ct) \cdot p_A(a_t)}{\sum_{a_t = 0}^{\infty} a_t \cdot p_A(a_t)}, \quad Ct = 0, 1, 2, ..., S+1 . \quad (6.26)$$

Dieser Schranke liegt die Annahme zugrunde, daß alle innerhalb eines Zeitintervalls der Dauer t eintreffenden Zellen auf das Ende des betrachteten Zeitintervalls verschoben werden. Bei der Bearbeitungsreihenfolge DF können von den  $a_t$  ankommenden Zellen mindestens Ct Zellen in das System übernommen werden, so daß maximal  $\max(a_t-Ct,0)$  Zellen verloren gehen (es gilt die Randbedingung  $Ct \leq S+1$ ). Die mit der Ankunftswahrscheinlichkeit  $p_A(a_t)$  gewichtete Summation dieser Beiträge ergibt eine obere Grenze für die mittlere Anzahl der innerhalb des betrachteten Zeitintervalls verlorenen Zellen. Den angegebenen Zusammenhang erhält man, indem diese Summe auf die mittlere Anzahl der Zellankünfte innerhalb eines Zeitintervalls der Dauer t bezogen wird.

Dieses Verfahren sichert eine vorgegebene Netzgüte, da unter keinen Umständen zu viele Verbindungen angenommen werden. Ferner ist dieser Ansatz sehr allgemein anwendbar, da er keinerlei Voraussetzungen für die Verkehrscharakteristik der einzelnen Quellen trifft. Andererseits ergeben sich in Bezug auf die numerische Komplexität dieses Algorithmus dieselben Probleme wie beim Faltungsansatz. Hier könnte wiederum die Normalverteilungs- oder die "Large Deviation"-Approximation Abhilfe schaffen.

#### 6.2.3 Indirekte Verfahren

#### 6.2.3.1 Linearer Ansatz

Die einfachste Verbindungsannahmefunktion reserviert für jede Verbindung eine feste Bedienrate, die nur von den Systemparametern und den Verkehrsparametern der betrachteten Verbindung abhängt [5, 77, 102, 126, 160, 168, 172, 229]. Die zugewiesene Bedienrate muß gewährleisten, daß die Netzgütevorgaben aller Verbindungen eingehalten werden können. Die beschriebene lineare Kapazitätszuteilung kann damit motiviert werden, daß die Grenzfläche, die zulässige von unzulässigen Verbindungszuständen trennt, nur eine geringe Krümmung aufweist.

Die zugewiesene Bedienrate wird vorab bestimmt und in geeigneten Tabellen festgehalten. Dazu muß der Verkehr entsprechend seinen Verkehrs- und Netzgüteparameterwerten in K verschiedene Verkehrsklassen eingeteilt werden. Für eine Verbindung aus Verkehrsklasse k, k=1,2,...,K, wird eine effektive Bedienrate  $C_{E,k}$  definiert, die zwischen deren mittlerer und maximaler Zellrate liegt. Eine neue Verbindung wird nur dann angenommen, wenn die Summe der effektiven Bedienraten aller bestehenden und der neu hinzukommenden Verbindung unter der Bedienrate C liegt:

$$\sum_{k=1}^{K} N_k \cdot C_{E,k} \le C \ . \tag{6.27}$$

Diese Annahmeentscheidung entspricht dem in durchschaltevermittelnden Netzen verwendeten Prinzip, das jeder Verbindung einen durchgehenden Kanal fester Bitrate zur exklusiven Nutzung zur Verfügung stellt. Die dabei angewandte Spitzenbitratenzuweisung entspricht dem Sonderfall  $C_{E,k} = \hat{r}_k$ .

Bei der Anwendung dieser linearen Kapazitätszuweisung spielt die Bestimmung der effektiven Bedienrate, die für alle möglichen Verkehrszusammensetzungen Gültigkeit besitzen soll, eine Schlüsselrolle. Die effektive Bedienrate von Verkehrsklasse k kann beispielsweise dadurch bestimmt werden, daß die bei homogener Verkehrszusammensetzung maximal zulässige Zahl von Verbindungen  $\widehat{N}_k$  berechnet wird. Die effektive Bedienrate von Verkehrsklasse k ergibt sich dann zu

$$C_{E,k} = \frac{C}{\widehat{N}_k} \,. \tag{6.28}$$

Die maximale Zahl der von Verkehrsklasse k zulässigen Verbindungen kann mit Hilfe von Analysen, Simulationen oder Messungen ermittelt werden. In der Literatur wurden, aus den in Unterkapitel 3.3 diskutierten Gründen, ausschließlich analytische Methoden verwendet:

• "Fluid Flow"-Approximation: [168]

• Faltungsmethode: [126, 160, 229]

#### • "Large Deviation"-Approximation: [5, 77, 102]

Die beschriebene Vorgehensweise unterschätzt die effektiven Bedienraten, da bei homogener Verkehrszusammensetzung der größte Multiplexgewinn erzielt wird (vgl. die in den Bildern 6.11 und 6.12 gezeigten Ergebnisse). Für eine korrekte Bestimmung der effektiven Bedienrate müßte allerdings die gesamte Annahmegrenzfläche bestimmt werden, was bei einer realistischen Anzahl von Verkehrsklassen nicht möglich ist.

Aus diesem Grund wurde eine Reihe indirekter Verfahren vorgeschlagen, welche die effektive Bedienrate in Abhängigkeit von der momentanen Verkehrszusammensetzung definieren. Die wichtigsten Vertreter werden in den folgenden Unterabschnitten behandelt. Weitere Verfahren sind in [51, 126, 174] beschrieben. Das in [51] vorgestellte Verfahren versucht den Einfluß des Hintergrundverkehrs durch einfache Kenngrößen zu erfassen. Allerdings zeigen die in [247] durchgeführten Untersuchungen, daß auch dieser Algorithmus zu viele Verbindungen annimmt. In [126] wird angeregt, die effektive Bedienrate für verschiedene Verkehrszusammensetzungen zu tabellieren. Für die zwischen diesen Stützstellen liegenden Werte müssen geeignete Interpolationsverfahren eingesetzt werden. Eine sinnvolle Umsetzung dieser Idee erweist sich angesichts der großen Zahl von Verkehrsklassen und der engen zeitlichen Rahmenbedingungen als schwierig. Bei dem in [174] gemachten Vorschlag werden die Verbindungen in zwei Gruppen mit niederem bzw. hohem Multiplexgewinn aufgeteilt. Für Verbindungen aus der ersten Gruppe wird die maximale Zellrate reserviert, so daß sich die für die restlichen Verbindungen zur Verfügung stehende Bedienrate entsprechend reduziert. Innerhalb der zweiten Gruppe werden effektive Bedienraten zugewiesen, die sich an dieser reduzierten Bedienrate orientieren. In ungünstigen Fällen kann auch hier die geforderte Netzgüte verletzt werden.

#### 6.2.3.2 "Class Related Rule"

In [65, 66, 67] wird unter der Bezeichnung "Class Related Rule" ein Verbindungsannahmealgorithmus vorgeschlagen, der die effektive Bedienrate einer Verkehrsklasse an den momentanen Belegungszustand dieser Klasse anpaßt. Dazu wird jede Verkehrsklasse k gesondert
betrachtet, und es wird die zum Erreichen einer bestimmten Zellverlustwahrscheinlichkeit
benötigte effektive Bedienrate  $C_k$  mit Hilfe von analytischen Untersuchungen, Simulationen
oder Messungen bestimmt. Die daraus gewonnenen Ergebnisse werden als Funktion der von
dieser Klasse produzierten mittleren Ankunftsrate  $N_k \cdot \overline{\tau}_k$  in einer geeigneten Tabelle festgehalten. Die Summe der effektiven Bedienraten aller Verkehrsklassen  $(\sum_{k=1}^K C_k(N_k \cdot \overline{\tau}_k))$ liefert eine obere Schranke für die effektive Gesamtbedienrate, da nur der innerhalb der jeweiligen Verkehrsklasse erzielte Multiplexgewinn ausgeschöpft wird. Der Multiplexgewinn,
der sich aus der Überlagerung verschiedenartiger Verkehrsströme ergibt, wird bei dieser Betrachtungsweise vernachlässigt.

Deshalb wird ein zweiter Schätzwert  $C_W$  für die effektive Gesamtbedienrate abgeleitet, wobei angenommen wird, daß der gesamte Verkehr von der Verkehrsklasse mit dem höchsten Kapazitätsbedarf erzeugt wird:

$$C_W\left(\sum_{k=1}^K N_k \cdot \overline{r}_k\right) = \max_j \left[C_j\left(\sum_{k=1}^K N_k \cdot \overline{r}_k\right)\right]. \tag{6.29}$$

Die getroffene Annahme liefert eine vorsichtige Abschätzung des zwischen den verschiedenen Verkehrsklassen möglichen Multiplexgewinns, so daß auch diese Beziehung eine obere Schranke für den momentanen Bedarf liefert.

Da jede dieser Näherungen die tatsächlich benötigte Bedienrate überschätzt, ergibt sich folgendes Kriterium für die Annahme einer neuen Verbindung [65]:

$$\min\left[\sum_{k=1}^{K} C_k(N_k \cdot \overline{r}_k), C_W\left(\sum_{k=1}^{K} N_k \cdot \overline{r}_k\right)\right] \le C.$$
(6.30)

Der geschilderte Ansatz sichert die vorgegebene Netzgüte und kann die Annahmeentscheidung in Echtzeit treffen. Andererseits reduziert sich die Effizienz dieses Verfahrens mit zunehmender Anzahl von Verkehrsklassen. Dies ist darauf zurückzuführen, daß der Verkehrsanteil einer einzelnen Klasse abnimmt, was sich ungünstig auf den innerhalb dieser Verkehrsklasse möglichen Multiplexgewinn auswirkt (vgl. Abschnitt 4.1.3). Ferner erhöht sich mit zunehmender Klassenanzahl der für die Tabellierung der effektiven Bedienraten erforderliche Speicherungsaufwand.

#### 6.2.3.3 "plaNET"-Verbindungsannahmealgorithmus

Ein weiterer Annahmealgorithmus, der innerhalb des experimentellen Hochgeschwindigkeits-Kommunikationsnetzes "plaNET" eingesetzt wird, ist in [83, 84] beschrieben. Dieser Ansatz berücksichtigt bei der Abschätzung des Multiplexgewinns sowohl die gegenseitige Kompensation der Verkehrsschwankungen, als auch den Ausgleich von kurzzeitigen Überlastsituationen durch die Netzpuffer. Die erste Komponente wird mit Hilfe der in Unterabschnitt 6.2.2.2 beschriebenen Normalverteilung erfaßt. Im Unterschied zu dem dort beschriebenen Verfahren verwendet der vorliegende Algorithmus die Blockierungswahrscheinlichkeit, die aus den Gleichungen 6.9 und 6.13 berechnet werden kann:

$$p_C \approx \int_{r=C}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_R} \exp\left[-\frac{1}{2} \cdot \left(\frac{r-\overline{r}}{\sigma_R}\right)^2\right] dr$$

$$p_C \approx \frac{1}{2} \cdot \operatorname{erfc}\left(\frac{C-\overline{r}}{\sqrt{2} \cdot \sigma_R}\right), \qquad (6.31)$$

mit

$$\overline{r} = \sum_{n=1}^{N} \overline{r}_n \quad \text{und} \quad \sigma_R^2 = \sum_{n=1}^{N} \overline{r}_n \cdot (\hat{r}_n - \overline{r}_n) .$$
 (6.32)

Dabei wurde angenommen, daß der über Verbindung n transportierte Verkehr durch eine sporadische Quelle beschrieben werden kann, die durch die maximale Zellrate  $\hat{r}_n$ , die mittlere Zellrate  $\overline{r}_n$  und die mittlere Anzahl der innerhalb eines Büschels gesendeten Zellen  $b_n$  charakterisiert ist.

Der Algorithmus bestimmt nun eine effektive Bedienrate  $C_I$ , die für eine vorgegebene Blockierungswahrscheinlichkeit  $\hat{p}_C$  zur Verfügung stehen muß. Dazu wird das  $\epsilon$ -Fraktil  $r_F(\epsilon)$  der Standard-Normalverteilung definiert:

$$\epsilon = \frac{1}{\sqrt{2\pi}} \int_{r=r_F(\epsilon)}^{\infty} \exp\left(-\frac{r^2}{2}\right) dr = \frac{1}{2} \operatorname{erfc}\left(\frac{r_F(\epsilon)}{\sqrt{2}}\right) . \tag{6.33}$$

In [83] ist die folgende Näherung für diese Kenngröße angegeben:

$$r_F(\epsilon) \approx \sqrt{-2\ln(\epsilon) - \ln(2\pi)}$$
 (6.34)

Ein Vergleich der Gleichungen 6.31 und 6.33 liefert den folgenden Schätzwert  $C_I$  für die effektive Bedienrate [83]:

$$\frac{C_I - \overline{r}}{\sigma_R} = r_F(\hat{p}_C) \iff C_I = \overline{r} + r_F(\hat{p}_C) \cdot \sigma_R.$$
 (6.35)

Die Netzpuffer können kurzzeitige Überlastsituationen überbrücken, was bei dem eben beschriebenen Ansatz vollkommen vernachlässigt wird. Der durch die Zwischenpufferung erzielte Auslastungsgewinn kann mit Hilfe des in Anhang A ausführlich erörterten "Fluid Flow"-Modells abgeschätzt werden. Dabei wird angenommen, daß der Multiplexer ausschließlich von der betrachteten Verbindung genutzt wird. Die Bedienrate des Multiplexers wird so eingestellt, daß eine zulässige Verlustwahrscheinlichkeit  $\hat{B}_B$  eingehalten werden kann. Die so definierte Bedienrate entspricht der effektiven Bedienrate der Verbindung, wofür in [83] die folgende einfache Beziehung angegeben ist:

$$C_{E,n} = \frac{\widehat{r}_n}{2} \left( 1 - \frac{S - \sqrt{[\gamma_n - S]^2 + 4 \cdot S \cdot \gamma_n \cdot \overline{r}_n / \widehat{r}_n}}{\gamma_n} \right) , \qquad (6.36)$$

mit

$$\gamma_n = \ln(1/\widehat{B}_B) \cdot b_n \cdot \left(1 - \frac{\overline{r}_n}{\widehat{r}_n}\right). \tag{6.37}$$

Die Summation dieser effektiven Bedienraten liefert einen zweiten Schätzwert für die effektive Gesamtbedienrate:

$$C_{II} = \sum_{n=1}^{N} C_{E,n} . {(6.38)}$$

Da die effektiven Bedienraten  $C_I$  und  $C_{II}$  verschiedene Ausgleichseffekte berücksichtigen, kann die neu hinzukommende Verbindung angenommen werden, falls die folgende Bedingung erfüllt ist:

$$\min(C_I, C_{II}) \le C \ . \tag{6.39}$$

Dieser Algorithmus kann sehr einfach implementiert werden, da er nur auf elementare Rechenoperationen zurückgreift. Insbesondere können die Kenngrößen  $\overline{r}$ ,  $\sigma_R$  und  $C_{II}$  in entsprechenden Registern zwischengespeichert werden [84], womit für einen Verbindungsaufbzw. abbau nur wenige Rechenschritte erforderlich sind. Dagegen wirkt sich die detaillierte Quellbeschreibung nachteilig auf die Flexibilität des Algorithmus auf. Reale Dienste weichen von der zugrundegelegten sporadischen Quellcharakteristik mehr oder weniger stark ab.

#### 6.2.3.4 Verfahren basierend auf neuronalen Netzen

Eine ganze Reihe von Verfahren zur Verbindungsannahme setzen neuronale Netze ein, die den komplexen nichtlinearen Zusammenhang zwischen der Netzbelastung und der daraus resultierenden Netzgüte erlernen können. Neuronale Netze versuchen die in der Natur vorkommenden Lernvorgänge nachzuahmen und bestehen aus einzelnen Neuronen, die untereinander verbunden sind. Über diese Verbindungen werden "Signale" ausgetauscht, die in den Neuronen verarbeitet werden: Die dem betrachteten Neuron i von den angeschlossenen Neuronen j (j=1,2,...) zugeführten Eingangssignale  $e_j$  werden zunächst mit einem Gewichtungsfaktor  $w_{ij}$  multipliziert. Anschließend werden diese gewichteten Eingangssignale aufsummiert und mit Hilfe einer Aktivierungsfunktion f auf das Ausgangssignal  $a_i$  umgesetzt. Der Ausgang des Neurons kann wiederum mit weiteren Neuronen verknüpft sein. Die prinzipielle Struktur und Funktionsweise eines vorwärtsgerichteten neuronalen Netzes ("Feed Forward Neural Network") ist in Bild 6.2 veranschaulicht. Wie angedeutet können in diesem Fall verschiedene Schichten unterschieden werden, wobei keine Verbindungen innerhalb einer Schicht oder zu vorangehenden Schichten erlaubt sind. Diese Beschränkung wird bei rekurrenten neuronalen Netzen aufgehoben, bei denen auch Rückkopplungen möglich sind.

Das gelernte "Wissen" des neuronalen Netzes ist auf alle Gewichtungsfaktoren  $w_{ij}$  verteilt. Diese Gewichtungsfaktoren werden während des Trainings des neuronalen Netzes verändert, wobei normalerweise eine Reihe von Eingangssignalen mit den dazugehörigen Ausgangssignalen (Zielwerten) an das neuronale Netz angelegt wird. Mit Hilfe eines Lernalgorithmus werden die internen Gewichte  $w_{ij}$  so verändert, daß die Differenz zwischen dem vom neuronalen Netz gelieferten Ausgabewert und dem Zielwert minimiert wird. Dabei handelt es sich um eine Optimierungsaufgabe, für die — abhängig vom gewählten Netztyp — eine Vielzahl von Methoden und Algorithmen bekannt sind, auf die allerdings nicht näher eingegangen werden soll. Für eine fundierte Einführung in dieses Gebiet sei auf [89] verwiesen.

Der Einsatz neuronaler Netze für die Verbindungsannahme in ATM-Netzen scheint vielversprechend, da ihre Lernfähigkeit eine langfristige Optimierung der Verbindungsannahmefunktion ermöglicht. In der Literatur wurden daher verschiedene Ansätze vorgeschlagen, die das neuronale Netz während des Wirkbetriebs trainieren [95, 96, 171]. In [95, 96] wird die Anzahl der bestehenden Verbindungen und die diesem Belegungszustand zugeordnete

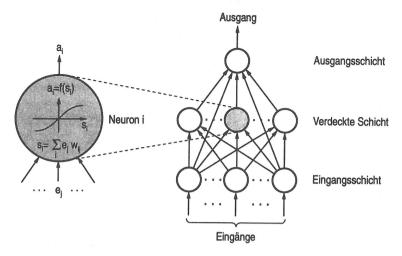


Bild 6.2: Prinzipielle Struktur und Funktionsweise eines vorwärtsgerichteten neuronalen Netzes

Zellverlustwahrscheinlichkeit während des Wirkbetriebs gemessen. Die gewonnenen Meßdaten dienen als Vorgabewerte für das parallel zum Netzbetrieb ablaufende Training des neuronalen Netzes. Der vorgeschlagene Ansatz ist allerdings ungeeignet für die vorgesehene Anwendung, da die extrem kleinen Zellverlustwahrscheinlichkeiten nur in langfristigen Zeiträumen gemessen werden können, in denen sich der Lastzustand durch Verbindungsaufund -abbauvorgänge vollkommen ändert (die zuverlässige Erfassung einer Zellverlustwahrscheinlichkeit von  $10^{-10}$  setzt eine Meßdauer von über 32 Tagen voraus). Der in [171] vorgestellte Ansatz erfaßt daher die mittlere Verzögerungszeit anstatt der Zellverluste. Die in Unterkapitel 4.2 durchgeführten Untersuchungen zeigen allerdings, daß diese Meßgröße kein aussagekräftiger Netzgüteparameter ist. Ein weiterer Nachteil dieses Ansatzes besteht darin, daß das neuronale Netz nur aus Fehlern lernen kann. Dies führt zu Verletzungen der vereinbarten Netzgüte während des laufenden Netzbetriebs. Eine ausführliche Kritik dieser Ansätze ist in [252] zu finden.

Die in [176, 204, 234] beschriebenen Algorithmen trainieren das neuronale Netz vor dessen Inbetriebnahme innerhalb des ATM-Netzes. Die für das Training des neuronalen Netzes verwendeten "Verbindungsmuster" und die dazugehörigen Zellverlustwahrscheinlichkeiten werden mit Hilfe von geeigneten analytischen Methoden berechnet. Da während des Wirkbetriebs keine Fortführung des Trainings stattfindet, geht die Anpassungsfähigkeit des neuronalen Netzes verloren, d.h. das neuronale Netz kann nur die während des Trainings erlernten Zusammenhänge reproduzieren und in gewissen Grenzen verallgemeinern. Durch die in neuronalen Netzen angewandte Parallelverarbeitung kann allerdings eine sehr viel

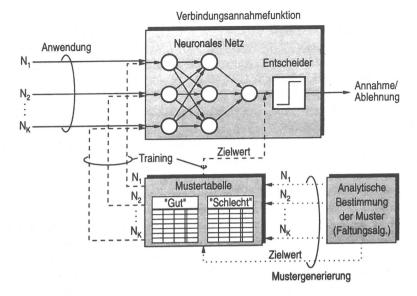


Bild 6.3: Training und Anwendung eines vorwärtsgerichteten neuronalen Netzes für die Verbindungsannahme in ATM-Netzen

schnellere Annahmeentscheidung getroffen werden.

Bild 6.3 zeigt die den im folgenden Unterkapitel beschriebenen Untersuchungen zugrundeliegende Konfiguration, die im Rahmen einer Diplomarbeit [204] ausführlich untersucht wurde. Den Kern der Verbindungsannahmefunktion bildet ein ausschließlich in Vorwärtsrichtung betriebenes neuronales Netz, das aus dem am Eingang anliegenden Verbindungsvektor  $(N_1, N_2, ..., N_K)$  die Zellverlustwahrscheinlichkeit ableitet. Die Neuronen sind in drei Schichten — Eingangsschicht, verdeckte Schicht und Ausgangsschicht — angeordnet (vgl. Bild 6.2). Die Anzahl der Neuronen in der Eingangsschicht entspricht der Anzahl der Verkehrsklassen. Diese Neuronen sind vollvermascht mit den in der verdeckten Schicht befindlichen Neuronen. Erfahrungswerte belegen, daß die Neuronenzahl der verdeckten Schicht zumindest proportional zur Anzahl der Verkehrsklassen zunehmen sollte. Die Ausgänge der verdeckten Neuronen führen auf das Ausgangsneuron, das mit einem Entscheider verbunden ist. Der Entscheider vergleicht den Ausgabewert mit einer Schwelle (zulässige Zellverlustwahrscheinlichkeit bzw. Zellbedienrate) und entscheidet über die Annahme bzw. Ablehnung des Verbindungswunschs.

Für das Training des neuronalen Netzes werden zunächst geeignete Lernmuster mit Hilfe des in Unterabschnitt 6.2.2.1 beschriebenen Faltungsansatzes erzeugt. Ein Lernmuster umfaßt die Anzahl der von den einzelnen Verkehrsklassen aufgebauten Verbindungen  $N_k$ 

und den zu diesem Verbindungsvektor gehörenden Zielwert (Zellverlustwahrscheinlichkeit bzw. effektive Bedienrate). Alle Muster werden in einer Mustertabelle abgelegt, die in "Gut"-und "Schlechtmuster" unterteilt ist (ein "Gutmuster" kennzeichnet einen zulässigen Verbindungszustand, während bei einem "Schlechtmuster" eine unzulässige Verschlechterung der Netzgüte auftritt). Beim Training des neuronalen Netzes werden diese Muster in zufälliger Reihenfolge und einem fest vorgegebenen Verhältnis zwischen "Gut"- und "Schlechtmustern" aus der Mustertabelle entnommen. Das Training des neuronalen Netzes erfolgt mit Hilfe eines modifizierten "Backpropagation"-Algorithmus. Detaillierte Informationen zur Lernmustergenerierung und zum Training des neuronalen Netzes (Grundeinstellungen, Lernparameter, Trainingszyklen …) können aus [204] entnommen werden.

Ein Nachteil dieses Ansatzes ist die überproportionale Zunahme der Neuronenzahl und des Trainingsaufwands, falls die Anzahl der Verkehrsklassen anwächst. Die Simulation des Netzes mußte daher auf drei verschiedene Verkehrsklassen eingeschränkt werden. Diesen Nachteil vermeidet der in Bild 6.4 dargestellte Ansatz, der auf ein zeitdiskretes rekurrentes neuronales Netz zurückgreift. Dieser Netztyp beinhaltet Rückkopplungen, die eine Speicherung des ATM-Netzzustandes ermöglichen. Der Ausgabewert des neuronalen Netzes hängt somit nicht nur von den momentan anliegenden Eingabewerten, sondern von der gesamten zurückliegenden Eingabesequenz ab. Diese Eigenschaft ermöglicht die Zwischenspeicherung des Belegungszustands des betrachteten Verbindungsabschnitts. Aus diesem Grund werden dem neuronalen Netz nur noch Änderungen des Belegungszustands mitgeteilt. Dies erfolgt durch das Anlegen der entsprechenden Verbindungsparameter, beispielsweise der mittleren und der maximalen Zellrate. Genauer gesagt werden für einen Verbindungswunsch die Verbindungsparameter  $\bar{r}_i$  und  $\hat{r}_i$  angelegt, während bei einem Verbindungsabbau bzw. einer Verbindungsablehnung die ensprechenden negativen Werte zugeführt werden.

Für das Training des neuronalen Netzes müssen verschiedene Trainingssequenzen erstellt werden, welche die typische Reihenfolge der Verbindungsauf- und -abbauvorgänge wiedergeben. Diese Trainingssequenzen werden wiederum mit Hilfe des Faltungsansatzes erstellt, müssen in diesem Falle aber direkt an das neuronale Netz angelegt werden, da eine zufällige Auswahl aus einer Mustertabelle aufgrund der Reihenfolgeabhängigkeit ausscheidet. Dies führt zu einer deutlichen Erhöhung des Trainingsaufwands.

Auf eine detaillierte Beschreibung dieses Ansatzes und der damit erzielten Ergebnisse soll aus Platzgründen verzichtet werden. Die in Studien- und Diplomarbeiten [63, 176] durchgeführten Untersuchungen bestätigen die prinzipielle Anwendbarkeit dieses Konzepts. Allerdings zeigte sich, daß das Training von rekurrenten neuronalen Netzen große Probleme aufwirft. Insbesondere muß sehr viel heuristisches Wissen in die Struktur des Netzes und eine geeignete Initialisierung der Gewichte eingebracht werden, um das Netz erfolgreich trainieren zu können.

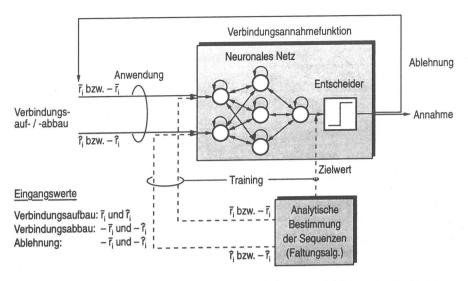


Bild 6.4: Training und Anwendung eines rekurrenten neuronalen Netzes für die Verbindungsannahme in ATM-Netzen

# 6.3 Vergleichende Leistungsuntersuchung der Annahmealgorithmen

# 6.3.1 Bestimmung der Verbindungsannahme-Grenzkurven

Zunächst sollen die verschiedenen Verfahren auf der Basis der in Abschnitt 5.3.2 eingeführten Verbindungsannahme-Grenzkurven verglichen werden. Alle unterhalb dieser Grenzkurven liegenden Verbindungszustände werden von den zugeordneten Algorithmen zugelassen, wogegen darüberliegende Verbindungszustände nicht auftreten können. Dem Leistungsvergleich liegen vier unterschiedliche Verkehrsklassen zugrunde, die verschiedene Dienste — Daten-, Video-, Sprach- und Festbildübertragung — repräsentieren. Dazu wurde aus Tabelle 2.2 jeweils ein typischer Vertreter dieser Dienste ausgewählt und entsprechend gekennzeichnet. Diese Verkehrsklassen decken ein breites Spektrum von Verkehrsparametern ab, so daß der Einfluß des Quellverhaltens erfaßt werden kann.

Verbindungen aus der Verkehrsklasse 3 senden mit konstanter Bitrate, wogegen die Klassen 1 und 4 sporadischen Quellverkehr erzeugen, der durch die maximale Zellrate, den Büschelfaktor und die mittlere Büschelgröße charakterisiert werden kann. Eine Quelle aus Verkehrsklasse 2 setzt sich aus 20 sogenannten "Miniquellen" zusammen, die wiederum sporadischen Verkehr generieren (vgl. Bild 4.19). Die Verkehrsparameter aller Verkehrsklassen sind in Ta-

Verkehrs- klasse	Dienst	Anzahl "Miniquellen"	Max. Zellrate [Zellen/s] <sup>a</sup>	Büschel- faktor <sup>a</sup>	Mittlere Büschelgröße <sup>a</sup>
1	Datenübertragung	1	26041,7	10	338,5
2	Videokommunikation	20	2520,8	5	809,2
3	Sprachkommunikation	1	166,7	1	00
4	Festbildübertragung	1	5208,3	23	2604,2

<sup>&</sup>lt;sup>a</sup> Die Angaben beziehen sich auf eine "Miniquelle"

Tabelle 6.1: Charakteristische Kenngrößen der gewählten Verkehrsklassen

belle 6.1 zusammengefaßt. Ferner wurde den Untersuchungen eine Puffergröße von 64 Zellen und eine Bedienrate von  $3,532\cdot 10^5$  Zellen/s, entsprechend einer Übertragungsbitrate von 149,76 Mbit/s, zugrundegelegt.

Aus Gründen der Übersichtlichkeit werden bei den nachfolgenden Untersuchungen nur Verkehrsmischungen aus den Verkehrsklassen 1 und 2 bzw. den Verkehrsklassen 3 und 4 betrachtet. Die in den Bildern 6.5 und 6.6 gezeigten Ergebnisse vergleichen die für den Faltungsalgorithmus definierten Annahmekriterien. Für die den verschiedenen Kriterien zugeordneten Grenzwerte wurden folgende Annahmen getroffen:

- Überlastwahrscheinlichkeit  $p_C \leq 10^{-10}$
- Gesamte Verlustwahrscheinlichkeit  $B_B \leq 10^{-10}$
- Individuelle Verlustwahrscheinlichkeit  $B_{B,n} \leq 10^{-10}, \forall n = 1, 2, ..., N$
- $\bullet$  Unterlast dauer  $t_U > 10^7~\mathrm{s} \approx 116~\mathrm{Tage}$

Um die Unterschiede zwischen den einzelnen Kriterien zu betonen, wurde das in Gleichung 6.1 definierte Lastkriterium vernachlässigt, d.h. es gilt  $\hat{A}_Z = 1$ . Diese Annahme wird auch für alle weiteren Untersuchungen beibehalten.

Die eingezeichneten Grenzkurven geben die Trennlinie zwischen zulässigen und unzulässigen Verbindungszuständen wieder, was zu dem gezeigten stufenförmigen Verlauf führt. Weiterhin wurden die Annahmekurven gegeneinander verschoben, um den Gesamtverlauf hervorzuheben. Die Ergebnisse bestätigen, daß das Überlastkriterium die zulässige Verbindungszahl am stärksten einschränkt. Dies läßt sich damit erklären, daß die Überlastwahrscheinlichkeit eine obere Grenze für die Verlustwahrscheinlichkeit  $B_B$  liefert. Da die verschiedenen Verkehrsklassen unterschiedliche Verlustwahrscheinlichkeiten aufweisen, verläuft die durch die individuelle Verlustwahrscheinlichkeit definierte Grenzkurve unterhalb der für die Gesamtverlustwahrscheinlichkeit gültigen Kurve.

Der Grenzwert für die mittlere Unterlastdauer wurde so gewählt, daß die hierdurch definierte Annahmekurve einen ähnlichen Verlauf aufweist wie die anderen Grenzkurven. Insbesondere ist der enge Zusammenhang zwischen der Überlastwahrscheinlichkeit und der mittleren Unterlastdauer zu erkennen. Insgesamt zeigt sich, daß die einzelnen Kriterien zu ähnlichen Verbindungsannahme-Grenzkurven führen. Dieses Ergebnis läßt sich dahingehend interpretieren, daß eines der genannten Kriterien ausreicht, um alle relevanten Netzgüteparameter zu kontrollieren. Damit keine Verletzungen der Netzgüte auftreten, sollte das restriktivste

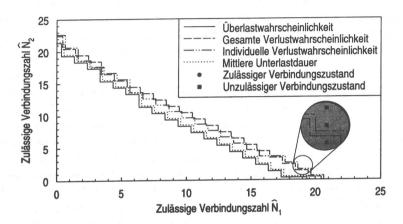


Bild 6.5: Vergleich der für den Faltungsalgorithmus definierten Annahmekriterien (Verkehrsklassen 1 und 2)

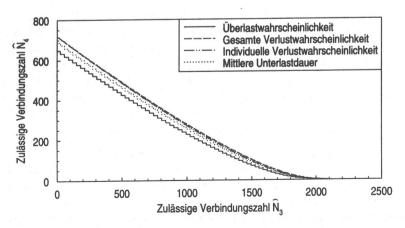


Bild 6.6: Vergleich der für den Faltungsalgorithmus definierten Annahmekriterien (Verkehrsklassen 3 und 4)

Kriterium gewählt werden. Den weiteren Untersuchungen liegt die gesamte Zellverlustwahrscheinlichkeit zugrunde, da dies der einzige Netzgüteparameter ist, der von allen oben beschriebenen Verbindungsannahmealgorithmen kontrolliert werden kann.

Die folgende Studie vergleicht die in Abschnitt 6.2.2 beschriebenen direkten Verbindungsannahme<br/>algorithmen. Für die unterhalb der Annahmekurven liegenden Verbindungszustände ist die durch den Annahme<br/>algorithmus abgeschätzte Gesamtzellverlustwahrscheinlichkei<br/>t $\hat{B}_B$ kleiner als  $10^{-10}$ . Zu Vergleichszwecken wurde ein idealer Algorithmus definiert, der die individuelle Zellverlustwahrscheinlichkeiten unter dem beschriebenen Grenzwert hält. Dieser Algorithmus basiert auf der für das "Fluid Flow"-Modell abgeleiteten asymptotischen Näherung nach Gleichung A.51. Diese Näherung überschätzt die tatsächlich auftretenden Zellverluste nur leicht, wie die in Bild 5.17 dargestellten Ergebnisse bestätigen.

Die in den Bildern 6.7 und 6.8 wiedergegebenen Ergebnisse zeigen, daß der Faltungsansatz und die "Large Deviation"-Approximation nahezu identische Grenzkurven liefern, die zudem sehr dicht an der idealen Grenzkurve liegen. Nur im mittleren Kurvenbereich treten geringe Abweichungen von den Idealwerten auf, die darauf zurückgeführt werden können, daß der ideale Algorithmus auf einem anderen Annahmekriterium basiert. Diese Diskrepanzen können vermieden werden, indem auch dem Faltungsalgorithmus und der "Large Deviation"-Approximation die individuellen Zellverlustwahrscheinlichkeiten zugrundegelegt werden. Demgegenüber unterschätzt die Normalverteilung die tatsächlich auftretende Zellverlustwahrscheinlichkeit deutlich, weshalb zu viele Verbindungen angenommen werden. Selbst bei der Verkehrsklasse 4, die nur eine verhältnismäßig geringe Spitzenbitrate von 2 Mbit/s aufweist, kommt es zu einer Überschreitung der Idealkurve. Dies kann durch den relativ großen Büschelfaktor begründet werden.

In Bild 6.9 wird das beschriebene Verhalten anhand der komplementären Verteilungsfunktion der Ankunftsrate näher untersucht, wobei 20 Quellen aus Verkehrsklasse 1 überlagert werden. Das exakte Ergebnis gibt die durch den Faltungsalgorithmus errechneten Verteilungswerte wieder. Die Normalverteilung weist im Bereich der mittleren Ankunftsrate  $(\overline{r}=2\,\hat{r}_1)$  ihre höchste Genauigkeit auf, während für große Ankunftsraten eine deutliche Unterschätzung der komplementären Verteilungsfunktion zu erkennen ist. Die "Large Deviation"-Approximation weist die höchste Genauigkeit in der Umgebung der Bedienrate  $C=13,564\,\hat{r}_1$  auf. Dieser Bereich ist für die Bestimmung der Zellverlustwahrscheinlichkeit ausschlaggebend.

Der auf dem Zählprozeß basierende Algorithmus liefert eine obere Schranke für die Zellverlustwahrscheinlichkeit, weshalb die dafür geltende Annahmekurve unter der idealen Grenzkurve liegt. Allerdings schränkt dieser Algorithmus die zulässige Verkehrslast stark ein, falls ein Großteil der Verbindungen aus der Verkehrsklasse 3 stammt, die eine sehr niedrige Spitzenbitrate aufweist. In diesem Fall werden sogar weniger Verbindungen als bei der Spitzenbitratenzuweisung angenommen. Dies wird durch den in Bild 6.10 gezeigten Vergleich zwischen der exakten Zellverlustwahrscheinlichkeit (vgl. Unterkapitel 4.1) und den

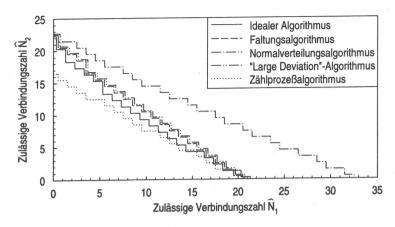


Bild 6.7: Vergleich der Verbindungsannahme-Grenzkurven von direkten Annahmealgorithmen für die Verkehrsklassen 1 und 2

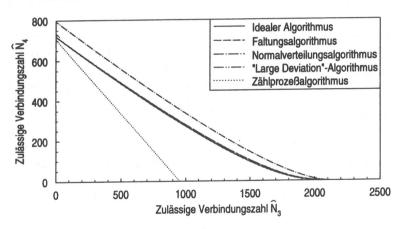
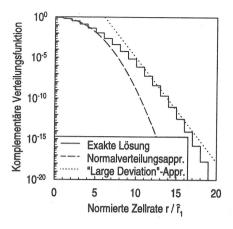


Bild 6.8: Vergleich der Verbindungsannahme-Grenzkurven von direkten Annahmealgorithmen für die Verkehrsklassen 3 und 4

aus dem Zählprozeß berechneten Werten bestätigt: Die Überlagerung von sporadischen Verkehrsströmen wird relativ gut approximiert, wogegen die Zellverlustwahrscheinlichkeit von CBR-Verkehr signifikant überschätzt wird. Da der untersuchte Algorithmus auf der Bearbeitungsreihenfolge DF beruht, wurde diese auch den gezeigten Ergebnissen zugrundegelegt. Die restlichen Quellparameter können aus dem Bild entnommen werden.

Die in Abschnitt 6.2.3 behandelten indirekten Verbindungsannahmealgorithmen werden in den Bildern 6.11 und 6.12 verglichen. Dabei sind für alle Algorithmen — mit Ausnahme



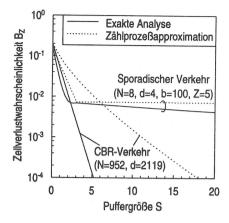


Bild 6.9: Vergleich zwischen exakter Verteilung, Normalverteilung und "Large Deviation"-Verteilung

Bild 6.10: Vergleich zwischen der exakten Lösung und der aus dem Zählprozeß berechneten Näherung

des neuronalen Netzes — deutliche Abweichungen von der idealen Grenzkurve zu erkennen. Die für die Spitzenbitratenzuweisung und die "Class Related Rule" erhaltenen Grenzkurven liegen in allen Fällen unterhalb der idealen Grenzkurve, weshalb diese Verfahren die spezifizierten Zellverlustwahrscheinlichkeiten einhalten können. Allerdings ist die Spitzenbitratenzuweisung äußerst ineffizient, da der durch statistisches Multiplexen mögliche Gewinn nicht ausgeschöpft wird.

Die für die "Class Related Rule" ermittelten Ergebnisse liegen sehr viel näher an der idealen Grenzkurve. Mit zunehmender Anzahl von Verkehrsklassen sinkt die Effizienz dieses Verfahrens weiter ab, da die Anzahl der in einer Verkehrsklasse befindlichen Verbindungen und damit der innerhalb dieser Verkehrsklasse erzielte Multiplexgewinn abnimmt. Dies wird auch aus den in Bild 6.13 dargestellten Kurven deutlich, welche die effektive Bedienrate als Funktion der mittleren Ankunftsrate wiedergeben. Die Krümmung der Kurven bestätigt, daß mit abnehmender Anzahl von Verbindungen das Verhältnis aus effektiver Bedienrate und mittlerer Ankunftsrate ungünstiger wird und somit der innerhalb einer Verkehrsklasse erzielte Multiplexgewinn abnimmt. Die effektiven Bedienraten wurden mit Hilfe der für das "Fluid Flow"-Modell abgeleiteten asymptotischen Approximation berechnet, die auch die Grundlage des idealen Algorithmus bildet. Das relative Verhältnis der effektiven Bedienraten der unterschiedlichen Verkehrsklassen läßt sich mit den in Unterkapitel 4.1 aufgezeigten Zusammenhängen erklären.

Der lineare Algorithmus und der "plaNET"-Algorithmus nehmen mehr Verbindungen an als der ideale Algorithmus, d.h. diese Verfahren sind nicht in der Lage, die vorgeschriebe-

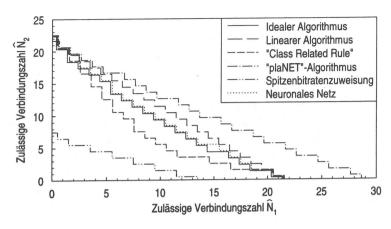


Bild 6.11: Vergleich der Verbindungsannahme-Grenzkurven von indirekten Annahmealgorithmen für die Verkehrsklassen 1 und 2

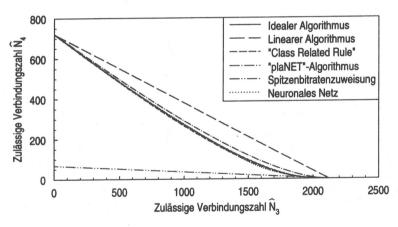


Bild 6.12: Vergleich der Verbindungsannahme-Grenzkurven von indirekten Annahmealgorithmen für die Verkehrsklassen 3 und 4

ne Zellverlustwahrscheinlichkeit zu sichern. Beim linearen Algorithmus wurde die effektive Bedienrate einer Verbindung nach Gleichung 6.28 bestimmt, wobei die bei homogener Verkehrszusammensetzung zulässige Verbindungszahl aus dem idealen Algorithmus hervorgeht. Bei heterogener Verkehrszusammensetzung werden zu viele Verbindungen angenommen, da die ideale Grenzkurve konvex verläuft. Prinzipiell könnte die lineare Annahmekurve in Richtung des Ursprungs so verschoben werden, daß die ideale Grenzkurve nur tangiert wird. Dazu müßte allerdings die (K-1)-dimensionale Verbindungsannahme-Grenzfläche vollständig be-

stimmt werden, was für eine realistische Anzahl von Verkehrsklassen K>10 nicht möglich ist. Ferner würde dies zu Einschränkungen in Bezug auf die Netzauslastung führen.

Die durch den "plaNET"-Algorithmus bestimmten effektiven Bedienraten  $C_I$  und  $C_{II}$  sind in Bild 6.14 für die Verkehrsklasse 1 in Abhängigkeit von der Verbindungszahl aufgetragen. Die aus der "Fluid Flow"-Approximation berechnete effektive Bedienrate  $C_{II}$  entspricht nahezu der Summe der Spitzenbitraten, so daß dieses Kriterium mit der Spitzenbitratenzuweisung vergleichbar ist. Dies würde sich ändern, wenn der Puffer wesentlich größer wäre als die mittlere Büschelgröße (den in [83] beschriebenen Untersuchungen liegt beispielsweise eine Puffergröße von 3 Mbit/s, entsprechend 7812 Zellen, zugrunde). Dies bedeutet jedoch einen erheblichen Implementierungsaufwand und erschwert die Verkehrssteuerung zusätzlich, da dies zu einer nicht zu vernachlässigenden Verformung der Verkehrsströme führen würde. In den gezeigten Beispielen entscheidet deshalb ausschließlich die aus der Normalverteilung bestimmte effektive Bedienrate  $C_I$  über die Annahme einer neuen Verbindung. Für den "plaNET"-Algorithmus gelten daher ähnliche Bemerkungen wie für den Normalverteilungsalgorithmus.

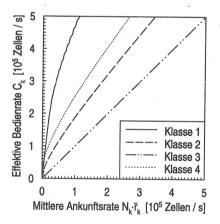


Bild 6.13: Effektive Bedienrate als Funktion der mittleren Ankunftsraten der Verkehrsklassen 1-4

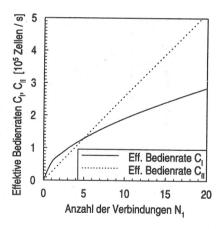


Bild 6.14: Effektive Bedienrate als Funktion der Verbindungszahl für den "plaNET"-Algorithmus

Der auf einem neuronalen Netz basierende Algorithmus liefert die genaueste Verbindungsannahme-Grenzkurve aller indirekten Verfahren. Den gezeigten Ergebnissen liegt die in Bild 6.3 skizzierte Anordnung zugrunde, wobei in der verdeckten Schicht 7 Neuronen verwendet wurden. Der für das Training des Netzes verwendete Mustersatz bestand aus 100 "Schlecht-" und 50 "Gutmustern", die jeweils 15000 mal an das neuronale Netz angelegt wurden. Einzelheiten zur Erzeugung des Mustersatzes und zum genauen Ablauf des Trainings können aus [204] entnommen werden.

Die primäre Aufgabe der Verbindungsannahmefunktion ist die Sicherung einer spezifizierten Netzgüte [7, 121]. Im vorliegenden Fall sollte eine Zellverlustwahrscheinlichkeit von  $10^{-10}$  eingehalten werden. Die bei den zugelassenen Verbindungszuständen auftretenden Zellverlustwahrscheinlichkeiten wurden mit Hilfe der für das "Fluid Flow"-Modell abgeleiteten asymptotischen Näherung nach Gleichung A.51 abgeschätzt. In Tabelle 6.2 sind die maximal auftretenden Verlustwahrscheinlichkeiten zusammengefaßt, wobei die vorher gezeigten Verbindungsannahme-Grenzkurven zugrundegelegt wurden.

	Verkehrsmischung 1		Verkehrsmischung 2	
	Klasse 1	Klasse 2	Klasse 3	Klasse 4
Idealer Algorithmus	$1,0\cdot 10^{-10}$	$1,0\cdot 10^{-10}$	$1,0\cdot 10^{-10}$	$1,0\cdot 10^{-10}$
Spitzenbitratenzuweisung	0	0	0	0
Faltungsalgorithmus	$2,6\cdot 10^{-10}$	$6,0\cdot 10^{-11}$	$8, 2 \cdot 10^{-11}$	$1,7\cdot 10^{-9}$
Normalverteilungsalg.	$3, 1 \cdot 10^{-7}$	$6, 3 \cdot 10^{-8}$	$1,0\cdot 10^{-5}$	$2, 4 \cdot 10^{-4}$
"Large Deviation"-Alg.	$2,6\cdot 10^{-10}$	$6,0\cdot 10^{-11}$	$1,4\cdot 10^{-10}$	$2,8 \cdot 10^{-9}$
Zählprozeßalgorithmus	$3,0\cdot 10^{-11}$	$4,9 \cdot 10^{-12}$	$1,9\cdot 10^{-11}$	$4,4\cdot 10^{-11}$
Linearer Algorithmus	$1,6\cdot 10^{-9}$	$2, 3 \cdot 10^{-10}$	$1,8 \cdot 10^{-3}$	$1, 3 \cdot 10^{-2}$
"Class Related Rule"	$8, 3 \cdot 10^{-11}$	$6,0\cdot 10^{-11}$	$4, 4 \cdot 10^{-11}$	$9,9 \cdot 10^{-11}$
"plaNET"-Algorithmus	$2,5\cdot 10^{-8}$	$5,0\cdot 10^{-9}$	$5, 3 \cdot 10^{-7}$	$1, 2 \cdot 10^{-5}$
Neuronales Netz	$1, 1 \cdot 10^{-10}$	$6,0\cdot 10^{-11}$	$4,9 \cdot 10^{-11}$	$9,5\cdot 10^{-10}$

Tabelle 6.2: Maximale Zellverlustwahrscheinlichkeiten verschiedener Annahmealgorithmen

Für den Normalverteilungsansatz, den linearen Algorithmus und den in "plaNET" eingesetzten Algorithmus treten sehr große Zellverlustwahrscheinlichkeiten auf. Diese Beeinträchtigungen der Netzgüte können nicht toleriert werden, so daß diese Algorithmen für den praktischen Einsatz nicht geeignet sind. Der Faltungsalgorithmus und die "Large Deviation"-Approximation überschreiten den angestrebten Grenzwert von 10<sup>-10</sup> nur knapp. Diese geringfügige Einschränkung der Netzgüte kann sogar ganz vermieden werden, falls das verbindungsindividuelle Verlustkriterium nach Gleichung 6.6 eingesetzt wird. Das neuronale Netz liefert eine Approximation der idealen Annahmekurve, die sowohl nach oben als auch nach unten abweichen kann. Alle weiteren Algorithmen können die Zellverlustwahrscheinlichkeit unter dem angestrebten Grenzwert halten, da sie auf einer oberen Grenze für die Zellverlustwahrscheinlichkeit bzw. die effektive Bedienrate basieren.

### 6.3.2 Leistungsvergleich auf Verbindungsebene

In diesem Abschnitt sollen die besprochenen Algorithmen in Bezug auf das zulässige Verkehrsangebot auf Verbindungsebene bewertet werden. Diese Untersuchung basiert auf dem in Abschnitt 3.2.1 beschriebenen allgemeinen Verlustsystem. Es wird angenommen, daß der Verkehr in K verschiedene Verkehrsklassen unterteilt werden kann, womit sich der momentane Verbindungszustand durch den Zustandsvektor  $N = (N_1, N_2, ..., N_K)$  beschreiben läßt. Dabei bezeichnet  $N_k$  die Anzahl der Verbindungen aus Verkehrsklasse k, k = 1, 2, ..., K. Die Verbindungswünsche innerhalb einer Verkehrsklasse treffen entsprechend einem Poisson-Prozeß ein. Die Verbindungsdauer wird durch eine allgemeine Verteilung beschrieben und das Verkehrsangebot von Klasse k betrage  $A_{V,k}$ .

Da sich die Zustandsübergänge im lokalen Gleichgewicht [150] befinden, ergibt sich die folgende Produktlösungsform für die Zustandswahrscheinlichkeiten (siehe z.B. [133, 245]):

$$p_{V}(N) = p_{V}(N_{1}, N_{2}, ..., N_{K}) = \frac{\frac{A_{V,1}^{N_{1}}}{N_{1}!} \cdot \frac{A_{V,2}^{N_{2}}}{N_{2}!} \cdot ... \cdot \frac{A_{V,K}^{N_{K}}}{N_{K}!}}{\sum_{N^{*} \in \Omega} \frac{A_{V,1}^{N_{1}^{*}}}{N_{1}^{*}!} \cdot \frac{A_{V,2}^{N_{2}^{*}}}{N_{2}^{*}!} \cdot ... \cdot \frac{A_{V,K}^{N_{K}^{*}}}{N_{K}^{*}!}}.$$
(6.40)

Dabei umfaßt die Menge  $\Omega$  alle zulässigen Verbindungszustände. Für die Bestimmung der Verbindungsverlustwahrscheinlichkeit von Verkehrsklasse k müssen die Zustandswahrscheinlichkeiten aller Blockierungszustände aufsummiert werden. Die Menge der Blockierungszustände von Verkehrsklasse k ergibt sich aus den oben beschriebenen Grenzkurven und ist folgendermaßen definiert [133, 245]:

$$V_k = \{(N_1, N_2, ..., N_k, ..., N_K) \in \Omega | (N_1, N_2, ..., N_k + 1, ..., N_K) \notin \Omega \}, \quad k = 1, 2, ..., K.$$
(6.41)

Da die Verbindungswünsche entsprechend einem Poisson-Prozeß ankommen, ergibt sich die Verbindungsverlustwahrscheinlichkeit aus [133, 245]

$$B_{V,k} = \sum_{N \in V_k} p_V(N) . \tag{6.42}$$

Im folgenden wird das auf einem Verbindungsabschnitt zugelassene Verkehrsangebot unter der Voraussetzung ermittelt, daß weniger als 1% der ankommenden Verbindungswünsche jeder Verkehrsklasse verloren gehen. Es werden wiederum zwei verschiedene Verkehrsmischungen aus den Verkehrsklassen 1 und 2 bzw. aus den Verkehrsklassen 3 und 4 betrachtet. Ferner wird angenommen, daß das Verkehrsangebot der Verkehrsklassen 1 und 2 gleich ist, während das Verkehrsangebot der Verkehrsklasse 3 dreimal höher ist als das Verkehrsangebot von Klasse 4.

Die in Tabelle 6.3 wiedergegebenen Werte für das zulässige Verkehrsangebot spiegeln das qualitative Verhalten der Verbindungsannahme-Grenzkurven wider. Insbesondere ist zu erkennen, daß die Spitzenbitratenzuweisung, der auf dem Zählprozeß basierende Algorithmus und die "Class Related Rule" zum Teil deutliche Leistungseinschränkungen gegenüber dem idealen Algorithmus aufweisen. Für die "Class Related Rule" reduziert sich das zulässige Verkehrsangebot weiter, falls die Anzahl der Verkehrsklassen zunimmt. Diese Algorithmen sind zwar prinzipiell geeignet für die Verbindungsannahme in ATM-Netzen, führen allerdings nur zu einer relativ geringen Netzauslastung. Alle anderen Verfahren übertreffen das vom idealen Algorithmus zugelassene Verkehrsangebot mit den für die Netzgüte beschriebenen Folgen.

	Verkehrsmischung 1		Verkehrsmischung 2	
	Klasse 1	Klasse 2	Klasse 3	Klasse 4
Idealer Algorithmus	5,62	5,62	881,92	293, 97
Spitzenbitratenzuweisung	3, 26	3,26	147,88	49, 29
Faltungsalgorithmus	6,02	6,02	890, 30	296, 77
Normalverteilungsalg.	7, 95	7,95	968, 91	322,97
"Large Deviation"-Alg.	6,02	6,02	890,11	296, 70
Zählprozeßalgorithmus	4,82	4,82	631,82	210,61
Linearer Algorithmus	6,42	6,42	1030,48	343,95
"Class Related Rule"	4,12	4,12	874, 40	291,47
"plaNET"-Algorithmus	6,99	6,99	906,01	302,00
Neuronales Netz	5,63	5,63	886,40	295, 47

Tabelle 6.3: Zulässiges Verkehrsangebot auf Verbindungsebene in Erlang

# 6.3.3 Diskussion der Ergebnisse

Die beschriebenen qualitativen und quantitativen Ergebnisse dokumentieren, daß keines der beschriebenen Verfahren alle in Unterkapitel 6.1 dargelegten Anforderungen erfüllen kann. Der Normalverteilungsansatz, der lineare Algorithmus und der in "plaNET" eingesetzte Algorithmus sind für die Verbindungsannahme ungeeignet, da sie nicht in der Lage sind, die vordefinierte Netzgüte zu sichern. Dabei kann es zu drastischen Überschreitungen der vorgeschriebenen Zellverlustwahrscheinlichkeit kommen.

Andere Algorithmen verhalten sich dagegen sehr zurückhaltend bei der Annahme neuer Verbindungen und führen deshalb zu einer unbefriedigenden Netzauslastung. Die Spitzenbitratenzuweisung vernachlässigt den durch das statistische Multiplexen möglichen Gewinn völlig, weshalb die Netzauslastung bei zunehmender Büschelförmigkeit des Verkehrs abnimmt. Der auf dem Zählprozeß basierende Algorithmus führt zu einer niedrigen Netzauslastung, falls die Verbindungen eine geringe Spitzenbitrate aufweisen. Darüberhinaus stellt dieser Algorithmus außerordentlich hohe Ansprüche in Bezug auf die Prozessorleistung und den Speicherplatzbedarf. Die "Class Related Rule" arbeitet nur bei wenigen verschiedenen Verkehrsklassen effizient. Für eine realistische Anzahl von Verkehrsklassen ist die Leistungsfähigkeit mit der Spitzenbitratenzuweisung vergleichbar.

Bei den verbleibenden Algorithmen ergeben sich Schwierigkeiten bei der praktischen Umsetzung. Der Faltungsalgorithmus ist aufgrund seiner Komplexität nicht in der Lage die Verbindungsannahmeentscheidung in Echtzeit zu treffen. Die Komplexität des "Large Deviation"-Ansatzes ist zwar wesentlich geringer, beinhaltet aber immer noch eine Nullstellensuche, die unter Umständen sehr aufwendig sein kann. Beim neuronalen Netz liegt das Hauptproblem darin, daß der Implementierungs- und Trainingsaufwand überproportional zunimmt, falls die Anzahl der Verkehrsklassen ansteigt. Daher muß dieser Ansatz auf wenige Verkehrsklassen beschränkt werden. Demgegenüber benötigt die in [176] untersuchte Anordnung keine Klassifizierung des Verkehrs, worunter wiederum die Leistungsfähigkeit leidet.

Zusammenfassend läßt sich feststellen, daß Algorithmen mit geringem Implementierungsaufwand deutliche Einschränkungen in Bezug auf die Leistungsfähigkeit aufweisen, während die Komplexität leistungsfähiger Algorithmen eine reale Implementierung verteuert oder ganz verhindert. Im folgenden Unterkapitel wird ein zweistufiger Algorithmus vorgeschlagen, der je ein Verfahren aus beiden Gruppen so kombiniert, daß die Vorteile beider Ansätze weitgehend erhalten bleiben.

# 6.4 Zweistufiger Annahmealgorithmus

In verschiedenen Veröffentlichungen wurde eine zweistufige Implementierung der Verbindungsannahmefunktion angeregt [26, 27, 166, 193, 215]. In der ersten Stufe wird ein einfacher Algorithmus eingesetzt, der eine rasche Entscheidung über die Annahme bzw. Ablehnung eines neuen Verbindungswunschs treffen kann. Die von diesem Algorithmus verwendeten Parameter werden durch einen im Hintergrund ablaufenden Algorithmus dynamisch an die aktuelle Verkehrszusammensetzung angepaßt. Der nachfolgend vorgestellte Ansatz basiert auf diesem Grundkonzept.

#### 6.4.1 Beschreibung

In der ersten Stufe wird ein einfacher indirekter Algorithmus eingesetzt, der den für die Verbindungsannahme geltenden Echtzeitbedingungen gerecht wird. Dieser Algorithmus weist effektive Bedienraten zu, die entweder einen festen Wert haben oder von dem in der zweiten Stufe implementierten Faltungsalgorithmus dynamisch an die aktuelle Verkehrszusammensetzung angepaßt werden. Der Faltungsalgorithmus wurde deshalb zugrundegelegt, weil er die höchste Leistungsfähigkeit aller im vorhergehenden Unterkapitel besprochenen Verfahren aufweist. Andere analytische Ansätze zur Bestimmung der effektiven Bedienrate sind ebenfalls denkbar. Bild 6.15 verdeutlicht die Funktionsweise des zweistufigen Ansatzes anhand einer Stichprobe des zeitlichen Verlaufs der effektiven Gesamtbedienrate  $C_E$ , die zur Sicherung der vorgeschriebenen Netzgüte zur Verfügung stehen muß. Die effektive Gesamtbedienrate kann sich aufgrund der folgenden Ereignisse ändern:

- Verbindungsaufbau: Die effektive Gesamtbedienrate wird durch den in der ersten Stufe ablaufenden Algorithmus um den Wert  $C_{S,k}(j)$  erhöht, falls eine Verbindung aus Verkehrsklasse k aufgebaut wird. Die effektive Bedienrate  $C_{S,k}(j)$  wird durch den Hintergrundalgorithmus bestimmt und hängt vom betrachteten Zeitintervall  $(j \cdot T, (j+1) \cdot T]$  ab. Ein neuer Verbindungswunsch wird abgelehnt, falls die effektive Gesamtbedienrate  $C_E + C_{S,k}(j)$  die Zellbedienrate C überschreiten würde.
- Verbindungsabbau: Die effektive Gesamtbedienrate wird um den Wert  $C_{R,k}(j)$  erniedrigt, falls eine Verbindung aus Verkehrsklasse k abgebaut wird. Im allgemeinen Fall ist die effektive Bedienrate  $C_{R,k}(j)$  und die beim Verbindungsaufbau zugewiesene Bedienrate  $C_{S,k}(j)$  verschieden. Deshalb wird zusätzlich abgeprüft, ob die effektive Gesamtbedienrate unter der maximalen Zellrate  $\hat{r}$  bleibt. Die für den Verbindungsabbau

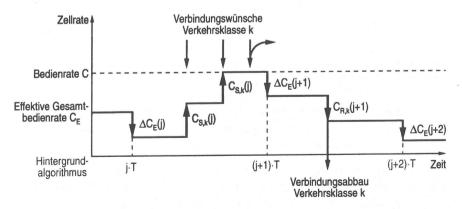


Bild 6.15: Stichprobe des Zeitverlaufs der effektiven Gesamtbedienrate

gültige effektive Bedienrate  $C_{R,k}(j)$  wird ebenfalls durch den Hintergrundalgorithmus festgelegt.

• Parameteranpassung durch den Hintergrundalgorithmus: Da der in der ersten Stufe ablaufende Algorithmus die effektive Gesamtbedienrate  $C_E$  überschätzt, bestimmt der in der zweiten Stufe angesiedelte Hintergrundalgorithmus die Differenz  $\Delta C_E$  zur tatsächlich benötigten Gesamtbedienrate. Ebenso werden auch die effektiven Bedienraten  $C_{S,k}$  und  $C_{R,k}$  zu periodischen Zeitpunkten  $j \cdot T$ , j = 0, 1, ..., an die tatsächliche Verkehrszusammensetzung angepaßt.

Bild 6.16 zeigt einen Gesamtüberblick über die Funktionsweise und das Zusammenwirken der in der ersten und zweiten Stufe implementierten Algorithmen. Diese Beschreibung basiert auf der Spezifikations- und Beschreibungssprache SDL (Specification and Description Language), die sich zur formalen Spezifikation von vermittlungstechnischen Prozessen eignet [149].

Der für die erste Stufe zuständige Prozeß wickelt die eigentliche Verbindungsannahme ab. Ein neuer Verbindungswunsch wird nur dann angenommen, falls die durch kurz- und langfristige Überlastsituationen verursachten Zellverluste unter den entsprechenden Grenzwerten bleiben. Für die Annahme einer neuen Verbindung müssen daher folgende Annahmekriterien erfüllt sein:

- Damit die durch kurzzeitige Zellkollisionen verursachten Zellverluste unter dem zugeordneten Grenzwert bleiben, darf das Verkehrsangebot einen bestimmten Maximalwert  $\widehat{A}_Z$  nicht überschreiten, der sich aus dem M/D/1/(S+1)-System ergibt. Dies führt zu dem in Gleichung 6.1 definierten Annahmekriterium.
- Die langfristigen Verkehrsschwankungen werden mit Hilfe der effektiven Gesamtbedienrate  $C_E$  kontrolliert, die unter der Zellbedienrate C liegen muß, damit die vereinbarte Verlustwahrscheinlichkeit  $\hat{B}_B$  eingehalten werden kann.

Falls die Verbindung angenommen wird, müssen zu den für die Verkehrsbeschreibung verwendeten Variablen  $N_k$ ,  $A_Z$ ,  $\hat{r}$  und  $C_E$  die entsprechenden Verkehrsparameter der neuen Verbindung hinzuaddiert werden. Beim Verbindungsabbau werden diese Änderungen wieder rückgängig gemacht. Eine Besonderheit ergibt sich dadurch, daß die effektive Gesamtbedienrate unter die Summe der Spitzenbitraten sinken könnte. In diesem Fall wird sie gleich der Summe der Spitzenbitraten gesetzt.

Schließlich findet zu periodischen Zeitpunkten ein Datenaustausch mit dem Hintergrundalgorithmus statt. Dabei übergibt die erste Stufe den zum Zeitpunkt  $j \cdot T$  gültigen Verbindungsvektor N und die dazugehörige effektive Gesamtbedienrate  $C_E$  an die zweite Stufe. In der zweiten Stufe wird mit Hilfe des Faltungsalgorithmus ein genauerer Schätzwert

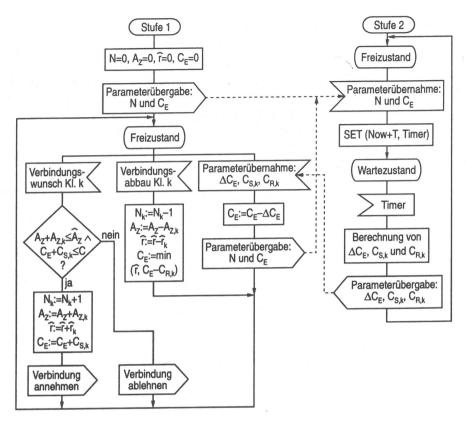


Bild 6.16: SDL-Diagramm des zweistufigen Verbindungsannahmealgorithmus

 $C_E^*$  für die tatsächlich benötigte Gesamtbedienrate ermittelt. Die zuviel zugewiesene Kapazität  $\Delta C_E = C_E - C_E^*$  wird zum folgenden Übergabezeitpunkt  $(j+1)\cdot T$  an die erste Stufe zurückgemeldet, welche die effektive Gesamtbedienrate entsprechend ändert. Ebenso paßt der Hintergrundalgorithmus die effektiven Bedienraten  $C_{S,k}$  und  $C_{R,k}$  dynamisch an die aktuelle Verkehrszusammensetzung an.

Zwischen dem Aufruf der zweiten Stufe und der Rückgabe der effektiven Bedienraten kann sich der den Berechnungen zugrundeliegende Verbindungszustand ändern, was zu einer Verfälschung der Ergebnisse führt. Um diese Inkonsistenzen zu minimieren, sollte die Zeitdauer T möglichst klein sein, d.h. sie sollte der Rechenzeit des Hintergrundalgorithmus entsprechen. In der Realität wird der Abstand zwischen zwei aufeinanderfolgenden Übergabezeitpunkten deshalb schwanken. Für die folgenden Betrachtungen wird jedoch von einer konstanten Berechnungsdauer ausgegangen, die innerhalb des SDL-Diagramms durch eine entsprechende

Zeitsteuerung symbolisiert wird.

Abschließend soll auf den Hintergrundalgorithmus näher eingegangen werden. Zunächst wird die zur Sicherung einer vorgegebenen Verlustwahrscheinlichkeit  $\hat{B}_B$  benötigte Gesamtbedienrate  $C_E^*$  bestimmt. Dabei wird angenommen, daß diese einem ganzzahligen Vielfachen der Grundrate  $\Delta r$  entspricht. Der in Gleichung 6.5 für die Gesamtverlustwahrscheinlichkeit angegebene Zusammenhang kann wie folgt umgeformt werden, indem der Term  $i-C/\Delta r$  durch die Summe  $\sum_{j=1}^{i-C/\Delta r} 1$  ersetzt wird:

$$B_B(C) = \frac{\Delta r}{\overline{r}} \sum_{i=C/\Delta r+1}^{\widehat{r}/\Delta r} (i - C/\Delta r) \cdot p_R(i)$$

$$B_B(C) = \frac{\Delta r}{\overline{r}} \sum_{i=C/\Delta r+1}^{\widehat{r}/\Delta r} \sum_{j=1}^{i-C/\Delta r} p_R(i)$$

$$B_B(C) = \frac{\Delta r}{\overline{r}} \sum_{j=1}^{\widehat{r}/\Delta r-1} \sum_{i=C/\Delta r+j}^{\widehat{r}/\Delta r} p_R(i)$$

$$B_B(C) = \frac{\Delta r}{\overline{r}} \sum_{j=C/\Delta r}^{\widehat{r}/\Delta r-1} [1 - F_R(j)]$$

$$B_B(C) = B_B(C + \Delta r) + \frac{\Delta r}{\overline{r}} [1 - F_R(C/\Delta r)], \qquad (6.43)$$

wobei  $1 - F_R(j)$  der komplementären Verteilungsfunktion der Ankunftsrate entspricht:

$$1 - F_R(j) = P\{R > j \cdot \Delta r\} = \sum_{i=j+1}^{\widehat{r}/\Delta r} p_R(i) .$$
 (6.44)

Die effektive Gesamtbedienrate kann damit wie folgt bestimmt werden: Im ersten Schritt wird die Verteilung der Ankunftsrate mit Hilfe von Gleichung 6.2 berechnet. Daraus wird die komplementäre Verteilungsfunktion der Ankunftsrate nach Gleichung 6.44 bestimmt. Nun kann Gleichung 6.43 beginnend bei  $C = \hat{r}$  schrittweise ausgewertet werden. Diese Auswertung kann abgebrochen werden, wenn  $B_B(C)$  den Maximalwert  $\hat{B}_B$  übersteigt. Die gesuchte effektive Gesamtbedienrate  $C_E^*$  ergibt sich aus den folgenden Bedingungen:

$$B_B(C_E^*) \le \hat{B}_B \quad \text{und} \quad B_B(C_E^* - \Delta r) > \hat{B}_B.$$
 (6.45)

In den Aufgabenbereich der zweiten Stufe fällt auch die Anpassung der effektiven Bedienraten  $C_{S,k}$  und  $C_{R,k}$ . Der einfachste denkbare Ansatz weist diesen Kenngrößen einen festen, zeitunabhängigen Wert zu. Eine Verletzung der Dienstgüte kann ausgeschlossen werden, falls einer neuen Verbindung die maximale Zellrate zugewiesen wird. Der Hintergrundalgorithmus paßt diese zugewiesene Bedienrate dynamisch an den tatsächlichen Bedarf an. Um die Einhaltung der Netzgüte garantieren zu können, darf deshalb beim Verbindungsabbau nur die

mittlere Zellrate freigegeben werden, die eine untere Grenze für die tatsächlich benötigte Bedienrate darstellt:

$$C_{S,k} = \hat{r}_k \quad \text{und} \quad C_{R,k} = \overline{r}_k .$$
 (6.46)

Da diese Kenngrößen direkt aus den Quellparametern hervorgehen, könnte auf eine Klassifizierung des Verkehrs verzichtet werden.

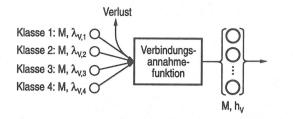
Die Leistungsfähigkeit kann verbessert werden, indem auch die effektiven Bedienraten  $C_{S,k}$  und  $C_{R,k}$  dynamisch an die aktuelle Verkehrszusammensetzung angepaßt werden. Die für einen Belegungszustand  $N=(N_1,N_2,...,N_K)$  benötigte Gesamtbedienrate wird im folgenden mit  $C_E^*(N)$  bezeichnet. Falls alle Verbindungen aus Verkehrsklasse k abgebaut werden, liefert der beschriebene Algorithmus eine reduzierte Bedienrate  $C_E^*(N_1,...,N_{k-1},0,N_{k+1},...,N_K)$ . Die effektive Bedienrate einer Verbindungen aus Verkehrsklasse k bezogen wird. Mit einer ähnlichen Überlegung für den Sonderfall  $N_k=0$  erhält man schließlich

$$C_{S,k} = C_{R,k} = \begin{cases} [C_E^*(N) - C_E^*(N_1, ..., N_{k-1}, 0, N_{k+1}, ..., N_K)]/N_k & \text{für } N_k > 0 \\ C_E^*(N_1, ..., N_{k-1}, 1, N_{k+1}, ..., N_K) - C_E^*(N) & \text{für } N_k = 0 \end{cases}$$
(6.47)

#### 6.4.2 Leistungsuntersuchung

Bei dem beschriebenen Annahmealgorithmus ergibt sich keine eindeutige Verbindungsannahme-Grenzkurve, da die effektive Gesamtbedienrate nicht nur vom momentanen Verbindungszustand, sondern auch von seinem Zustandekommen abhängt. Die Leistungsfähigkeit dieses Algorithmus muß deshalb auf simulativem Wege erfaßt werden. Bild 6.17 zeigt das zugrundeliegende Simulationsmodell, das die Abläufe auf Verbindungsebene nachbildet. Die Verbindungswünsche der oben eingeführten Verkehrsklassen werden durch Poisson-Ankunftsprozesse mit den Ankunftsraten  $\lambda_{V,k}$  (k=1,2,3,4) nachgebildet. Die Entscheidung über die Annahme bzw. Abweisung eines Verbindungswunschs wird von der Verbindungsannahmefunktion getroffen, wobei wiederum eine maximale Gesamtverlustwahrscheinlichkeit von  $10^{-10}$  eingehalten werden soll. Die Verbindungsdauern sind negativ-exponentiell verteilt und weisen für alle Verkehrsklassen denselben Mittelwert h=100 s auf. Dieser Wert entspricht der mittleren Dauer eines heutigen Telefongesprächs. Für Breitbanddienste wird dieser Wert höher liegen, was sich günstig auf die Effizienz des zweistufigen Annahmeverfahrens auswirkt.

Für die folgenden Untersuchungen werden wiederum Verkehrsmischungen aus den Verkehrsklassen 1 und 2 bzw. aus den Verkehrsklassen 3 und 4 ausgewählt. Um die gesamte Verbindungsannahme-Grenzkurve erfassen zu können, werden jeweils fünf verschiedene Verkehrszusammensetzungen untersucht, die sich hinsichtlich des Verkehrsangebots unterscheiden.



M: Markoff-Prozeß

 $\lambda_{V,k}$ : Ankunftsrate von Klasse k

h<sub>v</sub>: Mittlere Verbindungsdauer A<sub>v.k</sub>: Verkehrsangebot von

Klasse k

Bild 6.17: Verkehrsmodell der Abläufe auf Verbindungsebene

Für die Verkehrsklassen 1 und 2 wurden folgende Angebotspaare festgelegt:

$$A_{V,1} = \frac{10 \cdot (5-j)}{3}$$
 Erlang und  $A_{V,2} = \frac{10 \cdot (j-1)}{3}$  Erlang,  $j = 1, 2, ..., 5$ . (6.48)

Für die aus den Verkehrsklassen 3 und 4 bestehenden Verkehrsmischungen gelten analoge Überlegungen:

$$A_{V,3} = 500 \cdot (5-j) \text{ Erlang} \quad \text{ und } \quad A_{V,4} = \frac{500 \cdot (j-1)}{3} \text{ Erlang }, \ \ j=1,2,...,5. \ \ (6.49)$$

In den Bildern 6.18 und 6.19 wird die erste Variante des zweistufigen Algorithmus mit der Spitzenbitratenzuweisung und dem Faltungsalgorithmus verglichen. Der zeitliche Abstand zwischen zwei Parameteranpassungen wurde zu T=1 s gewählt. Für den zweistufigen Algorithmus sind alle Verbindungszustände markiert, bei denen eine Verbindung aus der Verkehrsklasse mit der kleineren Spitzenbitrate abgewiesen wurde. Die Ergebnisse bestätigen,

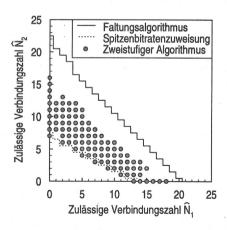


Bild 6.18: Verbindungsannahme-Grenzkurven für die Klassen 1 und 2 (Variante 1, T = 1 s)

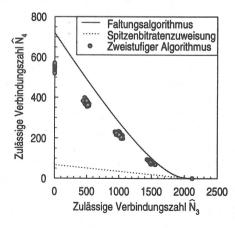


Bild 6.19: Verbindungsannahme-Grenzkurven für die Klassen 3 und 4 (Variante 1, T = 1 s)

daß die Leistungsfähigkeit des zweistufigen Algorithmus zwischen der Spitzenbitratenzuweisung und dem Faltungsalgorithmus liegt und die geforderte Netzgüte eingehalten wird. Andererseits läßt der zweistufige Algorithmus deutlich weniger Verbindungen zu als der in der zweiten Stufe implementierte Faltungsalgorithmus. Dies ist hauptsächlich auf die innerhalb der ersten Stufe angewandte großzügige Kapazitätszuweisung zurückzuführen.

Durch die dynamische Anpassung der innerhalb der ersten Stufe verwendeten effektiven Bedienraten läßt sich die Effizienz des zweistufigen Annahmeverfahrens deutlich verbessern (Bilder 6.20 und 6.21). Dies wird auch daran deutlich, daß der Zeitabstand zwischen zwei Parameteranpassungen gegenüber dem vorigen Beispiel verzehnfacht wurde, um den erhöhten Rechenaufwand dieser Variante zu berücksichtigen. Tabelle 6.4 dokumentiert, daß die ange-

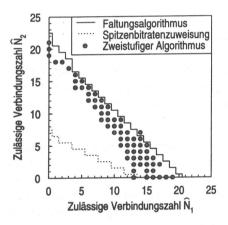


Bild 6.20: Verbindungsannahme-Grenzkurven für die Klassen 1 und 2 (Variante 2, T = 10 s)

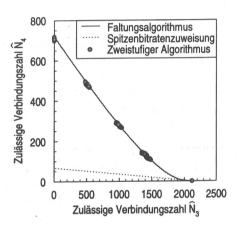


Bild 6.21: Verbindungsannahme-Grenzkurven für die Klassen 3 und 4 (Variante 2, T = 10 s)

	Verkehrsmischung 1		Verkehrsmischung 2	
	Klasse 1	Klasse 2	Klasse 3	Klasse 4
Maximale Zellverlust- wahrscheinlichkeit	$5,8 \cdot 10^{-10}$	$9,8\cdot 10^{-11}$	$1,3\cdot 10^{-10}$	$5, 2 \cdot 10^{-10}$
Zulässiges Verkehrs- angebot	5,4 Erlang	5,4 Erlang	878 Erlang	293 Erlang

Tabelle 6.4: Maximale Zellverlustwahrscheinlichkeit und auf der Verbindungsebene zugelassenes Verkehrsangebot für den zweistufigen Algorithmus (Variante 2,  $T=10\ s$ )

strebte Zellverlustwahrscheinlichkeit nur geringfügig überschritten wird und daß eine dem idealen Algorithmus vergleichbare Netzauslastung erzielt werden kann. Diese Resultate wurden unter denselben Voraussetzungen ermittelt, wie die in den Tabellen 6.2 und 6.3 festgehaltenen Ergebnisse der anderen Annahmealgorithmen.

## 6.5 Festlegung der Verkehrsparameter

Abschließend sollen verschiedene Alternativen zur Festlegung der Verkehrsparameter einer einzelnen Verbindung erläutert und qualitativ bewertet werden. Welcher Ansatz sich letztendlich für die Bestimmung der Verkehrsparameter durchsetzen wird, ist noch nicht abzusehen.

### 6.5.1 Statistische Beschreibung des Verkehrs

Da die meisten Verbindungsannahmefunktionen auf statistischen Verkehrskenngrößen wie beispielsweise der mittleren Zellrate basieren, liegt die Vereinbarung und Überwachung von statistischen Quellparametern nahe. Allerdings weist dieser Ansatz eine Reihe von Nachteilen auf [200, 201]:

- Ein fundamentales Problem stellt die präzise Vorhersage des erwarteten Verkehrsauf-kommens zu Beginn einer Verbindung dar, da der erzeugte Verkehr von einer Vielzahl von Einflußfaktoren abhängt, die in der Regel nicht vorhergesehen werden können. Dieser Unsicherheitsfaktor führt dazu, daß ein weit über dem tatsächlichen Bedarf liegendes Verkehrsaufkommen vereinbart werden muß, um die angestrebte Zellverlustwahrscheinlichkeit im Bereich von 10<sup>-10</sup> einhalten zu können.
- Die Überwachung statistischer Verkehrsparameter führt ebenfalls zu Schwierigkeiten, da für die Dimensionierung der Quellflußkontrollfunktion weitere Angaben benötigt werden. Beispielsweise reicht die mittlere Zellrate nicht für die Dimensionierung der Quellflußkontrollfunktion aus, da sie keinerlei Aussagen in Bezug auf das zeitliche Verhalten der Verkehrsschwankungen zuläßt. Ferner muß die Quellflußkontrollfunktion überdimensioniert werden, um auf eine Verletzung der vereinbarten Verkehrsparameter schnell reagieren zu können.
- Dies führt allerdings dazu, daß mehr Verkehr ins Netz gelangen kann, als ursprünglich vereinbart wurde. Damit eine bestimmte Netzgüte garantiert werden kann, sollte für die Annahme einer neuen Verbindung der in Bezug auf die Netzgüte ungünstigste Verkehr ("worst case traffic") zugrundegelegt werden. Die im folgenden Abschnitt beschriebenen Untersuchungen zeigen allerdings, daß sich keine eindeutige Quellcharakteristik für diesen ungünstigsten Verkehr angeben läßt.

Die genannten Nachteile haben dazu geführt, daß sich die Standardisierungsgremien auf eine algorithmische Beschreibung der Verkehrsparameter geeinigt haben.

## 6.5.2 Algorithmische Beschreibung des Verkehrs

Der in [121] definierte generische Zellratenalgorithmus ermöglicht die gleichzeitige Überwachung der maximalen Zellrate und der im Netzzugangsbereich eingefügten Zellverzugsschwankungen. Derselbe Algorithmus kann auch dazu eingesetzt werden, eine langzeitige Zellrate ("sustainable cell rate") mit einer gewissen Toleranz ("burst tolerance"), welche die statistischen Verkehrs- und Zellverzugsschwankungen berücksichtigt, zu erfassen [7]. Diese Verkehrsparameter können direkt mit Hilfe des generischen Zellratenalgorithmus überwacht werden, womit das angesprochene Dimensionierungsproblem entfällt.

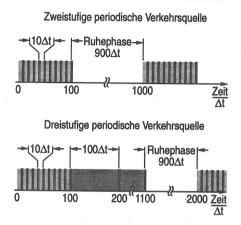
Die Deklaration der genannten Verkehrsparameter gestaltet sich ähnlich schwierig wie die Vereinbarung statistischer Verkehrsparameter, da das erwartete Verkehrsaufkommen zu Beginn einer Verbindung angegeben werden muß. Allerdings erlaubt die algorithmische Beschreibung der Verkehrsparameter eine entsprechende Verkehrsformung, indem der generische Zellratenalgorithmus innerhalb des Endgeräts nachgebildet wird. In diesem Fall kann der dem Netz zugeführte Verkehr spezifiziert und überwacht werden.

Um eine bestimmte Netzgüte garantieren zu können, muß für die Verbindungsannahme der Verkehr zugrundegelegt werden, der die niedrigste Netzgüte, d.h. die höchste Zellverlustwahrscheinlichkeit liefert. Häufig wird davon ausgegangen, daß eine sporadische Quelle mit fester Büschelgröße und konstanter Ruhedauer die höchste Zellverlustwahrscheinlichkeit verursacht, falls die maximale und die mittlere Zellrate fest vorgegeben wird. Neuere Untersuchungen stehen im Widerspruch zu dieser Annahme [48, 255].

In [255] wurde die komplementäre Verteilung des Pufferfüllstands für die in Bild 6.22 gezeigten periodischen Verkehrsquellen verglichen. Beide Verkehrsquellen weisen eine mittlere Zellrate von  $1/(100\cdot\Delta t)$  und eine maximale Zellrate von  $1/(10\cdot\Delta t)$  auf. Ebenso sind die oben eingeführten algorithmischen Verkehrsparameter für beide Quellen gleich groß. Die in [255] beschriebenen Ergebnisse zeigen, daß die dreistufige Quelle zu einem größeren Pufferfüllstand führt als die zweistufige Quelle. In [253] wurde gezeigt, daß sich diese Verhältnisse für die Zellverlustwahrscheinlichkeit umkehren können, d.h. die Zellverlustwahrscheinlichkeit der zweistufigen Quelle ist höher als die der dreistufigen Quelle.

Bild 6.23 zeigt die für die Überlagerung von 99 Verkehrsquellen erhaltenen Zellverlustwahrscheinlichkeiten, falls alle möglichen Phasenlagen der Verkehrsquellen mit den dazugehörigen Auftrittswahrscheinlichkeiten gewichtet werden. Die Ergebnisse wurden mit Hilfe einer Systemsimulation ermittelt und gelten für die Bearbeitungsreihenfolge AF. Die angegebenen Kurven bekräftigen die in Kapitel 4 ausführlich diskutierten Blockierungseffekte auf der Zell-

und Büschelebene. Für kleine Puffergrößen (S<10) ist ein steiler Anstieg der Zellverlustwahrscheinlichkeit zu erkennen, der auf die asynchronen Zellankünfte zurückzuführen ist. Für größere Puffergrößen werden die Zellverluste hauptsächlich durch die auf der Büschelebene auftretenden Überlastsituationen verursacht.



Zweistufige Quelle

Toelstufige Quelle

Toelstufige Quelle

Dreistufige Quelle

Puffergröße S

Bild 6.22: Periodische Verkehrsquellen mit gleicher maximaler und mittlerer Zellrate

Bild 6.23: Zellverlustwahrscheinlichkeiten periodischer Verkehrsquellen

Ferner ist aus diesen Ergebnissen zu ersehen, daß die Zellverlustwahrscheinlichkeit der zweistufigen Verkehrsquellen für kleine Puffergrößen über der Zellverlustwahrscheinlichkeit der dreistufigen Verkehrsquellen liegt, während sich die Verhältnisse für große Puffergrößen umkehren. Dies deutet an, daß der ungünstigste Verkehr von der Größe des Puffers abhängt. Zudem zeigen die in [48] beschriebenen Fallstudien, daß der ungünstigste Verkehr auch vom Hintergrundverkehr beeinflußt wird. Beispielsweise kann bei heterogener Verkehrszusammensetzung der Fall auftreten, daß eine Quelle mit konstanter Zellrate zu einer höheren Zellverlustwahrscheinlichkeit führt als eine Quelle mit variabler Zellrate und gleichem Mittelwert. Dies bedeutet, daß der ungünstigste Verkehr von sehr vielen Faktoren abhängt. Allerdings unterscheiden sich die Zellverlustwahrscheinlichkeiten des ungünstigsten Verkehrs und der zweistufigen periodischen Verkehrsquelle nur geringfügig, so daß der von dieser Quelle generierte Verkehr als Entscheidungsgrundlage für die Verbindungsannahme dienen kann [3].

#### 6.5.3 Messung des Verkehrs

Andere Ansätze bestimmen die der Verbindungsannahme zugrundeliegenden Verkehrsparameter über Verkehrsmessungen. Dies hat den Vorteil, daß nur die Spitzenzellrate einer

Verbindung vereinbart und überwacht werden muß. Andererseits bedeutet die Messung des Verkehrs einen nicht unerheblichen Zusatzaufwand, da diese Messungen in Echtzeit und getrennt für jeden Verbindungsabschnitt durchgeführt werden müssen. Ein weiteres Problem betrifft die Genauigkeit der Messung, da Meßfehler zu einer Verletzung der vereinbarten Netzgüte führen können. Die Genauigkeit der Messung reduziert sich mit zunehmender Korrelation des Verkehrs und abnehmender Meßdauer (vgl. auch Unterabschnitt 3.3).

Ein konkreter Verbindungsannahmelgorithmus, der sich auf Verkehrsmessungen abstützt, ist in [214] beschrieben. Bei diesem Ansatz wird die Anzahl der innerhalb eines Meßintervalls ankommenden Zellen gezählt. Aus den Meßergebnissen von mehreren aufeinanderfolgenden Meßintervallen kann der Zählprozeß, d.h. die Verteilung der Zellankünfte innerhalb des betrachteten Meßintervalls, abgeschätzt werden. Auf dieser Basis kann der in Unterabschnitt 6.2.2.4 beschriebene Verbindungsannahmealgorithmus über die Annahme bzw. Ablehnung einer neuen Verbindung entscheiden.

Weitere Methoden zur Bestimmung der Zellratenverteilung werden in [209] diskutiert. Die verallgemeinerte "Karlsson-Meßmethode" unterteilt die Messung ebenfalls in einzelne Meßintervalle konstanter Dauer, in denen die Anzahl der ankommenden Zellen gezählt wird. Am Ende der Messung wird aus den gemessenen Stichproben die relative Häufigkeit ermittelt, daß die Zellrate innerhalb eines bestimmten Intervalls liegt. Unter der Bezeichnung "logarithmic histogram counting" wird ein weiterer Ansatz vorgestellt, bei dem der zeitliche Abstand zwischen zwei aufeinanderfolgenden Zellen gemessen wird. Die gemessenen Zellabstände werden klassifiziert, wobei der Klasse i, i = 0, 1, ..., alle Zellabstände im Bereich von  $[2^{i-1}, 2^i)$  Zeitschlitzen zugeordnet werden. Am Ende der Messung wird die relative Häufigkeit der diesen Klassen zugeordneten Zellraten bestimmt.

# Kapitel 7

# Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit wurden verschiedene Aspekte des für ATM-Netze vorgesehenen Verkehrsmanagements untersucht. Zunächst wurde eine Einführung in den Asynchronen Transfer-Modus und das darauf aufbauende Breitband-ISDN gegeben. Insbesondere wurde die hierarchische Gesamtkonzeption des Verkehrsmanagements mit den darin eingebetteten Steuerungsfunktionen erläutert. Im Anschluß wurden die den nachfolgenden Leistungsuntersuchungen zugrundeliegenden verkehrstheoretischen Modelle und Methoden beschrieben. Insbesondere wurde auf die hierarchische Quell- und Systemmodellierung eingegangen, bei der zwischen Zell-, Büschel- und Verbindungsebene unterschieden wurde.

Auf dieser Basis wurden dann verschiedene analytische Ansätze zur Leistungsuntersuchung eines statistischen Multiplexers und einer ganzen ATM-Referenzverbindung vorgestellt, mit denen die generellen Zusammenhänge zwischen den Verkehrs- und Leistungskenngrößen systematisch untersucht wurden. Aus den Ergebnissen lassen sich im Hinblick auf die Verkehrssteuerung die folgenden allgemeinen Aussagen ableiten:

- Die auf der Zell- und Büschelebene auftretenden Verkehrsschwankungen führen zu kurz- und langfristigen Überlastsituationen, die mit Hilfe von unterschiedlichen analytischen Modellen erfaßt werden können. Die innerhalb eines kleinen Zeitintervalls ankommenden Zellen lassen sich mit Hilfe eines Poisson-Ankunftsprozesses beschreiben, weshalb die durch kurzzeitige Zellkollisionen verursachten Zellverluste bei einer Vergrößerung des Puffers rasch abklingen. Demgegenüber kann für die längerfristigen, durch das "Fluid Flow"-Modell nachgebildeten Überlastsituationen erst dann eine merkliche Abnahme der Zellverlustwahrscheinlichkeit festgestellt werden, wenn die Puffergröße über der mittleren Büschelgröße der Verkehrsquellen liegt.
- Die Netzpuffer sollten so dimensioniert werden, daß kurzzeitige Zellkollisionen, die sich aus der Asynchronität der Zellankünfte ergeben, ausgeglichen werden können. Eine weitere Vergrößerung des Puffers erscheint dagegen aus Aufwandsgründen nicht sinnvoll.

Darüberhinaus würde dies zu nicht vernachlässigbaren Verkehrsverformungen führen, welche die Komplexität des Verkehrsmanagements erhöhen. Beispielsweise müßten in diesem Fall Bedienprioritäten vorgesehen werden, damit die zeitlichen Anforderungen von Echtzeitdiensten befriedigt werden können. Alle nachfolgenden Aussagen gelten für kleine Puffer, die mit Hilfe des M/D/1/(S+1)-Modells dimensioniert werden können.

- Längerfristige Überlastsituationen, die sich dadurch ergeben, daß die Zellankunftsrate die Bedienrate des Multiplexers überschreitet, müssen durch die Verbindungsannahme (in Verbindung mit den auf der Büschelebene wirkenden Steuerungsfunktionen) begrenzt werden. Diese Funktion darf nur soviele Verbindungen zulassen, daß eine spezifizierte Netzgüte garantiert werden kann.
- Die Zellverlustwahrscheinlichkeit stellt den mit Abstand wichtigsten Netzgüteparameter dar. Die Zellverzugsschwankungen sind bei der gegebenen Pufferdimensionierung vernachlässigbar, während die Korrelation von aufeinanderfolgenden Zellverlusten nahezu unabhängig von der Verbindungszahl ist und deshalb von der Verbindungsannahmefunktion nicht beeinflußt werden kann.
- Das Quellverhalten bleibt innerhalb des Netzes praktisch unverändert, da nur sehr kleine Zellverzugsschwankungen und -verluste auftreten. Damit kann die Verbindungsannahme auf einem beliebigen Verbindungsabschnitt mit den am Netzzugang definierten Quellparametern erfolgen. Eine obere Grenze für die Zellverlustwahrscheinlichkeit ergibt sich aus der Addition der für die einzelnen Verbindungsabschnitte ermittelten Teilergebnisse.

In den folgenden Kapiteln wurden zwei Verkehrssteuerungsfunktionen, die Verlustprioritätsbehandlung und die Verbindungsannahme, ausführlich untersucht. Das Hauptaugenmerk liegt auf der Beschreibung, Klassifizierung und Bewertung unterschiedlicher Ansätze, wobei vor allem Leistungs- und Implementierungsgesichtspunkte einbezogen werden.

Die Untersuchungen belegen die durch Verlustprioritäten erzielten Leistungsvorteile: Einerseits genügen kleinere Puffer zum Ausgleich kurzzeitiger Überlastsituationen, andererseits kann die Auslastung des Netzes bei büschelförmigem Verkehr deutlich gesteigert werden. Darüberhinaus gestatten Verlustprioritäten eine graduelle Einführung des statistischen Multiplexens, indem dieses zunächst auf die niederprioritäre Verkehrsklasse beschränkt wird. Hochprioritäre Verbindungen werden davon nicht beeinflußt, da diesen nach wie vor die Spitzenbitrate zugewiesen wird.

Der qualitative und quantitative Leistungsvergleich der verschiedenen Verlustprioritätsmechanismen erbrachte für das Schwellwert- und das Verdrängungsverfahren die besten Ergebnisse. Die mit diesen Verfahren erzielte Netzauslastung liegt sehr nahe an der theoretischen Auslastungsgrenze. Bei der Auswahl eines dieser beiden Verfahren spielen hauptsächlich

Implementierungsgesichtspunkte eine Rolle. Falls der Pufferzugriff in FIFO-Reihenfolge erfolgt, ergeben sich deutliche Vorteile für das Schwellwertverfahren. Ansonsten macht sich der zusätzliche Implementierungsaufwand des Verdrängungsverfahrens weniger stark bemerkbar, so daß auch dieses Verfahren aufgrund seiner etwas höheren Leistungsfähigkeit attraktiv ist.

Für die verschiedenen aus der Literatur bekannten Verbindungsannahmeverfahren konnte ein enger Zusammenhang zwischen der Leistungsfähigkeit und dem Implementierungsaufwand aufgezeigt werden. Einfache Mechanismen sind entweder ineffizient oder führen zu nicht zu vernachlässigenden Verletzungen der vereinbarten Netzgüte, während genaue Verfahren eine sehr hohe Prozessorleistung erfordern. Das auf der "Large Deviation"-Approximation basierende Verfahren liefert den besten Kompromiß zwischen den angegebenen Anforderungen.

Weitere Verbesserungen konnten mit Hilfe eines neuartigen zweistufigen Verfahrens erzielt werden, das zwei unterschiedliche Ansätze kombiniert. In der ersten Stufe wird ein einfacher linearer Algorithmus eingesetzt, der eine schnelle Entscheidung über die Annahme eines neuen Verbindungswunschs treffen kann. Die zweite Stufe stützt sich auf den Faltungsalgorithmus, der für die langfristige Anpassung der innerhalb der ersten Stufe verwendeten effektiven Bedienraten sorgt. Mit Hilfe von Leistungsuntersuchungen konnte nachgewiesen werden, daß dieses Verfahren die vorgeschriebene Netzgüte sichern kann und eine zufriedenstellende Netzauslastung erzielt.

Die in dieser Arbeit beschriebenen Ergebnisse basieren auf Modellannahmen, deren Gültigkeit mit Hilfe von Verkehrsmessungen und -experimenten nachgewiesen werden muß. Als besonders kritisch erweist sich die Modellierung und Parametrisierung des dem Netz zugeführten Verkehrs. Einerseits sind die Bitratenanforderungen zukünftiger Dienste noch nicht absehbar, andererseits bereitet die Vielzahl der Ursachen und Abhängigkeiten große Schwierigkeiten bei der Beschreibung der Verkehrsströme. Gegenwärtig werden im Rahmen von Labor- und Feldversuchen Verkehrsmessungen durchgeführt, um die theoretischen Modelle und Ergebnisse zu validieren, aussagekräftige Verkehrs- und Netzgüteparameter festzulegen und die Eignung der Verkehrssteuerungsverfahren unter realen Lastbedingungen nachzuweisen.

Über diese Verkehrsbeobachtung hinausgehend erscheint die aktive Beeinflussung des Quellverhaltens erfolgversprechend. Beispielsweise kann der Quellverkehr an das beim Verbindungsaufbau vereinbarte Verkehrsaufkommen angepaßt werden. Das unkontrollierte Verwerfen von Zellen durch die Quellflußkontrolle kann vermieden werden, indem bei einer Überschreitung der vereinbarten Verkehrsparameter geeignete Gegenmaßnahmen innerhalb des Endgeräts ergriffen werden. Beispielsweise kann das Verkehrsaufkommen durch die Zwischenspeicherung von Zellen oder die Änderung von Codierparametern gesenkt werden. Der Entwurf und die Untersuchung entsprechender Verfahren erfordert eine ganzheitliche Betrachtung von Quelle und Netz.

# Literaturverzeichnis

Die ursprünglich vom CCITT (Comité Consultatif International Téléphonique et Télégraphique) verabschiedeten Empfehlungen sind bereits unter dem neuen Organisationsnamen ITU-TSS aufgeführt.

- M. ABDAT, M. BELLANGER: Towards full communication through multimedia personal terminals. European Transactions on Telecommunications and Related Technologies, vol. 3, no. 6, November-December 1992, pp. 547-554.
- [2] M. AKIYAMA: Variable communication network design. Proceedings of the 9th ITC, Torremolinos, October 1979, session no. 61.
- [3] J. Andrade: ATM source traffic descriptor based on the peak, mean and second moment of the cell rate. Proceedings of the 14th ITC, Antibes Juan-les-Pins, June 1994, pp. 223-232.
- [4] D. ANICK, D. MITRA, M.M. SONDHI: Stochastic theory of a data-handling system with multiple sources. The Bell System Technical Journal, vol. 61, no. 8, October 1982, pp. 1871-1894.
- [5] J. APPLETON: Modelling a connection acceptance control strategy for Asynchronous Transfer Mode networks. Proceedings of the 7th ITC Specialist Seminar, Morristown, October 1990, paper 5.1.
- [6] H. ARMBRÜSTER, H.J. ROTHAMEL: Breitbandanwendungen und -dienste: Qualitative und quantitative Anforderungen an künftige Netze. Nachrichtentechnische Zeitschrift, Band 43, Heft 3, März 1990, S. 150–159.
- [7] ATM FORUM: ATM user-network interface specification, The ATM Forum, version 3.0, September 1993.
- [8] F. AUTHENRIETH: Untersuchung verschiedener Realisierungsmöglichkeiten für Verlustprioritätsmechanismen in ATM-Netzen, Semesterarbeit Nr. 1078, IND, Universität Stuttgart, Stuttgart, 1991.
- [9] G.A. AWATER, F.C. SCHOUTE: Optimal queueing policies for fast packet switching on mixed traffic. IEEE JSAC, vol. 9, no. 3, April 1991, pp. 458-467.
- [10] J.J. BAE, T. SUDA: Survey of traffic control schemes and protocols in ATM networks. Proceedings of the IEEE, vol. 79, no. 2, February 1991, pp. 170-189.

- [11] J.J. BAE, T. Suda, R. Simha: Analysis of individual packet loss in a finite buffer queue with heterogeneous Markov modulated arrival processes: a study of traffic burstiness and priority packet discarding. *Proceedings of the IEEE INFOCOM '92*, Florence, May 1992, pp. 219–230.
- [12] A. BAIOCCHI, N.B. MELAZZI, M. LISTANI, A. ROVERI, R. WINKLER: Loss performance analysis of an ATM multiplexer loaded with high-speed ON-OFF sources. IEEE JSAC, vol. 9, no. 3, April 1991, pp. 388-393.
- [13] A. BAIOCCHI, N.B. MELAZZI, A. ROVERI: Buffer dimensioning criteria for an ATM multiplexer loaded with homogeneous on-off sources. *Proceedings of the 13th ITC*, vol. on queueing, performance and control in ATM, Copenhagen, June 1991, pp. 13–18.
- [14] A. BAIOCCHI, N.B. MELAZZI, A. ROVERI, F. SALVATORE: Stochastic fluid analysis of an ATM multiplexer loaded with heterogeneous ON-OFF sources: an effective computational approach. *Proceedings of the IEEE INFOCOM '92*, Florence, May 1992, pp. 405–414.
- [15] K. Bala, I. Cidon, K. Shoraby: Congestion control for high speed packet switched networks. *Proceedings of the IEEE INFOCOM '90*, San Francisco, June 1990, pp. 520– 526.
- [16] F. BASKETT, K.M. CHANDY, R.R. MUNTZ, F. PALACOIS: Open, closed and mixed networks of queues with different classes of customers. *Journal of the ACM*, vol. 22, no. 2, April 1975, pp. 248–260.
- [17] BELLCORE: Generic System Requirements in Support of Switched Multi-Megabit Data Service, Technical Reference TR\_TSV\_000772, issue 1, May 1991.
- [18] B. BENSAOU, J. GUIBERT, J.W. ROBERTS: Fluid queueing models for a superposition of on/off sources. Proceedings of the 7th ITC Specialist Seminar, Morristown, October 1990, paper 9.3.
- [19] F. BERNABEI, L. GRATTA, M. LISTANI, A. SARGHINI: Analysis of ON-OFF source shaping for ATM multiplexing. *Proceedings of the IEEE INFOCOM '93*, San Francisco, April 1993, pp. 1330–1336.
- [20] A. BHARGAVA, M.G. HLUCHYI: Frame loss due to buffer overflows in fast packet networks. Proceedings of the IEEE INFOCOM '90, San Francisco, June 1990, pp. 132– 139.
- [21] A. BHARGAVA, P. HUMBLET, M.G. HLUCHYI: Queueing analysis of continuous bit stream transport in packet networks. Proceedings of the IEEE GLOBECOM '89, Dallas, October 1989, pp. 903-907.
- [22] A. BIOCCA, E. CONTINI, G. FIORETTO, L. PIOVANO: Signalling for B-ISDN communications. European Transactions on Telecommunications and Related Technologies, vol. 2, no. 4, July-August 1991, pp. 381-389.
- [23] C. BISDIKIAN, J.S. LEW, A.N. TANTAWI: The analysis of the generalized D<sup>[X]</sup>/D/1 queue. Proceedings of the Performance '93, Roma, September 1993, pp. 375-379.

- [24]. C. BLONDIA, O. CASALS: Performance analysis of statistical multiplexing of VBR sources. *Performance Evaluation*, vol. 16, nos. 1–3, November 1992, pp. 5–20.
- [25] C. BLONDIA, T.H. THEIMER: A discrete-time model for ATM traffic. RACE 1022 Document PRLB\_123\_0018\_CD\_CC/UST\_123\_0022\_CD\_CC, October 1989.
- [26] R. Bolla, F. Davoli, A. Lombardo, S. Palazzo, D. Panno: Adaptive access control for multiple traffic classes in ATM networks. *Proceedings of the IEEE GLO-BECOM '91*, Phoenix, December 1991, pp. 331-338.
- [27] R. BOLLA, F. DAVOLI, A. LOMBARDO, S. PALAZZO, D. PANNO: Adaptive bandwidth allocation by hierarchical control of multiple ATM traffic classes. *Proceedings of the IEEE INFOCOM '92*, Florence, May 1992, pp. 30-38.
- [28] A.B. BONDI: On the bunching of cell losses in ATM-based networks. Proceedings of the IEEE GLOBECOM '91, Phoenix, December 1991, pp. 444-447.
- [29] F. BONOMI, L. FRATTA, S. MONTAGNA, R. PAGLINO: Priority on cell service and on cell loss in ATM switching. *Proceedings of the 7th ITC Seminar*, Morristown, October 1990, paper 7.2.
- [30] O.J. BOXMA, J.A.C. RESING: Tandem queues with deterministic service times. Technical Report Eindhoven University of Technology, Memorandum COSOR92-33, Eindhoven, August 1992.
- [31] P. BOYER: A congestion control for the ATM. Proceedings of the 7th ITC Specialist Seminar, Morristown, October 1990, paper 4.3.
- [32] P. BOYER, J. BOYER, J.R. LOUVION, L. ROMOEUF: Modelling the ATD transfer technique. ITC Seminar, Lake Como, May 1987.
- [33] P.E. BOYER, F.M. GUILLEMIN, M.J. SERVEL, J.-P. COUDREUSE: Spacing cells protects and enhances utilization of ATM network links. *IEEE Network*, vol. 6, no. 5, September 1992, pp. 38-49.
- [34] P. BOYER, D. TRANCHIER: A reservation principle with applications to the ATM traffic control. Computer Networks and ISDN Systems, vol. 24, no. 4, 1992, pp. 321– 334.
- [35] J. BRENNER, P. LESKY: Mathematik für Ingenieure und Naturwissenschaftler, Band III, Akademische Verlagsgesellschaft, Wiesbaden, 1976.
- [36] J.A. Bucklew: Large Deviation Techniques in Decision, Simulation, and Estimation, John Wiley & Sons, New York/Chichester/Brisbane/Toronto/Singapore, 1990.
- [37] J. Burgin: Dynamic capacity management in the BISDN. International Journal of Digital and Analog Communication Systems, vol. 3, no. 2, April-June 1990, pp. 161– 165.
- [38] M. BUTTÓ, E. CAVALLERO, A. TONIETTI: Effectiveness of the leaky bucket policing mechanism in ATM networks. IEEE JSAC, vol. 9, no. 3, April 1991, pp. 335-342.
- [39] P. CASTELLI, E. CAVALLERO, A. TONIETTI: Policing and call admission problems in ATM networks. *Proceedings of the 13th ITC*, Copenhagen, June 1991, pp. 847-852.

- [40] C. Chamzas, D.L. Duttweiler: Encoding facsimile images for packet-switched networks. *IEEE JSAC*, vol. 7, no. 5, June 1989, pp. 857–864.
- [41] J.-F. CHANG, T.-H. YANG: The effect of prioritization on the behaviour of a finite queue with batch Poisson inputs and a synchronous server. *Computer Networks and ISDN Systems*, vol. 21, no. 1, 1991, pp. 1-16.
- [42] H.J. Chao: A novel architecture for queue management in the ATM network. IEEE JSAC, vol. 9, no. 7, September 1991, pp. 1110-1118.
- [43] I. CIDON, M. SIDI: Performance analysis of asynchronous transfer mode (ATM) systems. IBM Research Report RC 14395, T.J. Watson Research Center, Yorktown Heights, 1989.
- [44] A. CISNEROS, C.A. BRACKETT: A large ATM switch based on memory switches and optical star couplers. *IEEE JSAC*, vol. 9, no. 8, October 1991, pp. 1348–1360.
- [45] R.B. COOPER: Introduction to Queueing Theory. MacMillan, New York/London, 1972.
- [46] J.N. DAIGLE, J.D. LANGFORD: Models for analysis of packet voice communications systems. *IEEE JSAC*, vol. 4, no. 9, September 1986, pp. 847–855.
- [47] B.T. Doshi: Performance of in-call buffer/window allocation schemes for short intermittent file transfers over broadband packet networks. *International Journal of Digital and Analog Communication Systems*, vol. 5, no. 1, January-March 1992, pp. 49-61.
- [48] B.T. Doshi: Deterministic rule based traffic descriptors for broadband ISDN: worst case behavior and connection acceptance control. *Proceedings of the 14th ITC*, Antibes Juan-les-Pins, June 1994, pp. 591-600.
- [49] B.T. Doshi, S. Dravida: Congestion control for bursty data in high speed wide area packet networks: in-call parameter negotiations. Proceedings of the 7th ITC Specialist Seminar, Morristown, October 1990, paper 12.4.
- [50] L.G. DRON, G. RAMAMURTHY, B. SENGUPTA: Delay analysis of continuous bit rate traffic over an ATM network. *IEEE JSAC*, vol. 9, no. 3, April 1991, pp. 402–407.
- [51] Z. DZIONG, J. CHOQUETTE, K.-Q. LIAO, L. MASON: Admission control and routing in ATM networks. Computer Networks and ISDN Systems, vol. 20, nos. 1-5, 1990, pp. 189-196.
- [52] M. EBERSPÄCHER: Mathematische Analyse der Durchlaufzeitverteilung in ATM-Netzen, Diplomarbeit Nr. 1066, IND, Universität Stuttgart, Stuttgart, 1991.
- [53] A.E. ECKBERG: The single server queue with periodic arrival process and deterministic service. *IEEE Transactions on Communications*, vol. 27, no. 3, March 1979, pp. 556– 562.
- [54] A.E. ECKBERG: B-ISDN/ATM traffic and congestion control. IEEE Network, vol. 6, no. 5, September 1992, pp. 28-37.

- [55] A.E. ECKBERG, B.T. DOSHI, R. ZOCCOLILLO: Controlling congestion in B-ISDN/ATM: issues and strategies. *IEEE Communications Magazine*, vol. 29, no. 9, September 1991, pp. 64-70.
- [56] F. EISENHARDT: Untersuchung von Verbindungsannahmestrategien in ATM-Netzen, Semesterarbeit Nr. 991, IND, Universität Stuttgart, Stuttgart, 1990.
- [57] A.I. ELWALID, D. MITRA: Fluid models for the analysis and design of statistical multiplexing with loss priorities on multiple classes of bursty traffic. Proceedings of the IEEE INFOCOM '92, Florence, May 1992, pp. 415-425.
- [58] A.I. ELWALID, D. MITRA: Effective bandwidth of general Markovian traffic sources and admission control of high speed networks. IEEE/ACM Transactions on Networking, vol. 1, no. 3, June 1993, pp. 329-343.
- [59] H. ESAKI, K. IWAMURA, T. KODAMA: A simple and effective admission control for an ATM network. Proceedings of the IEEE GLOBECOM '90, San Diego, December 1990, pp. 28-33.
- [60] W. FELLER: An Introduction to Probability Theory and its Applications, vol. I, Wiley, New York, 1966.
- [61] J. FILIPIAK: Accuracy of traffic modelling in fast packet switching. Proceedings of the IEEE GLOBECOM '88, Hollywood, November 1988, pp. 1623-1627.
- [62] J. FILIPIAK: Structured systems analysis methodology for design of an ATM network architecture. *IEEE JSAC*, vol. 7, no. 8, October 1989, pp. 1263-1273.
- [63] J. FINCKH: Untersuchungen zur Eignung von rekurrenten neuronalen Netzen für das Bitratenmanagement in ATM-Systemen, Semesterarbeit Nr. 1278, IND, Universität Stuttgart, Stuttgart, 1994.
- [64] L. FRATTA, L. MUSUMECI, G. GALLASSI, L. VERRI: Congestion control strategies in ATM networks. European Transactions on Telecommunications and Related Technologies, vol. 3, no. 2, March-April 1992, pp. 183-193.
- [65] G. GALLASSI, G. RIGOLIO, L. FRATTA: ATM: bandwidth assignment and bandwidth enforcement policies. Proceedings of the IEEE GLOBECOM '89, Dallas, November 1989, pp. 1788–1793.
- [66] G. GALLASSI, G. RIGOLIO, L. FRATTA: Bandwidth assignment in prioritized ATM networks. Proceedings of the IEEE GLOBECOM '90, San Diego, December 1990, pp. 852–856.
- [67] G. GALLASSI, G. RIGOLIO, L. VERRI: Resource management and dimensioning in ATM networks. IEEE Network, vol. 4, no. 3, May 1990, pp. 8-17.
- [68] J. GARCIA, O. CASALS: Statistical multiplexing gain using space priority mechanisms. Proceedings of the IEEE GLOBECOM '91, Phoenix, December 1991, pp. 929-933.
- [69] J. GARCIA, O. CASALS: Stochastic models of space priority mechanisms with Markovian arrival processes. Annals of Operations Research, vol. 35, nos. 1-4, April 1992, pp. 271-296.

- [70] M.W. GARRETT, M. VETTERLI: Joint source/channel coding of statistically multiplexed real-time services on packet networks. *IEEE/ACM Transactions on Networking*, vol. 1, no. 1, February 1993, pp. 71–80.
- [71] M. GERLA, T.-Y. TAI, G. GALLASSI: Internetting LAN's and MAN's to B-ISDN's for connectionless traffic support. *IEEE JSAC*, vol. 11, no. 8, October 1993, pp. 1145– 1159.
- [72] M. GHANBARI: Two-layer coding of video signals for VBR networks. IEEE JSAC, vol. 7, no. 5, June 1990, pp. 771-781.
- [73] P. GONET, P. ADAM, J.P. COUDREUSE: Asynchronous time-division switching: the way to flexible broadband communication networks. *Proceedings of the 1986 International Zurich Seminar on Digital Communications*, Zurich, March 1986, pp. 141-148.
- [74] C.W.B. GOODE: Broadband services and applications. Electrical Communication, vol. 64, no. 2/3, March 1990, pp. 124-131.
- [75] A. GRAVEY, G. HÉBUTERNE: Mixing time and loss priorities in a single server queue. Proceedings of the 13th ITC, Copenhagen, June 1991, pp. 141-152.
- [76] A. GRAVEY, G. HÉBUTERNE: Simultaneity in discrete-time single server queues with Bernoulli inputs. *Performance Evaluation*, vol. 14, no. 2, January 1992, pp. 123–131.
- [77] T.R. GRIFFITHS: Analysis of a connection acceptance strategy for Asynchronous Transfer Mode networks. Proceedings of the IEEE GLOBECOM '90, San Diego, December 1990, pp. 862–868.
- [78] D. GROSS, C.M. HARRIS: Fundamentals of Queueing Theory. J. Wiley and Sons, New York/London, 2nd Edition, 1985.
- [79] R. GRÜNENFELDER: Statistical modelling of the cell queueing delay in an ATM node. Proceedings of the ICCC '90, New Delhi, November 1990, pp. 284-288.
- [80] R. GRÜNENFELDER: A correlation based end-to-end cell queueing delay characterization in an ATM network. Proceedings of the 13th ITC, vol. on queueing, performance and control in ATM, Copenhagen, June 1991, pp. 59-64.
- [81] R. GRÜNENFELDER: Measurement techniques and equipment for B-ISDN. Proceedings of the IBCN&S, Copenhagen, April 1993, pp. 21.3-1-21.3-9.
- [82] R. GRÜNENFELDER, J.P. COSMAS, S. MANTHORPE, A. ODINMA-OKAFOR: Characterization of video codecs as autoregressive moving average process and related queueing system performance. *IEEE JSAC*, vol. 9, no. 3, April 1991, pp. 284–293.
- [83] R. GUÉRIN, H. AHMADI, M. NAGHSHINEH: Equivalent capacity and its applications to bandwidth allocation in high-speed networks. *IEEE JSAC*, vol. 9, no. 7, September 1991, pp. 968–981.
- [84] R. GUÉRIN, L. GÜN: A unified approach to bandwidth allocation and access control in fast packet-switched networks. Proceedings of the IEEE INFOCOM '92, Florence, May 1992, pp. 1-12.

- [85] F. GUILLEMIN, P. BOYER, A. DUPUIS, L. ROMOEUF: Peak rate enforcement in ATM networks. *Proceedings of the IEEE INFOCOM '92*, Florence, May 1992, pp. 753-758.
- [86] S. GUPTA, K.W. Ross, M. EL ZARKI: On routing in ATM networks. Proceedings on Modelling and Performance Evaluation of ATM Technology, La Martinique, January 1993.
- [87] R. HÄNDEL, M.N. HUBER: Integrated Broadband Networks, Addison-Wesley, Wokingham, 1991.
- [88] G. Hebuterne, A. Gravey: A space priority queueing mechanism for multiplexing ATM channels. Computer Networks and ISDN Systems, vol. 20, nos. 1-5, 1990, pp. 37-43.
- [89] R. HECHT-NIELSEN: Neurocomputing, Addison-Wesley, 1990.
- [90] H. HEEKE: Statistical multiplexing gain for variable bit rate video codecs in ATM networks. *International Journal of Digital and Analog Communication Systems*, vol. 4, no. 4, October-December 1991, pp. 261-268.
- [91] H. HEFFES, D.M. LUCANTONI: A markov modulated characterization of packetized voice and data traffic and related statistical multiplexer performance. *IEEE JSAC*, vol. 4, no. 6, September 1986, pp. 856-868.
- [92] H. HEISS: The impact of jitter on peak cell rate policing. Proceedings of the 2nd RACE Workshop on Traffic and Performance Aspects in IBCN, Aveiro, January 1992.
- [93] W.A. Held: Untersuchung von Prioritätsmechanismen für ATM-Netze, Diplomarbeit Nr. 964, IND, Universität Stuttgart, Stuttgart, 1989.
- [94] H. HEMMER, T.P. HUTH: Evaluation of policing functions in ATM networks. Proceedings of the 13th ITC, vol. on queueing, performance and control in ATM, Copenhagen, June 1991, pp. 111-116.
- [95] A. HIRAMATSU: ATM communications network control by neural network. IEEE Transactions on Neural Networks, vol. 1, no. 1, March 1990, pp. 122-130.
- [96] A. HIRAMATSU: Integration of ATM call admission control and link capacity control by distributed neural networks. *IEEE JSAC*, vol. 9, no. 7, September 1991, pp. 1131– 1138.
- [97] A. HUANG, S. KNAUER: STARLITE: a wideband digital switch. Proceedings of the IEEE GLOBECOM '84, Atlanta, November 1984, pp. 121-125.
- [98] M.N. HUBER: Ein Netzknotenkonzept für integrierte Durchschalte- und Paketvermittlung, Dissertationsschrift, Universität Stuttgart, Stuttgart, 1990.
- [99] M.N. HUBER, V. FRANTZEN, G. MAEGERL: Proposed evolutionary paths for B-ISDN signalling. Proceedings of the XIV ISS, Yokohama, October 1992, pp. 334-338.
- [100] F. HÜBNER: Analysis of a finite-capacity asynchronous multiplexer with deterministic traffic sources. *Proceedings of the 7th ITC Specialist Seminar*, Morristown, October 1990, paper 11.2.

- [101] F. HÜBNER, P. TRAN-GIA: Performance of a finite capacity asynchronous multiplexer with modulated input. *Telecommunication Systems*, vol. 1, no. 3, February 1993, pp. 263-278.
- [102] J.Y. Hui: Resource allocation for broadband networks. IEEE JSAC, vol. 6, no. 9, December 1988, pp. 1598–1608.
- [103] ISO, ISO 7498: Information processing systems open systems interconnection basic reference model. American National Standards Association, New York, 1984.
- [104] ITU-TSS, RECOMMENDATION E.800: Terms and definitions related to the quality of telecommunication services. Blue Book, Fascicle II.3, International Telecommunication Union, Geneva, 1989.
- [105] ITU-TSS, RECOMMENDATION G.114: Mean one-way propagation time. Blue Book, Fascicle III.1, International Telecommunication Union, Geneva, 1989.
- [106] ITU-TSS, RECOMMENDATION G.711: Pulse code modulation (PCM) of voice frequencies. Blue Book, Fascicle III.4, International Telecommunication Union, Geneva, 1989.
- [107] ITU-TSS, RECOMMENDATION G.721: 32 kbit/s adaptive differential pulse code modulation (ADPCM). Blue Book, Fascicle III.1, International Telecommunication Union, Geneva, 1989.
- [108] ITU-TSS, RECOMMENDATION G.821: Error performance of an international digital connection forming part of an integrated services digital network. Blue Book, Fascicle III.5, International Telecommunication Union, Geneva, 1989.
- [109] ITU-TSS, RECOMMENDATION H.261: Codec for audiovisual services at n×384 kbit/s. Blue Book, Fascicle III.6, International Telecommunication Union, Geneva, 1989.
- [110] ITU-TSS, RECOMMENDATION I.121: Broadband Aspects of ISDN, International Telecommunication Union, Geneva, 1991.
- [111] ITU-TSS, RECOMMENDATION I.140: Attribute technique for the characterization of telecommunication services supported by an ISDN and network capabilities of an ISDN. Blue Book, Fascicle III.7, International Telecommunication Union, Geneva, 1989.
- [112] ITU-TSS, RECOMMENDATION I.150: B-ISDN Asynchronous Transfer Mode Functional Characteristics, International Telecommunication Union, Geneva, 1992.
- [113] ITU-TSS, RECOMMENDATION I.211: B-ISDN Service Aspects, International Telecommunication Union, Geneva, 1991.
- [114] ITU-TSS, RECOMMENDATION I.311: B-ISDN General Network Aspects, International Telecommunication Union, Geneva, 1992.
- [115] ITU-TSS, RECOMMENDATION I.321: B-ISDN Protocol Reference Model and its Applications, International Telecommunication Union, Geneva, 1991.

- [116] ITU-TSS, RECOMMENDATION I.352: Network performance objectives for connection processing delays in an ISDN. Blue Book, Fascicle III.8, International Telecommunication Union, Geneva, 1989.
- [117] ITU-TSS, RECOMMENDATION I.356: B-ISDN ATM Layer Cell Transfer Performance, International Telecommunication Union, Geneva, March 1993.
- [118] ITU-TSS, RECOMMENDATION I.361: B-ISDN ATM Layer Specification, International Telecommunication Union, Geneva, 1992.
- [119] ITU-TSS, RECOMMENDATION I.362: B-ISDN ATM Adaptation Layer (AAL) Functional Description, International Telecommunication Union, Geneva, 1992.
- [120] ITU-TSS, RECOMMENDATION I.363: B-ISDN ATM Adaptation Layer (AAL) Specification, Temporary Document 60, International Telecommunication Union, Geneva, 1993.
- [121] ITU-TSS, RECOMMENDATION I.371: Traffic Control and Congestion Control in B-ISDN, International Telecommunication Union, Geneva, 1992.
- [122] ITU-TSS, RECOMMENDATION I.432: B-ISDN User-network Interface Physical Layer Specification, International Telecommunication Union, Geneva, 1991.
- [123] ITU-TSS, DRAFT RECOMMENDATION Q.SAAL1: Service Specific Connection Oriented Protocol (SSCOP) Specification, International Telecommunication Union, Geneva, 1993.
- [124] ITU-TSS, STUDY GROUP XVIII REPORT R 56: Report of Working Party XVIII/8 — General B-ISDN Aspects, International Telecommunication Union, Matsuyama, 1990.
- [125] V.B. IVERSEN, Y. LIU: The performance of convolution algorithms for evaluating the total load in an ISDN system. 9th Nordic Teletraffic Seminar, Kjeller, August 1990.
- [126] B. Jabbari: A connection control strategy for bursty sources in broadband packet networks. *International Journal of Digital and Analog Communication Systems*, vol. 3, no. 4, October-December 1990, pp. 351-356.
- [127] B. Jabbari, F. Yegenoglu: An efficient method for computing cell loss probability for heterogeneous bursty traffic in ATM networks. *International Journal of Digital and Analog Communication Systems*, vol. 5, no. 1, January-March 1992, pp. 39-48.
- [128] J.T. JOHNSON: Coping with public frame relay: a delicate balance. Data Communications, vol. 21, no. 2, January 1992, pp. 31-38.
- [129] P. JOOS, W. VERBIEST: A statistical bandwidth allocation and usage monitoring algorithm for ATM networks. Proceedings of the ICC '89, Boston, June 1990, pp. 415– 422.
- [130] B. KALTENMORGEN, R. KLEINEWILLINGHÖFER-KOPP: Verkehrskontrolle in ATM-Netzen. Der Fernmelde-Ingenieur, 46. Jahrgang, Heft 1, January 1992, pp. 1-33.

- [131] T. KAMITAKE, T. SUDA: Evaluation of an admission control scheme for an ATM network considering fluctuations in cell loss rate. Proceedings of the IEEE GLOBECOM '89, Dallas, November 1989, pp. 1774-1780.
- [132] M.J. KAROL, G. HLUCHYI: Using a packet switch for circuit-switched traffic: a queueing system with periodic input traffic. *Proceedings of the ICC '87*, Seattle, June 1987, pp. 1677-1682.
- [133] J.S. KAUFMAN: Blocking in a shared resource environment. IEEE Transactions on Communications, vol. 29, no. 10, October 1981, pp. 1474-1481.
- [134] M. KAWARASAKI, B. JABBARI: B-ISDN architecture and protocol. *IEEE JSAC*, vol. 9, no. 9, December 1991, pp. 1405–1415.
- [135] L.P. KHADJIIVANOV, B.T. TASKOV, A.A. ALIAZIDI, B.P. TSANKOV: Application of priority queueing mechanisms to ATM multiplexing and traffic control. *Proceedings* of the IBCN&S, Copenhagen, April 1993, pp. 33.3-1-33.3-11.
- [136] T. KIRSTE: Leistungsuntersuchung zweistufiger Verbindungsannahmestrategien für ATM-Netze, Semesterarbeit Nr. 1137, IND, Universität Stuttgart, Stuttgart, 1992.
- [137] R. KLEINEWILLINGHÖFER-KOPP, B. KALTENMORGEN: Connection acceptance control in ATM networks. Technical Report DBP Telekom, Darmstadt, June 1991.
- [138] R. KLEINEWILLINGHÖFER-KOPP, E. WOLLNER: Comparison of access control strategies for ISDN-traffic on common trunk groups. Proceedings of the 12th ITC, Torino, June 1988, pp. 873–879.
- [139] L. KLEINROCK: Queueing Systems, vol. 1: Theory, John Wiley & Sons, New York/Chichester/Brisbane/Toronto, 1975.
- [140] L. KOSTEN: Stochastic theory of data handling systems with groups of multiple sources. Proceedings on the 2nd International Symposium on the Performance of Computer-Communication Systems, Zurich, March 1984, pp. 321-331.
- [141] R. KOXHOLT: Die Rolle von ISDN in der Bürokommunikation. Datacom, 5. Jahrgang, Heft 7, Juli 1988, S. 80–88.
- [142] H. KRÖNER: Prioritäten in einem Netz auf ATM-Basis. DFG-Bericht, Stuttgart, April 1989.
- [143] H. KRÖNER, T. RENGER, R. KNOBLING: Performance modelling of an adaptive CAC strategy for ATM networks. Proceedings of the 14th ITC, Antibes Juan-les-Pins, June 1994, pp. 1077-1088.
- [144] H. KRÖNER, T.H. THEIMER, U. BRIEM: Queueing systems for ATM systems a comparison. Proceedings of the 7th ITC Specialist Seminar, Morristown, October 1990, paper 9.1.
- [145] P.J. KÜHN: Approximate analysis of general queueing networks by decomposition. *IEEE Transactions on Communications*, vol. 27, no. 1, January 1979, pp. 113-126.
- [146] P.J. KÜHN: Input process model GMDP. Race Document UST\_123\_003\_CD\_CC, June 1988.

- [147] P.J. KÜHN: Codierung zur Fehlererkennung und Fehlerkorrektur, Vorlesungsskript, IND, Universität Stuttgart, Stuttgart, 1993.
- [148] P.J. KÜHN: Datenverarbeitung II, Vorlesungsskript, IND, Universität Stuttgart, Stuttgart, 1993.
- [149] P.J. KÜHN: Nachrichtenvermittlung I, Vorlesungsskript, IND, Universität Stuttgart, Stuttgart, 1993.
- [150] P.J. KÜHN: Wartezeitprobleme der Daten- und Nachrichtenverkehrstheorie, Vorlesungsskript, IND, Universität Stuttgart, Stuttgart, 1993.
- [151] P.J. KÜHN, T. RAITH, P. TRAN-GIA: Methodik der stationären Systemsimulation (in [150]).
- [152] J.J. KULZER, W.A. MONTGOMERY: Statistical switching architectures for future services. *Proceedings of the XI ISS*, Florence, May 1984, paper 43-A-1.
- [153] K. KVOLS, S. BLAABJERG: Bounds and approximations for the periodic on/off queue with applications to ATM traffic control. Proceedings of the IEEE INFOCOM '92, Florence, May 1992, pp. 487-494.
- [154] W.-C. LAU, S.-Q. LI: Traffic analysis in large-scale high-speed integrated networks: validation of nodal decomposition approach. Proceedings of the IEEE INFOCOM '93, San Francisco, March 1993, pp. 1320-1329.
- [155] J.-Y. LE BOUDEC: An efficient solution method for Markov models of ATM links with loss priorities. IEEE JSAC, vol. 9, no. 3, April 1991, pp. 408-417.
- [156] S.-Q. LI: Study of information loss in packet voice systems. IEEE Transactions on Communications, vol. 37, no. 11, November 1989, pp. 1192-1202.
- [157] K.-Q. LIAO, L.G. MASON: A heuristic approach for performance analysis of ATM systems. *Proceedings of the IEEE GLOBECOM '90*, San Diego, December 1990, pp. 1931–1935.
- [158] W.P. LIDINSKY: Data communications needs. IEEE Network, vol. 4, no. 2, March 1990, pp. 28-33.
- [159] A.Y.M. LIN, J.A. SILVESTER: Priority queueing strategies and buffer allocation protocols for traffic control at an ATM integrated broadband switching system. IEEE JSAC, vol. 9, no. 9, December 1991, pp. 1524-1536.
- [160] K. LINDBERGER: Analytical methods for the traffical problems with statistical multiplexing in ATM-networks. Proceedings of the 13th ITC, Copenhagen, June 1991, pp. 807-813.
- [161] M. LOGOTHETIS, S. SHIODA: Centralized virtual path bandwidth allocation scheme for ATM networks. IEICE Transactions on Communications, vol. E75-B, no. 10, October 1992, pp. 1071-1080.
- [162] D.M. LUCANTONI, S.P. PAREKH: Selective cell discard mechanisms for a B-ISDN congestion control architecture. Proceedings of the 7th ITC Seminar, Morristown, October 1990, paper 10.3.

- [163] M. LUONI: ATM traffic characterization. Proceedings of the 2nd RACE Workshop on Traffic and Performance Aspects in IBCN, Aveiro, January 1992.
- [164] B. MAGLARIS, D. ANASTASSIOU, P. SEN, G. KARLSSON, J.D. ROBBINS: Performance models of statistical multiplexing in packet video communications. *IEEE Transactions on Communications*, vol. 36, no. 7, July 1988, pp. 834-844.
- [165] B.A. MAKRUCKI: On the performance of submitting excess traffic to ATM networks. Proceedings of the IEEE GLOBECOM '91, Phoenix, December 1991, pp. 281-288.
- [166] K. MASE, S. SHIODA: Real-time network management for ATM networks, Proceedings of the 13th ITC, Copenhagen, June 1991, pp. 133-140.
- [167] D. MITRA: Stochastic theory of a fluid model of producers and consumers coupled by a buffer. Advances in Applied Probability, vol. 20, 1988, pp. 646-676.
- [168] N.M. MITROU, K.P. KONTOVASSILIS: Fast analysis of on-off heterogeneous traffic multiplexing with fluid-flow models and its applications to ATM. Proceedings of the 2nd Race Workshop on Traffic and Performance Aspects in IBCN, Aveiro, January 1992.
- [169] N.M. MITROU, A.K. KOUKOS: An effective-rate enforcement algorithm for ATM traffic and its hardware implementation. *Proceedings of the IBCNES*, Copenhagen, April 1993, pp. 13.2-1-13.2.16.
- [170] N.M. MITROU, D.E. PENDARAKIS: Cell-level statistical multiplexing in ATM networks: analysis, dimensioning and call-acceptance control w.r.t. QOS criteria. Proceedings of the 13th ITC, vol. on queueing, performance and control in ATM, Copenhagen, June 1991, pp. 7-12.
- [171] R.J.T. Morris: Prospects for neural networks in broadband network resource management. Proceedings of the 13th ITC, Copenhagen, June 1991, pp. 335-340.
- [172] K. MURAKAMI, T. TAKASE, K. HAJIKANO, S. HATTORI: Considerations in ATM switching system traffic control. *International Journal of Digital and Analog Commu*nication Systems, vol. 3, no. 2, April-June 1990, pp. 137-142.
- [173] H. MURAKAMI, T. YOKOI, M. TAKA: Considerations on ATM network performance planning. *IEICE Transactions on Communications*, vol. E75-B, no. 7, July 1992, pp. 563-571.
- [174] T. MURASE, H. SUZUKI, S. SATO, T. TAKEUCHI: A call admission control scheme for ATM networks using a simple quality estimate. *IEEE JSAC*, vol. 9, no. 9, December 1991, pp. 1461–1470.
- [175] R. NAGARAJAN, J.F. KUROSE: Approximation techniques for computing packet loss in finite-buffered voice multiplexers. *IEEE JSAC*, vol. 9, no. 3, April 1991, pp. 368-377.
- [176] T. NECKER: Bitratenmanagement in ATM-Systemen mittels rekurrenter neuronaler Netze, Diplomarbeit Nr. 1198, IND, Universität Stuttgart, Stuttgart, 1993.
- [177] A.G. NEUMANN: Simulation der Transferzeit in ATM-Netzen, Diplomarbeit Nr. 1070, IND, Universität Stuttgart, Stuttgart, 1991.

- [178] M.F. Neuts: Matrix-Geometric Solutions in Stochastic Models, The John Hopkins University Press, Baltimore, 1981.
- [179] M.F. Neuts: Structured Stochastic Matrices of M/G/1 Type and their Applications, Marcel Dekker, New York, 1989.
- [180] G. NIESTEGGE: The 'leaky bucket' policing method in the ATM (Asynchronous Transfer Mode) network. International Journal of Digital and Analog Communication Systems, vol. 3, no. 2, April-June 1990, pp. 187-197.
- [181] I. NORROS, J.W. ROBERTS, A. SIMONIAN, J.T. VIRTAMO: The superposition of variable bit rate sources in an ATM multiplexer. *IEEE JSAC*, vol. 9, no. 3, April 1991, pp. 378–387.
- [182] I. NORROS, J.T. VIRTAMO: Importance sampling simulation studies on the discrete time nD/D/1 queueing system. Proceedings of the 8th Nordic Teletraffic Seminar, Otnäs, August 1989.
- [183] I. NORROS, J.T. VIRTAMO: Who looses cells in the case of burst scale congestion. Proceedings of the 13th ITC, Copenhagen, June 1991, pp. 829-833.
- [184] NTG, NTG-EMPFEHLUNG 0903: Nachrichtenverkehrstheorie Begriffe. Nachrichtentechnische Zeitschrift, Band 37, Heft 7, 1984, S. 465-474 und Heft 8, 1984, S. 539-552.
- [185] Y. Ohba, M. Murata, H. Miyahara: Analysis of interdeparture processes for bursty traffic. IEEE JSAC, vol. 9, no. 3, April 1991, pp. 468-476.
- [186] H. OHNISHI, T. OKADA, K. NOGUCHI: Flow control schemes and delay/loss tradeoff in ATM networks. IEEE JSAC, vol. 6, no. 9, December 1988, pp. 1609–1616.
- [187] H. Ohta, T. Kitami: A cell loss recovery method using FEC in ATM networks. IEEE JSAC, vol. 9, no. 9, December 1991, pp. 1471-1483.
- [188] S. Ohta, K.-I. Sato, I. Tokizawa: A dynamically controllable ATM transport network based on the virtual path concept. Proceedings of IEEE GLOBECOM '88, Ft. Lauderdale, November 1988, pp. 1272-1276.
- [189] O. ØSTERBØ: Duration of heavy load states in an ATM network. Proceedings of 13th ITC, vol. on queueing, performance and control in ATM, Copenhagen, June 1991, pp. 91-95.
- [190] P. PANCHA, M.E. ZARKI: A look at the MPEG video coding standard for variable bit rate video transmission. Proceedings of the IEEE INFOCOM '92, Florence, May 1992, pp. 85-94.
- [191] P. PANCHA, M.E. ZARKI: Bandwidth requirements of variable bit rate MPEG sources in ATM networks. *Proceedings of the IEEE INFOCOM '93*, San Francisco, March 1993, pp. 902-909.
- [192] K. PAWLIKOWSKI: Steady-state simulation of queueing processes: a survey of problems and solutions. ACM Computing Surveys, vol. 22, no. 2, June 1990, pp. 123-170.

- [193] F.T. PEDERSEN, J. WENG: The implementation of call set-up. RACE 1022 Document JTAS\_TG-IV'90\_0001\_CD\_CC, April 1990.
- [194] G.H. Petit, E. Desmet: Delay jitter in ATM switching networks: a closed-form expression applicable for planning purposes. Proceedings of the RACE Workshop on ATM Network Planning and Evolution, London, April 1991.
- [195] D.W. Petr, L.A. Dasilva, V.S. Frost: Priority discarding of speech in integrated packet networks. *IEEE JSAC*, vol. 7, no. 5, June 1989, pp. 644-656.
- [196] D.W. Petr, V.S. Frost: Nested threshold cell discarding for ATM overload control: optimization under cell loss constraints. Proceedings of the IEEE INFOCOM '91, Bal Harbour, April 1991, pp. 1403-1412.
- [197] M.N. RANSON, D.R. SPEARS: Applications of Public Gigabit Networks. IEEE Network Magazine, vol. 6, no. 2, March 1992, pp. 30-40.
- [198] C. RASMUSSEN, J.H. SØRENSEN, K.S. KVOLS, S.B. JACOBSEN: Source-independent call acceptance procedures in ATM networks. *IEEE JSAC*, vol. 9, no. 3, April 1991, pp. 351–358.
- [199] E.P. RATHGEB: Modeling and performance comparison of policing mechanisms for ATM networks. *IEEE JSAC*, vol. 9, no. 3, April 1991, pp. 325-334.
- [200] E.P. RATHGEB: Verkehrsflüsse in ATM-Netzen Modellierung und Analyse von Verkehrsquellen und Quellfluβkontrollverfahren, Dissertationsschrift, Universität Stuttgart, Stuttgart, 1991.
- [201] E.P. RATHGEB: Policing of realistic VBR video traffic a case study. Proceedings of the IFIP Workshop on Broadband Communications, Estoril, January 1992, pp. 267-280.
- [202] H. REBMANN: Verbindungsannahmealgorithmen für ATM-Netze Untersuchung von Zell- und Rufverlusten, Diplomarbeit Nr. 1071, IND, Universität Stuttgart, Stuttgart, 1991.
- [203] L.K. REISS, L.F. MERAKOS: Priority shaping of source traffic in ATM B-ISDN. Computer Communications, vol. 16, no. 12, December 1993, pp. 794-797.
- [204] T. RENGER: Verbindungsannahme in ATM-Systemen basierend auf neuronalen Netzen, Diplomarbeit Nr. 1131, IND, Universität Stuttgart, Stuttgart, 1991.
- [205] G. RIGOLIO, L. FRATTA: Input rate regulation and bandwidth assignment in ATM networks: an integrated approach. Proceedings of the 13th ITC, Copenhagen, June 1991, pp. 141-146.
- [206] J. RIORDAN: Stochastic Service Systems, J. Wiley and Sons, New York/London, 1962.
- [207] J.W. ROBERTS: Performance Evaluation and Design of Multiservice Networks (COST 224 Final Report), Commission of the European Communities, Brussels, 1991.
- [208] J.W. ROBERTS, F. GUILLEMIN: Jitter in ATM networks and its impact on peak rate enforcement. *Performance Evaluation*, vol. 16, nos. 1-3, November 1992, pp. 35-48.

- [209] J.W. ROBERTS, J.T. VIRTAMO: The superposition of periodic cell arrival processes in an ATM multiplexer. *IEEE Transactions on Communications*, vol. 39, no. 2, February 1991, pp. 298-303.
- [210] H. SAITO: New dimensioning concept for ATM networks. Proceedings of the 7th ITC Specialist Seminar, Morristown, October 1990, paper 15.3.
- [211] H. SAITO: Queueing analysis of cell loss probability control in ATM networks. Proceedings of the 13th ITC, vol. on queueing, performance and control in ATM, Copenhagen, June 1991, pp. 19–24.
- [212] H. SAITO: Call admission control in an ATM network using upper bound of cell loss probability. *IEEE Transactions on Communications*, vol. 40, no. 9, September 1992, pp. 1512-1521.
- [213] H. SAITO, K. KAWASHIMA, K.-I. SATO: Traffic control technologies in ATM networks. *IEICE Transactions*, vol. E-74, no. 4, April 1991, pp. 761-771.
- [214] H. SAITO, K. SHIOMOTO: Dynamic call admission control in ATM networks. *IEEE JSAC*, vol. 9., no. 7, September 1991, pp. 982–989.
- [215] S. SALMELA: A proposal for connection acceptance control of the R1022 demonstrator. RACE 1022 Document NOKIA\_SOFT\_00011\_CD\_CC, June 1990.
- [216] K. SÄLLBERG, B. STAVENOW, I. ANDERSEN: A resource allocation framework in B-ISDN. Proceedings of the XIII ISS, Stockholm, May 1990, pp. 111-118.
- [217] K.-I. SATO, S. OHTA, I. TOKIZAWA: Broad-band ATM network architecture based on virtual paths. *IEEE Transactions on Communications*, vol. 38, no. 8, August 1990, pp. 1212-1222.
- [218] M. SCHOPP: Analytische Behandlung allgemeiner Fluid-Flow-Modelle zur Modellierung von ATM-Systemen, Diplomarbeit Nr. 1194, IND, Universität Stuttgart, Stuttgart, 1992.
- [219] F.C. Schoute: Simple decision rules for acceptance of mixed traffic streams. Proceedings of the 12th ITC, Torino, June 1988, pp. 4.2A.5.1-4.2A.5.7.
- [220] P. Sen, B. Maglaris, N.-E. Rikli, D. Anastassiou: Models for packet switching of variable-bit-rate video sources. IEEE JSAC, vol. 7, no. 5, September 1989, pp. 865-869.
- [221] H.-D. SHENG, S-Q. LI: Second order effect of binary sources on characteristics of queue and loss rate. Proceedings of the IEEE INFOCOM '93, San Francisco, March 1993, pp. 18-27.
- [222] S. SHIODA, H. UOSE: Virtual path bandwidth control method for ATM networks: successive modification method. *IEICE Transactions*, vol. E-74, no. 12, December 1991, pp. 4061-4068.
- [223] M. SIEGEL, R. MEDOW, T.H. THEIMER: Design of a single-chip ATM switching element. Proceedings of the ICCC '90, New Dehli, November 1990, pp. 804-811.

- [224] P. SKELLY, S. DIXIT, M. SCHWARTZ: A histogramm-based model for video traffic behaviour in an ATM network node with application to congestion control. *Proceedings* of the IEEE INFOCOM '92, Florence, May 1992, pp. 95-104.
- [225] K. SRIRAM, W. WHITT: Characterizing superposition arrival processes in packet multiplexers for voice and data. *IEEE JSAC*, vol. 4, no. 6, September 1986, pp. 833– 846.
- [226] T.E. STERN, A.E. ELWALID: Analysis of separable Markov-modulated rate models for information-handling systems. Advances of Applied Probability, vol. 23, 1991, pp. 105– 139.
- [227] S. SUMITA, T. OZAWA: Achievability of performance objectives in ATM switching nodes. Proceedings of the International Seminar on Performance of Distributed and Parallel Systems, Kyoto, December 1988, pp. 45-56.
- [228] H. Suzuki, S. Sato: Temporal cell loss behavior in an ATM multiplexer with heterogeneous burst input. *IEICE Transactions on Communications*, vol. E75-B, no. 12, December 1992, pp. 1346-1353.
- [229] E.D. SYSKAS, I.C. PASCHALIDIS, G.K. MOURTZINOU, K.M. VLAKOS: Congestion avoidance in ATM networks. *Proceedings of the IEEE INFOCOM '92*, Florence, May 1992, pp. 905-914.
- [230] L. TASSIULAS, Y. HUNG, S.S. PANWAR: Optimal buffer control during congestion in an ATM network node. *Proceedings of IEEE INFOCOM '93*, San Francisco, April 1993, pp. 1059–1066.
- [231] T.H. THEIMER: How to compute the moments of a GMDP. RACE 1022 Document UST\_123\_0023\_CD\_CC, November 1989.
- [232] T.H. THEIMER: Vergleichende Untersuchungen an ATM-Koppelnetzstrukturen, Dissertationsschrift, Universität Stuttgart, Stuttgart, 1994.
- [233] A. THOMAS, J.P. COUDREUSE, M. SERVEL: Asynchronous time-division techniques: an experimental packet network integrating videocommunications. *Proceedings of the* XI ISS, Florence, May 1984, paper 32-C-2.
- [234] P. TRAN-GIA, O. GROPP: Structure and performance of neural nets in broadband system admission control. Research Report University of Würzburg, Institute of Computer Science, no. 37, December 1991.
- [235] A. TRIANTAFILIDIS: Modellierung und Leistungsuntersuchung einer adaptiven Verbindungsannahmestrategie für ATM-Netze. Diplomarbeit Nr. 1243, IND, Universität Stuttgart, Stuttgart, 1993.
- [236] R.C.F. Tucker: Accurate method for analysis of a packet-speech multiplexer with limited delay. *IEEE Transactions on Communications*, vol. 36, no. 4, April 1988, pp. 479-483.
- [237] J.S. TURNER: New directions in communications (or which way in the information age?). Proceedings of the 1986 International Zurich Seminar on Digital Communications, Zurich, March 1986, pp. 25-32.

- [238] J.S. TURNER: Managing bandwidth in ATM networks with bursty traffic. IEEE Network, vol. 6, no. 5, September 1992, pp. 50-58.
- [239] K. TUTUFOR: On admission control and policing in an ATM based network. Proceedings of the 7th ITC Specialist Seminar, Morristown, October 1990, paper 5.4.
- [240] F. VAKIL, H. SAITO: On congestion control in ATM networks. IEEE Lightwave Technology Systems, August 1991, pp. 55-65.
- [241] M. VILLEN-ALTAMIRANO, J. VILLEN-ALTAMIRANO: RESTART: a method for accelerating rare event simulation. *Proceedings of the 13th ITC*, vol. on queueing, performance and control in ATM, Copenhagen, June 1991, pp. 71-76.
- [242] J.T. VIRTAMO: An exact analysis of the  $D_1 + ... + D_K/D/1$  queue. Proceedings of the 7th ITC Specialist Seminar, Morristown, October 1991, paper 11.1.
- [243] J.T. VIRTAMO, J.W. ROBERTS: Evaluating buffer requirements in an ATM multiplexer. Proceedings of the IEEE GLOBECOM '89, Dallas, December 1989, pp. 1473-1477.
- [244] E. WALLMEIER: A connection acceptance algorithm for ATM networks based on mean and peak bitrate. *International Journal of Digital and Analog Communication Systems*, vol. 3, no. 2, April-June 1990, pp. 143-153.
- [245] E. WALLMEIER, C.M. HAUBER: Blocking probabilities in ATM pipes controlled by a connection acceptance algorithm based on mean and peak bit rates. *Proceedings of the* 13th ITC, vol. on queueing, performance and control in ATM, Copenhagen, June 1991, pp. 137-142.
- [246] E. WALLMEIER, T. WORSTER: The spacing policer, an algorithm for efficient peak bit rate control in ATM networks. *Proceedings of the XIV ISS*, Yokohama, October 1992, pp. 22–26.
- [247] J. WENZEL: Vergleich einfacher Verbindungsannahme-Algorithmen für ATM-Netze auf der Basis der effektiven Bandbreite, Semesterarbeit Nr. 1035, IND, Universität Stuttgart, Stuttgart, 1990.
- [248] M. WERNIK, O. ABOUL-MAGD, H. GILBERT: Traffic management for B-ISDN services. IEEE Network, vol. 6, no. 5, September 1992, pp. 10-19.
- [249] R.W. WOLFF: Poisson arrivals see time averages. Operations Research, vol. 30, no. 2, April 1982, pp. 223-231.
- [250] A.K. Wong: Queueing analysis for ATM switching of continuous-bit-rate traffic a recursion computation method. Proceedings of the IEEE GLOBECOM '90, San Diego, December 1990, pp. 1438–1444.
- [251] G.M. WOODRUFF, R. KOSITPAIBOON: Multimedia traffic management principles for guaranteed ATM network performance. *IEEE JSAC*, vol. 8, no. 3, April 1990, pp. 437– 446.
- [252] T. WORSTER: Neural network based controllers for connection acceptance. Proceedings of the 2nd Race Workshop on Traffic and Performance Aspects in IBCN, Aveiro, January 1992.

- [253] T. WORSTER: Modelling deterministic queues: the leaky bucket as an arrival process. Proceedings of the 14th ITC, Antibes Juan-les-Pins, June 1994, pp. 581-590.
- [254] Y. YAMAMOTO, T. WRIGHT: Error performance in evolving digital networks including ISDNs. IEEE Communications Magazine, vol. 27, no. 4, April 1989, pp. 12-18.
- [255] N. YAMANAKA, Y. SATO, K.-I. SATO: Performance limitation of leaky bucket algorithm for usage parameter control and bandwidth allocation methods. *IEICE Transactions on Communications*, vol. E75-B, no. 2, February 1992, pp. 82-86.
- [256] N. YAMANAKA, Y. SATO, K.-I. SATO: Traffic shaping for VBR traffic in ATM networks. *IEICE Transactions on Communications*, vol. E75-B, no. 10, October 1992, pp. 1105-1108.
- [257] Y.S. YEH, M.G. HLUCHYI, A.S. ACOMPARA: The knockout switch: a simple, modular architecture for high-performance packet switching. *IEEE JSAC*, vol. 5, no. 8, October 1987, pp. 1274-1283.
- [258] N. YIN, M.G. HLUCHYJ: A dynamic rate control mechanism for source coded traffic in a fast packet network. IEEE JSAC, vol. 9, no. 7, September 1991, pp. 1003-1012.
- [259] N. YIN, S.-Q. LI, T.E. STERN: Congestion control for packet voice by selective packet discarding. *IEEE Transactions on Communications*, vol. 38, no. 5, May 1990, pp. 674-683.
- [260] C. YUAN, J.A. SILVESTER: Queueing analysis of delay constrained voice traffic in a packet switching system. IEEE JSAC, vol. 7, no. 5, June 1989, pp. 729-738.
- [261] R. ZURMÜHL, S. FALK: Matrizen und ihre Anwendungen, Band 1: Grundlagen, Springer-Verlag, Berlin, 6. Auflage, 1992.

# Anhang A

# "Fluid Flow"-Analyse für allgemeine Markoff-modulierte Ratenprozesse

## A.1 System- und Quellcharakterisierung

Auf der Büschelebene kann das System durch das in Abschnitt 3.2.1 eingeführte "Fluid Flow"-Modell beschrieben werden, das den diskreten Charakter der Zellankünfte und -bedienungen durch einen kontinuierlichen Zu- bzw. Abfluß von Information ersetzt. Die Zuflußrate errechnet sich aus der Summe der Ankunftsraten aller N Verkehrsquellen, und die Abflußrate entspricht dem Zelltakt C des Multiplexers. Es wird angenommen, daß die Abflußrate immer ungleich der Zuflußrate ist und zwischen der mittleren und der maximalen Zuflußrate liegt. Falls diese Bedingungen nicht erfüllt sind, muß der analytische Lösungsansatz modifiziert werden [167]. Für die Zwischenspeicherung der Information steht ein Puffer mit endlicher Größe S zur Verfügung. Falls die Zuflußrate die Abflußrate übersteigt, ergibt sich ein Informationsüberschuß, der zu einem Anstieg des Pufferfüllstands führt. Beim Erreichen eines maximalen Pufferfüllstands geht die Differenz zwischen Zu- und Abflußrate verloren. Der Pufferfüllstand nimmt wieder ab, sobald die Zuflußrate unter die Abflußrate sinkt. Dieses Modell ist in Bild A.1 veranschaulicht.

Eine beliebige Verkehrsquelle n (n=1,2,...,N) wird, wie in Unterabschnitt 3.1.2.2 besprochen, durch einen endlichen Zustandsautomaten mit insgesamt  $L_n+1$  Zuständen beschrieben. Die Zustandsdauern seien negativ-exponentiell verteilt, womit das Quellverhalten durch eine zeitkontinuierliche Markoff-Kette mit  $L_n+1$  Zuständen beschrieben werden kann. Die Übergänge zwischen den Quellzuständen l und m  $(l, m=0,1,...,L_n)$  werden durch die Übergangsraten  $\lambda_n(m|l)$  beschrieben. Im Quellzustand l, der mit der Wahrscheinlichkeit  $p_{B,n}(l)$  auftritt, sendet die Verkehrsquelle mit einer konstanten aber zustandsabhängigen Zellrate  $r_n(l)$ . Da die Ankunftsrate durch eine Markoff-Kette gesteuert wird, spricht man von einem

Markoff-modulierten Ratenprozeß [226]. Bild A.2 veranschaulicht die für die Büschelebene geltende Quellbeschreibung eines solchen Prozesses mit drei Zuständen.

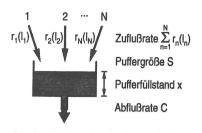


Bild A.1: "Fluid Flow"-Modell eines Multiplexers

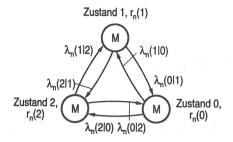


Bild A.2: Markoff-modulierter Ratenprozeß
mit drei Zuständen

Da die Überlagerung der beschriebenen Prozesse wiederum auf einen Markoff-modulierten Ratenprozeß führt, kann eine einzelne Verkehrsquelle auch den Überlagerungsverkehr mehrerer Verbindungen nachbilden. Insbesondere werden Verbindungen mit identischen statistischen Eigenschaften sinnvollerweise durch eine Ersatzverkehrsquelle beschrieben. Dies hat den Vorteil, daß Systemzustände, die sich nur durch eine Vertauschung der Quellzustände einzelner Verkehrsquellen unterscheiden, innerhalb eines globalen Systemzustands zusammengefaßt werden können. Beispielsweise kann die Überlagerung von L sporadischen Verkehrsquellen durch eine Ersatzquelle mit L+1 Zuständen vollständig charakterisiert werden (vgl. Bild 4.19). Der Zustand dieser Ersatzquelle ist durch die Anzahl der sich in der Sendephase befindlichen Einzelquellen bestimmt. Falls jede Quelle getrennt beschrieben wird, entsteht ein Zustandsraum der Größe  $2^L$ .

In der Literatur wurden zahlreiche Spezialfälle des oben eingeführten Verkehrsmodells behandelt. Die Überlagerung von sporadischen Quellen wurde sowohl für homogene [4, 156, 164, 236] als auch für heterogene Verkehrszusammensetzung [14, 140, 220] unter den in Tabelle 4.1 angeführten Modellannahmen eingehend untersucht. In [226] wurde die allgemeinere Klasse der zeitlich umkehrbaren Markoff-modulierten Ankunftsprozesse behandelt, die sich dadurch auszeichnet, daß sich ihr statistisches Verhalten bei einer Umkehrung des Zeitverlaufs nicht ändert. Dies ist genau dann der Fall, wenn sich die Übergänge zwischen zwei Zuständen im lokalen statistischen Gleichgewicht befinden, d.h. ein zukünftiger Zustandswechsel vom Zustand l in den Zustand l mit den Zustandswahrscheinlich auf wie ein vorangegangener Wechsel vom Zustand l in den Zustand l. Mit den Zustandswahrscheinlichkeiten  $p_{B,n}(l)$  kann diese Bedingung folgendermaßen formuliert werden [226]:

$$p_{B,n}(l) \cdot \lambda_n(m|l) = p_{B,n}(m) \cdot \lambda_n(l|m), \quad n = 1, 2, ..., N, \quad l, m = 0, 1, ..., L_n.$$
 (A.1)

In [58] wurde das asymptotische Systemverhalten allgemeiner Markoff-modulierter "Fluid Flow"-Modelle für große Puffergrößen untersucht. Für sporadische Quellen mit beliebig verteilten Sende- und Ruhedauern wurde in [18] ein approximativer Lösungsansatz vorgestellt.

Im folgenden wird ein exakter und ein approximativer Lösungsansatz für "Fluid Flow"Modelle mit allgemeinen Markoff-modulierten Ankunftsprozessen vorgestellt. Zunächst werden die wichtigsten Charakteristika des Ankunftsprozesses bestimmt. Der Zustand aller NVerkehrsquellen kann durch den Zustandsvektor  $\mathbf{l} = (l_1, l_2, ..., l_N), l_n = 0, 1, ..., L_n$ , beschrieben werden. Da alle Quellen statistisch unabhängig sind, ergibt sich die zugehörige Verbundwahrscheinlichkeit aus dem Produkt der Zustandswahrscheinlichkeiten  $p_{B,n}(l_n)$  der Einzelquellen, d.h.

 $p_B(\mathbf{l}) = p_B(l_1, l_2, ..., l_N) = \prod_{n=1}^{N} p_{B,n}(l_n)$  (A.2)

Die Zustandswahrscheinlichkeiten der Einzelquelle n können aus der entsprechenden zeitkontinuierlichen Markoff-Kette bestimmt werden. Im stationären Fall gelten dafür die Gleichungen des statistischen Gleichgewichts nach Gleichung 3.23:

$$p_{B,n}(l) \sum_{m=0,m\neq l}^{L_n} \lambda_n(m|l) = \sum_{m=0,m\neq l}^{L_n} p_{B,n}(m) \cdot \lambda_n(l|m) , \quad l=0,1,...,L_n , \quad n=1,2,...,N .$$
(A.3)

Dieses System von Zustands-Übergangsgleichungen kann in die folgende äquivalente Matrizenschreibweise umgeformt werden, die für die nachfolgenden Analysen hilfreich ist [139]:

$$\boldsymbol{P}_n \cdot \boldsymbol{Q}_n = \boldsymbol{0} \; , \tag{A.4}$$

mit

dem Wahrscheinlichkeitsvektor  $P_n = (p_{B,n}(0), p_{B,n}(1), ..., p_{B,n}(L_n))$ 

dem Generator

$$Q_n = \begin{bmatrix} \lambda_n(0|0) & \lambda_n(1|0) & \cdots & \lambda_n(L_n|0) \\ \lambda_n(0|1) & \lambda_n(1|1) & \cdots & \lambda_n(L_n|1) \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_n(0|L_n) & \lambda_n(1|L_n) & \cdots & \lambda_n(L_n|L_n) \end{bmatrix}$$

und dem Nullvektor

$$\mathbf{0} = (0, 0, ..., 0)$$
.

Aus Notationsgründen wurden die Diagonale<br/>lemente der Übergangsmatrix  $\boldsymbol{Q}_n$  als

$$\lambda_n(l|l) = -\sum_{m=0, m \neq l}^{L_n} \lambda_n(m|l) \tag{A.5}$$

definiert. Aus dieser Gleichung ist ersichtlich, daß für den Generator  $Q_n$  die Summe aller in einer Zeile befindlichen Elemente gleich Null ist, weshalb die Gleichungen des statistischen

Gleichgewichts linear abhängig sind. Für eine eindeutige Lösung muß eine dieser Gleichungen durch die Normalisierungsbedingung

$$\sum_{l=0}^{L_n} p_{B,n}(l) = 1 \tag{A.6}$$

ersetzt werden. Das entstehende lineare Gleichungssystem kann mit Hilfe von numerischen Verfahren aufgelöst werden. Häufig können auch spezielle Lösungsalgorithmen eingesetzt werden, da in den meisten Anwendungen nur wenige Zustandsübergänge existieren. Insbesondere ist für zeitlich umkehrbare Prozesse eine rekursive Berechnung der Zustandswahrscheinlichkeiten mit Hilfe der in Gleichung A.1 eingeführten lokalen Gleichgewichtsbeziehungen möglich (vgl. [150]).

Aus den Zustandswahrscheinlichkeiten können die mittleren Ankunftsraten  $\overline{r}_n$  der einzelnen Verkehrsquellen und die mittlere Gesamtankunftsrate  $\overline{r}$  bestimmt werden.

$$\overline{r}_n = \mathbb{E}[R_n] = \sum_{l=0}^{L_n} p_{B,n}(l) \cdot r_n(l)$$
 bzw.  $\overline{r} = \mathbb{E}[R] = \sum_{n=1}^{N} \overline{r}_n$ . (A.7)

## A.2 Grundlegender Lösungsansatz

Die adäquate Zustandsbeschreibung des aus den Verkehrsquellen und dem Puffer zusammengesetzten Gesamtsystems ist der Schlüssel für die analytische Behandlung des im vorigen Unterkapitel eingeführten Verkehrsmodells. Der Systemzustand zum Zeitpunkt t ist durch die Quellzustände  $L_n(t)$  (n=1,2,...,N) und den Pufferfüllstand X(t) vollständig beschrieben, da die Verweildauern aller Quellzustände negativ-exponentiell verteilt sind. Hierfür wird die folgende Verbundwahrscheinlichkeit definiert [226]:

$$P(l, x, t) = P(l_1, l_2, ..., l_N, x, t) = P\{L_1(t) = l_1, L_2(t) = l_2, ..., L_N(t) = l_N, X(t) \le x\},$$

$$0 \le x < S, l_n = 0, 1, ..., L_n, n = 1, 2, ..., N.$$
(A.8)

Die Zustandswahrscheinlichkeit zum Zeitpunkt  $t+\Delta t$  kann aus den Zustandswahrscheinlichkeiten zum Zeitpunkt t bestimmt werden, da es sich um ein Markoffsches System handelt. Innerhalb des dazwischenliegenden Zeitintervalls  $\Delta t \to 0$  können folgende Änderungen des Systemzustands  $(l_1, l_2, ..., l_N, x)$  auftreten (Terme der Ordnung  $O(\Delta t^2)$  und höher werden vernachlässigt):

- Die Verkehrsquelle n geht mit der Wahrscheinlichkeit  $\lambda_n(l|l_n) \cdot \Delta t$  vom Zustand  $l_n$  in den Zustand  $l \neq l_n$  über.
- Alle Verkehrsquellen bleiben mit der komplementären Wahrscheinlichkeit  $1 \sum_{n=1}^{N} \sum_{l=0, l \neq l_n}^{L_n} \lambda_n(l|l_n) \cdot \Delta t = 1 + \sum_{n=1}^{N} \lambda_n(l_n|l_n) \cdot \Delta t \text{ in ihrem Ausgangszustand } l_n.$

• Der Pufferfüllstand ändert sich unabhängig von dem Verhalten der Verkehrsquellen um den Wert  $\left(\sum_{j=1}^{N} r_{j}(l_{j}) - C\right) \cdot \Delta t$ .

Falls nun alle Zustandsübergänge, die zum Zustand  $(l_1, l_2, ..., l_N, x, t + \Delta t)$  führen, betrachtet werden, ergibt sich folgender Zusammenhang zwischen den Zustandswahrscheinlichkeiten zu den Zeitpunkten t und  $t + \Delta t$ :

$$\lim_{\Delta t \to 0} P(l_1, l_2, ..., l_N, x, t + \Delta t) = \lim_{\Delta t \to 0} P\left(l_1, l_2, ..., l_N, x - \left[\sum_{j=1}^N r_j(l_j) - C\right] \cdot \Delta t, t\right) \cdot \left[1 + \sum_{n=1}^N \lambda_n(l_n|l_n) \cdot \Delta t\right] + \sum_{n=1}^N \sum_{l=0, l \neq l_n}^{L_n} P\left(l_1, ..., l_{n-1}, l, l_{n+1}, ..., l_N, x - \left[\sum_{j=1, j \neq n}^N r_j(l_j) + r_n(l) - C\right] \cdot \Delta t, t\right) \cdot \lambda_n(l_n|l) \cdot \Delta t .$$
(A.9)

Die Umformung des ersten Terms der rechten Seite und eine Erweiterung der linken Seite um  $P(l_1, l_2, ..., l_N, x, t) - P(l_1, l_2, ..., l_N, x, t)$  führt auf

$$\lim_{\Delta t \to 0} \frac{P(l_1, l_2, \dots, l_N, x, t + \Delta t) - P(l_1, l_2, \dots, l_N, x, t)}{\Delta t} - \frac{P(l_1, l_2, \dots, l_N, x - \left[\sum_{j=1}^{N} r_j(l_j) - C\right] \cdot \Delta t, t) - P(l_1, l_2, \dots, l_N, x, t)}{\Delta t} = \lim_{\Delta t \to 0} \sum_{n=1}^{N} \sum_{l=0}^{L_n} P\left(l_1, \dots, l_{n-1}, l, l_{n+1}, \dots, l_N, x - \left[\sum_{j=1, j \neq n}^{N} r_j(l_j) + r_n(l) - C\right] \cdot \Delta t, t\right) \cdot \lambda_n(l_n|l) .$$
(A.10)

Daraus erhält man das folgende System partieller Differentialgleichungen, falls der Grenzübergang  $\Delta t \rightarrow 0$  durchgeführt wird:

$$\frac{\partial P(l_1, l_2, ..., l_N, x, t)}{\partial t} + \frac{\partial P(l_1, l_2, ..., l_N, x, t)}{\partial x} \cdot \left( \sum_{j=1}^N r_j(l_j) - C \right) = \sum_{n=1}^N \sum_{l=0}^{L_n} P(l_1, ..., l_{n-1}, l, l_{n+1}, ..., l_N, x, t) \cdot \lambda_n(l_n|l) .$$
(A.11)

Im stationären Fall werden die Systemgrößen unabhängig von der Zeit t und man erhält für  $F(l_1, l_2, ..., l_N, x) = \lim_{t\to\infty} P(l_1, l_2, ..., l_N, x, t)$  ein System linearer Differentialgleichungen erster Ordnung [58, 226]:

$$\frac{\mathrm{d}F(l_1, l_2, ..., l_N, x)}{\mathrm{d}x} \left( \sum_{j=1}^N r_j(l_j) - C \right) = \sum_{n=1}^N \sum_{l=0}^{L_n} F(l_1, ..., l_{n-1}, l, l_{n+1}, ..., l_N, x) \cdot \lambda_n(l_n|l) . \tag{A.12}$$

Üblicherweise wird für dieses Differentialgleichungssystem die folgende Matrizenschreibweise eingeführt:

$$\frac{\mathrm{d}\boldsymbol{F}(x)}{\mathrm{d}x} \cdot (\boldsymbol{R} - C \cdot \boldsymbol{E}) = \frac{\mathrm{d}\boldsymbol{F}(x)}{\mathrm{d}x} \cdot \boldsymbol{D} = \boldsymbol{F}(x) \cdot \boldsymbol{Q} \tag{A.13}$$

bzw.

$$\frac{\mathrm{d}F\left(x\right)}{\mathrm{d}x} = F\left(x\right) \cdot QD^{-1}, \tag{A.14}$$

mit

$$\boldsymbol{D} = \boldsymbol{R} - C\boldsymbol{E} \ . \tag{A.15}$$

An dieser Stelle soll angemerkt werden, daß in der Literatur auch häufig die transponierte Übergangsmatrix  $Q^T$  verwendet wird (siehe z.B. [4, 140, 226, 236]), was zu einer modifizierten Form dieser Gleichung führt. Die Matrix E bezeichnet die Einheitsmatrix, d.h. E=diag[1]. Der Vektor F(x) enthält alle Elemente  $F(l_1, l_2, ..., l_N, x)$ , wobei die N-dimensionale Zustandsbeschreibung  $(l_1, l_2, ..., l_N)$  auf einen eindimensionalen Zeilenvektor abgebildet wird. Für die Elemente des Zeilenvektors gilt folgende Abbildungsvorschrift:

Zustand 
$$(0,0,...,0)$$
  $(0,0,...,1)$   $\cdots$   $(0,0,...,L_N)$   $(0,0,...,1,0)$   $\cdots$   $(L_1,L_2,...,L_N)$ 

$$\downarrow \qquad \qquad \downarrow \qquad \qquad \downarrow \qquad \qquad \downarrow$$
Element 1 2  $\cdots$   $L_N+1$   $L_N+2$   $\cdots$   $\prod_{n=1}^N(L_n+1)$ .

Die Elemente der Matrizen Q und R müssen nach derselben Abbildungsvorschrift abgelegt werden. Dieses Ordnungsschema ergibt sich automatisch, falls die für die Einzelquellen gültigen Teilmatrizen  $Q_n$  und  $R_n$ , n=1,2,...,N, entsprechend den Rechenvorschriften der Kronecker-Summe verknüpft werden [58, 226], d.h.

$$R = R_1 \oplus R_2 \oplus ... \oplus R_N \quad \text{und}$$
 (A.16)

$$Q = Q_1 \oplus Q_2 \oplus ... \oplus Q_N$$
 (A.17)

Die Matrix  $\mathbf{R}_n$  enthält auf der Hauptdiagonale die Ankunftsraten der einzelnen Quellzustände, d.h.  $\mathbf{R}_n = \operatorname{diag}[r_n(0), r_n(1), ..., r_n(L_n)]$ . Die Matrix  $\mathbf{Q}_n$  wurde in Unterkapitel A.1 definiert.

Die Kronecker-Summe der  $a \times a$  Matrix A und der  $b \times b$  Matrix B ist definiert als [58, 226]

$$\mathbf{A} \oplus \mathbf{B} = \mathbf{A} \otimes \mathbf{E}_b + \mathbf{E}_a \otimes \mathbf{B} , \qquad (A.18)$$

wobei  $E_a$  und  $E_b$  Einheitsmatrizen der Größe  $a \times a$  bzw.  $b \times b$  darstellen. Das Symbol  $\otimes$  bezeichnet das Kronecker-Produkt zweier Matrizen. Für das Kronecker-Produkt der oben definierten Matrizen A und B ergibt sich eine Matrix der Größe  $(a \cdot b) \times (a \cdot b)$ , indem jedes Element  $a_{cd}$  der Matrix A durch die komplette Matrix  $a_{cd} \cdot B$  ersetzt wird.

# A.3 Allgemeine Lösung des Differentialgleichungssystems

Für das Differentialgleichungssystem nach Gleichung A.14 wird in der Literatur der folgende exponentielle Lösungsansatz gemacht [4, 14, 58, 140, 226, 236]:

$$F(x) = \sum_{\mathbf{m}} a(\mathbf{m}) \cdot \boldsymbol{\varphi}(\mathbf{m}) \cdot \exp[z(\mathbf{m}) \cdot x],$$

$$\mathbf{m} = (m_1, m_2, ..., m_N), \quad m_n = 0, 1, ..., L_n, \quad n = 1, 2, ..., N.$$
(A.19)

Mit diesem Ansatz und der zugehörigen Ableitung

$$\frac{\mathrm{d}\boldsymbol{F}(x)}{\mathrm{d}x} = \sum_{\boldsymbol{m}} a(\boldsymbol{m}) \cdot z(\boldsymbol{m}) \cdot \boldsymbol{\varphi}(\boldsymbol{m}) \cdot \exp[z(\boldsymbol{m}) \cdot x] \tag{A.20}$$

vereinfacht sich das in Gleichung A.14 angegebene Differentialgleichungssystem auf das folgende spezielle Eigenwertproblem:

$$z(\mathbf{m}) \cdot \varphi(\mathbf{m}) = \varphi(\mathbf{m}) \cdot QD^{-1}$$
 (A.21)

Die in der Lösung des Differentialgleichungssystems auftretenden Terme z(m) und  $\varphi(m)$  entsprechen somit den Eigenwerten bzw. den dazugehörigen linksseitigen Eigenvektoren der Matrix  $QD^{-1}$ . Die Koeffizienten a(m) müssen aus den Randbedingungen bestimmt werden.

Die in Gleichung A.19 eingeführte rein exponentielle Lösung entspricht allerdings nur dann der allgemeinen Lösung des Differentialgleichungssystems, falls eine der beiden nachfolgenden Bedingungen erfüllt ist [35]:

- Alle Eigenwerte der Matrix  $QD^{-1}$  sind verschieden.
- Für einen r-fachen Eigenwert z(m) der Matrix  $QD^{-1}$  sinkt der Rang der Matrix  $QD^{-1} z(m) \cdot E$  um r ab. Dies ist insbesondere dann der Fall, wenn die Matrix  $QD^{-1}$  symmetrisch ist bzw. symmetrisiert werden kann.

Diese Bedingungen gewährleisten, daß alle Eigenvektoren linear unabhängig sind, womit eine ausreichende Anzahl von Freiheitsgraden zur Erfüllung der Randbedingungen zur Verfügung steht. Für die in der Literatur behandelten zeitlich umkehrbaren Ankunftsprozesse [4, 14, 140, 226, 236] kann die Matrix  $QD^{-1}$  immer symmetrisiert werden [226], womit die oben angegebene Lösung ausreichend ist. In [58] wurden zwar allgemeinere Prozesse behandelt, jedoch wurde nur die asymptotische Verteilung des Pufferfüllstands betrachtet. Hierfür ergibt sich ein exponentieller Verlauf, da der dominante Eigenwert reell ist und nur einfach auftritt.

Bei den in diesem Anhang behandelten allgemeinen Markoff-modulierten "Fluid Flow"Modellen könnten prinzipiell Fälle auftreten, für die die obigen Bedingungen nicht erfüllt
sind. Dann müßten die Koeffizienten a(m) für r-fache Eigenwerte z(m) durch ein Polynom (r-1)-ten Grades ersetzt werden, um die allgemeingültige Lösung des Differentialgleichungssystems zu erhalten [35]. Für die in dieser Arbeit behandelten praktischen Anwendungsbeispiele traten keine Fälle auf, bei denen diese erweiterte Lösungsform erforderlich
gewesen wäre. Ferner zeigte sich, daß mehrfache Eigenwerte nur dann auftreten, wenn einige
Quellparameter exakt gleich gewählt werden. Kleinste Abweichungen der Quellparameter
zerstören dieses Gleichgewicht und führen auf verschiedene Eigenwerte.

Ein weiterer Unterschied zu den bisher in der Literatur behandelten Modellen besteht darin, daß nun auch komplexe Eigenwerte und Eigenvektoren auftreten können, während in den bisherigen Anwendungen nur reelle Werte möglich waren. Den weiteren Betrachtungen soll daher eine Zusammenfassung der wichtigsten Eigenschaften der Eigenwerte und Eigenvektoren reellwertiger Matrizen vorangestellt werden.

- 1. Die Eigenwerte der Matrix  $QD^{-1}$  sind entweder reell oder treten als konjugiert komplexe Paare auf (vgl. [261], S. 258).
- Der zu einem reellen Eigenwert gehörende linksseitige Eigenvektor besitzt ausschließlich reelle Elemente, da er aus einem linearen Gleichungssystem bestimmt wird.
- 3. Die zu konjugiert komplexen Eigenwerten gehörenden linksseitigen Eigenvektoren bestehen aus paarweise konjugiert komplexen Elementen (vgl. [261], S. 265).
- 4. Die zu reellen Eigenwerten gehörenden Koeffizienten a(m) sind reell, da sie aus einem linearen Gleichungssystem bestimmt werden.
- 5. Die zu konjugiert komplexen Eigenwerten gehörenden Koeffizienten a(m) sind konjugiert komplex.

Die letzte Eigenschaft ergibt sich aus der Tatsache, daß die Verteilungsfunktion nur reelle Werte annehmen kann. Dies ist dann erfüllt, wenn die den konjugiert komplexen Eigenwerten zugeordneten Koeffizienten ebenfalls konjugiert komplex sind, wie die folgende Umformung von Gleichung A.19 zeigt:

$$\begin{split} F(x) &= \sum_{\boldsymbol{m}} a(\boldsymbol{m}) \cdot \boldsymbol{\varphi}(\boldsymbol{m}) \cdot \exp[z(\boldsymbol{m}) \cdot x] \\ &= \sum_{\boldsymbol{m}: \ z_I(\boldsymbol{m}) = 0} a_R(\boldsymbol{m}) \cdot \boldsymbol{\varphi}_R(\boldsymbol{m}) \cdot \exp[z_R(\boldsymbol{m}) \cdot x] + \\ &= \sum_{\boldsymbol{m}: \ z_I(\boldsymbol{m}) = 0} \left\{ [a_R(\boldsymbol{m}) + i \cdot a_I(\boldsymbol{m})] \cdot [\boldsymbol{\varphi}_R(\boldsymbol{m}) + i \cdot \boldsymbol{\varphi}_I(\boldsymbol{m})] \cdot \exp[z_R(\boldsymbol{m}) \cdot x + i \cdot z_I(\boldsymbol{m}) \cdot x] + \\ &= \sum_{\boldsymbol{m}: \ z_I(\boldsymbol{m}) > 0} \left\{ [a_R(\boldsymbol{m}) - i \cdot a_I(\boldsymbol{m})] \cdot [\boldsymbol{\varphi}_R(\boldsymbol{m}) - i \cdot \boldsymbol{\varphi}_I(\boldsymbol{m})] \cdot \exp[z_R(\boldsymbol{m}) \cdot x - i \cdot z_I(\boldsymbol{m}) \cdot x] \right\} \end{split}$$

$$F(x) = \sum_{\boldsymbol{m}: \ z_{I}(\boldsymbol{m}) = 0} a_{R}(\boldsymbol{m}) \cdot \boldsymbol{\varphi}_{R}(\boldsymbol{m}) \cdot \exp\left[z_{R}(\boldsymbol{m}) \cdot x\right] + \sum_{\boldsymbol{m}: \ z_{I}(\boldsymbol{m}) = 0} \left\{ 2 \cdot \left[a_{R}(\boldsymbol{m}) \cdot \boldsymbol{\varphi}_{R}(\boldsymbol{m}) - a_{I}(\boldsymbol{m}) \cdot \boldsymbol{\varphi}_{I}(\boldsymbol{m})\right] \cdot \exp\left[z_{R}(\boldsymbol{m}) \cdot x\right] \cdot \cos\left[z_{I}(\boldsymbol{m}) \cdot x\right] - 2 \cdot \left[a_{R}(\boldsymbol{m}) \cdot \boldsymbol{\varphi}_{I}(\boldsymbol{m}) + a_{I}(\boldsymbol{m}) \cdot \boldsymbol{\varphi}_{R}(\boldsymbol{m})\right] \cdot \exp\left[z_{R}(\boldsymbol{m}) \cdot x\right] \cdot \sin\left[z_{I}(\boldsymbol{m}) \cdot x\right] \right\},$$
(A.22)

mit

$$z(\boldsymbol{m})=z_R(\boldsymbol{m})+i\cdot z_I(\boldsymbol{m}): \boldsymbol{m}$$
-ter (komplexer) Eigenwert der Matrix  $QD^{-1}$ ,  $\varphi(\boldsymbol{m})=\varphi_R(\boldsymbol{m})+i\cdot \varphi_I(\boldsymbol{m}):$  zugehöriger linksseitiger Eigenvektor und  $a(\boldsymbol{m})=a_R(\boldsymbol{m})+i\cdot a_I(\boldsymbol{m}):$  zugeordneter frei wählbarer Koeffizient.

Diese reelle Darstellungsform der Lösung erweist sich günstig für das in Unterkapitel A.5 abgeleitete lineare Gleichungssystem zur Bestimmung der Koeffizienten  $a(\mathbf{m}) = a_R(\mathbf{m}) + i \cdot a_I(\mathbf{m})$  sowie für die Bestimmung der charakteristischen Systemgrößen.

# A.4 Lösung des Eigenwertproblems

Die direkte Bestimmung der Eigenwerte und Eigenvektoren führt zu einem beträchtlichen numerischen Aufwand, falls die Anzahl der Quellzustände große Werte annimmt. Für zeitlich umkehrbare Ankunftsprozesse wurde in [226] ein effizientes Zerlegungsverfahren zur Bestimmung der Eigenwerte und Eigenvektoren vorgestellt. In [58] wurde bewiesen, daß sich dieser Ansatz auch auf die hier behandelten allgemeinen Markoff-modulierten Ratenprozesse übertragen läßt. Ein weiterer Vorteil dieser Berechnungsmethode liegt darin, daß gezielt einzelne Eigenwerte bestimmt werden können. Insbesondere ist damit die direkte Berechnung des dominanten Eigenwerts möglich, der das asymptotische Systemverhalten für große Puffergrößen charakterisiert [58]. Die verwendete Methode soll im folgenden kurz aufgearbeitet werden; für eine ausführliche formale Ableitung des Verfahrens sei auf [58, 218, 226] verwiesen.

Im ersten Schritt soll gezeigt werden, daß der erste Eigenwert z(0,0,...,0)=0 eine Lösung des beschriebenen Eigenwertproblems ist. Hierzu wird das spezielle Eigenwertproblem nach Gleichung A.21 in das folgende allgemeine Eigenwertproblem umgeformt:

$$z(\mathbf{m}) \cdot \varphi(\mathbf{m}) \cdot \mathbf{D} = \varphi(\mathbf{m}) \cdot \mathbf{Q}$$
 (A.23)

Für den Sonderfall z(0,0,...,0)=0 ergibt sich daraus ein homogenes lineares Gleichungssystem:

$$\varphi(0,0,...,0) \cdot \mathbf{Q} = (0,0,...,0) = \mathbf{0}$$
 (A.24)

Ein Vergleich mit Gleichung A.4 zeigt, daß dieser Zusammenhang den Gleichungen des statistischen Gleichgewichts entspricht. Der zum Eigenwert 0 gehörende Eigenvektor ist somit identisch mit dem Vektor der Zustandswahrscheinlichkeiten, d.h. es gilt

$$z(0,0,...,0) = 0$$
 und  $\{\varphi(0,0,...,0)\}_{l} = p_{B}(l)$ . (A.25)

Für die Bestimmung der restlichen Eigenwerte und Eigenvektoren wird das allgemeine Eigenwertproblem nach Gleichung A.23 weiter umgeformt, wobei der bereits abgehandelte Fall z(m) = 0 ausgenommen wird:

$$z(\mathbf{m}) \cdot \varphi(\mathbf{m}) \cdot \mathbf{D} = z(\mathbf{m}) \cdot \varphi(\mathbf{m}) \cdot [\mathbf{R} - C \cdot \mathbf{E}] = \varphi(\mathbf{m}) \cdot \mathbf{Q}$$

$$\varphi(\mathbf{m}) \cdot \left[ \mathbf{R} - \frac{1}{z(\mathbf{m})} \mathbf{Q} \right] = C \cdot \varphi(\mathbf{m}), \quad z(\mathbf{m}) \neq 0$$

$$\varphi(\mathbf{m}) \cdot \mathbf{A}(z(\mathbf{m})) = C \cdot \varphi(\mathbf{m}), \quad (A.26)$$

mit

$$A(z(m)) = R - \frac{1}{z(m)}Q. \tag{A.27}$$

Die Beziehung nach Gleichung A.26 kann wiederum als spezielles Eigenwertproblem aufgefaßt werden: C entspricht dem m-ten Eigenwert und  $\varphi(m)$  dem zugeordneten linksseitigen Eigenvektor der Matrix A(z(m)). Dies bedeutet aber umgekehrt, falls ein Wert z gefunden wird, für den der m-te Eigenwert g(m,z) der Matrix A(z) gleich der Bedienrate C ist, so entspricht dieser dem m-ten Eigenwert z(m) der ursprünglichen Matrix  $QD^{-1}$ . Den beiden Eigenwerten g(m,z(m)) und z(m) ist derselbe Eigenvektor  $\varphi(m)$  zugeordnet [58, 226].

Der Eigenwert z(m) entspricht somit der Nullstelle der Funktion g(m,z)-C, die mit Hilfe eines iterativen Verfahrens gesucht werden kann. Da in jedem Iterationsschritt der Eigenwert g(m,z) der Matrix A(z) für den jeweils aktuellen z-Wert ermittelt werden muß, scheint dieses Verfahren wesentlich aufwendiger zu sein als das ursprüngliche Eigenwertproblem. Allerdings können die Eigenwerte der Matrix A(z) aus den Eigenwerten der für die einzelnen Verkehrsquellen gültigen Matrizen  $A_n(z)$  berechnet werden. Da die Ordnung der Matrizen  $A_n(z)$  wesentlich kleiner als die Ordnung der Gesamtmatrix A(z) ist, ergeben sich deutliche Rechenzeitvorteile bei der Anwendung dieser Methode zur Bestimmung der Eigenwerte. Die Zerlegung des Eigenwertproblems erhält man durch das Einsetzen der Gleichungen A.16 und A.17 in Gleichung A.27.

$$\mathbf{A}(z) = [\mathbf{R}_1 \oplus \mathbf{R}_2 \oplus \dots \oplus \mathbf{R}_N] - \frac{1}{z} \cdot [\mathbf{Q}_1 \oplus \mathbf{Q}_2 \oplus \dots \oplus \mathbf{Q}_N]$$
$$= \mathbf{A}_1(z) \oplus \mathbf{A}_2(z) \oplus \dots \oplus \mathbf{A}_N(z) , \qquad (A.28)$$

mit

$$A_n(z) = \left[ R_n - \frac{1}{z} Q_n \right], \quad n = 1, 2, ..., N.$$
 (A.29)

D.h. die Matrix A(z) läßt sich ebenso wie die Matrizen R und Q aus einer Kronecker-Summation der entsprechenden Subsystemmatrizen ermitteln. Für die Kronecker-Summe gilt nun der folgende Satz, der sich bei der Bestimmung der Eigenwerte und Eigenvektoren als nützlich erweist [58, 226]:

Sind  $\lambda_c$  (c=1,2,...,a) die Eigenwerte der Matrix  $\mathbf{A}$  mit den Eigenvektoren  $\alpha_c$  und  $\mu_d$  (d=1,2,...,b) die Eigenwerte einer Matrix  $\mathbf{B}$  mit den Eigenvektoren  $\beta_d$ , dann sind die Eigenwerte der Matrix  $\mathbf{A} \oplus \mathbf{B}$  die  $a \cdot b$  Summen  $\lambda_c + \mu_d$  mit den dazugehörigen Eigenvektoren  $\alpha_c \otimes \beta_d$ .

Bezeichnet man den  $m_n$ -ten Eigenwert der Matrix  $A_n(z)$  mit  $g_n(m_n, z)$ , so ergibt sich damit der m-te Eigenwert der Matrix A(z) zu

$$g(\mathbf{m}, z) = \sum_{n=1}^{N} g_n(m_n, z)$$
, mit  $\mathbf{m} = (m_1, m_2, ..., m_N)$ . (A.30)

Nachdem mit Hilfe des oben beschriebenen numerischen Verfahrens der Eigenwert  $z(\boldsymbol{m})$  gefunden ist, kann der dazugehörige linksseitige Eigenvektor  $\varphi(\boldsymbol{m})$  aus den zu den Eigenwerten  $g_n(m_n, z(\boldsymbol{m}))$  gehörenden linksseitigen Eigenvektoren  $\varphi_n(m_n, z(\boldsymbol{m}))$  der Matrizen  $\boldsymbol{A}_n(z(\boldsymbol{m}))$  bestimmt werden:

$$\varphi(\mathbf{m}) = \varphi_1(m_1, z(\mathbf{m})) \otimes \varphi_2(m_2, z(\mathbf{m})) \otimes ... \otimes \varphi_N(m_N, z(\mathbf{m}))$$
. (A.31)

# A.5 Lösung des Randwertproblems

Für die Bestimmung der speziellen Lösung des Randwertproblems wird der in Gleichung A.22 gegebene Lösungsvektor wiederum in seine Einzelkomponenten  $F(\boldsymbol{l},x)$  aufgespalten:

$$F(\boldsymbol{l},x) = \begin{cases} 0 & \text{für } x < 0 \\ \sum_{\boldsymbol{m}: \ z_I(\boldsymbol{m}) = 0} a_R(\boldsymbol{m}) \cdot \{\varphi_R(\boldsymbol{m})\}_{\boldsymbol{l}} \cdot \exp\left[z_R(\boldsymbol{m}) \cdot x\right] + \\ \sum_{\boldsymbol{m}: \ z_I(\boldsymbol{m}) > 0} \{2 \cdot [a_R(\boldsymbol{m}) \cdot \{\varphi_R(\boldsymbol{m})\}_{\boldsymbol{l}} - a_I(\boldsymbol{m}) \cdot \{\varphi_I(\boldsymbol{m})\}_{\boldsymbol{l}}] \cdot \\ m: \ z_I(\boldsymbol{m}) > 0 & \exp\left[z_R(\boldsymbol{m}) \cdot x\right] \cdot \cos\left[z_I(\boldsymbol{m}) \cdot x\right] - \\ 2 \cdot [a_R(\boldsymbol{m}) \cdot \{\varphi_I(\boldsymbol{m})\}_{\boldsymbol{l}} + a_I(\boldsymbol{m}) \cdot \{\varphi_R(\boldsymbol{m})\}_{\boldsymbol{l}}] \cdot \\ \exp\left[z_R(\boldsymbol{m}) \cdot x\right] \cdot \sin\left[z_I(\boldsymbol{m}) \cdot x\right] \end{cases}$$
für  $x \ge S$ , (A.32)

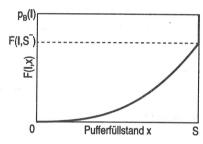
mit

$$z(\boldsymbol{m}) = z_R(\boldsymbol{m}) + i \cdot z_I(\boldsymbol{m})$$
:  $\boldsymbol{m}$ -ter (komplexer) Eigenwert der Matrix  $\boldsymbol{Q}\boldsymbol{D}^{-1}$ ,  $\{\varphi(\boldsymbol{m})\}_{\boldsymbol{l}} = \{\varphi_R(\boldsymbol{m})\}_{\boldsymbol{l}} + i \cdot \{\varphi_I(\boldsymbol{m})\}_{\boldsymbol{l}}$ :  $\boldsymbol{l}$ -tes Element des zugehörigen linksseitigen Eigenvektors und  $a(\boldsymbol{m}) = a_R(\boldsymbol{m}) + i \cdot a_I(\boldsymbol{m})$ : zugeordneter frei wählbarer Koeffizient.

Die gesuchte spezielle Lösung des Differentialgleichungssystems ergibt sich aus der Anpassung der Koeffizienten  $a_R(m)$  und  $a_I(m)$  an die für die verschiedenen Quellzustände gültigen Randbedingungen. Die Quellzustände werden dazu in Über- (OL) und Unterlastzustände (UL) unterteilt:

$$\mathbf{l} = (l_1, l_2, ..., l_N) \in \begin{cases} \text{OL} & \text{für } \sum_{n=1}^{N} r_n(l_n) > C \\ \text{UL} & \text{für } \sum_{n=1}^{N} r_n(l_n) < C \end{cases}$$
(A.33)

In den Überlastzuständen übersteigt die Ankunftsrate die Bedienrate, weshalb der Puffer niemals leer sein kann, d.h. die Verbundwahrscheinlichkeit  $F(\boldsymbol{l},x)$  verläuft an der Stelle x=0 stetig. Demgegenüber ist der Puffer mit einer endlichen Wahrscheinlichkeit voll belegt, was sich in einem Sprung an der Stelle x=S ausdrückt. Ebenso kann in Unterlastzuständen, für welche die Ankunftsrate kleiner als die Bedienrate ist, der Puffer niemals voll sein. Daher tritt in diesen Quellzuständen bei x=0 eine Unstetigkeitsstelle auf. Der typische Verlauf von  $F(\boldsymbol{l},x)$  für Über- und Unterlastzustände ist in den Bildern A.3 und A.4 qualitativ skizziert.



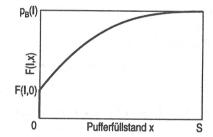


Bild A.3: Verbundwahrscheinlichkeit  $F(\boldsymbol{l},x)$  Bild A.4: Verbundwahrscheinlichkeit  $F(\boldsymbol{l},x)$  in Überlastzuständen in Unterlastzuständen

Mit den obigen Betrachtungen lassen sich folgende Randbedingungen formulieren:

$$F(\boldsymbol{l},0) = 0$$
 für  $\boldsymbol{l} \in \text{OL}$  und  $F(\boldsymbol{l},S^-) = p_B(\boldsymbol{l})$  für  $\boldsymbol{l} \in \text{UL}$ . (A.34)

Da die Gleichheit zwischen Ankunfts- und Bedienrate ausgenommen wurde, ergeben sich genau so viele Randbedingungen, wie zur eindeutigen Bestimmung der gesuchten Koeffizienten  $a_R(m)$  und  $a_I(m)$  notwendig sind. Durch die Auflösung des entsprechenden linearen Gleichungssystems erhält man die Lösung des Randwertproblems.

#### A.6 Charakteristische Leistungskenngrößen

Aus der in Gleichung A.32 angegebenen Verbundverteilung  $F(\boldsymbol{l},x)$  lassen sich nun die interessierenden Leistungskenngrößen bestimmen. Als besonders wichtiger Netzgüteparameter

erweist sich die Verlustwahrscheinlichkeit. Falls die Ankunftsrate die Bedienrate übersteigt  $(l \in OL)$  und der Puffer voll belegt ist, geht die Differenz aus der Ankunfts- und Bedienrate verloren (vgl. Bild A.5). Aus der mit der Pufferüberlaufwahrscheinlichkeit  $p_B(l) - F(l, S^-)$  gewichteten Summation der Verlustraten aller Überlastzustände ergibt sich die mittlere Verlustrate. Die Verlustwahrscheinlichkeit errechnet sich aus dem Verhältnis zwischen mittlerer Verlustrate und mittlerer Ankunftsrate  $\bar{r}$ :

$$B_B = \frac{1}{r} \sum_{\mathbf{l} \in \text{OL}} \left( \sum_{j=1}^N r_j(l_j) \right) \cdot \left( 1 - \frac{C}{\sum_{j=1}^N r_j(l_j)} \right) \cdot \left[ p_B(\mathbf{l}) - F\left(\mathbf{l}, S^-\right) \right] . \tag{A.35}$$

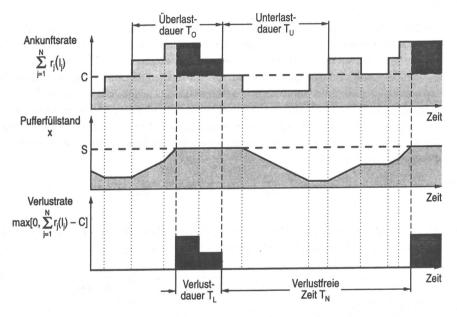


Bild A.5: Stichprobe für den Zustandsprozeß des "Fluid Flow"-Modells

Die Verluste verteilen sich auf die einzelnen Verbindungen proportional zu ihrem momentanen Verkehrsanteil. Die Verlustwahrscheinlichkeit von Verkehrsquelle n ergibt sich demnach zu

 $B_{B,n} = \frac{1}{\overline{r}_n} \sum_{\boldsymbol{l} \in \text{OL}} r_n(l_n) \cdot \left( 1 - \frac{C}{\sum_{j=1}^N r_j(l_j)} \right) \cdot \left[ p_B(\boldsymbol{l}) - F\left(\boldsymbol{l}, S^-\right) \right] . \tag{A.36}$ 

Aus den beiden letzten Gleichungen ist ersichtlich, daß der folgende Zusammenhang zwischen den Verlustwahrscheinlichkeiten der Einzelquellen und der gesamten Verlustwahrscheinlichkeit besteht:

 $\overline{r} \cdot B_B = \sum_{n=1}^N \overline{r}_n \cdot B_{B,n} . \tag{A.37}$ 

Abschließend soll das Korrelationsverhalten der Verluste bestimmt werden. Diese Korrelation drückt sich dadurch aus, daß das System zwischen Phasen unterschiedlicher Verlustwahrscheinlichkeit alterniert (vgl. Bild A.5). Falls der Puffer nicht voll belegt ist, treten überhaupt keine Verluste auf, wogegen bei vollem Puffer ein ganzes Büschel von Zellen verloren geht. Die Verlustphase umfaßt alle Systemzustände für die  $l \in OL$  und x = S gilt. In allen anderen Systemzuständen treten keine Verluste auf. Die Wahrscheinlichkeit  $p_L$ , daß sich das System in einer Verlustphase befindet, ergibt sich zu

$$p_L = \sum_{\mathbf{l} \in \text{OL}} p_B(\mathbf{l}) - F(\mathbf{l}, S^-) . \tag{A.38}$$

Von besonderer Bedeutung ist die mittlere Dauer der Verlustphase und der zeitliche Abstand zweier aufeinanderfolgender Verlustphasen. Im ersten Schritt soll die Übergangsrate  $\lambda(N|V)$  von der Verlustphase in die verlustfreie Phase bestimmt werden. Dieser Phasenübergang wird dadurch ausgelöst, daß eine Quelle ihre Zellrate so reduziert, daß ein Übergang in einen Unterlastzustand stattfindet. Da die Zustandsdauern der Quellen negativ-exponentiell verteilt sind, gelten für diese Übergänge die Kolmogoroffschen Gleichungen. Im statistischen Gleichgewicht können die Übergangsraten gewichtet mit der jeweiligen Auftrittswahrscheinlichkeit des Ausgangszustands aufsummiert werden, woraus sich der folgende Zusammenhang für die Übergangsrate  $\lambda(N|V)$  ableitet:

$$p_L \cdot \lambda(N|V) = \sum_{l \in OL} \left[ p_B(l) - F\left(l, S^-\right) \right] \cdot \sum_{m \in III.} Q(l, m) . \tag{A.39}$$

Die Größe Q(l,m) bezeichnet die Übergangsrate vom Überlastzustand l in den Unterlastzustand m. Die mittlere Dauer einer beliebigen Verlustphase ergibt sich damit zu

$$t_{L} = \mathbb{E}[T_{L}] = \frac{1}{\lambda(N|V)} = \frac{p_{L}}{\sum_{\boldsymbol{l} \in \text{OL}} \left[ p_{B}(\boldsymbol{l}) - F\left(\boldsymbol{l}, S^{-}\right) \right] \cdot \sum_{\boldsymbol{m} \in \text{UL}} \boldsymbol{Q}(\boldsymbol{l}, \boldsymbol{m})} . \tag{A.40}$$

Da die Verlustphase und die verlustfreie Phase im statistischen Gleichgewicht stehen, gilt

$$p_L \cdot \lambda(N|V) = (1 - p_L) \cdot \lambda(V|N) , \qquad (A.41)$$

woraus sich schließlich die mittlere Dauer der verlustfreien Phase zu

$$t_N = \mathbb{E}[T_N] = \frac{1}{\lambda(V|N)} = \frac{1 - p_L}{\sum_{\boldsymbol{l} \in \text{OL}} \left[ p_B(\boldsymbol{l}) - F\left(\boldsymbol{l}, S^-\right) \right] \cdot \sum_{\boldsymbol{m} \in \text{UL}} \boldsymbol{Q}(\boldsymbol{l}, \boldsymbol{m})}$$
(A.42)

errechnet.

Eine weitere interessante Leistungskenngröße ist die Verlustwahrscheinlichkeit  $B_L$  innerhalb der Verlustphase. Dazu muß die mittlere Verlustrate (vgl. Gleichung A.35) auf die von den

Verlustphasen erzeugte mittlere Ankunftsrate bezogen werden:

$$B_{L} = \frac{\sum_{\mathbf{l} \in OL} \left( \sum_{j=1}^{N} r_{j}(l_{j}) \right) \cdot \left( 1 - \frac{C}{\sum_{j=1}^{N} r_{j}(l_{j})} \right) \cdot \left[ p_{B}(\mathbf{l}) - F\left(\mathbf{l}, S^{-}\right) \right]}{\sum_{\mathbf{l} \in OL} \left( \sum_{j=1}^{N} r_{j}(l_{j}) \right) \cdot \left[ p_{B}(\mathbf{l}) - F\left(\mathbf{l}, S^{-}\right) \right]}$$
(A.43)

Aus ähnlichen Überlegungen kann die entsprechende Verlustwahrscheinlichkeit für eine beliebige Verkehrsquelle n hergeleitet werden:

$$B_{L,n} = \frac{\sum_{\boldsymbol{l} \in \text{OL}} r_n(l_n) \cdot \left(1 - \frac{C}{\sum_{j=1}^{N} r_j(l_j)}\right) \cdot \left[p_B(\boldsymbol{l}) - F\left(\boldsymbol{l}, S^{-}\right)\right]}{\sum_{\boldsymbol{l} \in \text{OL}} r_n(l_n) \cdot \left[p_B(\boldsymbol{l}) - F\left(\boldsymbol{l}, S^{-}\right)\right]}$$
(A.44)

Für die Berechnung weiterer charakteristischer Größen sei auf [218] verwiesen.

### A.7 Einfache Näherungsmodelle

Bei der numerischen Auswertung der exakten Lösung können die nachfolgend erörterten Probleme auftreten. Zum einen wächst die Größe des Zustandsraums exponentiell mit der Anzahl der Verkehrsklassen und polynomial mit der Anzahl der Quellzustände an. Mit vernünftigem numerischen Aufwand lassen sich daher nur homogene Verkehrsmischungen mit bis zu vier Quellzuständen bzw. heterogene Verkehrsmischungen mit jeweils bis zu zwei Quellzuständen behandeln. Ferner müssen für endliche Puffergröße auch die positiven Eigenwerte z(l) berücksichtigt werden. Dies wirkt sich ungünstig auf die numerische Stabilität der Lösung aus, da die Einzelterme  $\exp\left[z(l)\cdot x\right]$  stark variieren und bei großen Puffergrößen ein numerischer Überlauf auftreten kann [236]. Die Einführung sinnvoller Näherungsannahmen kann hier Abhilfe schaffen.

Da der Puffer hauptsächlich zum Ausgleich kurzzeitiger Verkehrsschwankungen dient, kann er für die Betrachtungen auf Büschelebene häufig ganz vernachlässigt werden (siehe z.B. [125, 137, 144, 160, 174]). Für dieses einfache Modell kann der Füllstand des Puffers nur den festen Wert  $x\equiv S=0$  annehmen. Der Systemzustand ist damit durch den Zustand der Verkehrsquellen vollkommen bestimmt. Mit den in Gleichung A.34 formulierten Randbedingungen ergibt sich die bei der Bestimmung der charakteristischen Größen verwendete Verbundwahrscheinlichkeit für den Quellzustand  $\boldsymbol{l}$  und den Pufferfüllstand S

$$p_{B}(\boldsymbol{l}) - F(\boldsymbol{l}, S^{-}) = \begin{cases} p_{B}(\boldsymbol{l}) & \text{für } \boldsymbol{l} \in \text{OL} \\ 0 & \text{für } \boldsymbol{l} \in \text{UL} \end{cases}$$
 (A.45)

Die weitere Auswertung dieser Lösung basiert auf den im vorigen Unterkapitel angegebenen Gleichungen für die charakteristischen Leistungskenngrößen. Eine direkte Ableitung dieser Größen folgt aus der Überlegung, daß nur dann Verluste auftreten, wenn die Ankunftsrate die Bedienrate übersteigt. In diesen Fällen geht die Differenz aus Zu- und Abflußrate verloren. Dieser direkte Ansatz wird in Unterabschnitt 6.2.2.1 verwendet, um einen effizienten Faltungsalgorithmus abzuleiten, der den Rechenaufwand weiter reduziert. Ferner wird dort gezeigt, daß diese Lösung auch für allgemein verteilte Zustandsdauern der Verkehrsquellen gültig ist.

In [4] wird der komplementäre Fall sehr großer Pufferfüllstände x diskutiert. Für unendliche Puffergröße verschwinden alle zu positiven Eigenwerten  $z(\boldsymbol{m})$  gehörenden Koeffizienten  $a(\boldsymbol{m})$ , so daß das asymptotische Systemverhalten durch den größten negativen Eigenwert  $z(\boldsymbol{m}^*)$  bestimmt ist, der daher auch als dominanter Eigenwert bezeichnet wird. In [58] wurde bewiesen, daß dieser Eigenwert reell ist und nur einfach auftritt. Mit diesen Überlegungen ergibt sich aus Gleichung A.32 ein negativ-exponentieller Abfall der komplementären Verteilungsfunktion für große Werte des Pufferfüllstands x (vgl. [4]):

$$p_B(l) - F(l, x) \approx -a(\mathbf{m}^*) \cdot \{\varphi(\mathbf{m}^*)\}_l \cdot \exp[z(\mathbf{m}^*) \cdot x] . \tag{A.46}$$

Diese asymptotische Approximation bringt für das vorliegende allgemeinere Modell allerdings kaum Vorteile, da der Koeffizient  $a(\boldsymbol{m}^*)$  aus dem durch die Randbedingungen definierten linearen Gleichungssystem bestimmt werden muß. Der hierfür erforderliche numerische Aufwand liegt in derselben Größenordnung wie die Auswertung der exakten Lösung. Andererseits scheint eine Näherungslösung basierend auf dem dominanten Eigenwert vielversprechend zu sein, da er mit Hilfe der in [58, 226] vorgeschlagenen Zerlegungsmethode (vgl. Anhang A.4) auf direktem Weg bestimmt werden kann.

Im folgenden wird ein einfacher Ansatz vorgeschlagen, bei dem das in Gleichung A.46 auftretende Produkt  $a(m^*) \cdot \{\varphi(m^*)\}_l$  durch einen frei wählbaren Faktor c(l) ersetzt wird:

$$F(l,x) \approx p_B(l) + c(l) \cdot \exp[z(\boldsymbol{m}^*) \cdot x]$$
 (A.47)

Die Faktoren c(l) werden nun so gewählt, daß die in Gleichung A.34 eingeführten Randbedingungen erfüllt sind, d.h. es gilt

$$p_B(\mathbf{l}) + c(\mathbf{l}) = 0 \quad \text{für } \mathbf{l} \in \text{OL}$$
 (A.48)

und

$$p_B(\mathbf{l}) + c(\mathbf{l}) \cdot \exp[z(\mathbf{m}^*) \cdot S] = p_B(\mathbf{l}) \quad \text{für } \mathbf{l} \in UL.$$
 (A.49)

Somit lautet die aus Gleichung A.47 abgeleitete Näherungslösung

$$F(\boldsymbol{l}, x) \approx \begin{cases} p_{B}(\boldsymbol{l}) \left(1 - \exp\left[z(\boldsymbol{m}^{*}) \cdot x\right]\right) & \text{für } \boldsymbol{l} \in \text{OL }, 0 \le x < S \\ p_{B}(\boldsymbol{l}) & \text{für } \boldsymbol{l} \in \text{UL }, 0 \le x \le S \end{cases}$$
(A.50)

Dieses Ergebnis kann in die in Unterabschnitt A.6 hergeleiteten Formeln zur Bestimmung der charakteristischen Leistungskenngrößen eingesetzt werden. Insbesondere lassen sich einfache Zusammenhänge mit dem vorher diskutierten pufferlosen System herstellen, welche den Einfluß der Netzpuffer auf die von der Büschelebene herrührenden Zellverluste aufzeigen. Beispielsweise ergibt sich aus Gleichung A.35 die folgende Näherung für die Verlustwahrscheinlichkeit:

$$B_{B} \approx \frac{1}{r} \sum_{l \in OL} \left( \sum_{j=1}^{N} r_{j}(l_{j}) - C \right) \cdot \left( p_{B}(l) - p_{B}(l) + p_{B}(l) \cdot \exp\left[ z(\boldsymbol{m}^{*}) \cdot S \right] \right)$$

$$B_{B} \approx B_{B}|_{S=0} \cdot \exp\left[ z(\boldsymbol{m}^{*}) \cdot S \right] , \qquad (A.51)$$

d.h. die Verlustwahrscheinlichkeit fällt — ausgehend vom pufferlosen System — bei einer Vergrößerung des Puffers exponentiell ab. Derselbe Zusammenhang wurde in [153] aus heuristischen Betrachtungen gewonnen. Mit Gleichung A.36 kann weiter gezeigt werden, daß dieses exponentielle Gesetz auch für die individuellen Verlustwahrscheinlichkeiten der einzelnen Verkehrsquellen gilt.

Für die weiteren Leistungskenngrößen gelten dagegen andere Zusammenhänge mit dem pufferlosen System. Nach einigen elementaren Umformungen ergeben sich aus den Gleichungen A.40, A.42 und A.43 die folgenden Approximationen für die mittlere Dauer der Verlustphase, für den mittleren zeitlichen Abstand zweier Verlustphasen und für die Verlustwahrscheinlichkeit innerhalb der Verlustphase:

$$t_L = t_L|_{S=0}$$
, (A.52)

$$t_N = [t_N|_{S=0} + t_L|_{S=0}] \cdot \exp\left[-z(\boldsymbol{m}^*) \cdot S\right] - t_L|_{S=0}$$
(A.53)

und

$$B_L = B_L|_{S=0} (A.54)$$

Diese Gleichungen verdeutlichen, daß die Verringerung der Verlustwahrscheinlichkeit allein durch die Vergrößerung des zeitlichen Abstands zweier Verlustphasen erzielt wird, d.h. der Puffer kann kurzzeitige Überlastsituationen durch eine Zwischenspeicherung der Information vollkommen ausgleichen (vgl. Bild A.5). Falls allerdings ein Pufferüberlauf auftritt, so bleibt die Dauer und die Intensität dieser Verlustsituation in erster Näherung unbeeinflußt von der Größe des Puffers. Dieses Verhalten wird in Unterkapitel 4.1.3 ausführlich erörtert.

# A.8 Untersuchung des Schwellwertverfahrens

In diesem Unterkapitel soll die oben beschriebene "Fluid Flow"-Analyse auf das in Abschnitt 5.2.2 beschriebene Schwellwertverfahren übertragen werden. Dies stellt eine Erweiterung der

in [57, 68] für diesen Prioritätsmechanismus beschriebenen Lösungsansätze dar, die sich auf die Überlagerung von zweistufigen Quellen mit identischen statistischen Eigenschaften beschränken (eine Diskussion dieser und weiterer analytischer Ansätze findet sich in Abschnitt 5.3.2).

Die folgende Analyse basiert auf den in Bild A.2 gezeigten allgemeinen Markoff-modulierten Ankunftsprozessen. Ferner wird angenommen, daß das Verhältnis zwischen hoch- und niederprioritären Zellen für jeden Quellzustand individuell vorgegeben werden kann. Der Anteil der von Verkehrsquelle n im Quellzustand l ( $l=0,1,...,L_n$ ) gesendeten hochprioritären Zellen wird mit  $v_n(l)$ ,  $0 \le v_n(l) \le 1$ , bezeichnet. Im Quellzustand l werden somit hoch- und niederprioritäre Zellen mit den Raten  $v_n(l) \cdot r_n(l)$  bzw.  $[1-v_n(l)] \cdot r_n(l)$  erzeugt. Ferner muß das in Bild A.1 dargestellte Systemmodell dahingehend erweitert werden, daß beim Überschreiten des Pufferfüllstands  $S_L$  der Zufluß von niederprioritärem Verkehr unterbunden wird. Somit entspricht die Zuflußrate für Pufferfüllstände  $S_L \le x < S$  der Ankunftsrate des hochprioritären Verkehrs.

Dies führt zu einem gegenüber Gleichung A.14 geänderten Differentialgleichungssystem innerhalb des angegebenen Pufferbereichs.

$$\frac{d\mathbf{F}(x)}{dx} = \mathbf{F}(x) \cdot \mathbf{Q} \mathbf{D}_{H}^{-1} \quad \text{für } S_{L} \le x < S .$$
(A.55)

Die Matrix  $D_H$  ergibt sich aus der Matrix D, indem die Gesamtankunftsraten  $\sum_{n=1}^{N} r_n(l_n)$  durch die Ankunftsraten  $\sum_{n=1}^{N} v_n(l_n) \cdot r_n(l_n)$  ersetzt werden. Da dieses Differentialgleichungssystem dieselbe Form hat wie das ursprüngliche Differentialgleichungssystem, läßt sich die Verbundverteilung F(l,x) direkt angeben:

$$F\left(\boldsymbol{l},\boldsymbol{x}\right) = \begin{cases} 0 & \text{für } \boldsymbol{x} < 0 \\ \sum\limits_{\boldsymbol{m}: \ z_{I}(\boldsymbol{m}) = \ 0} & a_{R}(\boldsymbol{m}) \cdot \{\varphi_{R}(\boldsymbol{m})\}_{\boldsymbol{l}} \cdot \exp\left[z_{R}(\boldsymbol{m}) \cdot \boldsymbol{x}\right] + \\ \sum\limits_{\boldsymbol{m}: \ z_{I}(\boldsymbol{m}) > \ 0} & \{2 \cdot [a_{R}(\boldsymbol{m}) \cdot \{\varphi_{R}(\boldsymbol{m})\}_{\boldsymbol{l}} - a_{I}(\boldsymbol{m}) \cdot \{\varphi_{I}(\boldsymbol{m})\}_{\boldsymbol{l}}] \cdot \\ \sum\limits_{\boldsymbol{m}: \ z_{I}(\boldsymbol{m}) > \ 0} & \exp\left[z_{R}(\boldsymbol{m}) \cdot \boldsymbol{x}\right] \cdot \cos\left[z_{I}(\boldsymbol{m}) \cdot \boldsymbol{x}\right] - \\ & 2 \cdot [a_{R}(\boldsymbol{m}) \cdot \{\varphi_{I}(\boldsymbol{m})\}_{\boldsymbol{l}} + a_{I}(\boldsymbol{m}) \cdot \{\varphi_{R}(\boldsymbol{m})\}_{\boldsymbol{l}}] \cdot \\ & \exp\left[z_{R}(\boldsymbol{m}) \cdot \boldsymbol{x}\right] \cdot \sin\left[z_{I}(\boldsymbol{m}) \cdot \boldsymbol{x}\right] + \\ \sum\limits_{\boldsymbol{n}: \ y_{I}(\boldsymbol{n}) = \ 0} & \{2 \cdot [b_{R}(\boldsymbol{n}) \cdot \{\gamma_{R}(\boldsymbol{n})\}_{\boldsymbol{l}} \cdot \exp\left[y_{R}(\boldsymbol{n}) \cdot \boldsymbol{x}\right] + \\ \sum\limits_{\boldsymbol{n}: \ y_{I}(\boldsymbol{n}) > \ 0} & \{2 \cdot [b_{R}(\boldsymbol{n}) \cdot \{\gamma_{R}(\boldsymbol{n})\}_{\boldsymbol{l}} - b_{I}(\boldsymbol{n}) \cdot \{\gamma_{I}(\boldsymbol{n})\}_{\boldsymbol{l}}] \cdot \\ & n: \ y_{I}(\boldsymbol{n}) > \ 0} & \exp\left[y_{R}(\boldsymbol{n}) \cdot \boldsymbol{x}\right] \cdot \cos\left[y_{I}(\boldsymbol{n}) \cdot \boldsymbol{x}\right] - \\ & 2 \cdot [b_{R}(\boldsymbol{n}) \cdot \{\gamma_{I}(\boldsymbol{n})\}_{\boldsymbol{l}} + b_{I}(\boldsymbol{n}) \cdot \{\gamma_{R}(\boldsymbol{n})\}_{\boldsymbol{l}}] \cdot \\ & \exp\left[y_{R}(\boldsymbol{n}) \cdot \boldsymbol{x}\right] \cdot \sin\left[y_{I}(\boldsymbol{n}) \cdot \boldsymbol{x}\right] \end{cases} \qquad \text{für } \boldsymbol{x} \geq \boldsymbol{S} ,$$

$$\left(\boldsymbol{A}.56\right)$$

mit

$$y(\boldsymbol{n}) = y_R(\boldsymbol{n}) + i \cdot y_I(\boldsymbol{n})$$
:  $\boldsymbol{n}$ -ter (komplexer) Eigenwert der Matrix  $\boldsymbol{Q}\boldsymbol{D}_H^{-1}$ ,  $\{\boldsymbol{\gamma}(\boldsymbol{n})\}_{\boldsymbol{l}} = \{\boldsymbol{\gamma}_R(\boldsymbol{n})\}_{\boldsymbol{l}} + i \cdot \{\boldsymbol{\gamma}_I(\boldsymbol{n})\}_{\boldsymbol{l}}$ :  $\boldsymbol{l}$ -tes Element des zugehörigen linksseitigen Eigenvektors und  $b(\boldsymbol{n}) = b_R(\boldsymbol{n}) + i \cdot b_I(\boldsymbol{n})$ : zugeordneter frei wählbarer Koeffizient.

Die Koeffizienten  $a(\boldsymbol{m})$  und  $b(\boldsymbol{n})$  müssen aus den entsprechenden Randbedingungen ermittelt werden. Die für die Pufferbereiche  $x \in [0, S_L)$  und  $x \in [S_L, S)$  geltenden Randbedingungen sind in Bild A.6 für verschiedene Lastzustände graphisch veranschaulicht (vgl. auch [57]).

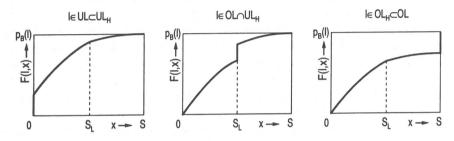


Bild A.6: Verbundverteilung  $F(\boldsymbol{l},x)$  für verschiedene Lastzustände

Zunächst werden die Pufferfüllstände  $x \in [0, S_L)$  betrachtet, für die keine Verluste auftreten. Falls die gesamte Ankunftsrate die Bedienrate übersteigt, kann das System nicht leer sein, d.h. es gilt

$$F(\boldsymbol{l},0) = 0 \quad \text{für} \quad \boldsymbol{l} \in \text{OL} . \tag{A.57}$$

In allen anderen Quellzuständen liegt die Gesamtankunftsrate unter der Bedienrate und der gemeinsam genutzte Teil des Puffers kann nicht voll belegt sein. Deshalb verläuft die Verbundverteilung F(l,x) stetig an der Stelle  $x=S_L$ :

$$F(\boldsymbol{l}, S_L^-) = F(\boldsymbol{l}, S_L)$$
 für  $\boldsymbol{l} \in \text{UL}$ . (A.58)

Im Teilbereich  $x \in [S_L, S)$  entspricht die Zuflußrate der Ankunftsrate der hochprioritären Zellen. Daher müssen die Über- und Unterlastzustände in diesem Pufferbereich neu definiert werden:

$$\mathbf{l} = (l_1, l_2, ..., l_N) \in \begin{cases} \text{OL}_H & \text{für } \sum_{n=1}^N v_n(l_n) \cdot r_n(l_n) > C \\ \text{UL}_H & \text{für } \sum_{n=1}^N v_n(l_n) \cdot r_n(l_n) < C \end{cases}$$
(A.59)

In den Überlastzuständen kann der für die hohe Priorität reservierte Pufferbereich nicht leer sein, weshalb die Verbundverteilung  $F(\mathbf{l}, x)$  an der Stelle  $x = S_L$  stetig ist:

$$F(\boldsymbol{l}, S_L) = F(\boldsymbol{l}, S_L^-) \quad \text{für} \quad \boldsymbol{l} \in \text{OL}_H.$$
 (A.60)

Im umgekehrten Fall kann der Puffer nicht voll belegt sein, da die Ankunftsrate des hochprioritären Verkehrs unter der Bedienrate liegt. Dieser Sachverhalt wird durch die folgende Gleichung ausgedrückt:

$$F(\boldsymbol{l}, S^{-}) = p_{\boldsymbol{B}}(\boldsymbol{l}) \quad \text{für} \quad \boldsymbol{l} \in \text{UL}_{\boldsymbol{H}}. \tag{A.61}$$

Da für jeden Quellzustand zwei unterschiedliche Randbedingungen existieren, ergeben sich genügend Gleichungen zur Bestimmung der Koeffizienten a(m) und b(n).

Aus der Verbundverteilung  $F(\boldsymbol{l},x)$  können alle relevanten Leistungskenngrößen abgeleitet werden. Die prinzipielle Vorgehensweise soll exemplarisch anhand der Verlustwahrscheinlichkeit erläutert werden. Zellen mit hoher Priorität gehen nur dann verloren, wenn der Puffer voll belegt ist und die Ankunftsrate der hochprioritären Zellen die Bedienrate übersteigt. Die entsprechende Verlustwahrscheinlichkeit  $B_{B,H}$  kann aus Gleichung A.35 berechnet werden, indem die gesamte Ankunftsrate durch die Ankunftsrate der hohen Priorität ersetzt wird:

$$B_{B,H} = \frac{1}{\overline{r}_H} \sum_{\boldsymbol{l} \in \text{OL}_H} \left( \sum_{n=1}^N v_n(l_n) \cdot r_n(l_n) - C \right) \cdot \left[ p_B(\boldsymbol{l}) - F\left(\boldsymbol{l}, S^-\right) \right] \ . \tag{A.62}$$

Darin bezeichnet  $\overline{r}_H$  die mittlere Ankunftsrate der hochprioritären Zellen, die sich aus Gleichung A.7 ergibt, indem die Gesamtankunftsrate durch die Ankunftsrate der hohen Priorität ersetzt wird:

$$\overline{r}_{H} = \sum_{n=1}^{N} \sum_{l=0}^{L_{n}} p_{B,n}(l) \cdot v_{n}(l) \cdot r_{n}(l) . \tag{A.63}$$

Die innerhalb der niederprioritären Klasse auftretenden Verluste setzen sich aus zwei verschiedenen Anteilen zusammen (vgl. auch [57]):

- Beim Pufferfüllstand  $S_L$  geht die Differenz aus der Ankunftsrate  $\sum_{n=1}^{N} r_n(l_n)$  und der Bedienrate C verloren. Dieser Pufferfüllstand tritt mit der Wahrscheinlichkeit  $F(l, S_L) F(l, S_L^-)$  auf.
- Falls der Pufferfüllstand den Wert  $S_L$  überschreitet, geht der gesamte niederprioritäre Verkehr verloren, d.h. die Verlustrate entspricht  $\sum_{n=1}^{N} [1 v_n(l_n)] \cdot r_n(l_n)$ . Diese Pufferfüllstände treten mit der Wahrscheinlichkeit  $p_B(l) F(l, S_L)$  auf.

Mit diesen Betrachtungen läßt sich die Verlustwahrscheinlichkeit des niederprioritären Verkehrs angeben:

$$B_{B,L} = \frac{1}{\overline{r}_L} \sum_{\boldsymbol{l} \in \text{OL} \cap \text{UL}_{\boldsymbol{H}}} \left( \sum_{n=1}^{N} r_n(l_n) - C \right) \cdot \left[ F(\boldsymbol{l}, S_L) - F(\boldsymbol{l}, S_L^-) \right] + \frac{1}{\overline{r}_L} \sum_{\boldsymbol{l}} \left( \sum_{n=1}^{N} \left[ 1 - v_n(l_n) \right] \cdot r_n(l_n) \right) \cdot \left[ p_B(\boldsymbol{l}) - F(\boldsymbol{l}, S_L) \right] , \tag{A.64}$$

mit

$$\overline{r}_L = \overline{r} - \overline{r}_H \ . \tag{A.65}$$

Abschließend sollen die im vorigen Unterkapitel besprochenen Approximationsmethoden auf das vorliegende Problem angewandt werden. Zunächst wird der Sonderfall  $S, S_L \to 0$  betrachtet, für welchen sich die zur Bestimmung der charakteristischen Leistungskenngrößen benötigten Verbundwahrscheinlichkeiten formal aus den obigen Randbedingungen ergeben. Dies führt zu den folgenden, auch aus der Anschauung ableitbaren Zusammenhängen:

$$\lim_{S_L \to 0} F(\boldsymbol{l}, S_L^-) = \begin{cases} 0 & \text{für } \boldsymbol{l} \in \text{OL} \\ p_B(\boldsymbol{l}) & \text{für } \boldsymbol{l} \in \text{UL} \end{cases}$$
(A.66)

und

$$\lim_{S_L \to 0} F(\boldsymbol{l}, S_L) = \lim_{S \to 0} F(\boldsymbol{l}, S^-) = \begin{cases} 0 & \text{für } \boldsymbol{l} \in \text{OL}_H \\ p_B(\boldsymbol{l}) & \text{für } \boldsymbol{l} \in \text{UL}_H \end{cases}$$
 (A.67)

Mit diesen Randwerten können die Gleichungen A.62 und A.64 ausgewertet werden, woraus sich die Verlustwahrscheinlichkeiten  $B_{B,H}|_{S=0}$  und  $B_{B,L}|_{S=0}$  des pufferlosen Systems errechnen.

Die Genauigkeit der eben beschriebenen Approximation kann signifikant verbessert werden, indem die dominanten Eigenwerte  $z(\boldsymbol{m}^*)$  und  $y(\boldsymbol{n}^*)$  einbezogen werden. Dazu muß der in Gleichung A.47 beschriebene Ansatz auf die vorliegende Problemstellung übertragen werden:

$$F(\boldsymbol{l}, x) \approx \begin{cases} p_B(\boldsymbol{l}) + c(\boldsymbol{l}) \cdot \exp\left[z(\boldsymbol{m}^*) \cdot x\right] & \text{für } 0 \le x < S_L \\ p_B(\boldsymbol{l}) + d(\boldsymbol{l}) \cdot \exp\left[y(\boldsymbol{n}^*) \cdot x\right] & \text{für } S_L \le x < S \end{cases}.$$
(A.68)

Die unbekannten Faktoren c(l) und d(l) können mit Hilfe der durch die Gleichungen A.57–A.61 gegebenen Randbedingungen eliminiert werden, womit sich folgende Näherungslösung angeben läßt:

$$F(\boldsymbol{l},x) \approx \begin{cases} p_{B}(\boldsymbol{l}) \cdot \{1 - \exp\left[z(\boldsymbol{m}^{*}) \cdot x\right]\} & \text{für } 0 \leq x < S_{L}, \ \boldsymbol{l} \in \text{OL} \\ p_{B}(\boldsymbol{l}) & \text{für } 0 \leq x < S_{L}, \ \boldsymbol{l} \in \text{UL} \\ p_{B}(\boldsymbol{l}) \cdot \{1 - \exp\left[z(\boldsymbol{m}^{*}) \cdot S_{L}\right] \cdot \exp\left[y(\boldsymbol{n}^{*}) \cdot (x - S_{L})\right]\} & \text{für } S_{L} \leq x < S, \ \boldsymbol{l} \in \text{OL}_{H} \\ p_{B}(\boldsymbol{l}) & \text{für } S_{L} \leq x < S, \ \boldsymbol{l} \in \text{UL}_{H}. \end{cases}$$

$$(A.69)$$

Falls diese Lösung in die Gleichungen A.62 und A.64 eingesetzt wird, erhält man nach einigen elementaren Umformungen die nachstehenden Zusammenhänge für die Verlustwahrscheinlichkeiten der hoch- und niederprioritären Verkehrsströme:

$$B_{B,H} \approx B_{B,H}|_{S=0} \cdot \exp\left[z(\boldsymbol{m}^*) \cdot S_L\right] \cdot \exp\left[y(\boldsymbol{n}^*) \cdot (S - S_L)\right]$$
 (A.70)

und

$$B_{B,L} \approx B_{B,L}|_{S=0} \cdot \exp\left[z(\boldsymbol{m}^*) \cdot S_L\right]$$
 (A.71)

