

Institut für Nachrichtenvermittlung und Datenverarbeitung
der Universität Stuttgart
Prof. Dr.-Ing. P. Kühn

43. Bericht über verkehrstheoretische Arbeiten

Leistungsuntersuchung
von Multi-Bus-Verbindungsnetzwerken
in lose gekoppelten Systemen

von
Thomas Raith

Institute of Communications Switching and Data Technics
University of Stuttgart
Prof. Dr.-Ing. P. Kühn

43rd Report on Studies in Congestion Theory

Performance Analysis
of Multi-Bus Interconnection Networks
of Loosely Coupled Systems

by
Thomas Raith

© 1988 Institut für Nachrichtenvermittlung und Datenverarbeitung Universität Stuttgart

Druck: E. Kurz & Co., Druckerei + Reprografie GmbH.

ISBN 3-922403-53-0

PERFORMANCE ANALYSIS OF MULTI-BUS INTERCONNECTION NETWORKS
OF LOOSELY COUPLED SYSTEMS

SUMMARY

The subject of this report is the performance analysis of message interchanging mechanisms via a multi-bus interconnection network which operates according to a cyclic scheduling strategy. The investigation is based on a multi-server polling system analysis with cyclic, nonexhaustive service, whereby the realistic assumptions of finite buffer capacity in the sending and the receiving direction of the stations, as well as port interference, are taken into account.

In general, the approximate performance analysis is based on decomposition methods and an imbedded Markov chain approach in conjunction with a two-moment approximation for the inter-visit time density function. The validation of the approximation is done by means of computer simulations.

The class of multi-bus interconnection networks considered in this report arises in modularized packet switching nodes, token-ring local area networks, and distributed systems of multiple interconnected computers.

Chapter 1 Introduction

In the first chapter the basic structure of networks used for tight or loose interconnection of several peripheral devices or processing units is presented. The performance degradation of the interconnection networks due to port or buffer blocking effects are briefly discussed.

Chapter 2 Architecture of Multi-Bus Systems

In Chapter 2 the architecture of multi-bus systems for loose interconnection of autonomous units in systems with centralized or decentralized control is considered. The units operate according to principles of load and function sharing and communicate with each other by addressed messages via their transmit and receive buffers which are connected to the multi-bus network through an individual full duplex port. The bus allocation to each particular unit is organized by means of a centralized or decentralized scheduling scheme under consideration of the buffer blocking conditions and port interference.

Chapter 3 Models with Cyclic Service

This chapter aims at the modelling of a communication subsystem consisting of a single- or multi-bus interconnection network and finite transmit and receive buffers per unit. The cyclic allocation of a bus to the various units is modelled by a multi-server polling system with ordinary non-exhaustive cyclic service.

The arrival process of messages generated by the units is assumed to be Poissonian with buffer-individual arrival rates. Each receive buffer will be served by a Markovian server. Message transmission times are considered to be generally distributed. After service of a transmit buffer the bus is allocated by the scheduler to the succeeding unit in a finite and generally distributed allocation or switchover time to model all overheads spent and procedures performed by the scheduler. In case of port interference at a transmit buffer the bus will be allocated to the following transmit buffer in an additional allocation time, i.e. the transmit buffer can only be served by one bus at a time; therefore, overtaking of bus grants may occur. If message blocking due to port interference or buffer limitation of the receiver occurs, different bus scheduling modes will be discussed.

Chapter 4 Modelling and Analysis Methods

In chapter 4 the basic features and functions of random variables and stochastic processes applied in the performance analysis of multi-bus interconnection networks are presented. The chapter concludes with a short introduction of the imbedded Markov chain approach which is used for the buffer state analysis, and by discussing some measures for the system characteristics and its performance.

Chapter 5 Analysis of Multi-Server Polling Models With Finite Buffers

In this chapter an approximate analytical method for multi-server polling systems with finite buffer capacity and non-exhaustive cyclic service is presented, whereby a numerical algorithm is developed.

The analytical method is based on an alternating iteration between transmit buffer state analysis (section 5.2) and a two-moment approximation of the intervisit time density function (section 5.3). For the calculation of the intervisit time density function the server interference at the transmit buffer ports is taken into account.

The accuracy of the method is illustrated and discussed by means of some numerical results which have been obtained for symmetrically as well as for nonsymmetrically loaded systems.

Chapter 6 Feedback Effect Caused By Receiver Constraints

In this chapter the effect of receiver constraints caused by port interference or buffer limitation has been investigated by combining the multi-server polling system analysis of the sending part (chapter 5) with a feedback caused by

the receiver constraints. In Chapter 6 one of the scheduling modes discussed above is investigated.

In the investigated scheduling mode a message is transmitted without considering any blocking conditions at the receiver, i.e. in case the transmission might fail due to receive buffer blocking or port interference the message keeps waiting within its transmit buffer until the next bus grant occurs. At the next server visit instant the message transmission will be repeated.

For systems with symmetrical load some numerical results are given in order to demonstrate the accuracy of the approximate method if receiver constraints are considered.

INHALTSVERZEICHNIS

Verzeichnis der wichtigsten Abkürzungen und Formelzeichen

	Seite
1 EINLEITUNG	9
1.1 Leistungssteigerung durch Parallelarbeit	9
1.2 Prinzipien der Rechnerkopplung	10
1.2.1 Die enge Kopplung	10
1.2.2 Die lose Kopplung	12
1.3 Strukturen von Verbindungsnetzwerken	13
1.4 Rechner-zu-Rechner Kommunikation	15
1.5 Verkehrstheoretische Problemstellungen	16
1.6 Aufgabenstellung und Übersicht über die Arbeit	17
2 ARCHITEKTUR VON MULTI-BUS-VERBINDUNGSNETZWERKEN	20
2.1 Allgemeines	20
2.2 Netzwerkstruktur	20
2.3 Buszugriffsverfahren	22
2.3.1 Zentralgesteuerte Zugriffsverfahren	22
2.3.1.1 Polling-Verfahren	23
2.3.1.2 Daisy-Chain-Verfahren	24
2.3.2 Verteilt gesteuerte Zugriffsverfahren	25
2.3.2.1 Konkurrenz-Verfahren	25
2.3.2.2 Zuteilungs-Verfahren	26
2.4 Einschränkungen der Verkehrsleistung	27
2.4.1 Blockierungen und Verluste bei endlichen Kommunikationsspeichern	28
2.4.2 Blockierung durch Halb-Duplex-Betrieb	28
2.4.3 Interferenzen am Netzwerkport	29
2.4.4 Interferenzen beim Buszugriff	29

3 SYSTEMMODELLE MIT ZYKLISCHEN ZUGRIFFSVERFAHREN	30
3.1 Allgemeines zur Modellierung	30
3.1.1 Der Modellbegriff	30
3.1.2 Elemente zur Modellbildung	30
3.1.3 Beschreibung des Ablaufgeschehens	31
3.2 Modellierung von zyklischen Zugriffsverfahren	32
3.2.1 Modellbeschreibung	32
3.2.2 Bedienstrategien	33
3.2.3 Zuteilungsstrategien	34
3.3 Modell des Multi-Bus-Verbindungsnetzwerkes	35
3.3.1 Modellbeschreibung	35
3.3.2 Blockierungen und Interferenzen des Verbindungsnetzwerkes	37
3.3.3 Ablaufsteuerung des Verbindungsnetzwerkes	38
3.4 Anwendungen der Systemmodelle	39
3.4.1 Rechner/Terminal-Kommunikation	39
3.4.2 Steuerungen in Fernsprech-Vermittlungssystemen	41
3.4.3 Lokale Netze mit Token-gesteuertem Kanal-Zugriff	42
3.4.4 Echtzeitbetriebssysteme mit Mehrprozeßbetrieb	43
3.4.5 Multi-Bus-Verbindungssysteme für modulare Rechensysteme	44
4 WAHRSCHEINLICHKEITSTHEORETISCHE MODELLANALYSE	45
4.1 Methoden der Systemanalyse	45
4.1.1 Messung	45
4.1.2 Simulation	46
4.1.3 Mathematische Analyse	47
4.2 Statistische Merkmale und Eigenschaften zufallsabhängiger Größen	48
4.2.1 Zufallsvariable	48
4.2.2 Verteilungsfunktion	49
4.2.2.1 Momente und Maßzahlen	50

4.2.2.2 Phasenmodelle nach Cox	52
4.2.2.3 Momentenapproximation	52
4.2.3 Funktionen von Zufallsvariablen	54
4.2.3.1 Statistische Unabhängigkeit	55
4.2.3.2 Summation von Zufallsvariablen	56
4.2.3.3 Differenz von unabhängigen Zufallsvariablen	57
4.3 Stochastische Prozesse	58
4.3.1 Definition	58
4.3.2 Eigenschaften	59
4.3.3 Markoff-Prozesse	60
4.3.4 Punktprozesse	61
4.3.5 Poisson-Prozesse	62
4.3.6 Erneuerungsprozesse	63
4.3.6.1 Rekurrenzzeiten	63
4.3.6.2 Überlagerung und Zerlegung	64
4.4 Methode der eingebetteten Markoff-Kette	67
4.5 Bewertung von Systemmodellen	68
4.5.1 Charakteristische Größen	68
4.5.2 Das Little'sche Theorem	70
5 ANALYSE VON MULTI-BUS-VERBINDUNGSNETZWERKEN MIT ENDLICHEN SENDEWARTESCHLANGEN	72
5.1 Literaturübersicht	72
5.2 Modellannahmen und Modellparameter	73
5.3 Wahrscheinlichkeitsverteilung der Systemzustände	74
5.3.1 Verteilung zu den Regenerationszeitpunkten	75
5.3.2 Verteilung zu beliebigen Zeitpunkten	78
5.3.3 Momentenapproximation der Besuchszeitverteilungsfunktion	79
5.4 Analyse der Besuchzeitverteilungsfunktion	81
5.4.1 Momente des Zykluszeitsegments	82
5.4.2 Der Zykluszeitprozeß	83

5.4.3	Der Zuteilungsprozeß	84
5.4.4	Sendeportinterferenzen	86
5.4.5	Der Besuchszeitprozeß	87
5.4.6	Prinzip der bedingten Zykluszeiten	88
5.4.6.1	Momente der bedingten Zykluszeiten	89
5.4.6.2	Einfluß der bedingten Zykluszeiten auf die Verteilung der Systemzustände	90
5.5	Ergebnisse und Approximationsgenauigkeit	92
5.5.1	Einfluß der Stationszahl	92
5.5.2	Dimensionierung der Sendewarteschlangen	97
5.5.3	Multi-Bus-Verbindungsnetzwerke	98
5.5.4	Systeme mit unsymmetrischen Verkehrsangebot	100
6	ANALYSE EINES MULTI-BUS-SYSTEMS UNTER EINSCHRÄNKUNG DER VERKEHRSLAUFLEISTUNG DURCH DEN EMPFÄNGER	103
6.1	Modellierungsaspekte	103
6.2	Analyse des Empfängers einer Station	105
6.2.1	Wahrscheinlichkeitsverteilung der Systemzustände zu den Regenerationszeitpunkten	105
6.2.2	Nachrichtenantrittsprozeß am Empfänger	107
6.3	Wahrscheinlichkeit für Blockierung oder Interferenz	109
6.3.1	Empfangsspeicherblockierung	109
6.3.2	Empfangsportinterferenz	112
6.4	Der Besuchszeitprozeß bei Portinterferenz und Speicherblockierung	113
6.5	Ergebnisse	114
6.5.1	Vergleich von Uni- und Doppel-Bus-System	114
6.5.2	Leistungsverhalten des Paketvermittlungsknotens EWSP	117
6.5.2.1	Knotenarchitektur	117
6.5.2.1	Leistung des Verbindungsnetzwerkes	119

7	ZUSAMMENFASSUNG	123
8	LITERATURVERZEICHNIS	124
	ANHÄNGE	134
	Anhang 1: Algorithmus für die Analyse nach Kapitel 5	134
	Anhang 2: Algorithmus für die Analyse nach Kapitel 6	135
	LEBENS LAUF	136

VERZEICHNIS DER WICHTIGSTEN ABKÜRZUNGEN, FORMELZEICHEN UND ZUFALLSVARIABLEN

Abkürzungen

- ZV : Zufallsvariable
- VF : Verteilungsfunktion
- VDF : Verteilungsdichtefunktion
- LT : Laplace-Transformation
- LST : Laplace-Stieltjes-Transformation
- D : Konstant verteilte Phase
- M : Negativ-exponentiell verteilte Phase
- s_j : Speicherkapazität der Sendewarteschlange j
- r_j : Speicherkapazität der Empfangswarteschlange j
- g : Anzahl der Stationen
- n : Anzahl der Busse
- $t^{(n)}$: Regenerationszeitpunkt
- B_{ps} : Sendeportinterferenzwahrscheinlichkeit
- B_{pe} : Empfangsportinterferenzwahrscheinlichkeit
- B_m : Empfangsspeicherblockierwahrscheinlichkeit

Formelzeichen

- λ : Ankunftsrate
- ϵ : Bedien- oder Enderate
- ρ : Verkehrsangebot

Zufallsvariable

- T_A : Ankunftsabstand (Inter Arrival Time)
- T_C : Zykluszeit (Cycle Time)
- T_D : Ausgangsabstand (Inter Departure Time)
- T_F : Durchlaufzeit (Flow Time)

- T_G : Zuteilungsabstand (Inter Grant Time)
- T_H : Busbelegungsdauer (Holding Time)
- T_I : Eingangsabstand (Inter Input Time)
- T_O : Verwaltungsdauer (Change Over Time)
- T_P : Prozessorbelegungsdauer (Processing Time)
- T_S : Segmentdauer (Cycle Segment Time)
- T_V : Besuchszeitabstand (Inter Visit Time)
- T_W : Wartezeit (Waiting Time)

Definitionen zur Zufallsvariablen T_X

- $F_X(t)$: Verteilungsfunktion
- $f_X(t)$: Verteilungsdichtefunktion
- $\Phi_X(s)$: Laplace-Transformierte von $f_X(t)$
- t_X : Erwartungswert oder 1. gewöhnliches Moment
- $t_X^{(2)}$: 2. gewöhnliches Moment
- σ_X^2 : Varianz
- c_X : Variationskoeffizient

1 EINLEITUNG

1.1 Leistungssteigerung durch Parallelarbeit

Mit der Entwicklung leistungsstarker Mikrocomputer zeichnet sich der Trend ab, hohe Rechenleistung nicht nur auf der Basis einer schnellen und aufwendigen Zentraleinheit zu erreichen, sondern auch durch die Kopplung mehrerer Recheneinheiten und die Nutzung von expliziter und impliziter Parallelität der Aufgabenstruktur. Expliziter Parallelismus einer Aufgabenstruktur tritt typischerweise beim Verarbeiten von strukturierten Datenmengen auf (Felder, Vektoren). Implizite Parallelität ist dagegen in Aufgabenstrukturen gegeben, in denen eine Reihe von Anweisungen oder gar ganze Programmblöcke voneinander unabhängig ablauffähig sind und keiner fest vorgegebenen Ablaufreihenfolge unterliegen. Im Gegensatz zur expliziten Parallelität ist die implizite Parallelität nur durch entsprechende Programmkonstrukte des Programmierers, durch einen speziellen Compiler oder durch ein spezielles Operationsprinzip (Datenflußprinzip) zu erkennen. Die Nutzung von Parallelität durch Rechnerarchitekturen, die nach dem von Neumann-Operationsprinzip arbeiten, ist auf verschiedenen Ebenen möglich. Rechnerarchitekturen der Klasse SIMD (Single Instruction-Stream, Multiple Data-Stream), wie Vektorrechner oder Pipeline-Rechner, nutzen die explizite Parallelität auf Anweisungsebene. Multi-Mikrocomputersysteme der Klasse der MIMD-Architekturen (Multiple Instruction-Stream, Multiple Data-Stream) nutzen neben expliziter auch die implizite Parallelität auf Prozeß-Ebene. Neben leichter Erweiterbarkeit von Komponenten (Recheneinheiten) und Diensten (Prozesse) zeichnen sich MIMD-Systeme, bei redundanter Auslegung der Komponenten, auch durch hohe Verfügbarkeit aus. Des Weiteren ergibt die gemeinsame Benutzung von Betriebsmitteln wie Programme (Funktionsverbund), Datenbestände (Datenverbund) und Verarbeitungskapazität (Lastverbund) eine zusätz-

liche Steigerung der Leistungsfähigkeit und Effektivität. Eine weitere Klassifizierung der Rechensysteme des Typs MIMD kann nach folgenden Kriterien vorgenommen werden:

Hardware-Struktur einzelner Komponenten

- homogen, alle Komponenten besitzen dieselbe Hardware-Struktur (z.B. Multiprozessorsystem),
- heterogen, System setzt sich aus Komponenten unterschiedlicher Struktur zusammen (z.B. Polyprozessorsystem),

Verteilung der Prozesse auf die Komponenten

- symmetrisch, alle Komponenten bearbeiten gleiche Aufgaben,
- asymmetrisch, Komponenten bearbeiten verschiedene, spezialisierte Aufgaben.

Neben dem Problem der Erkennung und Nutzung der Parallelität besteht auch noch das Problem, die Recheneinheiten durch ein kostengünstiges und leistungsfähiges Kommunikationssystem zu koppeln. Durch diese Rechnerkopplung wird eine Parallelarbeit der Einheiten erst möglich.

1.2 Prinzipien der Rechnerkopplung

1.2.1 Die enge Kopplung

Das Prinzip der engen Kopplung findet vor allem in homogenen Multiprozessorsystemen seine Anwendung. Mehrere Prozessoren und ein ihnen gemeinsamer globaler Speicher sind über eine Zugangslogik (Port) mit dem Verbindungsnetzwerk verbunden, (Bild 1.1). Der globale Speicher ist im allgemeinen in mehrere Module unterteilt, auf die parallel zugegriffen werden kann (Multi-Port-Speicher). Mit der Anzahl der gekoppelten Recheneinheiten steigt das Kommunikationsaufkommen im gesamten System. Es entstehen Speicherzugriffskonflikte. Durch

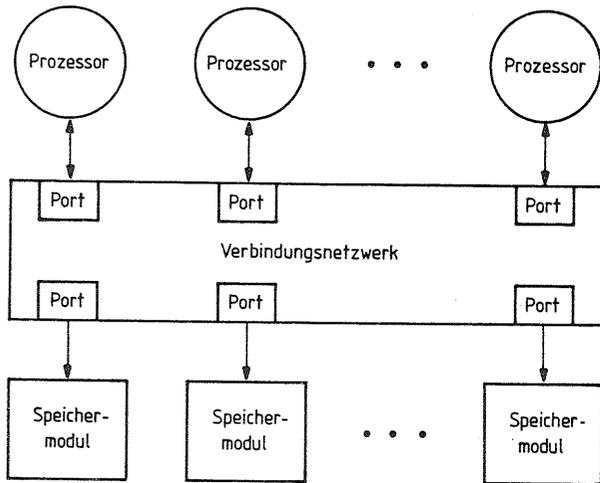


Bild 1.1: Eng gekoppeltes Rechensystem

eine Aufteilung des Speichers in mehrere unabhängige Module sowie durch eine horizontal organisierte Speicherung (Speicherverschränkung) werden Zugriffskonflikte reduziert, jedoch nicht vollständig vermieden. Im Konfliktfall warten die Prozesse (Prozessoren) bis eine Verbindung zu dem gewünschten Modul aufgebaut ist, d.h. sie sind für die Dauer des Zugriffskonfliktes blockiert. Auftretende Wartezeiten bei Blockierungen führen deshalb zu einer Einschränkung der Parallelität. Die Effizienz des Multiprozessorsystems nimmt dadurch mit steigender Anzahl der gekoppelten Prozessoren stark ab.

1.2.2 Die lose Kopplung

Die Verbindung autonomer Recheneinheiten über ein Kommunikationssystem wird als lose Kopplung bezeichnet. Als autonome Recheneinheit soll dabei ein Rechnerkern, der aus einem Prozessor mit lokalem Speicher besteht und in dem ein Software-Prozess autonom ablaufen kann, verstanden werden.

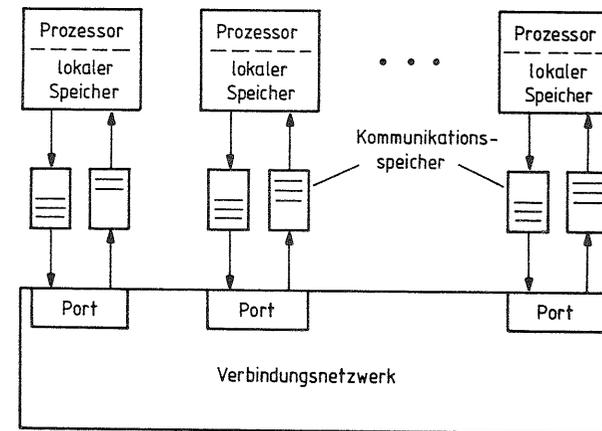


Bild 1.2: Lose gekoppeltes Rechensystem

Das Kommunikationssystem besteht aus Kommunikationsspeichern, die über eine Zugangslogik (Port) mit einem Verbindungsnetzwerk verbunden sind (siehe Bild 1.2). Die Kommunikationsspeicher dienen zur Wandlung von Datenformaten (parallel / seriell) und zur Anpassung der Geschwindigkeiten von Recheneinheit und Verbindungsnetzwerk. Weiterhin ermöglichen sie eine Zwischenspeicherung von Nachrichten und damit die zeitliche Entkopplung der simultan ablaufenden Prozesse, die auf verschiedenen Recheneinheiten aktiviert sind. Im Gegensatz zur engen Kopplung können die Recheneinheiten deshalb auch

bei temporären Engpässen des Verbindungsnetzwerks weiterarbeiten.

Eine lose Kopplung von Recheneinheiten, dezentral organisierte Systemkontrolle durch verteilte Zugriffssteuerung des Verbindungsnetzwerkes sowie die verteilte Kontrolle der Prozesskooperation kennzeichnen ein verteiltes System.

Basierend auf der Entfernung der einzelnen Recheneinheiten untereinander können verteilte Systeme prinzipiell in drei Klassen unterteilt werden:

- geographisch verteilt (z.B. Rechnerverbundnetz),
- räumlich verteilt (z.B. Lokales Netz),
- räumlich konzentriert (z.B. Polyprozessor-System).

1.3 Strukturen von Verbindungsnetzwerken

Enge oder lose Kopplung einer größeren Anzahl von Recheneinheiten sind durch verschiedene Strukturen von Verbindungsnetzwerken möglich, siehe Bild 1.3. Bei enger Kopplung werden durch das Verbindungssystem Prozessoren und Speichermodule (Bild 1.1), bei loser Kopplung die Kommunikationsspeicher der autonomen Recheneinheiten (Bild 1.2) miteinander verbunden.

Ein Verbindungsnetzwerk kann strukturell wie folgt realisiert sein:

- durch ein mehrstufiges Koppelnetz, in dem temporäre Verbindungen zwischen den Einheiten auf- bzw. abgebaut werden können. Typische Netzwerke sind z.B. Delta-, Banyan- oder Clos-Netzwerke. Mehrstufige Koppelnetze werden hauptsächlich in Systemen mit enger Kopplung eingesetzt.
- durch ein Lokales Netz (LAN, Local Area Network) mit Bus- oder Ring-Topologie oder ein globales Datennetz (WAN, Wide Area Network). Derartige Netze werden in Systemen verwendet, in denen heterogene, autonome Einheiten lose gekoppelt werden sollen.

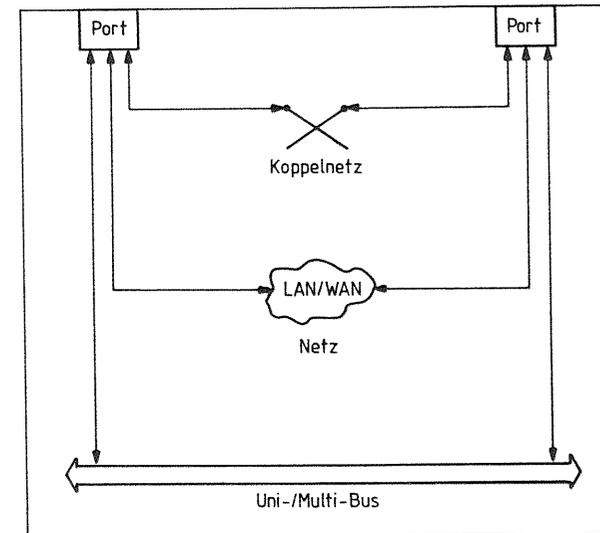


Bild 1.3: Strukturen von Verbindungsnetzwerken

- durch ein Uni- bzw. Multi-Bus-System, in dem durch parallele (wortweise) Übertragung höhere Bandbreiten wirtschaftlich realisiert werden können. Uni- bzw. Multi-Bus-Systeme eignen sich sowohl für lose als auch für eng gekoppelte Systeme.

In allen Verbindungsnetzwerken können Konflikte beim Zugriff auf das Kommunikationsmedium oder die Kommunikationsspeicher auftreten.

1.4 Rechner-zu-Rechner Kommunikation

Im folgenden wird Parallelarbeit auf Prozeß-Ebene in einem Polyprozessorsystem mit loser Kopplung der Recheneinheiten betrachtet. Die simultan ablauffähigen Prozesse werden auf die einzelnen Recheneinheiten verteilt und laufen dort autonom ab. Ein Prozeß soll dabei als ein zeitinvariantes Programm verstanden werden, das mit einem Satz von Daten aktiviert wird und zeitlich variante Zustände annehmen kann /1,2/.

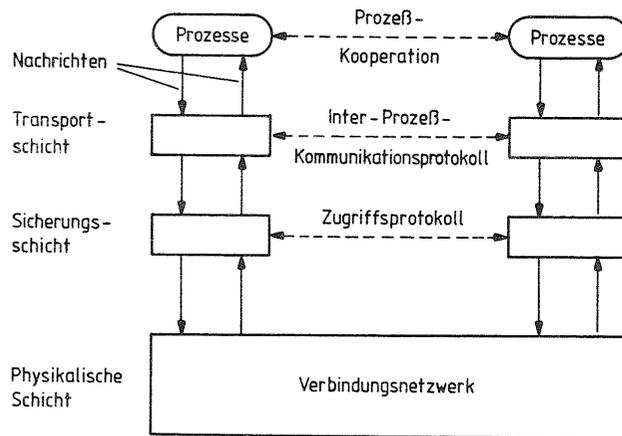


Bild 1.4: Rechner-zu-Rechner Kommunikation

Die im Gesamtsystem aktiven Prozesse arbeiten an gemeinsamen Aufgaben und kooperieren miteinander, indem sie indirekt, d.h. mittels vereinbarter Prozeduren oder Funktionen über gemeinsame Daten (Prozedur-orientiert) kommunizieren, oder direkt durch den Austausch von Nachrichten über ein No-Wait-Send- oder ein Rendezvous-Konzept /3/ (Nachrichten-orientiert) miteinander kommunizieren. Die Prozeßkooperation erzwingt somit auch die Kommunikation der Recheneinheiten untereinander. Um dies zu ermöglichen, sind

alle Einheiten durch ein Kommunikationssystem gekoppelt. Der Nachrichten- und Datentransfer zwischen den Recheneinheiten erfolgt über das Kommunikationssystem und ist durch eine geschichtete Protokollhierarchie festgelegt, welche ähnlich den Prinzipien des Architekturmodells für offene Kommunikationssysteme aufgebaut ist (siehe Bild 1.4). Das Inter-Prozeß-Kommunikationsprotokoll regelt den Nachrichtentransport zwischen den kooperierenden Prozessen. Das Zugriffsprotokoll greift über ein Zugriffsverfahren auf das Verbindungsnetzwerk des Kommunikationssystems zu und stellt damit Verbindungen bereit, über die Nachrichten oder Daten auch in Konfliktsituationen sicher transportiert werden können. Das Verbindungsnetzwerk ermöglicht die mechanische und physikalische Verbindung zwischen den Recheneinheiten. Die Struktur des Verbindungsnetzwerks, das verwendete Zugriffsverfahren und die Eigenschaften des Zugriffsprotokolls beeinflussen wesentlich die Leistungsfähigkeit des Kommunikationssystems.

1.5 Verkehrstheoretische Problemstellungen

Sowohl der Nachrichtenaustausch zwischen kooperierenden Prozessen, die in lose gekoppelten Recheneinheiten aktiv sind, als auch die Zugriffswünsche eng gekoppelter Prozessoren auf gemeinsame Datenbereiche werden durch das Kommunikationssystem bzw. Verbindungsnetzwerk verzögert. Verzögerungen bewirken Einschränkungen der Parallelarbeit auf Prozeßebene und führen damit zu einer Minderung der Leistungsfähigkeit des Gesamtsystems. Weiterhin können durch Verzögerungen Nachrichten in den endlichen Kommunikationsspeichern verloren gehen oder Prozeßabläufe in den Einheiten blockiert werden. Die Charakteristik der Verzögerung (Mittelwert und Verteilung) hängt von dem Kommunikationsaufkommen (Verkehrsintensität), sowie von Struktur und Betriebsweise des Kommunikationssystems bzw. Verbindungsnetzwerkes

ab. Ziel der Untersuchungen in dieser Arbeit ist es, abhängig von der Verkehrsintensität sowie einer bestimmten Struktur und Betriebsweise des Kommunikationssystems, charakteristische Größen zu bestimmen, die Aussagen über die Leistungsfähigkeit und das Verkehrsverhalten des Kommunikationssystems möglich machen. Charakteristische Größen sind z.B. Mittelwert und Verteilung der Zeiten für den Zugriff auf das Verbindungsnetzwerk, die mittlere Wartezeit, die mittlere Warteschlangenlänge und die Wahrscheinlichkeit für Nachrichtenverluste in den Kommunikationsspeichern. Alle charakteristischen Größen sind abhängig von der Struktur des Verbindungsnetzwerks, siehe Bild 1.3, und von dem verwendeten Zugriffsprotokoll. Mit Hilfe der charakteristischen Größen ist es möglich, das Verhalten des Systems, z.B. bei unsymmetrischer Last, bei Überlast oder bei Ausfall einer redundanten Komponente zu beschreiben. Dadurch ist eine optimale Dimensionierung des Kommunikationssystems für eng und lose gekoppelte Recheneinheiten möglich.

1.6 Aufgabenstellung und Übersicht über die Arbeit

In der Arbeit wird das Leistungsverhalten eines Kommunikationssystems für lose gekoppelte Rechensysteme untersucht. Als Verbindungsnetzwerk wird ein Uni- bzw. Multi-Bus-System mit zyklischem Buszugriffsprotokoll betrachtet. Im besonderen wird die Leistungsfähigkeit des Kommunikationssystems im Falle von Blockierungen des Nachrichtentransfers, bedingt durch endliche Kommunikationsspeicher oder Interferenzen an den Ports des Multi-Bus-Verbindungsnetzwerks, untersucht.

In Kapitel 2 wird zunächst der strukturelle Aufbau von Multi-Bus-Verbindungsnetzwerken erläutert. Dann werden verteilt oder zentral gesteuerte Buszugriffsverfahren diskutiert, speziell für Systeme, bei denen sich aufgrund des Zugriffsverfahrens eine zyklische (Polling-Verfahren,

Token-Verfahren), priorisierte (Daisy-Chain-Verfahren) oder teilweise zufallsabhängige (CSMA-CD-Verfahren) Buszuteilung ergibt. Am Ende des Kapitels wird noch auf mögliche Einschränkungen der Verkehrsleistung eingegangen, die durch Zugriffsinterferenzen im Verbindungsnetzwerk oder bei Blockierungen der Kommunikationsspeicher entstehen können.

Die verkehrstheoretische Modellierung des Kommunikationssystems, welche die in der Realität ablaufenden Vorgänge möglichst genau nachbilden soll, wird in Kapitel 3 durchgeführt. Nachdem zunächst die Problematik der Modellierung kurz erläutert wurde und die verwendeten Modellelemente eingeführt wurden, wird anschließend die Modellierung eines Uni-Bus- sowie eines Multi-Bus-Verbindungsnetzwerkes unter Berücksichtigung endlicher Kommunikationsspeicher durchgeführt.

Das Modell des Kommunikationssystems wird mit Hilfe der wahrscheinlichkeitstheoretischen Modellanalyse untersucht. In Kapitel 4 werden zunächst drei prinzipielle Methoden der Systemanalyse vorgestellt. Im weiteren Verlauf des Kapitels wird näher auf die wahrscheinlichkeitstheoretische Modellanalyse eingegangen, indem einige Merkmale und Eigenschaften stochastischer Prozesse und ihre Verwendung zur Bewertung von Systemmodellen dargestellt und behandelt werden.

In Kapitel 5 wird das Systemmodell des Sendeteils eines Uni-Bus- bzw. Multi-Bus-Verbindungsnetzwerkes analytisch, unter Anwendung der in Kapitel 4 eingeführten Methoden, behandelt. Dabei werden erstmals Nachrichtenverluste durch endliche Sendespeicher bzw. Konflikte beim Zugriff auf den Sendeport einer Station (Recheneinheit) berücksichtigt. Die approximative analytische Lösung kombiniert bisher bekannte Verfahren zur Analyse zyklischer Systeme mit der Methode der eingebetteten Markoff-Kette, wobei die endliche Zustandsverteilung der Sendespeicher berücksichtigt wird.

Ziel der Ausführungen in Kapitel 6 ist es, die Einschränkung der Leistungsfähigkeit des gesamten Kommunikationssystems zu bestimmen, falls Nachrichtenübertragungen am Empfänger der Stationen durch Portinterferenzen oder Speicherblockierung verfälscht werden oder verloren gehen. Dabei werden Systemparameter wie die Empfangsspeicherkapazität oder die Leistungsfähigkeit des Empfängers berücksichtigt.

Die analytisch gewonnenen Ergebnisse werden anhand einiger Fallstudien in graphischer Form dargestellt und mittels der zeitreuen Simulation validiert. Die Genauigkeit der approximativ gefundenen Lösung steigt mit zunehmender Anzahl der durch das Verbindungsnetzwerk gekoppelten Stationen und mit zunehmender Verwaltungsdauer für eine Buszuteilung. Unter anderem werden Ergebnisse über die Dimensionierung der Sendespeicher, den Einsatz eines Multi-Bus-Verbindungsnetzwerkes sowie das Verhalten des Verbindungsnetzwerkes bei unsymmetrischer Belastung dargestellt und diskutiert.

Die Ergebnisse dieser Arbeit finden Eingang in die Dimensionierung von Kommunikationsstrukturen wie sie in Multi-Rechnersystemen mit loser Kopplung oder Systemen mit verteilter Vermittlungssteuerung (Fernsprechvermittlungssysteme, Lokale Netzwerke) verwendet werden.

Als konkrete Anwendungsbeispiele seien an dieser Stelle die Vernetzung von Arbeitsplatzrechnern mit leistungsfähigen Verarbeitungseinheiten in einer Büroumgebung, die Dimensionierung eines modularen Paketvermittlungsknotens für das Dutex-P Netz der Deutschen Bundespost sowie die Leistungsfähigkeit von Token-gesteuerten Lokalen Netzwerken genannt.

2 ARCHITEKTUR VON MULTI-BUS-VERBINDUNGSNETZWERKEN

2.1 Allgemeines

Multi-Bus-Verbindungsnetzwerke werden zur engen Kopplung von Prozessoren und Speichermodulen sowie auch zur losen Kopplung autonomer Recheneinheiten eingesetzt (vergleiche Kapitel 1). Im allgemeinen bilden sie die kostengünstigere und flexiblere Alternative zu den Crossbar-Verbindungsnetzwerken, in denen z.B. bei enger Kopplung m Prozessoren mit n Speichermodulen über n Busse verbunden werden. Jeder der n Busse verbindet damit einen Speichermodul mit allen m Prozessoren. Die Kosten eines Crossbar-Netzwerks steigen deshalb mit $O(m \cdot n)$. Zwar sind Multi-Bus-Verbindungsnetzwerke im Vergleich zu Crossbar-Netzwerken intern nicht mehr blockierungsfrei, da die Anzahl der Busse $k < n$ gewählt wird, im allgemeinen aber für grössere Systeme kostengünstiger $O(k \cdot (m+n))$ und flexibler, da ein Bus mit allen Systemeinheiten verbunden werden kann. Darüber hinaus eignen sich Multi-Bus-Verbindungsnetzwerke noch für den Aufbau von redundanten bzw. ausfallsicheren Systemen mit enger oder loser Kopplung.

2.2 Netzwerkstruktur

Die typische Struktur eines lose gekoppelten Rechensystems mit einem Multi-Bus-Verbindungsnetzwerk ist in Bild 2.1 dargestellt. Die Verbindung zwischen den Einheiten (Prozessoren, Speichermodule oder Recheneinheiten) und dem Multi-Bus-Verbindungsnetzwerk wird durch den Netzwerkport (Netzwerkzugang einer Einheit) durchgeführt. Die Buszugriffswünsche werden über den Netzwerkport der Einheit an das Verbindungsnetzwerk weitergeleitet. Der Netzwerkport steuert den Netzwerkzugang, d.h. Busanforderung, Busfreigabe und Buszugriff,

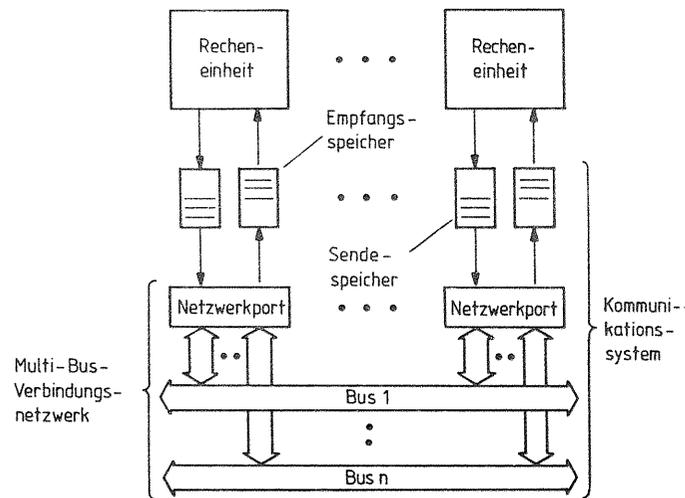


Bild 2.1: Struktur eines Kommunikationssystems mit Multi-Bus-Verbindungsnetzwerk

sowie den Übertragungsvorgang sendend wie empfangend. Weitere Funktionen, z.B. Paritätsprüfung oder Leitungscodierung bzw. -decodierung, können ebenfalls dem Netzwerkport zugeordnet werden. Ein Bus soll als Medium verstanden werden, über das Information bitseriell (Koaxialkabel, Glasfaser) oder bitparallel (Busrückwandverdrahtung) übertragen werden kann. Die Übertragungsgeschwindigkeit und damit die Bandbreite ist durch die physikalischen Eigenschaften des Mediums, wie Impulsanstiegszeit oder Signallaufzeit, begrenzt. Weiterhin beeinflussen noch Organisation und Geschwindigkeit des Buszugriffsverfahrens die Leistungsfähigkeit bzw. die real verfügbare Bandbreite des Multi-Bus-Verbindungsnetzwerks.

2.3 Buszugriffsverfahren

Die Abläufe eines Buszugriffs werden durch das Zugriffsverfahren festgelegt. Die Funktionen eines Zugriffs sind für jeden Bus entweder in Form einer zentralen Einheit (Arbiter) realisiert oder dezentral auf die verschiedenen mit dem Verbindungsnetzwerk verbundenen Einheiten verteilt. Eine zentrale Organisation des Buszugriffs führt im allgemeinen zu den einfacheren abfragegesteuerten Zugriffsverfahren. Die dezentrale Organisation des Buszugriffs bietet dagegen ein hohes Maß an Ausfallsicherheit, da die Regelung des Netzwerkzugangs auf alle Einheiten verteilt ist. Sie führt aber zu komplexeren Zugriffsprotokollen, die außer dem Buszugriff auch Konfliktsituationen in der Zugriffssteuerung verteilt regeln. Diese Zugriffsprotokolle sind Bestandteil einer hierarchisch geschichteten Protokollarchitektur /4/, (ISO-Referenzmodell) und Gegenstand von Standardisierungsbemühungen /5,6/ für die Kommunikation in offenen Systemen. In Bild 2.2 werden einige Merkmale der Buszugriffsverfahren tabellarisch dargestellt und miteinander verglichen.

2.3.1 Zentralgesteuerte Zugriffsverfahren

In Systemen mit zentralgesteuerten Zugriffsverfahren entscheidet eine zentrale Instanz über die Zuteilung des Übertragungsmediums (auch als abfragegesteuerte Zugriffsverfahren bezeichnet).

Merkmal	Abfrageverfahren	Zuteilungsverfahren	Konkurrenzverfahren
Steuerungsprinzip	zentral	dezentral	dezentral
Verzögerung			
Niederlast	klein	klein	null
Hochlast	endlich	endlich	unendlich
Prioritärer Buszugriff	Abfrage-reihenfolge	Prioritäts-token	Back-off Algorithmus
Zyklischer Buszugriff	möglich	möglich	Back-off Algorithmus
Realisierte Verfahren	Polling Daisy-Chain	Token-passing	CSMA-CD ALOHA
Implementierung	Terminal-Multiplexer	Tokenbus	Ethernet
Erweiterbarkeit	mittel	gut	gut
Fehlererkennung und -behebung	nicht möglich	Verlust oder Duplizierung des Tokens	exzessive Kollisionen

Bild 2.2: Vergleich der Buszugriffsverfahren

2.3.1.1 Polling-Verfahren

Die Kommunikation zwischen mehreren gleichberechtigten Einheiten und einer steuernden zentralen Einheit (z.B. Terminal-Multiplexer) kann durch das Polling-Verfahren geregelt werden. Die zentrale steuernde Einheit (Rechner) frägt in zyklischer Reihenfolge über den gemeinsamen Bus alle zu steuernden Einheiten (Terminals) ab. Die Abfrage (Poll) ent-

spricht einem Steuerblock, der die Adresse der abgefragten Einheit und die Sendeaufforderung enthält. Nach Erhalt der Sendeaufforderung antwortet die abgefragte Einheit, indem sie Daten zum Rechner überträgt oder der steuernden Einheit mitteilt, daß momentan keine Daten zu übertragen sind. Anschließend gibt die abgefragte Einheit die Sendeberechtigung wieder in Form eines Steuerblocks zurück. Durch Abfrage und Antwort sowie die doppelten Signallaufzeiten entsteht ein Verwaltungsaufwand, der zu einer zusätzlichen Busbelastung führt. Über die Reihenfolge der Abfrage läßt sich zum einen eine faire Buszuteilung einstellen, indem jede Einheit im Zyklus genau einmal abgefragt wird, zum anderen ist eine priorisierte Behandlung der Einheiten durch die mehrfache Abfrage innerhalb eines Zykluses möglich.

2.3.1.2 Daisy-Chain-Verfahren

Ein weiteres zentral gesteuertes Zugriffsverfahren ist das Daisy-Chain-Verfahren. Dieses Verfahren wird hauptsächlich in Polyprozessor-Systemen sowie bei Einzelprozessorsystemen mit Peripheriegeräten eingesetzt. Die Busanforderung, Busfreigabe und Buszuteilung wird über zusätzliche Steuerleitungen durchgeführt, dadurch wird der Bus nicht durch das Zugriffsverfahren belastet. Während die Signale für die Busanforderung und Busfreigabe direkt mit der Zugriffssteuerung verbunden sind, durchläuft das Buszuteilungssignal von der Zugriffssteuerung aus sequentiell alle Netzwerkports, die mit dem Bus verbunden sind (Daisy-Chain). Die Reihenfolge ergibt sich durch die Anordnung der Einheiten am Bus und entspricht damit einer fest eingestellten Priorität. Die Geschwindigkeit der Buszuteilung und damit die Dauer erzwungener Busfreizeiten (Verwaltungszeit) hängt auch von der Anzahl der Einheiten (Gatter) ab, die das Buszuteilungssignal zu durchlaufen hat. Eine strikte prioritäre Buszuteilung, basierend auf der Anordnung der Einheiten am Bus,

kann durch die Einführung von Busanforderungen mit unterschiedlicher Priorität vermieden werden. Für Anforderungen mit gleicher Priorität gilt nach wie vor die strikte prioritäre Buszuteilung. Eine faire Buszuteilung (zyklische Zuteilung) kann dadurch erreicht werden, daß die Prioritäten für Busanforderungen rotierend vergeben werden /7/.

2.3.2 Verteilt gesteuerte Zugriffsverfahren

In verteilt gesteuerten Zugriffsverfahren wird die Zuteilung des Busses durch die angeschlossenen Einheiten dezentral geregelt.

2.3.2.1 Konkurrenz-Verfahren

Die Zugriffsregelung mit Hilfe des Konkurrenz-Verfahrens wurde ursprünglich zur Verbindung von mehreren Terminals mit einem Zentralrechner über Rundfunkkanäle angewendet (ALOHA-Verfahren). Eine Einheit ist zu jedem Zeitpunkt berechtigt, einen Zugriff auf das Kommunikationsmedium durchzuführen und dieses zu belegen. Durch dieses Zugriffsverfahren werden laufende Übertragungsvorgänge durch Kanal-Zugriffe anderer Stationen gestört. Die Zugriffskonflikte bzw. die Kanalstörungen führen deshalb zu einem additiven Verkehr durch wiederholte Übertragungen. Das ALOHA-Verfahren erreicht, bedingt durch die hohe Konfliktwahrscheinlichkeit, nur einen geringen Durchsatz. Die Zugriffskonflikte (Übertragungs-Kollisionen) können eingeschränkt werden, wenn der Zugriff davon abhängig ist, ob der Kanal momentan durch eine andere Einheit belegt ist. Diese Bedingung wird bei den CSMA-Verfahren (Carrier-Sense-Multiple-Access), speziell für die Zugriffssteuerung von Bussystemen, für den Buszugriff gestellt. Aufgrund der endlichen Laufzeit zwischen

den Einheiten können bei dem CSMA-Verfahren zu Beginn bzw. zum Ende eines Übertragungsvorganges, speziell bei der Satellitenkommunikation, weiterhin Zugriffskonflikte auftreten. Für das Erkennen und Auflösen dieser Zugriffskonflikte wurde das CSMA-Verfahren zum CSMA-CD-Verfahren (Carrier-Sense-Multiple-Access-Collision-Detect) erweitert, wobei verschiedene Verfahren zur Kollisionserkennung und Kollisionauflösung (Backoff-Algorithmen) angewendet werden können. Die Ausführung eines Backoff-Algorithmus bedeutet wiederum Verwaltungsaufwand in den Einheiten und erzwungene Busfreizeiten. Die Konflikthäufigkeit steigt überproportional mit dem Verkehrsaufkommen an.

2.3.2.2 Zuteilungs-Verfahren

Während eine Einheit bei den Konkurrenz-Verfahren zu beliebigen Zeitpunkten einen Buszugriff durchführen kann, falls dieser nicht belegt ist (CSMA), wartet die Einheit bei den Zuteilungs-Verfahren auf die Zugriffsberechtigung. Die Buszuteilung kann in Form einer Steuernachricht (Token) oder bei regulärer Aufteilung der Übertragungskapazität (Time-Division-Multiplex) in belegbare Zeitscheiben konstanter Länge erfolgen. Während die Zeitscheiben den Einheiten entweder fest zugeordnet sind oder ein Kennzeichen über ihren Zustand (frei/belegt) mitführen, enthält der Token die Adresse der Einheit, die als nächste die Buszugriffsberechtigung erhält. Nach dem Empfang des Tokens durch eine Einheit steht ihr ausschließlich der Bus zur Verfügung. Am Ende des Übertragungsvorgangs wird der Token durch die Adresse der logisch folgenden Einheit ergänzt und weitergegeben. Die Vergabe der Zugriffsberechtigungen liegt damit durch das Adressierungsschema fest und bildet einen logischen Ring. Die Übertragung des Tokens bzw. das Warten bis zum Beginn einer Zeitscheibe sowie die Verzögerungen durch Adressdecodierung führen bei den Zuteilungsverfahren

ebenfalls zu Verwaltungsaufwand und Busfreizeiten. Das zeitliche Ablaufgeschehen des Token-Verfahrens entspricht für Betrachtungen zur Verkehrsleistung dem zeitlichen Ablauf eines verteilt realisierten Polling-Verfahrens. Die Zuteilungs-Verfahren bewirken eine faire zyklische Buszuteilung und garantieren auch bei hohem Verkehrsaufkommen einen bestimmten Durchsatz, der im wesentlichen nur durch die Bandbreite des Übertragungsmediums begrenzt wird, da Zugriffskonflikte vermieden werden.

2.4 Einschränkungen der Verkehrsleistung

Abhängig vom strukturellen Aufbau des Verbindungsnetzwerks, der Netzwerkports, vom Verfahren der Buszuteilung sowie von der Dimensionierung der Kommunikationsspeicher treten im Kommunikationssystem Konfliktsituationen auf, welche die real verfügbare Bandbreite des Multi-Bus-Verbindungsnetzwerks einschränken und damit die Leistungsfähigkeit des gesamten Rechensystems reduzieren.

Der Einfluß von Speicherzugriffs- und Buszugriffskonflikten auf die Leistungsfähigkeit eng gekoppelter Multiprozessor-systeme mit einem Uni- bzw. Multi-Bus-Verbindungsnetzwerk wurde in der Literatur bereits ausgiebig untersucht /8-13/. Neuere Aufsätze behandeln Leistungsuntersuchungen zu lose gekoppelten Rechensystemen mit Multi-Bus-Verbindungsnetzwerken /14-17/ und berücksichtigen bereits Blockierungen auch am Netzwerkport. Im folgenden werden die möglichen Einschränkungen der Verkehrsleistung in einem Multi-Bus-Verbindungsnetzwerk durch Speicherbegrenzungen, Zugriffsverfahren und Portinterferenzen diskutiert.

2.4.1 Blockierungen und Verluste bei endlichen Kommunikationsspeichern

Vom Verbindungsnetzwerk zu übertragende bzw. übertragene Daten oder Nachrichten werden in den Kommunikationsspeichern (Sendespeicher und Empfangsspeicher) zwischengespeichert. Bei zu geringer Dimensionierung der Speicherkapazität können Blockierungen oder Verluste von Nachrichten im Kommunikationssystem entstehen. Ein Nachrichtenverlust hat im allgemeinen eine Wiederholung des Übertragungsvorganges, eine Blockierung erzwungene Busfreizeiten zur Folge. Die Wahrscheinlichkeit für Verluste bzw. Blockierungen hängt von Fluß und Intensität des Datenverkehrs im Kommunikationssystem ab. Verluste können durch großzügige Dimensionierung der Kommunikationsspeicher weitgehendst vermieden, die Auswirkungen von Blockierungen durch eine möglichst effiziente Buszuteilung reduziert werden.

2.4.2 Blockierung durch Halb-Duplex-Betrieb

Eine Recheneinheit, deren Port sowohl für das Empfangen als auch für das Senden von Daten verwendet werden kann, deren Empfangsteil und Sendeteil aber nicht unabhängig voneinander arbeiten (Halb-Duplex-Betriebsweise), ist beim Senden von Daten für den Empfang von Daten über einen der anderen Busse blockiert. Entsprechendes gilt für die Recheneinheit auch während einer Empfangsphase; sie ist in dieser Zeit für das Senden von Daten blockiert. Durch den Aufbau des Ports in einen unabhängigen Sende- und Empfangsteil wird ein Voll-Duplex-gesteuerter Datenaustausch über das Multi-Bus-Verbindungsnetzwerk ermöglicht. Die Halb-Duplex-Blockierung kann damit durch einen komplexeren Aufbau des Netzwerkports vermieden werden. Diese Art der Blockierung kann grundsätzlich nur bei Multi-Bus-Systemen auftreten.

2.4.3 Interferenzen am Netzwerkport

Das Multi-Bus-Verbindungsnetzwerk ist in der Lage, parallele Übertragungen von Nachrichten oder Daten durchzuführen. Für den Fall, daß sich Übertragungswünsche von verschiedenen Einheiten zu einer bestimmten Zieleinheit zeitlich überlappen, treten am Netzwerkport der Zieleinheit Zugriffsinterferenzen auf, da dieser nicht in der Lage ist, mehr als eine Nachricht simultan zu empfangen. Entsprechendes gilt auch für den sendenden Teil des Netzwerkports, der nur einen Bus anfordern und belegen kann, auch wenn momentan mehrere Übertragungswünsche in einer Einheit vorhanden sind und warten. Interferenzen, die am Port entstehen, können vermieden werden, indem eine Einheit für jeden Bus des Verbindungsnetzwerks einen individuellen Netzwerkport besitzt. Diese Lösung stellt jedoch einen grossen schaltungstechnischen Aufwand dar und ist deshalb sehr kostenintensiv.

2.4.4 Interferenzen beim Buszugriff

Die Abfrage- und Zuteilungsverfahren arbeiten im allgemeinen kollisionsfrei, allerdings mit einer Grundverzögerung, die durch die zyklische Reihenfolge der Buszuteilung begründet werden kann. Die Konkurrenzverfahren sind dagegen kollisionsbehaftet, ein Buszugriff wird im Kollisionsfall für eine bestimmte Zeit zurückgestellt, welche durch einen sogenannten Backoff-Algorithmus bestimmt wird. Die Zugriffs-wünsche der Einheiten werden damit mindestens für eine Bus-Verwaltungsdauer bei zyklischer Buszuteilung oder durch Wartezeiten, die bei Kollisionen entstehen, verzögert.

3 SYSTEMMODELLE MIT ZYKLISCHEN ZUGRIFFSVERFAHREN

3.1 Allgemeines zur Modellierung

3.1.1 Der Modellbegriff

Ein Systemmodell ist ein abstraktes Gebilde, das die Struktur, den Zusammenhang und die Ablaufsteuerung aller Komponenten eines Systems möglichst realitätsnah darstellt. Modelliert wird in erster Linie die Inanspruchnahme der Betriebsmittel (Prozessoren, Speicher, Datenwege) durch die an das System gestellten Anforderungen (Datenpakete, Signale, Programme) funktionell und dem korrekten logischen Ablauf entsprechend. Die Tiefe der Modellierung, bzw. der Detaillierungsgrad des Modells, kann unterschiedlich sein und wird zum einen durch die Bedeutung der verschiedenen Vorgänge in dem realen System bezüglich der Untersuchung und zum anderen durch die zur Verfügung stehenden Methoden zur Modellanalyse (siehe Kapitel 4) bestimmt.

3.1.2 Elemente zur Modellbildung

Die Modellierung soll mit Hilfe von wenigen Grundbausteinen durchgeführt werden. Typische Grundbausteine sind z.B.

- Verkehrsquellen, die in zufälligen Abständen Anforderungen erzeugen, welche das Systemmodell belasten, indem sie Betriebsmittel belegen;
- Bedieneinheiten (verarbeitende Betriebsmittel des realen Systems), die von einer Anforderung für die Dauer ihrer Bearbeitung belegt werden;
- Warteschlangen (speichernde Betriebsmittel des realen Systems), in denen Anforderungen bis zu ihrer Bedienung gespeichert werden können.

Darüber hinaus können noch abhängig vom modellierten System weitere Modellelemente verwendet werden, z.B. Koppelvielfache, Schalter, Laufzeitglieder usw.

3.1.3 Beschreibung des Ablaufgeschehens

Der zeitliche Ablauf von Anforderungen durch das Systemmodell wird im wesentlichen durch die Ankunftsabstände und die Bediendauer der Anforderungen bestimmt, die mit Hilfe von zufallsabhängigen Prozessen statistisch beschrieben werden. Die realen Verteilungsfunktionen sind im allgemeinen nicht vollständig bekannt. Oft genügt für verkehrstheoretische Betrachtungen der Mittelwert (erstes Moment) und die Streuung der Zufallsvariablen um den Mittelwert. Damit ist es möglich, eine Verteilungsfunktion zu finden, die in ihren ersten beiden Momenten exakt mit der unbekannteren Verteilungsfunktion übereinstimmt.

Der logische Ablauf der Anforderungen durch das Modell wird durch die Verknüpfung der Modellelemente bzw. durch gerichtete Graphen vorgegeben.

Die Auswahl einer Anforderung aus mehreren wartenden innerhalb einer Warteschlange kann durch unterschiedliche Abfertigungsstrategien erfolgen. Typische Strategien sind FIFO (first-in, first-ot), LIFO (last-in, first-ot) oder RANDOM (zufällig).

In Systemen, in denen sich die Anforderungen in mehreren Warteschlangen eine Bedieneinheit teilen, muß die angewendete Strategie für die Zuteilung der Bedieneinheit zu den Warteschlangen ebenfalls im Modellablauf berücksichtigt werden. Die Zuteilung kann zum Beispiel über Prioritäten geregelt sein oder in zyklischer Reihenfolge durchgeführt werden. Im folgenden wird auf die Modellierung zyklischer Zuteilungsverfahren näher eingegangen.

3.2 Modellierung von zyklischen Zugriffsverfahren

3.2.1 Modellbeschreibung

In Bild 3.1 ist das Systemmodell eines zyklischen Zugriffsverfahrens auf eine Bedieneinheit dargestellt. Im Modell erzeugen die g Verkehrsquellen Anforderungströme, die durch die Bedieneinheit abgefertigt werden sollen. Die Anforderungen einer Quelle j treffen zufällig mit dem Ankunftsabstand T_{Aj} ein und werden bis zu ihrer Bedienung in der Warteschlange j zwischengespeichert. Durch die endliche Anzahl der Warteplätze (s_j) gehen eintreffende Anforderungen bei voll belegter Warteschlange verloren. Die Zuteilung der Bedieneinheit zu den einzelnen Warteschlangen erfolgt in zyklischer Weise nach unterschiedlichen Strategien.

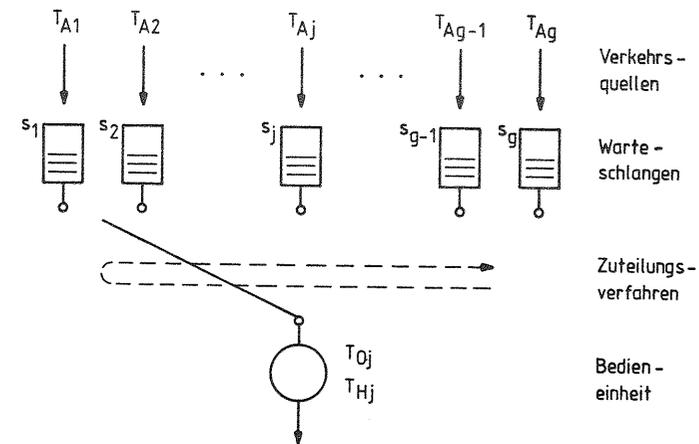


Bild 3.1: Systemmodell mit einer Bedieneinheit

Die Dauer des Verwaltungsvorgangs zur Buszuteilung für die Warteschlange j sei T_{Oj} . Nach erfolgter Zuteilung einer War-

teschlange zu der Bedieneinheit werden die wartenden Anforderungen nach unterschiedlichen Abfertigungsstrategien (FIFO, LIFO, oder RANDOM) ausgewählt und anschließend mit der Bediendauer T_{Hj} abgefertigt. Ankunftsabstand, Bus-Verwaltungsvorgang und Bediendauer sind zufallsabhängige Größen, die im allgemeinen vorgegeben sind und durch ihre Verteilungsfunktionen vollständig beschrieben werden.

Neben der Abfertigungsstrategie der Anforderungen innerhalb einer Warteschlange wird in Modellen mit zyklischen Zugriffsverfahren noch in der Bedienstrategie einer Warteschlange unterschieden.

3.2.2 Bedienstrategien

Die Bedienstrategie bestimmt im wesentlichen nach welchen Kriterien die Bedienung einer Warteschlange endet, d.h. ob nach Abfertigung einer Anforderung die Bedieneinheit freigegeben wird und einer der anderen Warteschlangen zugeteilt werden kann, oder ob mit der Abfertigung weiterer wartender Anforderungen fortgefahren wird. Prinzipiell sind folgende grundlegende Bedienstrategien denkbar:

- Erschöpfende Bedienung (exhaustive service)
Alle wartenden Anforderungen einer Warteschlange werden abgefertigt. Kriterium für das Ende der Bedienung ist eine leere Warteschlange.
- Torgesteuerte Bedienung (gated service)
Alle zum Zeitpunkt der Zuteilung wartenden Anforderungen werden abgefertigt. Anforderungen, die während der Bedienung der Warteschlange eintreffen, müssen bis zur nächsten Zuteilung der Bedieneinheit warten.
- Begrenzte Bedienung (limited service)
Die Anzahl der Anforderungen, die in Folge abgefertigt werden können, ist begrenzt.

- Zeitgesteuerte Bedienung (time-limited service)

Die Belegungsdauer der Bedieneinheit wird über eine Zeitsteuerung begrenzt. Falls mehrere Anforderungen warten, wird in jedem Fall eine Anforderung bedient. Weitere Anforderungen, die nicht innerhalb der vorgegebenen Belegungsdauer übertragen werden können, warten bis zur nächsten Zuteilung der Bedieneinheit.

Für den Fall, daß keine Anforderung zum Zuteilungszeitpunkt in der Warteschlange wartet, verhalten sich alle Strategien gleich; die Bedieneinheit wird sofort freigegeben und einer anderen Warteschlange zugeteilt.

3.2.3 Die Zuteilungsstrategien

Die Zuteilung der Bedieneinheit zu den g Warteschlangen wird zyklisch durchgeführt. Die Zuteilungsreihenfolge der Warteschlangen wird in einer endlichen Folge von Warteschlangennummern (Zyklus) festgelegt.

Die Zuteilung der Bedieneinheit benötigt im allgemeinen mehr oder weniger Verwaltungsaufwand (T_0), der zu einer zusätzlichen Belastung der Bedieneinheit führt (Betriebssystemaufruf) oder Freizeiten der Bedieneinheit erzwingt (Bus-Verwaltungsdauer). Basierend auf der Zuteilungsreihenfolge der Bedieneinheit zu den Warteschlangen sind folgende Strategien denkbar:

- Einfache zyklische Zuteilung
Die Bedieneinheit wird einer Warteschlange innerhalb eines Zyklus nur einmal zugeteilt.
Beispiel : 1-2-3- ... - g
- Priorisierte zyklische Zuteilung
Eine Warteschlange wird über den Zyklus priorisiert behandelt. Die Zuteilung erfolgt wechselweise zwischen der priorisierten und einer der nicht priorisierten Warteschlangen.

Beispiel : 1-2-1-3-1- ... -1-g
 (Priorität für Warteschlange 1)

- Beliebige zyklische Zuteilung

Die im Zyklus vorgegebene Reihenfolge der Zuteilung ist beliebig und nur durch die geforderte Endlichkeit der Zuteilungsreihenfolge begrenzt. Die Anzahl der Besuche einer bestimmten Warteschlange durch die Bedieneinheit kann beliebig groß und beliebig im Zyklus verteilt sein.

Beispiel : 1-2-1-2-3-1-2-4-1-2-5- ... -1-2-g

3.3 Modell des Multi-Bus-Verbindungsnetzwerkes

3.3.1 Modellbeschreibung

Ausgehend von der Struktur des Kommunikationssystems für lose gekoppelte Rechensysteme (Bild 2.1) kann ein Modell dieses Kommunikationssystems entwickelt werden, siehe Bild 3.2. In dem Verkehrsmodell werden g Stationen oder Recheneinheiten betrachtet, die durch ein Multi-Bus-Verbindungsnetzwerk verbunden sind.

Eine Recheneinheit j, (j=1,...,g), wird nachgebildet durch

- ihre Sende- bzw. Empfangswarteschlange, die jeweils eine endliche Anzahl von Warteplätzen besitzen (Sendewarteschlange s_j , Empfangswarteschlange r_j),
- eine Verkehrsquelle, die Anforderungen (Nachrichten) erzeugt, welche mit den Ankunftsabständen T_{Aj} eintreffen und anschließend in die Sendewarteschlange der Station eingetragen werden,
- eine Bedieneinheit, die empfangene Anforderungen in der Zeit T_{pj} bearbeitet.

Im Gegensatz zum bereits erläuterten Modell mit einer Bedieneinheit (Bild 3.1), existieren im Multi-Bus-Modell,

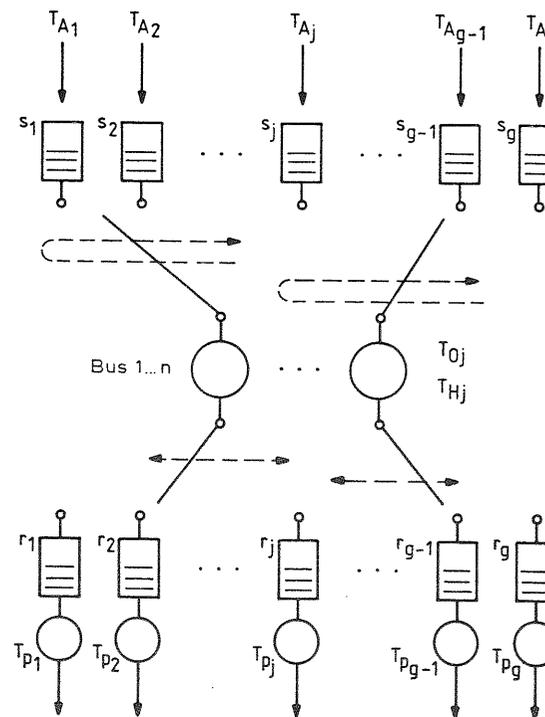


Bild 3.2: Systemmodell des Multi-Bus-Verbindungsnetzwerkes

entsprechend der Anzahl der Busse, mehrere (n) Bedieneinheiten. Darüber hinaus werden neben den Sendewarteschlangen die Empfangswarteschlangen der Recheneinheiten im Modell berücksichtigt. Die Bearbeitungszeiten der Nachrichten oder Anforderungen (T_{pj}) in der Empfangswarteschlange werden durch eine Bedieneinheit für jede Station individuell nachgebildet. Ein Bus wird zyklisch den g Sendewarteschlangen zugeteilt. Die Steuerung der Bus-Zuteilung kann zentral oder dezentral organisiert sein. Wartende Anforderungen werden nach erfolgter Bus-Zuteilung innerhalb der Übertragungszeit T_{Hj} von der Sendewarteschlange der Ursprungsein-

heit zu der Empfangswarteschlange der Zieleinheit übertragen. Bei Bedienungsende (Übertragungsende) der Sendewarteschlange j wird aufgrund der vorgegebenen Zuteilungsstrategie die nächste im Zyklus vorgesehene Recheneinheit bzw. deren Sendewarteschlange bestimmt und, falls dort Anforderungen warten, bedient.

Die Dauer der Bus-Verwaltung für die Zuteilung des Busses zur Station j sei T_{0j} . Der Ankunftsabstand, die Dauer eines Verwaltungsvorgangs sowie die Übertragungsdauer der Anforderungen und die Bediendauer der Empfangswarteschlange sind zufallsabhängige Größen, die im allgemeinen vorgegeben sind und durch ihre Verteilungsfunktion vollständig beschrieben werden.

Es wird angenommen, daß sich der Verkehrsfluß gleichmäßig auf alle Empfangswarteschlangen verteilt und vorausgesetzt, daß die Recheneinheiten keine Nachrichten über das Verbindungsnetzwerk an sich selbst senden. Die Wahrscheinlichkeit, daß Nachrichten der Recheneinheit i an die Recheneinheit j gesendet werden, ist somit wie folgt definiert:

$$r_{ij} = \begin{cases} 0 & , \quad j = i \\ \frac{1}{(g-1)} & , \quad j = 1, 2, \dots, g, j \neq i. \end{cases} \quad (3.1)$$

3.3.2 Blockierungen und Interferenzen des Verbindungsnetzwerkes

Durch die Möglichkeit, parallel Nachrichten über mehrere Busse zu übertragen, können an den Ports der Send- und Empfangswarteschlangen Zugriffs-Interferenzen entstehen. Weiterhin treten mit zunehmender Belastung des Systems durch die endliche Speicherkapazität Nachrichtenblockierungen in den Empfangswarteschlangen und Nachrichtenverluste in den Sendewarteschlangen auf (siehe auch Kapitel 2.4). Es wird angenommen, daß Interferenzen und Blockie-

rungen durch die Buszugriffssteuerung während der Busverwaltungsdauer erkannt werden. Dabei wird vorausgesetzt, daß keine Kommunikation unter den einzelnen Buszugriffssteuerungen stattfindet und jede Steuerung in der Lage ist, den Zustand einer Empfangswarteschlange (blockiert oder nicht blockiert) festzustellen. Nachrichtenverluste werden im Modell durch die endliche Kapazität der Sendewarteschlangen erfaßt.

Auftretende Interferenzen sollen momentan ablaufende Übertragungsvorgänge nicht stören. Bei einer durch die Buszugriffssteuerung erkannten Interferenz an der Sendewarteschlange wird der Bus entsprechend der vorgegebenen Zuteilungsreihenfolge in einem weiteren Verwaltungsvorgang der nächsten Recheneinheit zugeteilt. Interferenzen an den Sendewarteschlangen bewirken, daß die zur Verfügung stehende Übertragungskapazität des Multi-Bus-Verbindungsnetzwerkes nicht vollständig durch die Recheneinheiten genutzt werden kann und sich der Verwaltungsaufwand für die Buszugriffssteuerung erhöht.

3.3.3 Ablaufsteuerungen des Verbindungsnetzwerkes

Im Falle von Blockierungen und Interferenzen an den Empfangswarteschlangen der Zieleinheit kann die Buszugriffssteuerung nach folgenden Kriterien verfahren:

- A) Die Nachrichtenübertragung wird in jedem Falle durchgeführt. Nachrichten die durch Blockierung und Interferenz verfälscht oder zerstört wurden, werden zu einem späteren Zeitpunkt wiederholt übertragen. Diese Steuerungsvariante wird in Systemen mit verteilter Zugriffssteuerung angewendet, da die Steuerung keine Kenntnis über den Systemzustand besitzt, d.h. ob Empfangswarteschlange der Zieleinheit blockiert oder der Sendeport belegt ist. Nachrichtenverluste führen durch Wiederholungen zu einer zu-

sätzlichen Belastung des Kommunikationssystems.

- B) Der Beginn der Nachrichtenübertragung wird verzögert bis die Blockierung endet oder der Empfangsport wieder frei ist. Die Dauer der Verzögerung wird bei Speicherblockierung durch die Verarbeitungskapazität der Zieleinheit und bei Portblockierung durch die Nachrichtenlänge und die Übertragungsrage der Busse bestimmt.
- C) Die Übertragung der Nachricht wird zurückgestellt, der Bus freigegeben. Die Übertragungskapazität des Busses bleibt nicht für die Dauer der Speicher- oder Portblockierung ungenutzt, sondern wird der nächsten Einheit zugeteilt. Die Nachricht wartet in der Sendewarteschlange bis zur nächsten Buszuteilung.

3.4 Anwendungen der Systemmodelle

Systemmodelle mit zyklischen Zugriffsverfahren sind für Leistungsuntersuchungen oft grundlegend und vielseitig in realen Systemen anwendbar. Im folgenden werden einige Anwendungen der behandelten Systemmodelle vorgestellt.

3.4.1 Rechner/Terminal-Kommunikation

Die Konzentration des Verkehrs zwischen mehreren Terminals und einem Rechner wird durch einen Terminal-Multiplexer durchgeführt. Der Multiplexer steuert den Datenfluß auf den Anschlußleitungen und entlastet durch Vorverarbeitung die Zentraleinheit.

Die Modellierung des Kommunikationsablaufs zwischen Terminals und ihrer Zentraleinheit führt auf ein Systemmodell, welches in Bild 3.3 dargestellt ist. Die Leitungsspeicher der Terminals werden in zyklischer Weise durch den Termi-

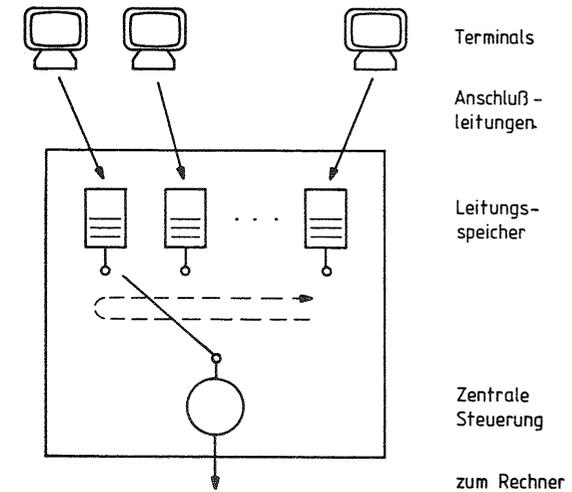


Bild 3.3: Systemmodell eines Terminal-Multiplexers

nal-Multiplexer abgefragt. Falls Daten zum Rechner zu übertragen sind, erfolgt entweder ein Datentransfer vom Leitungsspeicher zum Rechner oder die empfangenen Daten werden durch den Multiplexer verarbeitet. Durch die unbedingte Abfrage der Leitungsspeicher sowie wegen durchzuführender Prozeßwechsel in der Steuerung entsteht ein Verwaltungsaufwand in der zentralen Steuerung, der in der Modellierung nicht vernachlässigt werden darf. Die Bearbeitungsdauer eines Leitungsspeichers durch die zentrale Steuerung hängt vom Inhalt und von der Menge der empfangenen Daten sowie von der Bedienstrategie der Leitungsspeicher ab. Die Kommunikation zwischen dem Rechner und den Terminals kann auch durch Datenübertragungsprotokolle gesteuert werden. Derartige Protokolle enthalten im allgemeinen Mechanismen zur Flußsteuerung und zur Erkennung und Behebung von Übertragungsfehlern. In der Modellierung solcher Protokolle muß ebenfalls der Aufwand für die Verwaltung bzw. Steuerung der Punkt-zu-Mehrpunktverbindung berücksichtigt werden. Ver-

wartungszeiten entstehen hauptsächlich durch Signallaufzeiten auf den Anschlußleitungen und durch zusätzliche Übertragung von Steuerinformationen von bzw. zu den Terminals. Im Zusammenhang mit der Entwicklung von Time-Sharing Rechnersystemen wurden Terminal-Multiplexer und Datenübertragungsprotokolle dimensioniert und auf ihre Leistungsfähigkeit unter Berücksichtigung unterschiedlicher Bedienstrategien, Überlauf der Leitungsspeicher oder Übertragungsstörungen untersucht /18-24/.

3.4.2 Steuerungen in Fernsprech-Vermittlungssystemen

In zentral oder teilzentral gesteuerten Vermittlungssystemen werden periphere Steuerungen durch eine zentrale Einheit über einen gemeinsamen Bus zyklisch abgefragt (Bild 3.4).

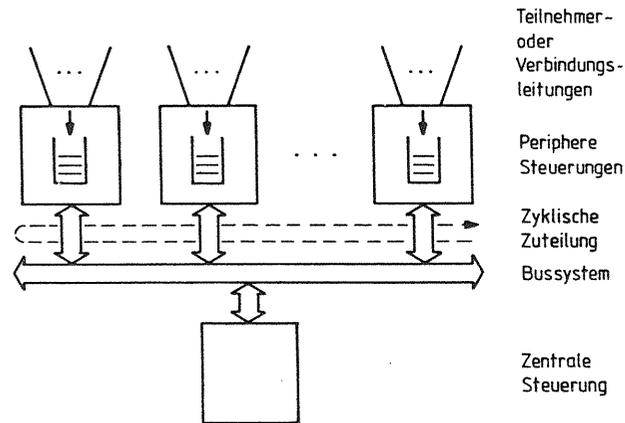


Bild 3.4: Systemmodell zentral oder teilzentral gesteuerter Vermittlungssysteme

Die peripheren Steuerungen verrichten vorverarbeitende Funktionen sowie Signalisierungsaufgaben an der Schnittstelle zwischen Teilnehmer- bzw. Verbindungsleitungen und

zentraler Steuerung. Änderungen des Leitungs- bzw. Teilnehmerzustandes werden durch sie erkannt und der zentralen Steuerung des Systems gemeldet. Die Abfrage der peripheren Steuerungen unterliegt Echtzeitbedingungen, damit eine geforderte Güte des Vermittlungssystems garantiert werden kann. Die zentrale Steuerung wird durch die Abfragen der peripheren Einheiten zusätzlich belastet. Untersuchungen der Leistungsfähigkeit dieses Steuerungskonzepts ohne /25,26/ oder mit /27-30/ Berücksichtigung von Verwaltungszeiten wurden u.a. im Zusammenhang mit der Entwicklung des Vermittlungssystems No.1 ESS durchgeführt. Um mit der zentralen Steuerung auf Zustandsänderungen der Peripherie entsprechend schnell reagieren zu können, muß der Abfragezyklus zum einen statisch unterschiedlichen Lastbedingungen in den peripheren Steuerungen (beliebige zyklische Zuteilung /31/) angepaßt werden und zum anderen in der Lage sein, durch mehrfache Abfrage einige der peripheren Steuerungen bevorzugt zu behandeln (priorisierte zyklische Zuteilung /32,33/).

3.4.3 Lokale Netze mit Token-gesteuertem Kanal-Zugriff

Die Untersuchung der Token-gesteuerten Kanal-Zuteilung in lokalen Netzen führt auf ein Modell nach Bild 3.5. In dem Modell werden die Stationen des lokalen Netzes durch ihre Sende- bzw. Empfangswarteschlange sowie durch den von ihnen erzeugten Verkehr (Ankunftsprozeß) und die Bearbeitung der empfangenen Nachrichten dargestellt. Das Kanalzugriffsverfahren entspricht der zyklischen Weitergabe des Tokens (Sendeberechtigung) von Station zu Station. Es wird durch die zyklische Zuteilungsreihenfolge der Bedieneinheit (Kanal) zu den Stationen modelliert. Die Verwaltungsdauer beinhaltet Stationsverzögerung, Signallaufzeit sowie eventuell durchzuführende Adreßmodifikationen (Token-Bus).

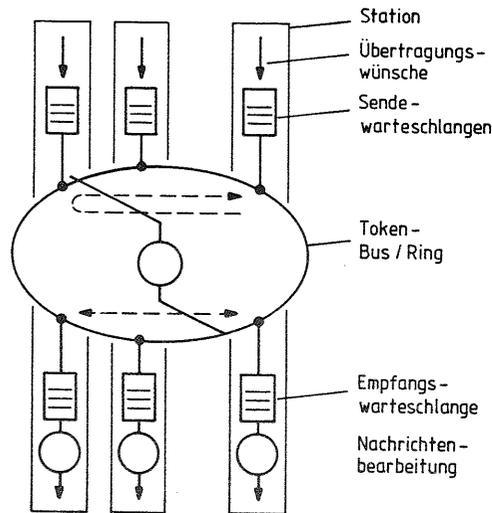


Bild 3.5: Systemmodell eines Lokalen Netzwerkes mit Token-gesteuertem Kanal-Zugriff

Die Belegungsdauer des Kanals durch eine Station hängt im wesentlichen von der Länge der zu übertragenden Datenpakete und ihrer Bedienstrategie ab.

In neuerer Zeit wurden auf dem Hintergrund der Leistungsuntersuchung und Dimensionierung lokaler Netze mit zyklischer Kanal-Zuteilung bisher bekannte analytische Verfahren angewendet, erweitert und modifiziert /34-45/.

3.4.4 Echtzeit-Betriebssysteme mit Mehrprozeßbetrieb

Beim Einsatz von Mikroprozessoren zur simultanen Bearbeitung mehrerer gleichberechtigter Prozeß-Klassen entsteht das Problem, die Prozessorkapazität fair auf die verschiedenen Klassen der Anforderungen zu verteilen. Die strikte priorisierte Zuteilung des Prozessors über Unterbrechungs-

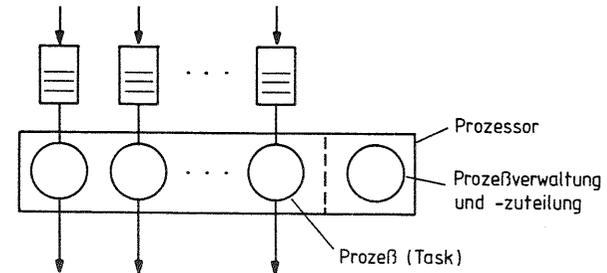


Bild 3.6: Systemmodell eines Multi-Tasking Betriebssystems

routinen führt im allgemeinen zu einer unfairen Behandlung der Prozeß-Klassen mit niederer Priorität. Die Zuteilung zu den Prozeß-Klassen durch das Betriebssystem, in einer fest vorgegebenen zyklischen Reihenfolge, stellt Bild 3.6 dar. Durch eine entsprechende Zuteilungsstrategie des Prozessors /31,46/, können bestimmte Prozeß-Klassen priorisiert behandelt werden, um damit Echtzeitbedingungen zu genügen. Der Verwaltungsaufwand des Betriebssystems wird durch eine separate Bedienphase im Modell berücksichtigt.

3.4.5 Multi-Bus-Verbindungsnetzwerke für modulare Rechensysteme

Die Modellierung modularer Rechensysteme mit hoher Verfügbarkeit und loser Kopplung durch kostengünstige und leistungsfähige Multi-Bus-Verbindungsnetzwerke entspricht dem Modell nach Bild 3.2. Mit Hilfe des Modells kann das Verhalten des Kommunikationssystems unter einer Zugriffsregelung nach dem Konkurrenz-Verfahren /15,16/ oder unter einer zyklischer Zuteilung /47,48/ speziell bei inneren Blockierungen oder auftretenden Interferenzen untersucht werden, um damit mögliche Auswirkungen auf die Leistungsfähigkeit des Gesamtsystems berechnen zu können.

4 WAHRSCHEINLICHKEITSTHEORETISCHE MODELLANALYSE

4.1 Methoden der Systemanalyse

Zur Beurteilung der Leistungsfähigkeit von Systemen wie Datennetze, Rechnerarchitekturen usw. ist es notwendig, die systeminternen Verkehrsabläufe, das zu verarbeitende Lastprofil sowie die Systemstruktur zu analysieren. Im allgemeinen soll durch eine Systemanalyse festgestellt werden,

- ob die gestellten Anforderungen bezüglich Güte oder Verarbeitungskapazität durch das System erfüllt werden;
- wie das System auf Überlast oder schiefe Last reagiert und wo Systemengpässe zu erwarten sind;
- wie sich Alternativen der Systemstruktur auf die Leistungsfähigkeit des Gesamtsystems auswirken können.

Eine Systemanalyse kann prinzipiell durch Messung am realen System sowie durch Verkehrssimulation oder mathematische Analyse eines Systemmodells durchgeführt werden.

4.1.1 Messungen

Messungen werden an einem Prototyp vorgenommen unter Verwendung von Meßwerkzeugen (Monitore), die in Software oder Hardware realisiert sind oder zum Teil erst implementiert werden müssen. Das Lastprofil, unter dem das Verhalten des Systems gemessen werden soll, wird durch synthetische Programme (Benchmarks) erzeugt, die eine reale Last modellieren und gleichzeitig zu Vergleichszwecken benutzt werden können.

Simulation oder mathematische Analyse sind unabhängig von einer Systemimplementierung einsetzbar. Beide Analyseverfahren erfordern allerdings die Abbildung des realen Sy-

stems in ein Verkehrsmodell (siehe Kapitel 3). Die Modellbildung kann bei konkreten Systemuntersuchungen durch Messungen am realen System verifiziert werden.

4.1.2 Simulation

Bei der Systemsimulation werden Zustandsänderungen des Simulationsmodells, die durch das Eintreffen von definierten Ereignissen hervorgerufen werden, zeittreu vorausgeplant und behandelt. Das dynamische Ablaufgeschehen im Modell wird somit in den Zeitpunkten der Zustandsübergänge (Ereigniszeitpunkte) betrachtet, die Verweildauern des Systems in den verschiedenen Zuständen werden zeitlich nicht nachgebildet; es wird in einem Rechnerprogramm von Ereigniszeitpunkt zu Ereigniszeitpunkt gesprungen (Event-by-Event Simulation). Typische Ereignisse sind z.B. Anforderungskünfte, das Belegende einer Bedieneinheit oder das Ende eines Taktintervalls.

Die Implementierung eines Simulationsprogramms erfolgt im allgemeinen modellindividuell und kann deshalb sehr zeitaufwendig sein. Einfache Modelle mit begrenzter Komplexität können allerdings auch durch allgemein verwendbare Simulationswerkzeuge behandelt werden. In diesem Falle beschränkt sich der Zeitaufwand auf die Einarbeitung in das Werkzeug und die Erstellung des Modells. Die Eingabe erfolgt bei den herkömmlichen Simulationswerkzeugen in einer programmiersprachenähnlichen Form (GPSS, General Purpose Simulation System, SIMULA, usw.). Neuere Simulationssysteme bieten eine interaktive Modellerstellung in graphischer Form (HIT, Hierarchical Interactive Tool, PAW, Performance Analysis Workstation, usw.).

Die Methodik der Systemsimulation stößt bei komplexen Modellen oder extremen Parameterkonstellationen an die Grenzen sinnvoller Programmlaufzeiten.

Die Systemsimulation wird oft auch zur Validierung von ap-

proximativen Annahmen in einer mathematischen Analyse eingesetzt.

4.1.3 Mathematische Analyse

Die Leistungsuntersuchung mittels der mathematischen Analyse basiert, wie in der Simulation, auf einem Modell des realen Systems. Der Detaillierungsgrad des Modells kann durch die angewendeten mathematischen Methoden oder die zur Verfügung stehenden analytischen Werkzeuge eingeschränkt werden. Prinzipiell kann zwischen zwei Klassen der Modellanalyse unterschieden werden:

- deterministisch-kombinatorische Analyse, mit welcher das Systemverhalten unter einer definierten und bekannten Last bestimmt und unter Umständen optimiert werden kann,
- wahrscheinlichkeitstheoretische Analyse, mit welcher das Ablaufgeschehen im System mit Hilfe zufallsabhängiger (stochastischer) Prozesse beschrieben werden kann.

Komplizierte und stark korrelierte Abläufe in den Verkehrsmodellen sind mathematisch zum Teil nur sehr schwierig zu beschreiben. Bei der Modellbildung für eine mathematische Analyse besteht die Möglichkeit, komplizierte Modelle zu einfacheren Ersatzmodellen zu aggregieren oder in einfachere Teilmodelle zu zerlegen. Die Ersatzmodelle werden dann mathematisch exakt oder über approximative Betrachtungen der Ablaufvorgänge näherungsweise gelöst. Typische approximative Annahmen sind zum Beispiel:

- die Unabhängigkeit von Ereigniszeitpunkten (Zufallsvariablen),
- Form und Eigenschaft (Erneuerungseigenschaft) der Wahrscheinlichkeits-Verteilungsfunktionen der Zufallsgrößen,
- die Auftrennung komplexer Systemmodelle unter Anwendung von Überlagerungs- und Zerlegungsverfahren für die Verkehrsströme.

Im folgenden werden einige der wichtigsten Grundlagen und Methoden der wahrscheinlichkeitstheoretischen Modellanalyse eingeführt und behandelt. Ein Schwerpunkt wird dabei auf die Methoden und Verfahren gelegt, die für die Analyse von Multi-Bus-Verbindungsnetzwerken von Bedeutung sind und im Rahmen dieser Arbeit verwendet werden.

4.2 Statistische Merkmale und Eigenschaften zufallsabhängiger Größen

Das Ablaufgeschehen in Systemmodellen kann prinzipiell mittels stochastischer Prozesse, wie zum Beispiel Ankunftsprozesse, Bedienprozesse oder Systemzustandsprozesse, charakterisiert werden. Mit Hilfe dieser stochastischen Prozesse ist es möglich, das Verhalten zufallsabhängiger Größen über der Zeit zu beschreiben und damit Aussagen über das statistische Verhalten der Zufallsgrößen zu erhalten. Eine zufallsabhängige Größe wird wahrscheinlichkeitstheoretisch durch eine diskrete oder kontinuierliche Zufallsvariable und ihrer Verteilungsfunktion definiert und beschrieben.

4.2.1 Zufallsvariable

Betrachtet werde ein Zustandsraum S mit der Menge seiner möglichen Elementarereignisse w . Die Zufallsvariable Z kann Werte aus der Menge der möglichen Elementarereignisse annehmen und bildet mit Hilfe der Funktion $g(Z)$ die unterschiedlichen Realisierungen der Elementarereignisse w auf die Zahlenwerte der Menge der reellen Zahlen ab. Hängt der Wert der Zufallsvariablen Z für die Realisierung eines bestimmten Elementarereignisses von der Zeit t ab, dann wird die Zufallsvariable durch die Funktion $Z(t)$ definiert. Exi-

stieren die Realisierungen einer Zufallsvariablen $Z(t)$ nur zu den diskreten Werten des Parameters $t=t_n$ ($n=0,1,2,\dots$), so wird $Z(t)$ als diskrete Zufallsvariable bezeichnet. Ist die Zufallsvariable $Z(t)$ für alle kontinuierlichen Werte des Parameters t definiert, so wird sie als kontinuierliche Zufallsvariable bezeichnet.

Die Wahrscheinlichkeitsverteilungsfunktion ($F_Z(t)$) für alle positiven Realisierungen einer Zufallsvariablen $Z(t)$ über dem Parameter t entsteht durch die Integration der Verteilungsdichtefunktion $f_Z(t)$ der Zufallsvariablen und errechnet sich wie folgt:

$$F_Z(t) = \int_{u=0}^t f_Z(u) du . \tag{4.1}$$

Aufeinanderfolgende Realisierungen der Zufallsvariablen $Z(t)$ können miteinander korreliert sein oder statistisch voneinander abhängen. Die Abhängigkeiten werden im allgemeinen durch die Verbundverteilungsfunktion (statistische Abhängigkeit) bzw. nur durch die ersten beiden Momente der Verbundverteilungsfunktion (Korrelation) beschrieben.

4.2.2 Verteilungsfunktion

Im folgenden werden am Beispiel einer kontinuierlichen Zufallsvariablen T_X einige Eigenschaften und Maßzahlen der zugehörigen Verteilungsfunktion $F_X(t)$ zusammengestellt und erläutert.

4.2.2.1 Momente und Maßzahlen

Die Verteilungsfunktion einer kontinuierlichen Zufallsvariablen T_X ist definiert durch

$$F_X(t) = 1 - F_X^C(t) = P(T_X \leq t) . \tag{4.2}$$

Sie gibt die Wahrscheinlichkeit an, daß die Zufallsvariable T_X Werte annimmt, die sich in dem Intervall $[0, t]$ befinden. $F_X^C(t)$ entspricht dabei der komplementären Verteilungsfunktion von T_X .

Durch Differentiation der Verteilungsfunktion entsteht die Verteilungsdichtefunktion

$$f_X(t) = \frac{dF_X(t)}{dt} . \tag{4.3}$$

Eine Bewertung der einzelnen Realisierungen der Zufallsvariablen T_X mit der Funktion $g(T_X)$ führt zu der mathematischen Erwartung der Funktion $g(T_X)$. Sie wird in Operator-schreibweise definiert zu

$$E[g(T_X)] = \int_{t=0}^{\infty} g(t) \cdot f_X(t) \cdot dt . \tag{4.4}$$

Für den speziellen Fall,

$$g(T_X) = t^k , \tag{4.5}$$

entsteht das k-te gewöhnliche Moment der Verteilungsfunktion

$$t_X^{(k)} = \int_{t=0}^{\infty} t^k \cdot f_X(t) \cdot dt . \tag{4.6}$$

Neben der Berechnung der k-ten Momente im Originalbereich besteht die Möglichkeit, die Momente im Bildbereich mit Hilfe

der Laplace-Stieltjes-Transformierten (LST) der Verteilungsfunktion zu bestimmen. Die LST der Verteilungsfunktion $F_X(t)$ ist definiert als

$$\Phi_X(s) = \int_{t=0^-}^{\infty} e^{-st} \cdot dF_X(t) . \quad (4.7)$$

Durch die k-te Ableitung der Laplace-Stieltjes-Transformierten an der Stelle $s=0$ ist es ebenfalls möglich, die k-ten Momente einer Verteilungsfunktion zu bestimmen

$$t_X^{(k)} = (-1)^k \cdot \left. \frac{d^k \Phi(s)}{ds^k} \right|_{s=0} . \quad (4.8)$$

Mittelwert und Varianz sind Maßzahlen der Verteilungsfunktion, die sich mit Hilfe der gewöhnlichen Momente einer Verteilungsfunktion berechnen lassen. Der Mittelwert ist identisch mit dem ersten Moment der Verteilungsfunktion ($k=1$). Zur Berechnung der Varianz wird neben dem ersten Moment noch das zweite Moment benötigt. Die Varianz entspricht dem Quadrat der Standardabweichung (σ) und ist definiert als

$$\sigma^2 = t_X^{(2)} - t_X^2 . \quad (4.9)$$

Neben der Varianz, die prinzipiell ein Maß angibt, wie die Werte um den Mittelwert einer Verteilungsfunktion streuen, wird oft auch der Variationskoeffizient einer Verteilungsfunktion verwendet, welcher sich durch die Normierung der Standardabweichung mit dem Mittelwert ergibt zu

$$c_X = \frac{\sigma_X}{t_X} . \quad (4.10)$$

4.2.2.2 Phasenmodelle nach Cox

Zur Behandlung beliebiger Verteilungsfunktionen in einer mathematischen Analyse ist es notwendig, eine geeignete analytische Darstellung im Originalbereich zu finden.

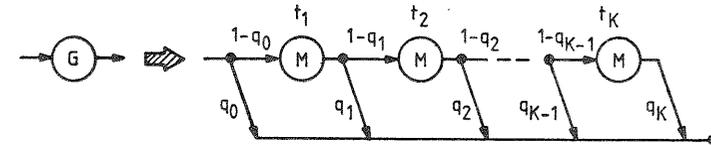


Bild 4.1: Phasenmodell nach Cox

Eine Möglichkeit der Darstellung allgemeiner Verteilungsfunktionen (G) besteht durch sogenannte Phasenmodelle, die aus negativ-exponentiell verteilten Teilphasen (M) bestehen, welche mit bestimmter Wahrscheinlichkeit durchlaufen werden. Ein allgemeines Phasenmodell für beliebige Verteilungsfunktionen wurde erstmals von Cox /49/ angegeben, vergleiche Bild 4.1. Durch Phasenmodelle ist es möglich, allgemeine Verteilungsfunktionen mit Hilfe negativ-exponentiell verteilter Teilphasen darzustellen und damit einer mathematischen Analyse zugänglich zu machen. Der Einsatz von Phasenmodellen hat allerdings die Erweiterung des Zustandsraumes der Zufallsvariablen um mindestens eine Dimension zur Folge.

4.2.2.3 Momentenapproximation

Häufig sind Verteilungsfunktionen in einer mathematischen Analyse im Bildbereich (Laplace-Stieltjes-Transformierte) oder nur in den ersten Momenten (z.B. Mittelwert und Varianz) bekannt. Liegt die Verteilungsfunktion im Bildbereich vor, so kann die Originalfunktion im allgemeinen durch Rück-

transformation gefunden werden. Abhängig von der Form der Verteilungsfunktion ist eine Rücktransformation oft nur numerisch möglich. Dagegen ist die Bestimmung der gewöhnlichen Momente erster, zweiter oder dritter Ordnung durch die Differentiation der Laplace-Stieltjes-Transformierten der Verteilungsfunktion (Gleichung (4.8)) wesentlich einfacher durchzuführen. Basierend auf dem ersten und zweiten Moment einer unbekannt beliebigen Verteilungsfunktion ist es möglich, ähnlich dem Verfahren nach Cox, Phasenmodelle anzugeben, deren Verteilungsfunktionen in den ersten beiden Momenten exakt mit der unbekannt Verteilungsfunktion übereinstimmen.

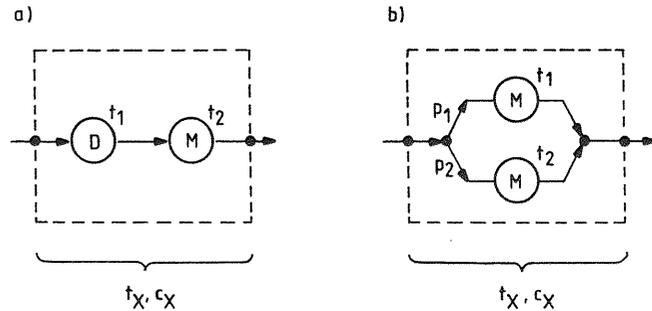


Bild 4.2: Phasenmodelle der Ersatz-Verteilungsfunktionen

Abhängig vom Variationskoeffizienten der unbekannt Verteilungsfunktion wird zwischen zwei "Ersatz"-Verteilungsfunktionen und damit zwei Phasenmodellen unterschieden:

- Die unbekannt Verteilungsfunktion $F_X(t)$ hat einen hypoexponentiellen Charakter ($0 \leq c_X \leq 1$). Sie wird dann ersetzt durch das Phasenmodell nach Bild 4.2 (Teil a).

$$F_X(t) = \begin{cases} 0 & , 0 \leq t \leq t_1 \\ 1 - e^{-(t-t_1)/t_2} & , t \geq t_1 \end{cases} \quad (4.11)$$

wobei $t_1 = t_X(1-c_X)$ und $t_2 = t_X c_X$.

- Die Verteilungsfunktion hat einen hyperexponentiellen Charakter ($c_X \geq 1$) und wird ersetzt durch das Phasenmodell nach Bild 4.2 (Teil b).

$$F_X(t) = 1 - p_1 \cdot e^{-t/t_1} - p_2 \cdot e^{-t/t_2} \quad (4.12)$$

$$\text{wobei } t_{1,2} = t_X \left(1 \pm \sqrt{\frac{c_X^2 - 1}{c_X^2 + 1}} \right)^{-1} \quad , \quad p_{1,2} = \frac{t_X}{2t_{1,2}}$$

$$\text{und } p_1 t_1 = p_2 t_2 \quad .$$

4.2.3 Funktionen von Zufallsvariablen

Zufallsabhängige Größen können auch durch Funktionen von Zufallsvariablen beschrieben werden. Typische elementare Funktionen sind beispielsweise das Minimum, das Maximum, die Differenz oder die Summe zweier Zufallsgrößen.

Gegeben seien zwei kontinuierliche Zufallsvariable T_X und T_Y mit ihren Verteilungsdichtefunktionen $f_X(x)$ und $f_Y(y)$ sowie ihrer Verbunddichtefunktion $f_{XY}(x,y)$. Die Verbunddichtefunktion entsteht durch partielle Differentiation der Verbundverteilungsfunktion, die wie folgt definiert ist:

$$F_{XY}(x,y) = P(T_X \leq x, T_Y \leq y) \quad . \quad (4.13)$$

Definiert sei im folgenden die Zufallsvariable T_Z als

$$T_Z = a \cdot T_X + b \cdot T_Y \quad . \quad (4.14)$$

T_Z stellt somit je nach Wahl der Koeffizienten a und b eine allgemeine Funktion zur Summation oder Differenzbildung der Zufallsvariablen T_X und T_Y dar, wobei die Verteilungsfunk-

tion $F_Z(z)$ sich allgemein ergibt zu

$$F_Z(z) = P(aT_X + bT_Y \leq z) = \iint_{ax+by \leq z} f_{XY}(x,y) \cdot dx dy. \quad (4.15)$$

4.2.3.1 Statistische Unabhängigkeit

Läßt sich die Verbunddichtefunktion $f_{XY}(x,y)$ der Zufallsvariablen T_X und T_Y durch das Produkt der Dichtefunktionen $f_X(x)$ und $f_Y(y)$ darstellen, d.h.

$$f_{XY}(x,y) = f_X(x) \cdot f_Y(y), \quad (4.16)$$

dann sind die Zufallsvariablen voneinander statistisch unabhängig. Die Dichtefunktion der Zufallsvariablen T_Z ergibt sich im Originalbereich durch die Faltung der Dichtefunktionen $f_X(x)$ und $f_Y(y)$:

$$f_Z(z) = \frac{1}{|a \cdot b|} \int_{y^* = -\infty}^{\infty} f_X\left(\frac{z-y}{a}\right) \cdot f_Y\left(\frac{y}{b}\right) \cdot dy^*, \quad (4.17)$$

wobei

$$x^* = ax, \quad y^* = by \quad \text{und} \quad z = ax+by.$$

Alternativ dazu kann die Dichtefunktion von T_Z über den Bildbereich durch Multiplikation der LST von $F_X(t)$ und $F_Y(t)$ und anschließender Rücktransformation gewonnen werden.

4.2.3.2 Summation von Zufallsvariablen

Die Verteilungsdichtefunktion der Summe zweier statistisch unabhängiger Zufallsvariablen T_X und T_Y errechnet sich durch die Faltung der Dichtefunktionen $f_X(x)$ und $f_Y(y)$. Aus Gleichung (4.17) entsteht dann speziell für $a=b=1$:

$$f_Z(z) = \int_{y=-\infty}^{\infty} f_X(z-y) \cdot f_Y(y) \cdot dy. \quad (4.18)$$

Für beliebig abhängige Zufallsvariable ergibt sich durch die additive Eigenschaft des Operators der mathematischen Erwartung (Erwartungswert) für das 1. Moment und 2. Moment der Verteilungsfunktion der Zufallsvariablen T_Z

$$t_Z = E[T_Z] = E[T_X + T_Y] = E[T_X] + E[T_Y] = t_X + t_Y \quad (4.19)$$

$$t_Z^{(2)} = E[T_Z^2] = E[(T_X + T_Y)^2] = E[T_X^2] + 2E[T_X T_Y] + E[T_Y^2]. \quad (4.20)$$

Durch Einsetzen der Gleichungen (4.19) und (4.20) in die Gleichung (4.9) errechnet sich die Varianz der Summenverteilungsfunktion zu

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 + 2 \cdot \mu_{11}. \quad (4.21)$$

Dabei entspricht μ_{11} der Kovarianz, dem ersten zentralen Moment der Verbundverteilungsfunktion $F_{XY}(x,y)$. Für den Fall, daß die Kovarianz verschwindet:

$$\mu_{11} = E[T_X T_Y] - E[T_X] \cdot E[T_Y] = 0, \quad (4.22)$$

sind die Zufallsvariablen T_X und T_Y nicht miteinander korreliert. Unabhängige Zufallsvariable sind in jedem Falle nicht miteinander korreliert; dagegen können unkorrelierte Zufallsvariable statistisch voneinander abhängig sein.

4.2.3.3 Differenz von unabhängigen Zufallsvariablen

Die Verteilungsdichtefunktion der Differenz zweier statistisch unabhängiger Zufallsvariablen T_X und T_Y errechnet sich ebenfalls mit Hilfe der Faltung der Dichtefunktionen $f_X(x)$ und $f_Y(y)$. Unter der Voraussetzung, daß die Koeffizienten a und b der Gleichung (4.14) wie folgt definiert sind,

$$a = 1 \quad \text{bzw.} \quad b = -1,$$

entsteht unter Verwendung der Gleichung (4.17) /50/ die Dichtefunktion der Differenz von T_X und T_Y

$$f_Z(z) = \int_{y=-\infty}^{\infty} f_X(z+y) \cdot f_Y(y) \cdot dy. \quad (4.23)$$

Mit Hilfe der Gleichung (4.23) kann dann unter der Voraussetzung, daß die Zufallsvariablen T_X und T_Y positive Werte annehmen, die Wahrscheinlichkeit berechnet werden, daß die Differenz, beschrieben durch die Zufallsvariable T_Z , positiv wird. Es entsteht für $T_X \geq 0$ und $T_Y \geq 0$

$$P(T_Z \geq 0) = P(T_X - T_Y \geq 0) = \int_{t=0}^{\infty} f_Z(t) \cdot dt$$

und somit (4.24)

$$P(T_Z \geq 0) = \int_{t=0}^{\infty} F_X^C(t) \cdot f_Y(t) \cdot dt = \int_{t=0}^{\infty} F_Y(t) \cdot f_X(t) \cdot dt.$$

4.3 Stochastische Prozesse

Stochastische Prozesse beschreiben mit Hilfe von Zufallsvariablen und ihren zugehörigen Verteilungsfunktionen das Verhalten zufallsabhängiger Größen über der Zeit t .

4.3.1 Definition

Die Familie der Funktionen $Z(t)$, wobei $t \geq 0$ ist, bildet einen stochastischen Prozeß. Er wird beschrieben durch

- Art und Größe des Zustandsraumes S , d.h. die Werte aller möglichen Realisierungen der Zufallsvariablen $Z(t)$;
- Parametertyp der Zufallsvariablen; existiert die Zufallsvariable $Z(t)$ nur zu diskreten Werten des Parameters t , so entsteht ein diskreter Prozeß. Die Folge der Realisierungen einer Zufallsvariablen Z_n zu den diskreten Zeitpunkten $t_n, n=0,1,2,\dots$, wird dann auch als stochastische Kette bezeichnet. Existiert die Zufallsvariable $Z(t)$ für kontinuierliche Werte des Parameters $t, (t \geq 0)$, so entsteht ein kontinuierlicher stochastischer Prozeß.
- die statistischen Abhängigkeiten der einzelnen Realisierungen der Zufallsvariablen $Z(t)$. Die Abhängigkeit von n aufeinanderfolgenden Realisierungen der Zufallsvariablen $Z(t)$ werden im allgemeinsten Fall durch die mehrdimensionale Verteilungsfunktion (Statistik n -ter Ordnung)

$$F(\underline{z}, \underline{t}) = P(Z(t_1) \leq z_1, Z(t_2) \leq z_2, \dots, Z(t_n) \leq z_n) \quad (4.25)$$

angegeben, dabei ist

$$\underline{z} = (z_1, z_2, \dots, z_n) \quad \text{und} \quad \underline{t} = (t_1, t_2, \dots, t_n).$$

4.3.2 Eigenschaften

Unabhängigkeit

Unter der Voraussetzung, daß Realisierungen aufeinanderfolgender Werte der Zufallsvariablen $Z(t_n)$, $n=0,1,\dots$ voneinander unabhängig sind, läßt sich die mehrdimensionale Verteilungsdichtefunktion eines stochastischen Prozesses durch das Produkt der Verteilungsdichtefunktionen der einzelnen Realisierungen darstellen

$$f(\underline{z}, \underline{t}) = f(z_1, \dots, z_n; t_1, \dots, t_n) = f(z_1, t_1) \cdot \dots \cdot f(z_n, t_n). \tag{4.26}$$

Stationarität

Ist die mehrdimensionale Verteilungsfunktion der Zufallsvariablen $Z(t)$ gegenüber dem Parameter t invariant:

$$F(\underline{z}, \underline{t}) = F(\underline{z}, \underline{t}+T), \tag{4.27}$$

dann ist sie um T frei verschiebbar und der stochastische Prozeß ist stationär. Die Wahrscheinlichkeit, daß die Zufallsvariable einen bestimmten Wert (Zustand) annimmt, ist dann unabhängig von dem Parameter t . Für einen stationären Prozeß ist es möglich, den Erwartungswert für die Verteilungsfunktion der Zufallsvariablen $Z(t)$ anzugeben.

Homogenität

Hängt die Übergangswahrscheinlichkeit $P(Z(t_2)=z_2 | Z(t_1)=z_1)$ zwischen aufeinanderfolgenden Realisierungen einer Zufallsvariablen $Z(t)$ nur von der Differenz der Zeitpunkte $t_2 - t_1$ ab, so ist der zugrundeliegende Prozeß homogen.

Erneuerungseigenschaft

Ein stochastischer Prozeß, der die Erneuerungseigenschaft besitzt (Erneuerungsprozeß), ist ein stationärer und unabhängiger Prozeß. Aufeinanderfolgende Realisierungen der Zufallsvariablen sind voneinander unabhängig und identisch

verteilt. Die Zufallsvariable wird deshalb auch als iid-Variable (independent and identically distributed variable) bezeichnet. Erneuerungsprozesse bilden das grundlegende Element der Erneuerungstheorie.

Markoff-Eigenschaft

Ein stochastischer Prozeß mit Markoff-Eigenschaft ist dadurch gekennzeichnet, daß sein zukünftiger Prozeßverlauf nur von dem gegenwärtigen Zustand abhängt und nicht von seiner Vergangenheit. Der Prozeß besitzt somit keine Kenntnis über den vergangenen Prozeßverlauf - er ist gedächtnisfrei. Für kontinuierlich verteilte Zufallsvariable wird die Markoff-Eigenschaft nur durch die negativ-exponentielle Verteilungsfunktion erfüllt. Für diskret verteilte Zufallsvariable erfüllt nur die geometrische Verteilung diese Eigenschaft. Stochastische Prozesse mit Markoff-Eigenschaft werden als Markoff-Prozesse bezeichnet.

Im folgenden werden einige wichtige und in dieser Arbeit verwendeten Klassen stochastischer Prozesse eingeführt. Die Prozeßklassen werden im wesentlichen an Hand ihrer Eigenschaften unterschieden.

4.3.3 Markoff-Prozesse

Die Abhängigkeit der zukünftigen Entwicklung eines Prozeßverlaufs von seiner Vergangenheit bildet die Grundlage der Klassifizierung in Markoff-Prozesse und Nicht-Markoff-Prozesse. Die Kenntnis über den vergangenen Prozeßverlauf kann als "Gedächtnis" interpretiert werden. Dementsprechend werden Prozesse ohne Gedächtnis allgemein als Markoff-Prozesse und Prozesse mit Gedächtnis als Nicht-Markoff-Prozesse bezeichnet.

Nicht-Markoff-Prozesse können somit beliebig komplex von ihrem vergangenen Prozeßverlauf abhängen. Wird das Gedächtnis

eines Nicht-Markoff-Prozesses auf den bestehenden Zustand beschränkt und hängt seine zukünftige Entwicklung außerdem noch vom Alter des bestehenden Zustandes ab, so handelt es sich um einen Semi-Markoff-Prozeß. Der Zustand $Z(t)$ bleibt zwischen zwei Zustandsübergängen unverändert und ändert sich in den Realisierungszeitpunkten sprunghaft.

4.3.4 Punktprozesse

Die Klasse der Punktprozesse beschreibt allgemein Ereigniszeitpunkte über der Zeitachse, bzw. den Strom von Anforderungen an ein Verkehrsmodell über der Zeit /51/.

Der Punktprozeß ist somit nur zu diskreten Werten des Parameters t (Zeit) definiert und zeichnet sich dadurch aus, daß das Bestehen eines Zustandes oder die Spanne zwischen zwei Prozeßpunkten, beliebig vom Parameter t sowie von der Vergangenheit des Prozeßverlaufs abhängen kann. Ein Punktprozeß ist damit im allgemeinen gedächtnisbehaftet.

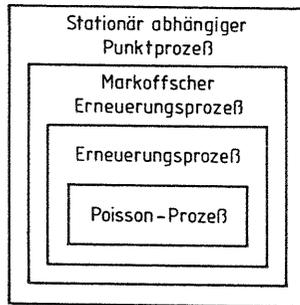


Bild 4.3: Klassifikation von Punktprozessen

Punktprozesse können, basierend auf den Eigenschaften wie Markoff-Eigenschaft, Erneuerungs-Eigenschaft und unter Berücksichtigung von Abhängigkeiten einzelner aufeinanderfol-

gender Realisierungen weiterhin unterschieden werden, siehe Bild 4.3.

Sind die Abstände von aufeinanderfolgenden Realisierungen der Zufallsvariablen negativ-exponentiell und identisch verteilt sowie unabhängig voneinander, dann kann der Punktprozeß durch einen Poisson-Prozeß beschrieben werden. Rekurrente Punktprozesse besitzen die Erneuerungseigenschaft. Im Gegensatz zum Poisson-Prozeß kann der Abstand zwischen zwei Prozeßpunkten beliebig verteilt sein. Der Markoffsche Erneuerungsprozeß beschreibt den Prozeßverlauf eines allgemeinen Prozesses in seinen regenerativen Punkten mit Hilfe einer Markoff-Kette, d.h. einem Markoff-Prozeß, der nur in den Regenerationspunkten definiert ist. Das Verfahren zur Berechnung der Charakteristika des allgemeinen Prozesses wird deshalb auch als "Eingebettete Markoff-Kette" bezeichnet. Während beim Markoffschen Erneuerungsprozeß in den Regenerationszeitpunkten zusätzlich die Eigenschaft der Unabhängigkeit gefordert wird, sind mit Hilfe stationär abhängiger Punktprozesse beliebig komplexe Abhängigkeiten prinzipiell behandelbar /51/.

4.3.5 Poisson-Prozesse

Der Poisson-Prozeß ist ein Punktprozeß, der die Erneuerungs- und Markoff-Eigenschaft besitzt. Die Abstände zwischen zwei Ereigniszeitpunkten des Poisson-Prozesses sind negativ-exponentiell und identisch verteilt. Die Wahrscheinlichkeitsverteilung für die Anzahl N eintreffender Ereignisse eines Poisson-Prozesses in dem Zeitintervall $(0, t)$, siehe Bild 4.4, entspricht der Poisson-Verteilung

$$p_n(t) = P(N = n) = \frac{(\lambda \cdot t)^n}{n!} \cdot e^{-\lambda t}, \quad n = 0, 1, \dots, \quad (4.28)$$

dabei ist λ die Rate, welche die mittlere Anzahl eintreffender Ereignisse pro Zeiteinheit beschreibt.

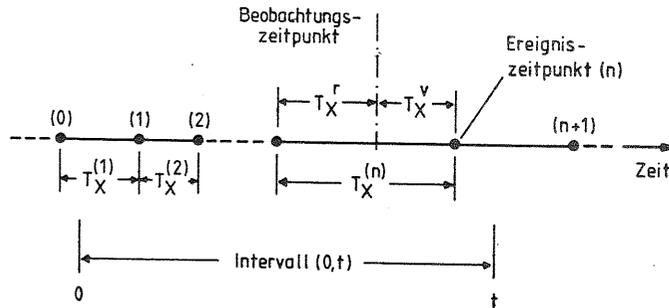


Bild 4.4: Punktprozeß mit Erneuerungseigenschaft

4.3.6 Erneuerungsprozesse

4.3.6.1 Rekurrenzenzeiten

Gegeben sei ein stochastischer Prozeß, der die Erneuerungseigenschaft besitzt und durch die Zufallsvariable T_X mit ihrer Verteilungsfunktion $F_X(t)$ beschrieben wird. Bild 4.4 stellt den Prozeßverlauf bis zur $(n+1)$ -ten Realisierung der Zufallsvariablen T_X über der Zeit dar. Betrachtet werde ein zufälliger Zeitpunkt innerhalb der n -ten Realisierung von T_X , siehe Bild 4.4. Die Zeitdauer vom letzten Ereigniszeitpunkt $(n-1)$ bis zum Zeitpunkt der Beobachtung wird als Rückwärtsrekurrenzenzeit T_X^r der Zufallsvariablen T_X definiert; entsprechend wird die Zeitdauer bis zum Ereigniszeitpunkt (n) als Vorwärtsrekurrenzenzeit T_X^v definiert.

Erneuerungsprozesse sind bezüglich des Parameters t reversibel, die Verteilungsfunktionen von Vorwärts- und Rückwärtsrekurrenzenzeit sind dann identisch. Im folgenden können damit die Ausführungen auf die Vorwärtsrekurrenzenzeit beschränkt

werden. Unter den obigen Voraussetzungen läßt sich die Verteilungsdichtefunktion der Vorwärtsrekurrenzenzeit $f_X^v(t)$ wie folgt angeben

$$f_X^v(t) = \frac{1 - F_X(t)}{t_X} \quad (4.29)$$

Daraus errechnen sich die k -ten Momente der Vorwärtsrekurrenzenzeit-Verteilungsfunktion entsprechend Gleichung (4.6), zu

$$t_X^{v(k)} = \frac{t_X^{(k+1)}}{(k+1) \cdot t_X} \quad (4.30)$$

Speziell für das erste Moment ergibt sich

$$t_X^v = \frac{1}{2} \cdot t_X \cdot (1 + c_X^2) \quad (4.31)$$

4.3.6.2 Überlagerung und Zerlegung

Eine mathematische Analyse komplexer Systemmodelle kann oft nur durch das Auftrennen des komplexen Modells in verschiedene Teilmodelle durchgeführt werden, die dann eigenständig betrachtet werden können. Zur isolierten Behandlung der Teilmodelle ist es notwendig, deren Verkehrsströme, die zum Beispiel durch Ankunftsprozesse oder Ausgangsprozesse gegeben sind, zu bestimmen. Beim Auftrennen des Modells muß beachtet werden, daß Abhängigkeiten zwischen den Teilsystemen verloren gehen können. Um den Fehler, der durch die Modelltrennung entsteht, so gering wie möglich zu halten, sollte der sich an der Schnittstelle zwischen den Teilsystemen einstellende Prozeß Erneuerungseigenschaft besitzen. Mit der Annahme, daß es sich an den Schnittstellen um stationäre Punktprozesse mit Erneuerungseigenschaft handelt, ist es

möglich, beliebig verteilte Ankunftsprozesse und Ausgangsprozesse der Teilmodelle durch Überlagerung bzw. Zerlegung von stochastischen Prozessen zu bestimmen /52/.

Die Erneuerungsannahme ist prinzipiell eine approximative Annahme. Die Zerlegung eines Erneuerungsprozesses führt im Falle unabhängiger Verzweigungswahrscheinlichkeiten wieder auf einen Punktprozeß mit Erneuerungseigenschaft. Dagegen führt die Überlagerung von Erneuerungsprozessen im allgemeinen zu einem Punktprozeß ohne Erneuerungseigenschaft. Besitzen die Erneuerungsprozesse die Markoff-Eigenschaft (Gedächtnisfreiheit), dann ist eine Überlagerung bzw. eine Zerlegung solcher Prozesse exakt.

Zerlegung

Im folgenden wird die Zerlegung eines gegebenen stationären Erneuerungsprozesses $F_X(t)$ /52/ in g Komponentenprozesse $F_{Xj}(t)$, ($j=1,2,\dots,g$) erläutert. Die Zerlegung von $F_X(t)$ führt im Bildbereich für einen Komponentenprozeß auf das Phasenmodell nach Bild 4.5.

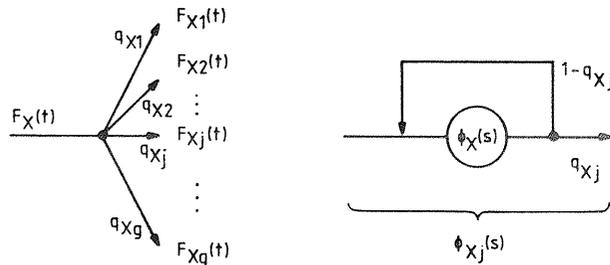


Bild 4.5: Zerlegung von stationären Erneuerungsprozessen

Die Wahrscheinlichkeit, daß ein Prozeßpunkt dem Komponentenprozeß j zugeordnet wird, sei q_{Xj} und für alle j ($j=1,\dots,g$) voneinander unabhängig angenommen. Weiterhin sei Y gleich der Anzahl der Phasendurchläufe von $F_X(t)$ bis ein Prozeßpunkt

dem Komponentenprozeß j zugeordnet werden kann, wobei Y wegen oben angenommener Unabhängigkeit der einzelnen q_{Xj} geometrisch verteilt ist. Daraus ergibt sich für die Laplace-Stieltjes-Transformierte von $F_{Xj}(t)$:

$$\Phi_{Xj}(s) = \sum_{y=0}^{\infty} q_{Xj} (1-q_{Xj})^{y-1} \cdot \Phi_X(s)^y = \frac{q_{Xj} \Phi_X(s)}{1-(1-q_{Xj}) \cdot \Phi_X(s)} \quad (4.32)$$

Unter Anwendung von Gleichung (4.8) ist es möglich, die ersten zwei Momente von $F_{Xj}(t)$ aus Gleichung (4.32) und damit den Mittelwert und den Variationskoeffizienten des Komponentenprozesse $F_{Xj}(t)$, $j=1,2,\dots,g$, zu bestimmen:

$$t_{Xj} = \frac{t_X}{q_{Xj}}, \quad c_{Xj}^2 = q_{Xj} \cdot c_X^2 + (1-q_{Xj}) \quad (4.33)$$

Überlagerung

Neben der Zerlegung findet die Überlagerung von Erneuerungsprozessen eine häufige Anwendung in der mathematischen Modellanalyse. An Hand des Beispiels der Überlagerung von zwei Erneuerungsprozessen, $F_{X1}(t)$ bzw. $F_{X2}(t)$, soll das wesentlichste des in /52/ dargestellten Kompositionsprinzips kurz erläutert werden. Nach /52/ führt die Überlagerung der Komponentenprozesse mit Hilfe einiger Ergebnisse der Erneuerungstheorie auf folgende Lösung für die Verteilungsfunktion des Summenprozesses $F_X(t)$:

$$F_X(t) = 1 - \frac{1}{t_{X1} + t_{X2}} \cdot (F_{X1}^C(t) \cdot \int_{u=t}^{\infty} F_{X2}^C(u) du + F_{X2}^C(t) \cdot \int_{u=t}^{\infty} F_{X1}^C(u) du) \quad (4.34)$$

Durch die Flußerhaltungsforderung ergibt sich für das erste Moment des Summenprozesses $F_X(t)$:

$$t_X = \frac{t_{X1} \cdot t_{X2}}{t_{X1} + t_{X2}} \quad (4.35)$$

Die Berechnung des zweiten Momentes bzw. des Variationskoeffizienten von $F_X(t)$ ist in /52/ angegeben, speziell für den Fall, daß die Komponentenprozesse nur in ihren ersten beiden Momenten bekannt sind. Die unbekanntenen Verteilungsfunktionen $F_{X1}(t)$ bzw. $F_{X2}(t)$ werden dabei durch das in Kapitel 4.2.2.3 beschriebene Verfahren zur Momentenapproximation angenähert.

4.4 Methode der eingebetteten Markoff-Kette

Weist ein Prozeß Zeitpunkte auf, von denen sich der zukünftige Prozeßverlauf unabhängig von der Vergangenheit entwickelt (Regenerationszeitpunkte), in denen der Prozeß also die Markoff-Eigenschaft besitzt, dann läßt sich zu diesen Regenerationszeitpunkten eine Markoff-Kette einbetten. Damit ist das Zustandsübergangsverhalten zwischen zwei benachbarten Regenerationszeitpunkten über ein Gleichungssystem vollständig bestimmbar, wobei die Zeit zwischen zwei Regenerationszeitpunkten beliebig verteilt sein kann. Die Methode der eingebetteten Markoff-Kette wird unter anderem zur Analyse von Wartesystemen oder Warte-Verlustsystemen mit allgemeinen Verteilungsfunktionen eingesetzt. Durch Auflösen des Gleichungssystems kann die Verteilung der Systemzustände zu den eingebetteten Zeitpunkten bestimmt werden. Ausgehend von der Zustandsverteilung ist es möglich, die Verteilung der Systemzustände zu beliebigen Zeitpunkten durch Zeitmittelung zu bestimmen (siehe Kapitel 5).

Weiterführende und detailliertere Ausführungen zur Berechnung eingebetteter Markoff-Ketten sind im Kapitel 5 dieser

Arbeit zu finden.

Neben der Methode der eingebetteten Markoff-Kette existieren noch weitere Möglichkeiten zur Analyse von Markoffschen Erneuerungsprozessen /53/.

4.5 Bewertung von Systemmodellen

Die Methoden einer Systemanalyse erbringen im allgemeinen Ergebnisse wie die Verteilung von Systemzuständen oder die Momente von Verteilungsfunktionen, d.h. Ergebnisse über die statistischen Eigenschaften und Merkmale der zufallsabhängigen Größen bzw. der stochastischen Prozesse. Diese Ergebnisse entsprechen zum Teil bereits den charakteristischen Größen des untersuchten Systemmodells oder ermöglichen deren Berechnung. Durch die Auswertung und Interpretation der charakteristischen Größen ist die Bewertung der Leistungsfähigkeit eines untersuchten Systems möglich.

4.5.1 Charakteristische Größen

Bild 4.6 zeigt als Beispiel das Modell eines Systems mit zyklischem Zugriffsverfahren einschließlich einiger stochastischer Prozesse, mit Hilfe derer charakteristische Eigenschaften und Merkmale des Systems beschrieben werden können. Aufgrund der im allgemeinen vorgegebenen Ankunfts- und Bedienprozesse werden unter Berücksichtigung des Systemzustandsprozesses noch weitere stochastische Prozesse betrachtet. Im folgenden werden einige typische, in dieser Arbeit verwendete Prozesse und die daraus abgeleiteten charakteristischen Größen für das Systemverhalten eingeführt:

Warteprozeß:

- mittlere Warteschlangenlänge Ω (Wartebelastung) - ent-

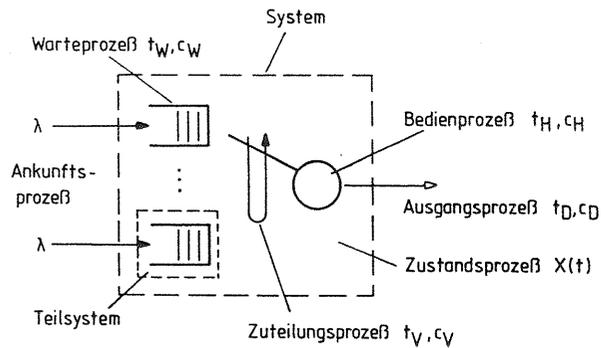


Bild 4.6: Charakteristische Größen und Prozesse

spricht der mittleren Anzahl von Anforderungen, die in der Warteschlange auf ihre Bedienung warten;

- Wartezeitverteilungsfunktion $F_W(t) = P(T_W \leq t)$;
- mittlere Wartezeit t_W - Wartezeit bezogen auf die wartenden Anforderungen, d.h. Wartezeit der Anforderungen, die beim Eintreffen nicht sofort bedient werden. Die Wartezeit kann auch auf alle Anforderungen bezogen werden;
- Variationskoeffizient der mittleren Wartezeit c_W bezüglich der wartenden Anforderungen;
- Wartewahrscheinlichkeit W - Wahrscheinlichkeit, daß eine eintreffende Anforderung nicht sofort bedient werden kann und somit warten muß;
- Verlust- oder Blockierungswahrscheinlichkeit B - Wahrscheinlichkeit, daß eine eintreffende Anforderung vom System nicht aufgenommen wird und damit verloren geht;

Zuteilungsprozeß:

- Besuchszeitverteilungsfunktion $F_V(t) = P(T_V \leq t)$;
- mittlere Besuchszeit t_V - mittlerer Abstand zwischen den Besuchen durch die Bedieneinheiten an einer Warteschlange;

- Variationskoeffizient der Besuchszeit c_V ;
- Zykluszeitverteilungsfunktion $F_C(t) = P(T_C \leq t)$;
- mittlere Zykluszeit t_C - mittlere Dauer für das Durchlaufen der Zuteilungsreihenfolge einer Bedieneinheit;
- Variationskoeffizient der Zykluszeit c_C ;

Bedien- und Ausgangsprozess:

- Belastung (Auslastung) Y - Wahrscheinlichkeit, daß die Bedieneinheit belegt ist;
- Ausgangsabstandsverteilungsfunktion $F_D(t) = P(T_D \leq t)$;
- mittlerer Ausgangsabstand t_D - mittlerer Abstand zwischen zwei Bedienungsenden einer Bedieneinheit;
- Variationskoeffizient des Ausgangsabstandes c_D .

Neben den vorgestellten Prozessen und charakteristischen Größen sind noch darüber hinaus denkbar: Zeitdauer für das Durchlaufen der Anforderungen durch das gesamte Systemmodell (flow-time), Dauer der Betriebsperioden bzw. der Freizeiten von Bedieneinheiten (busy-time, idle-time), höhere Momente (z.B. drittes Moment) der Verteilungsfunktion einer Zufallsgröße, usw..

4.5.2 Das Little'sche Theorem

Zwischen einigen charakteristischen Größen eines beliebigen Systems existiert eine exakte mathematische Beziehung, welche im sogenannten Theorem nach Little /54/ beschrieben wird. Dieses Theorem besagt, daß sich die mittlere Anzahl $E[X]$ von Anforderungen in einem System aus dem Produkt der mittleren Durchlaufzeit $E[T_P]$ einer Anforderung und der im Mittel eintreffenden Anzahl von Anforderungen pro Zeiteinheit λ ergibt:

$$E[X] = \lambda \cdot E[T_F] . \quad (4.36)$$

Für das Teilsystem "Warteschlange" nach Bild 4.6 gilt somit die Beziehung

$$\Omega = \lambda \cdot t_w . \quad (4.37)$$

5 ANALYSE VON MULTI-BUS-VERBINDUNGSNETZWERKEN MIT ENDLICHEN SENDEWARTESCHLANGEN

5.1 Literaturübersicht

Die Analyse basiert im wesentlichen auf der Berechnung der Zykluszeitverteilungsfunktion ähnlich wie sie in /18-20/ verwendet wurde. Um die Abhängigkeit aufeinanderfolgender Zyklen in einem gewissen Maße zu berücksichtigen, wurde das in /29/ vorgestellte Prinzip der bedingten Zykluszeiten angewendet. Für den Fall, daß nur eine Bedieneinheit mehrere Stationen in zyklischer Reihenfolge bedient, existiert bereits eine umfangreiche Literatur. Einige wesentliche Arbeiten sind für diesen Fall in den Literaturstellen /18-47/ angegeben. In den Arbeiten /55-60/ wurden erstmals unter der Verwendung von mehrdimensionalen erzeugenden Funktionen exakte Lösungen von Systemen mit zyklischer Zuteilung einer Bedieneinheit vorgestellt.

Die Analyse von Wartesystemen mit der parallelen zyklischen Bedienung durch mehrere Bedieneinheiten wurde erstmals in /61/ durchgeführt. Dabei wurden Zugriffsinterferenzen der diversen Bedieneinheiten, Speicherblockierungen der Warteschlangen sowie der Einfluß der empfangenden Stationen auf die Leistungsfähigkeit des Gesamtsystems nicht betrachtet. Eine Analyse für Multi-Bus-Verbindungsnetzwerke unter einer Zugriffsregelung nach dem Konkurrenz-Verfahren und unter der Berücksichtigung von Portinterferenzen wurde in /15,16/ für Systeme mit negativ-exponentiellen Busbelegungszeiten durchgeführt.

Im folgenden wird eine wahrscheinlichkeitstheoretische Modellanalyse eines Multi-Bus-Verbindungsnetzwerks durchgeführt. Abgesehen von dem Sonderfall der Station mit einem einzelnen Warteplatz (Puffer) und endlicher Quellenzahl (kein Poissonankunftsprozeß) /60, 62, 63/, wird in dieser Arbeit erstmals

- die endliche Anzahl der Speicherplätze in Sende- und Empfangswarteschlange einer Station sowie
- die parallele Bedienung mehrerer voneinander unabhängiger Bedieneinheiten (Busse) unter Berücksichtigung auftretender Interferenzen und Blockierungen beim Nachrichtentransfer

betrachtet.

5.2 Modellannahmen und Modellparameter

Das Modell nach Bild 3.2 wird zwischen den Ausgangsprozessen der Bedieneinheiten und den Eingangsprozessen der Empfangswarteschlangen aufgetrennt, damit wird die Einschränkung der Verkehrsleistung durch Port- oder Speicherblockierungen der Empfangswarteschlangen zunächst nicht betrachtet. Die im Modell betrachteten zufallsabhängigen Prozesse einer Station (Recheneinheit) j , $j = 1, \dots, g$, unterliegen folgenden Voraussetzungen und Vereinbarungen:

- Der Ankunftsprozeß entspricht einem Poisson-Prozeß, d.h. die Ankunftsabstände sind negativ-exponentiell verteilt mit dem Mittelwert t_{Aj} bzw. der Rate λ_j .
- Die Verwaltungsdauer für die Zuteilung eines Busses zu einer beliebigen Station wird durch den Verwaltungsprozeß beschrieben. Der Verwaltungsprozeß ist charakterisiert durch die Zufallsvariable T_{Oj} . Die Verteilungsfunktion von T_{Oj} ist beliebig und wird durch Mittelwert t_{Oj} und Variationskoeffizient c_{Oj} vorgegeben.
- Die Übertragungsdauer der Anforderungen, charakterisiert durch die Zufallsvariable T_{Hj} , ist ebenfalls beliebig verteilt und durch den Mittelwert t_{Hj} bzw. den Variationskoeffizienten c_{Hj} vorgegeben.

Ein Bus wird als eine Bedieneinheit betrachtet, welche stationsindividuell in der Zeit T_{Oj} zugeteilt wird und für die Übertragungszeit einer Anforderung (T_{Hj}) belegt wird. Die

Reihenfolge der Zuteilung einer Bedieneinheit zu den Sendewarteschlangen erfolgt gemäß der einfachen zyklischen Zuteilung nach Kapitel 3.2.3. Die verschiedenen Bedieneinheiten (Busse) werden voneinander unabhängig den Stationen zugeteilt. Die Bus-Zuteilungsprozesse überlagern sich an den Ports der Sendewarteschlangen.

Die Abfertigungsstrategie der wartenden Anforderungen innerhalb der Sendewarteschlangen ist FIFO. Nach erfolgter Buszuteilung wird eine Warteschlange nach der Strategie "Begrenzte Bedienung" bedient, damit wird, falls vorhanden, nur eine der wartenden Anforderungen übertragen (limited service 1). Die Zuteilung erfolgt unabhängig davon, ob Anforderungen in einer Warteschlange warten und stets mit dem Verwaltungsaufwand (T_{Oj}).

5.3 Wahrscheinlichkeitsverteilung der Systemzustände

Der Zustandsprozeß einer Sendewarteschlange (Bild 5.1) bildet eine stochastische Kette, welche durch die diskrete Zufallsvariable X_n ($n = 0, 1, 2, \dots$) beschrieben wird. Die Zustandsverteilung der Zufallsvariablen X_n soll mit Hilfe der Methode der eingebetteten Markoff-Kette berechnet werden.

Aus Übersichtlichkeitsgründen wird im folgenden auf den Index j , welcher die betrachtete Station kennzeichnet, verzichtet.

5.3.1 Verteilung zu den Regenerationszeitpunkten

Die Abstände zwischen zwei aufeinanderfolgenden Buszuteilungen einer betrachteten Station sind beliebig verteilt. Es wird eine Markoff-Kette in den Zeitpunkten kurz vor dem Ende der gedächtnisbehafteten Phase (Besuchszeitverteilungs-

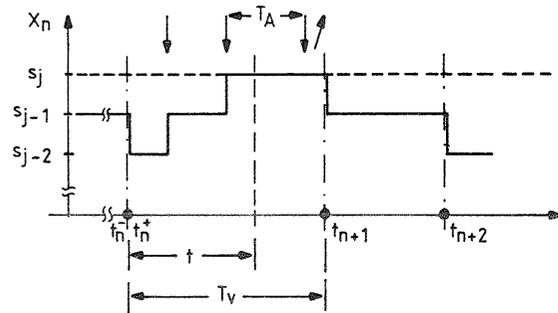


Bild 5.1: Zustandsprozeß einer Sendewarteschlange

funktion) eingebettet. Die Regenerationszeitpunkte (t_n^- , $n=0,1,\dots$) entsprechen dann den Zeitpunkten kurz vor Beginn der Buszuteilung. Alle Zustandswahrscheinlichkeiten der Zufallsvariablen X_n^- bilden gemeinsam den Wahrscheinlichkeitsvektor $\underline{p}^{(n)} = (p_0, \dots, p_s)^{(n)}$, dessen Komponenten die Wahrscheinlichkeit für k wartende Anforderungen in der Sendewarteschlange beschreiben und zu den Zeitpunkten t_n^- wie folgt definiert sind:

$$p_k^{(n)} = P(X_n^- = k), \quad k = 0, \dots, s. \quad (5.1)$$

Die Bedienung einer Anforderung hat eine Zustandsänderung der Sendewarteschlange zur Folge. Die Komponenten des Wahrscheinlichkeitsvektors kurz nach Bedienungsbeginn, zu der Zeit t_n^+ , bzw. die Zustandsübergänge in den Regenerationszeitpunkten, können durch folgende Gleichungen angegeben werden:

$$\begin{aligned} p_0^{(n)+} &= p_0^{(n)} + p_1^{(n)}, \\ p_k^{(n)+} &= p_{k+1}^{(n)} \quad k = 1, 2, \dots, s-1, \\ p_s^{(n)+} &= 0. \end{aligned} \quad (5.2)$$

Bild 5.1 zeigt einen exemplarischen Prozeßverlauf zwischen zwei Regenerationszeitpunkten. Der Prozeßverlauf entspricht einem Geburtsprozeß und wird durch den Ankunftsprozeß der Anforderungen in die betrachtete Sendewarteschlange (Poisson-Prozeß), sowie durch deren endliche Speicherkapazität bestimmt. Der Prozeßverlauf nach den Regenerationszeitpunkten (t_n^+) ist damit zeitabhängig, die Zustandswahrscheinlichkeiten sind von der Zeitspanne t in der folgenden Form abhängig

$$\begin{aligned} p_k^{(n)}(t) &= \sum_{h=0}^k p_h^{(n)+} \cdot a_{k-h}(t), \quad k = 0, 1, \dots, s-1 \\ p_s^{(n)}(t) &= \sum_{h=0}^s p_h^{(n)+} \cdot \sum_{m=s-h}^{\infty} a_m(t); \end{aligned} \quad (5.3)$$

dabei ist

$$a_i(t) = \frac{(\lambda t)^i}{i!} e^{-\lambda t}$$

die Wahrscheinlichkeit, daß in dem Zeitintervall (t_n, t_n+t) genau i Ereignisse (Ankünfte) des Poisson-Prozesses auftreten (siehe Gleichung 4.28).

Durch das Einsetzen der Gleichungen für Zustandsübergänge im Regenerationszeitpunkt (5.2) in das Gleichungssystem für die zeitabhängigen Zustandswahrscheinlichkeiten (5.3) entsteht somit

$$p_k^{(n)}(t) = p_0^{(n)} a_k(t) + \sum_{h=0}^k p_{h+1}^{(n)} a_{k-h}(t), \quad (5.4)$$

$$p_s^{(n)}(t) = p_0^{(n)} \cdot \sum_{k=s}^{\infty} a_k(t) + \sum_{h=0}^{s-1} p_{h+1}^{(n)} \cdot \sum_{m=s-h}^{\infty} a_m(t),$$

wobei $k = 0, 1, 2, \dots, s-1$.

Entspricht die Zeitspanne t dem Abstand zwischen zwei Regenerationszeitpunkten der eingebetteten Markoff-Kette, so ist es möglich, das Übergangsverhalten zwischen den benachbarten Regenerationszeitpunkten (t_n) und (t_{n+1}) zu bestimmen. Die Zeitdauer zweier aufeinanderfolgender Buszuteilungen wird durch die Zufallsvariable T_V bzw. durch ihre Verteilungsdichtefunktion $f_V(t)$ analytisch erfaßt. Damit ist es möglich, durch Bildung der totalen Wahrscheinlichkeit und unter Berücksichtigung der Verteilungsdichtefunktion der Besuchszeit die Zeitabhängigkeit der Zustandswahrscheinlichkeiten zu eliminieren, mit der Beziehung

$$p_k^{(n+1)} = \int_{t=0}^{\infty} p_k^{(n)}(t) f_V(t) dt, \quad (5.5)$$

entsteht ein Gleichungssystem, welches die Berechnung der Zustandswahrscheinlichkeiten einer Sendewarteschlange zu den eingebetteten Zeitpunkten ermöglicht:

$$p_k^{(n+1)} = p_0^{(n)} q_k + \sum_{h=1}^{k+1} p_h^{(n)} q_{k-h+1}, \quad k = 0, 1, \dots, s-1$$

$$p_s^{(n+1)} = p_0^{(n)} \cdot \sum_{k=s}^{\infty} q_k + \sum_{h=1}^s p_h^{(n)} \cdot \sum_{m=s-h+1}^{\infty} q_m. \quad (5.6)$$

Durch das Gleichungssystem (5.6) wird der Vektor der Zustandswahrscheinlichkeiten $\underline{p}^{(n)}$ berechnet und der Übergang zwischen zwei benachbarten Regenerationspunkten der Markoff-Kette im stationären Falle beschrieben, wobei

$$q_i = \int_{t=0}^{\infty} a_i(t) f_V(t) dt, \quad i = 0, 1, 2, \dots, \quad (5.7)$$

der Wahrscheinlichkeit entspricht, daß zwischen zwei Regenerationszeitpunkten i Ankunftsereignisse des Poisson-Prozesses auftreten.

5.3.2 Verteilung zu beliebigen Zeitpunkten

Aufgrund der Eigenschaften des Poisson-Ankunftsprozesses treten die Ankunfts-Ereignisse unabhängig voneinander und unabhängig vom Zustand der Sendewarteschlange auf. Zur Berechnung der charakteristischen Größen wie mittlere Wartezeit oder Verlustwahrscheinlichkeit für eine eintreffende Anforderung muß die Verteilung der Zufallsvariablen X_n zu beliebigen Zeitpunkten bekannt sein.

Im folgenden soll ein beliebiger Zeitpunkt $t_n + t$ zwischen den aufeinanderfolgenden Regenerationszeitpunkten t_n und t_{n+1} betrachtet werden. Die Zeitspanne t entspricht dann der Rückwärtsrekurrenzzeit der Besuchszeit mit der Dichtefunktion

$$f_V^r(t) = \frac{1}{t_V} (1 - F_V(t)). \quad (5.8)$$

Daraus folgt entsprechend der Gleichung (5.7) die Wahrscheinlichkeit für i Ankunftsereignisse des Poisson-Prozesses während der Zeitspanne t zu

$$q_i^* = \int_{t=0}^{\infty} a_i(t) f_V^r(t) dt. \quad (5.9)$$

Damit ist es möglich, entsprechend zu Gleichung (5.6) die Zustandswahrscheinlichkeiten der Sendewarteschlange zu beliebigen Zeitpunkten durch folgende Gleichung anzugeben

$$p_k^* = p_0^{(n)} \cdot q_k^* + \sum_{h=1}^{k+1} p_h^{(n)} \cdot q_{k-h+1}^*, \quad k = 0, 1, \dots, s-1$$

$$p_s^* = p_0^{(n)} \cdot \sum_{k=s}^{\infty} q_k^* + \sum_{h=1}^s p_h^{(n)} \cdot \sum_{m=s-h+1}^{\infty} q_m^* \quad (5.10)$$

Da sich die Wahrscheinlichkeit q_i^* zwischen den Regenerationszeitpunkten nicht ändert, entsprechen die Zustandswahrscheinlichkeiten der Zufallsvariablen X_n nach Gleichung (5.10) zu beliebigen Zeitpunkten einem zeitgemittelten Wert, welcher zwischen zwei Regenerationszeitpunkten gültig ist, nicht aber in diesen selbst.

5.3.3 Momentenapproximation der Besuchszeitverteilungsfunktion

Die Dichtefunktion der Besuchszeit kann für den Fall negativ-exponentiell verteilter Bedienzeiten, konstant verteilten Verwaltungsdauern und einer Bedieneinheit durch ein Phasenmodell nach Cox angegeben werden. Die Markoff-Phasen der Besuchszeitverteilungsfunktion entsprechen den Stationsbediendauern und werden jeweils mit der Wahrscheinlichkeit, daß Anforderungen in der stationszugeordneten Warteschlange warten, durchlaufen. Die gesamte Verteilungsfunktion ist außerdem noch um die Summe aller Zuteilungsdauern, die beim Durchlaufen der Zuteilungsreihenfolge der Bedieneinheit auftreten, verschoben.

Sind die Bedienzeiten und Verwaltungsdauern beliebig verteilt und nur erstes und zweites Moment bzw. Mittelwert und Variationskoeffizient der Besuchszeitverteilungsfunktion bekannt, so kann durch Momentenapproximation eine Verteilungsfunktion gefunden werden, die in den ersten beiden Momenten exakt mit der unbekannteren Besuchszeitverteilungsfunktion übereinstimmt (siehe Kapitel 4.2.2.3).

Damit ist es möglich, die Wahrscheinlichkeit für i Ankünfte während der Besuchszeit T_V , abhängig von bekannten Größen

t_V und c_V , zu bestimmen. Aus Gleichung (5.7) entsteht damit unter Verwendung der Gleichung (4.11) und mit Hilfe der Beziehung

$$\int_0^T x^n e^{-ax} dx = \frac{n!}{a^{n+1}} \left(1 - \sum_{k=0}^n \frac{(aT)^k e^{-aT}}{k!} \right),$$

(a und n positive und ganzzahlige Parameter)

für $0 \leq c_V \leq 1$:

$$q_i = \frac{(\lambda t_2)^i}{(1+\lambda t_2)^{i+1}} \cdot e^{-\lambda t_1} \cdot \sum_{k=0}^i \frac{1}{k!} \left(\frac{t_1}{t_2} (1+\lambda t_2) \right)^k,$$

für $c_V \geq 1$:

(5.11)

$$q_i = p_1 \cdot \frac{(\lambda t_1)^i}{(1+\lambda t_1)^{i+1}} + p_2 \cdot \frac{(\lambda t_2)^i}{(1+\lambda t_2)^{i+1}}.$$

Entsprechend kann bei der Berechnung der Wahrscheinlichkeit für i Ankunftsereignisse des Poisson-Prozesses bis zu einem beliebigen Betrachtungszeitpunkt (Gleichung 5.9) vorgegangen werden. Die approximierte Verteilungsdichtefunktion $f_V^r(t)$ für die Rückwärtsrekurrenzzeit der Besuchszeit errechnet sich ausgehend von der Verteilungsfunktion nach Gleichung (4.11) mit Hilfe der Gleichung (5.8) für $0 \leq c_V \leq 1$ zu

$$f_V^r(t) = \frac{1}{t_1 + t_2} \begin{cases} 1 & , \quad 0 \leq t \leq t_1 \\ e^{-(t-t_1)/t_2} & , \quad t \geq t_1 \end{cases}$$

und für $c_V \geq 1$ zu

(5.12)

$$f_V^r(t) = \frac{1}{p_1 t_1 + p_2 t_2} (p_1 e^{-t/t_1} + p_2 e^{-t/t_2}).$$

Die Wahrscheinlichkeit für i Ankünfte während der Zeit T_V^r errechnet sich dann ausgehend von den Gleichungen (5.8), (5.9) und (5.12) für $0 \leq c_V \leq 1$ zu

$$q_i^* = \frac{1}{\lambda(t_1 + t_2)} \cdot \left(1 - \sum_{k=0}^i \frac{(\lambda t_1)^k}{k!} e^{-\lambda t_1}\right) + \frac{t_2}{t_1 + t_2} q_i$$

bzw. für $c_V \geq 1$ zu (5.13)

$$q_i^* = \frac{(\lambda t_1)^i}{2(1 + \lambda t_1)^{i+1}} + \frac{(\lambda t_2)^i}{2(1 + \lambda t_2)^{i+1}}$$

5.4 Analyse der Besuchszeitverteilungsfunktion

In der Analyse der Zustandsverteilung der Sendewarteschlange wurde davon ausgegangen, daß erstes und zweites Moment der Besuchszeitverteilungsfunktion bekannt sind. Da Besuchszeitverteilung und Zustandswahrscheinlichkeiten direkt voneinander abhängig sind und der Zustandsvektor eine endliche Länge besitzt, kann die Analyse nur durch eine Iteration zwischen Besuchszeitverteilungsfunktion und Zustandsverteilung erfolgen (siehe Anhang 1).

Die Zeit für das Durchlaufen der Zuteilungsreihenfolge einer Bedieneinheit (Zykluszeit) wird im folgenden durch die Zufallsvariable T_C beschrieben. Zur Berechnung ihrer Verteilungsfunktion wird die Zykluszeit, entsprechend der Stationszahl, in g Segmente unterteilt. Die Zykluszeitsegmentdauer T_{Sj} entspricht der Aufenthaltsdauer einer Bedieneinheit an der Station j . Die Faltung aller Dichtefunktionen der Zykluszeitsegmentdauern ergibt die Verteilungsdichtefunktion des Zykluszeitprozesses einer Bedieneinheit. Für alle Zykluszeitprozesse wird die Erneuerungseigenschaft angenommen. Durch die Überlagerung aller Zykluszeitprozesse entsteht der

Zuteilungsprozeß - ein Punktprozeß, ebenfalls mit angenommener Erneuerungseigenschaft, der durch die Zufallsvariable T_G beschrieben wird. Eine Zerlegung des Zuteilungsprozesses unter Berücksichtigung auftretender Interferenzen ergibt dann den Besuchszeitprozeß.

5.4.1 Momente des Zykluszeitsegments

Das Phasenmodell des Zykluszeitsegments (siehe Bild 5.2) für die betrachteten Warteschlange j enthält als Parameter die zufallsabhängigen Größen T_{Oj} (Dauer des Verwaltungsvorgangs) und T_{Hj} (Bediendauer), die Wahrscheinlichkeiten p_{0j} (keine Anforderung wartet zum Zeitpunkt der Buszuteilung) und B_{pj} (eine Zuteilung interferiert am Port mit einer momentanen Bedienung durch eine andere Bedieneinheit).

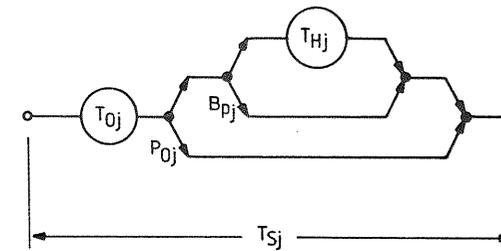


Bild 5.2: Phasenmodell des Zykluszeitsegmentes

Die Zykluszeitsegmentdauer (T_{Sj}) entspricht somit der Zeit, die eine Bedieneinheit (Bus) benötigt, um bei Durchlaufen der vorgegebenen Zuteilungsreihenfolge von der Station j zu der Station $j+1$ zu gelangen.

Die ersten beiden Momente der Verteilungsfunktion des Zykluszeitsegments (T_{Sj}) werden mit Hilfe des in Bild 5.2 dargestellten Phasenmodells über die Laplace-Stieltjes-Transformierte des Zykluszeitsegments

$$\Phi_{Sj}(s) = \Phi_{0j}(s) \cdot ((1-p_{0j}) \cdot ((1-B_{pj}) \Phi_{Hj}(s) + B_{pj}) + p_{0j}), \quad (5.14)$$

bestimmt. Der Mittelwert des Zykluszeitsegmentes errechnet sich dann nach Glg. (4.8) bzw. (4.9) zu

$$t_{Sj} = t_{0j} + (1-p_{0j})(1-B_{pj})t_{Hj}. \quad (5.15)$$

Entsprechend ergibt sich die Varianz zu

$$\sigma_{Sj}^2 = \sigma_{0j}^2 + (1-p_{0j})(1-B_{pj})(t_{Hj}^{(2)} - (1-p_{0j})(1-B_{pj})t_{Hj}^2). \quad (5.16)$$

5.4.2 Der Zykluszeitprozeß

Die Zeit für das Durchlaufen der Zuteilungsreihenfolge einer bestimmten Bedieneinheit (Zykluszeit) wird mit Hilfe der Zykluszeitsegmente T_{Sj} aller Stationen berechnet.

Der Zykluszeitprozeß wird im folgenden als stationärer Erneuerungsprozeß betrachtet, der für jede Station dieselbe Charakteristik aufweist. Der Index j für die betrachtete Station entfällt damit für den Zykluszeitprozeß.

Unter der Annahme, daß die einzelnen Zykluszeitsegmente statistisch voneinander unabhängig sind, kann die Verteilungsdichtefunktion der Zykluszeit $f_C(t)$ einer Bedieneinheit im Bildbereich durch das Produkt der einzelnen Laplace-Transformierten der Zykluszeitsegmente dargestellt werden:

$$\Phi_C(s) = \prod_{j=1}^g \Phi_{Sj}(s). \quad (5.17)$$

Für die Berechnung der ersten beiden Momente bzw. Mittelwert und Varianz der Zykluszeit genügt die Annahme, daß die

einzelnen Zyklussegmentdauern T_{Sj} nicht miteinander korreliert sind. Mittelwert und Varianz der Zykluszeit errechnen sich dann zu

$$t_C = \sum_{j=1}^g t_{Sj}, \text{ bzw. } \sigma_C^2 = \sum_{j=1}^g \sigma_{Sj}^2. \quad (5.18)$$

Die Zykluszeit einer Bedieneinheit hängt im allgemeinen davon ab, ob die besuchten Stationen Nachrichten in dem betrachteten Zyklus übertragen. Dieser Einfluß der Stationen auf die betrachtete Realisierung der Zykluszeit wird durch die ersten beiden Momente der Zykluszeitverteilungsdichtefunktion erfaßt. Unter der Voraussetzung der Unabhängigkeit zwischen den Zykluszeitsegmenten bleiben allerdings Abhängigkeiten zwischen den einzelnen Stationen unberücksichtigt. Diese zum Teil starken Unabhängigkeitsannahmen werden durch die Betrachtung der Abhängigkeit zweier aufeinanderfolgender Zyklen (bedingte Zykluszeiten) reduziert.

5.4.3 Der Zuteilungsprozeß

Während bisher immer Bedieneinheiten bzw. deren Zykluszeitverteilungsfunktion betrachtet wurden, wird im folgenden der Blick auf die Überlagerung der Zykluszeitprozesse am Sendeport einer ausgewählten Warteschlange j (Bild 5.3) gerichtet.

Die Zykluszeitprozesse der einzelnen Bedieneinheiten werden als voneinander unabhängige, stationäre Punktprozesse mit Erneuerungseigenschaft angenommen, welche sich an den Ports der Warteschlangen überlagern. Es entsteht ein stationärer Punktprozeß, der die Zuteilungszeitpunkte aller Bedieneinheiten bei der betrachteten Warteschlange darstellt.

Der Prozeß aufeinanderfolgender Buszuteilungen am Sendeport entspricht dem Bus-Zuteilungsprozeß und wird durch die beliebig verteilte Zufallsvariable T_C beschrieben (Bild 5.3).

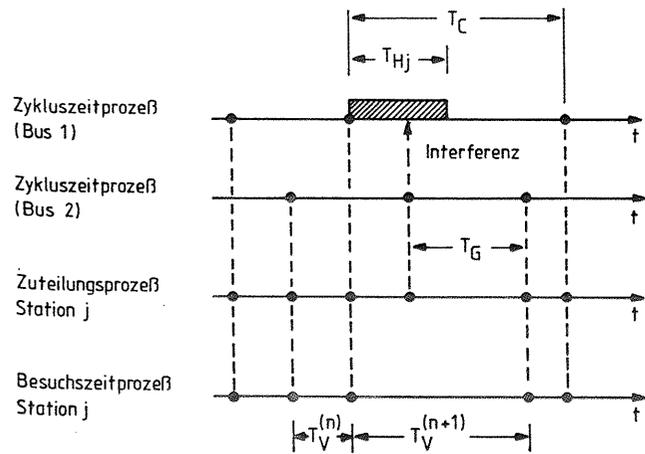


Bild 5.3: Punktprozesse am Sendeport

Der Zuteilungsprozeß ist, wie in Kapitel 4.3.6 bzw. in /52/ beschrieben, im allgemeinen Falle nicht mehr ein Erneuerungsprozeß. Unter Annahme der Erneuerungseigenschaft ist, gemäß Kapitel 4.3.6, durch einfache Flußbetrachtungen das erste Moment des Zuteilungsprozesses einfach aus Gleichung (4.35) zu bestimmen. Unter der Voraussetzung, daß die Zykluszeitprozesse aller Bedieneinheiten identisch sind, ergibt sich somit

$$t_G = \frac{t_C}{n} \quad (5.19)$$

Das zweite Moment des Zuteilungsprozesses bzw. der Variationskoeffizient c_G kann durch Momentenapproximation der Zykluszeitverteilungen entsprechend dem in /52/ angegebenen Verfahren ebenfalls berechnet werden. Der Zuteilungsprozeß existiert im Modell nur einmal und besitzt für jede Station dieselbe Charakteristik.

5.4.4 Sendeportinterferenzen

Existieren im Modell mehrere Bedieneinheiten, die unabhängig voneinander der Warteschlange einer betrachteten Station j zugeteilt werden, dann kann eine Buszuteilung am Sendeport der Station j mit einer momentanen Bedienung der Station j durch einen der anderen Busse interferieren, da nicht mehr als eine Bedieneinheit den Port der Station j belegen kann. Eine Interferenz wird innerhalb der Verwaltungszeit für die Buszuteilung erkannt. Die interferierende Bedieneinheit wird der nächsten in der Zuteilungsreihenfolge vorgesehenen Warteschlange bzw. Station zugeteilt. Prinzipiell entstehen Port-Interferenzen durch die Koinzidenz von Bedienphasen der betrachteten Station j mit dem Zuteilungsprozeß der Bedieneinheiten, siehe Bild 5.4.

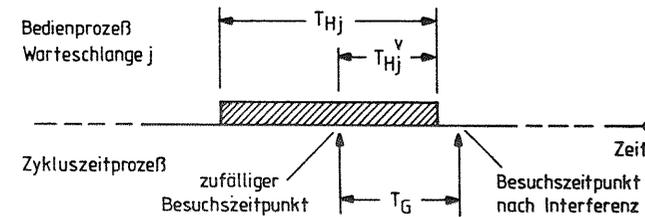


Bild 5.4: Interferenz am Sendeport

Die Wahrscheinlichkeit einer Sendeportinterferenz (B_{pj}) kann näherungsweise in folgender Weise errechnet werden:

$$B_{pj} \approx \frac{(n-1)}{n} \beta_j (1 + P(T_{Hj}^v - T_G \geq 0)), \quad (5.20)$$

dabei entspricht

$$\rho_j = \lambda_j (1 - p_{sj}) t_{Hj} \quad (5.21)$$

der durch das Multi-Bus-Verbindungsnetzwerk aufgenommenen bzw. akzeptierten Last.

Die Wahrscheinlichkeit, daß die Differenz von Vorwärtsrekurrenzzeit der Bediendauer und der Zuteilungsdauer positiv wird, errechnet sich mit Hilfe von Gleichung (4.24) zu

$$P(T_{Hj}^V - T_G \geq 0) = \frac{1}{t_{Hj}} \cdot \int_{t=0}^{\infty} F_{Hj}^C(t) \cdot F_G(t) \cdot dt. \quad (5.22)$$

5.4.5 Der Besuchszeitprozeß

Durch die am Sendeport der Warteschlange j auftretenden Interferenzen konkurrierender Buszuteilungen, d.h. durch den Verzicht von möglicher Parallelität am Sendeport, kann die vom gesamten Multi-Bus-Verbindungsnetzwerk zur Verfügung stehende Kapazität nicht vollständig genutzt werden. Damit ist der Prozeß, der Zustandsänderungen in der Warteschlange j bewirkt, nicht der Zuteilungsprozeß, da interferierende Buszuteilungen nicht als Besuch bei Warteschlange j interpretiert werden dürfen. Der Zuteilungsprozeß wird deshalb so in zwei Prozesse zerlegt, daß ein Komponentenprozeß (der Besuchszeitprozeß) die Abstände der Besuche beschreibt, die nicht mit momentan ablaufenden Bedienungen an Warteschlange j konkurrieren (Bild 5.3) und somit keiner Interferenz unterliegen.

Entsprechend der Zerlegung von Erneuerungsprozessen (Gleichung 4.33), berechnen sich Mittelwert und Variationskoeffizient der Besuchszeitverteilungsfunktion für die Warteschlange j wie folgt

$$t_{Vj} = \frac{t_G}{1 - B_{pj}}, \quad c_{Vj}^2 = (1 - B_{pj}) \cdot c_G^2 + B_{pj}, \quad (5.23)$$

wobei $j=1,2,\dots,g$.

Im Gegensatz zu Zykluszeitprozeß und Zuteilungsprozeß existiert der Besuchszeitprozeß individuell für jede der am Multi-Bus-Verbindungsnetzwerk angeschlossenen Stationen.

5.4.6 Prinzip der bedingten Zykluszeiten

Die Abhängigkeiten zweier unmittelbar aufeinanderfolgenden Zyklen (Besuche) einer Bedieneinheit an Warteschlange j können bezüglich der Wahrscheinlichkeit, ob die betrachtete Bedieneinheit beim letzten Besuch Warteschlange j bediente oder nicht, berücksichtigt werden. Im folgenden werden deshalb zwei Typen von Zykluszeiten einer Bedieneinheit betrachtet:

- T_{Cj}^- , Zykluszeit ohne Bedienung bei einem Besuch der Warteschlange j ;
- T_{Cj}^+ , Zykluszeit mit Bedienung bei einem Besuch der Warteschlange j ;

Das Prinzip der bedingten Zykluszeiten berücksichtigt damit Abhängigkeiten des Zykluszeitprozesses, in dem die Vergangenheit des Prozeßverlaufes teilweise mitberücksichtigt wird. Da die bedingte Zykluszeit mit Bedienung immer um eine Bedienzeit länger ist als die Zykluszeit ohne Bedienung, kann daraus gefolgert werden, daß eine Zykluszeit mit Bedienung mit größerer Wahrscheinlichkeit wiederum eine Zykluszeit mit Bedienung als eine Zykluszeit ohne Bedienung zur Folge hat. Entsprechendes gilt auch im umgekehrten Falle; eine Zykluszeit ohne Bedienung wird mit größerer Wahrscheinlichkeit eine Zykluszeit ohne Bedienung zur Folge haben. Im wesentlichen führt die explizite Betrachtung von bedingten Zykluszeiten, d.h. ob bei Station j während der letzten, vorausgegangenen

Zykluszeit eine Bedienung erfolgte oder nicht, zu einer Erhöhung der Varianz der berechneten Zykluszeitverteilungsfunktion.

5.4.6.1 Momente der bedingten Zykluszeiten

Unter Berücksichtigung der oben eingeführten Bedingung ist es möglich, entsprechend Gleichung (5.17), die Laplace-Stieltjes-Transformierte der bedingten Zykluszeiten anzugeben. Für die Zykluszeit ohne Bedienung an der betrachteten Warteschlange j ergibt sich

$$\bar{\Phi}_{Cj}(s) = \Phi_{Oj}(s) \cdot \prod_{\substack{i=1 \\ i \neq j}}^g \bar{\Phi}_{Si}(s); \quad (5.24a)$$

weiterhin ergibt sich für die Zykluszeit mit Bedienung an der betrachteten Warteschlange j

$$\bar{\Phi}_{Cj}^{\prime\prime}(s) = \Phi_{Oj}(s) \cdot (1 - B_{pj}) \Phi_{Hj}(s) + B_{pj} \cdot \prod_{\substack{i=1 \\ i \neq j}}^g \bar{\Phi}_{Si}(s). \quad (5.24b)$$

Mit Hilfe der Gleichungen (5.24a,b) ist es möglich, analog zu den Gleichungen (5.18), für die bedingten Zykluszeiten den Mittelwert:

$$\bar{t}_{Cj} = t_{Oj} + \sum_{\substack{i=1 \\ i \neq j}}^g t_{Si}, \quad (5.25a)$$

$$\bar{t}_{Cj}^{\prime\prime} = t_{Oj} + (1 - B_{pj}) t_{Hj} + \sum_{\substack{i=1 \\ i \neq j}}^g t_{Si}, \quad (5.25b)$$

sowie die Varianz:

$$\sigma_{Cj}^2 = \sigma_{Oj}^2 + \sum_{\substack{i=1 \\ i \neq j}}^g \sigma_{Si}^2, \quad (5.26a)$$

$$\sigma_{Cj}^{\prime\prime 2} = \sigma_{Oj}^2 + (1 - B_{pj}) \cdot (t_{Hj}^{(2)} - (1 - B_{pj}) t_{Hj}^2) + \sum_{\substack{i=1 \\ i \neq j}}^g \sigma_{Si}^2, \quad (5.26b)$$

anzugeben.

5.4.6.2 Einfluß der bedingten Zykluszeiten auf die Verteilung der Systemzustände

Das Prinzip der bedingten Zykluszeiten führt auf bedingte Zuteilungszeit- bzw. Besuchszeitverteilungsdichtefunktionen und damit auf bedingte Übergangswahrscheinlichkeiten der eingebetteten Markoff-Kette (Gleichung 5.7). Im wesentlichen bedeutet dies, daß die Gleichungen (5.6) bzw. (5.7) für die Berechnung der Zustandswahrscheinlichkeiten der eingebetteten Markoff-Kette sich in folgender Weise ändern (aus Übersichtlichkeitsgründen wird auf den Stationsindex j verzichtet):

$$p_k^{(n+1)} = p_0^{(n)} \cdot \bar{q}_k + \sum_{h=1}^{k+1} p_h^{(n)} \cdot \bar{q}_{k-h+1}^{\prime\prime}, \quad k = 0, 1, \dots, s-1$$

$$p_s^{(n+1)} = p_0^{(n)} \cdot \sum_{k=s}^{\infty} \bar{q}_k + \sum_{h=1}^s p_h^{(n)} \cdot \sum_{m=s-h+1}^{\infty} \bar{q}_m^{\prime\prime}, \quad (5.27)$$

wobei die Wahrscheinlichkeiten, daß zwischen zwei Regenerationszeitpunkten (Besuchszeit mit Bedienung bzw. Besuchszeit ohne Bedienung) i Ankunftsereignisse in der betrachteten Station auftreten, wie folgt berechnet werden ($i=0,1,2,\dots$):

$$q_i^- = \int_{t=0}^{\infty} a_i(t) \cdot f_V^-(t) dt, \quad (5.28a)$$

$$q_i^{--} = \int_{t=0}^{\infty} a_i(t) \cdot f_V^{--}(t) dt. \quad (5.28b)$$

Das Prinzip der bedingten Zykluszeiten beeinflusst darüber hinaus noch das Gleichungssystem zur Berechnung der Zustandswahrscheinlichkeiten zu beliebigen Zeitpunkten (Gleichung (5.9)) und die Berechnung der Übergangswahrscheinlichkeiten (Gleichung (5.10)) in folgender Weise:

$$q_i^{-*} = \int_{t=0}^{\infty} a_i(t) \cdot f_V^{-r}(t) dt, \quad (5.29a)$$

$$q_i^{--*} = \int_{t=0}^{\infty} a_i(t) \cdot f_V^{--r}(t) dt, \quad (5.29b)$$

wobei wiederum $i=0,1,2,\dots$.

$$p_k^* = p_0^{(n)} \cdot q_k^{-*} + \sum_{h=1}^{k+1} p_h^{(n)} q_{k-h+1}^{--*}, \quad k = 0,1,\dots,s-1 \quad (5.30)$$

$$p_s^* = p_0^{(n)} \cdot \sum_{k=s}^{\infty} q_k^{-*} + \sum_{h=1}^s p_h^{(n)} \cdot \sum_{m=s-h+1}^{\infty} q_m^{--*}.$$

5.5 Ergebnisse und Approximationsgenauigkeit

Die vorgestellte Methode zur Analyse von Multi-Bus-Verbindungsnetzwerken kann sowohl zur Untersuchung von symmetrischen Systemen, d.h. alle stationsindividuellen Systemparameter sind identisch, als auch zur Untersuchung von Systemen mit unterschiedlichem Verkehrsaufkommen in den Stationen oder unterschiedlichen Stationsparametern angewendet werden. Zur Validierung der analytisch gefundenen Ergebnisse wurde die Methode der zeitreuen Simulation angewendet (vergleiche Kapitel 4.1.2). Die Simulationsergebnisse sind mit ihrem 95% Vertrauensintervall angegeben.

Im folgenden wird das Verhalten einiger typischer charakteristischer Größen an unterschiedlichen Systemmodellen vorgestellt und diskutiert. Allen Ergebnissen liegt die einfache zyklische Zuteilung der Bedieneinheiten (Busse) zu den Stationen zugrunde. Die Verwaltungsdauer für die Buszuteilung besitze eine deterministische Verteilungsfunktion. Alle zeitlichen charakteristischen Größen sind normiert auf die mittlere Busbelegungsdauer einer Nachricht. Alle Parameterstudien wurden für negativ-exponentielle sowie hyperexponentielle ($c_H=5$) Verteilungsfunktionen der Busbelegungsdauer durchgeführt. Wenn nicht anders angegeben, besitzen die Warteschlangen der Stationen eine Kapazität von $S=5$ Warteplätzen.

5.5.1 Einfluß der Stationszahl

Alle Ergebnisse dieses Abschnittes sind über dem Gesamtangebot

$$\rho_0 = \sum_{j=1}^g \lambda_j t_{Hj}, \quad (5.31)$$

aufgetragen und betreffen ein symmetrisch belastetes System mit 4 bzw. 20 identisch aufgebauten Stationen.

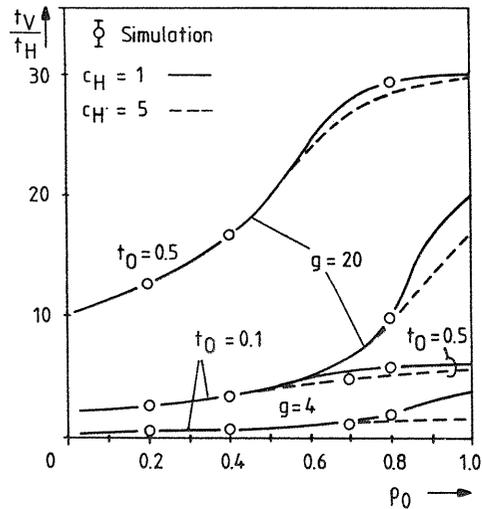


Bild 5.5: Einfluß von Stationszahl g und Verwaltungsdauer t_0 auf die mittlere Besuchszeit t_v

Bild 5.5 zeigt den Verlauf der mittleren Besuchszeit der Bedieneinheit zu einer Station für zwei verschiedene Stationszahlen sowie für die Verwaltungsdauer einer Buszuteilung (t_0) von 10% bzw. 50% der mittleren Busbelegungsdauer einer Nachricht (t_H). Die mittlere Besuchszeit ergibt sich bei geringem Verkehrsangebot durch die Summe aller Verwaltungszeiten, sie entspricht somit dem Leerzyklus, welcher sich beim Durchlaufen der Zuteilungsreihenfolge ergibt, unter der Bedingung, daß keine Nachricht übertragen wird.

Mit zunehmender Last nähern sich die mittleren Besuchszeiten einem weiteren Grenzwert, an jeder Station wird dann beim Durchlaufen der Zuteilungsreihenfolge eine Nachrichtenübertragung durchgeführt (Vollzyklus).

Für den Fall von hyperexponentiell verteilten Übertragungszeiten erhöht sich die Sendespeicherblockierung, die vom System akzeptierte Last ist dann im Vergleich zu negativ-expo-

nentiell verteilten Übertragungszeiten geringer. Dadurch ergeben sich speziell in höheren Lastbereichen geringere mittlere Besuchszeiten.

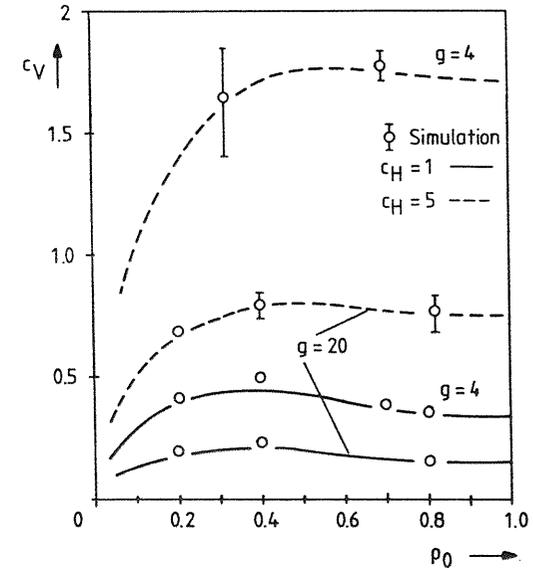


Bild 5.6: Einfluß der Stationszahl g auf den Variationskoeffizienten c_v der Besuchszeit

Durch die Zwei-Momente-Approximation der Besuchszeitverteilungsfunktion wird das zweite Moment bzw. der Variationskoeffizient der Besuchszeit benötigt. Bild 5.6 zeigt den Verlauf des Variationskoeffizienten der Besuchszeit über dem Gesamtangebot für die mittlere Bus-Verwaltungsdauer $t_0 = 0.5t_H$. Durch die Annahme, daß die Zykluszeitsegmente voneinander unabhängig sind, werden die Kovarianzterme bei der Berechnung der Besuchszeitvarianz nicht berücksichtigt (Gleichung 4.21). Dies führt im wesentlichen zu einer Unterschätzung der Besuchszeitvarianz und damit zu einer Unterschätzung des Variationskoeffizienten der Besuchszeitverteilungsfunktion. Für negativ-exponentiell verteilte Bedienzeiten und hohe

Angebotswerte nähert sich die Besuchszeitverteilungsfunktion einer Erlang-g-Verteilungsfunktion.

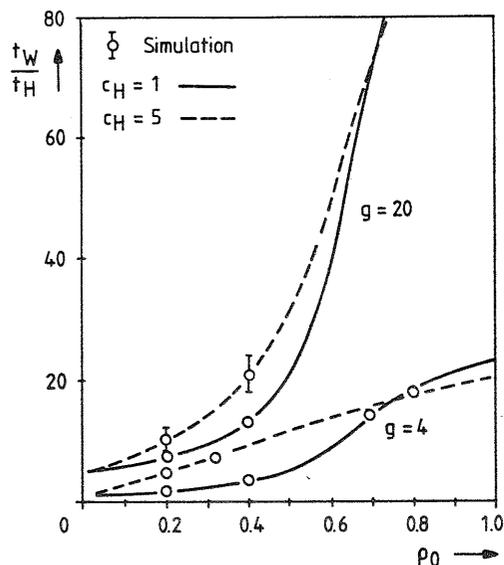


Bild 5.7: Einfluß der Stationszahl g auf die mittlere Wartezeit t_w in der Sendewarteschlange

Bedienzeiten mit hoher Varianz führen im Vergleich zu negativ-exponentiellen Bedienzeiten in niedrigen bis mittleren Lastbereichen zu einer Erhöhung der Wartezeiten in der Sendewarteschlange. In den höheren Lastbereichen führen Bedienzeiten mit hoher Varianz zu größeren Nachrichtenverlusten bzw. Blockierungen. Damit reduziert sich die aufgenommene Last und impliziert geringere Wartezeiten im Sendespeicher. Dieser Effekt ist in Bild 5.7 für den Fall $t_0 = 0.5t_H$ dargestellt.

Durch den relativ großen Verwaltungsaufwand steigt mit zunehmendem Angebot und großer Stationszahl die Besuchszeit und damit die mittlere Wartezeit in der Sendewarteschlange stark an.

Bild 5.8 zeigt den Einfluß der Stationszahl auf die Blockierwahrscheinlichkeit der Sendewarteschlange in Abhängigkeit des

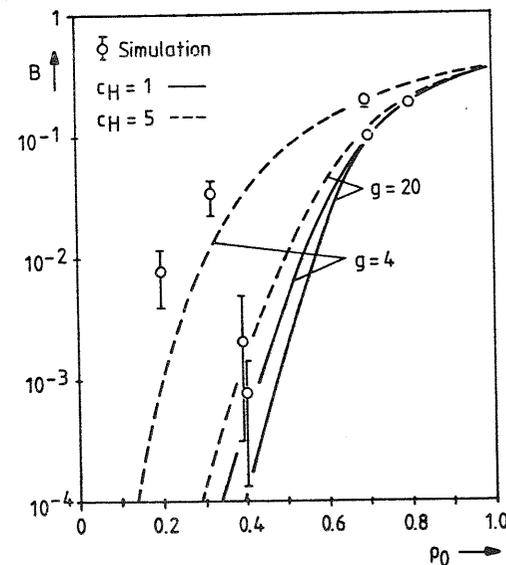


Bild 5.8: Einfluß der Stationszahl g auf die Sendespeicherblockierung B

Angebots und für unterschiedliche Variationskoeffizienten der Bedienzeit, wiederum für den Fall $t_0 = 0.5t_H$. Die Ergebnisse der Analyse weichen zum Teil von den Simulationsergebnissen ab, siehe Bild 5.8. Der Fehler der Analyse ist auf die Unterschätzung der Besuchszeitvarianz zurückzuführen. Generell weisen die Ergebnisse der Analyse immer dieselbe Tendenz auf wie Ergebnisse, die auf simulative Weise gewonnen wurden.

5.5.2 Dimensionierung der Sendewarteschlangen

Im folgenden Abschnitt wird als Anwendungsbeispiel der Analyse die Dimensionierung des Sendespeichers von peripheren Steuerungseinrichtungen einer digitalen Nebenstellenanlage (siehe auch Bild 3.4, Kapitel 3) für ein System mit 8 Stationen und einem Uni-Bus-Verbindungsnetzwerk betrachtet.

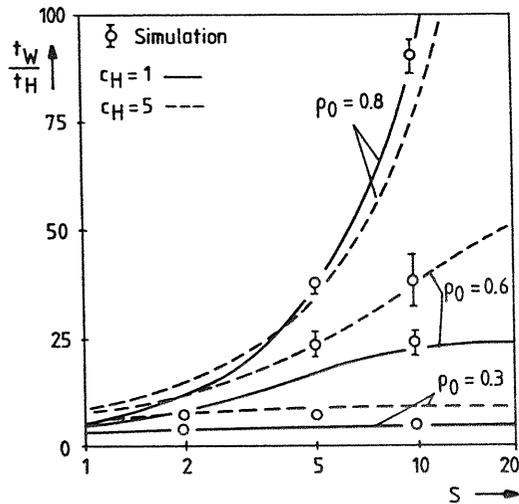


Bild 5.9: Einfluß der Anzahl von Warteplätzen S im Sendespeicher auf die mittlere Wartezeit t_w

Bild 5.9 und Bild 5.10 zeigen den Verlauf der mittleren Wartezeit und der Blockierwahrscheinlichkeit für Nachrichten im Sendespeicher einer Station über der Anzahl der Warteplätze. Es wird das Systemverhalten für drei typische Angebotswerte (Niederlast, Mittellast, Hochlast) sowie hyperexponentiell und negativ-exponentiell verteilte Übertragungszeiten untersucht.

Abhängig von den an das Gesamtsystem gestellten Anforderungen bezüglich Wartezeit oder Blockierung kann mit Hilfe der in

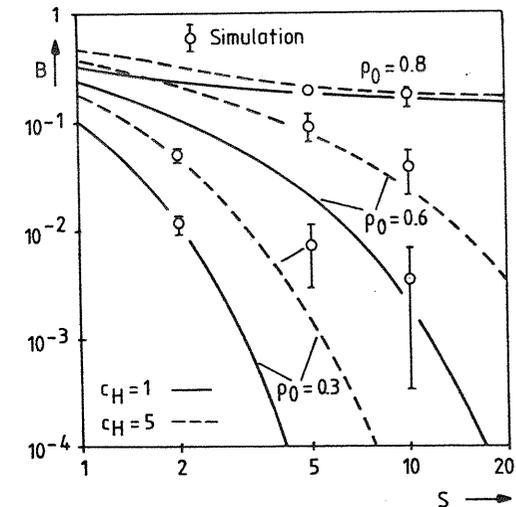


Bild 5.10: Einfluß der Anzahl von Warteplätzen S im Sendespeicher auf Blockierwahrscheinlichkeit B

Bild 5.9 bzw. Bild 5.10 dargestellten Ergebnisse eine Dimensionierung der Sendespeicher der Stationen erfolgen. Schwankungen der Nachrichtenlängen bzw. der Busbelegungszeiten werden durch die unterschiedlichen Verteilungsfunktionen der Busbelegungsdauer $c_H=1$ bzw. $c_H=5$ berücksichtigt.

5.5.3 Multi-Bus-Verbindungsnetzwerke

Nachfolgend wird der Einsatz mehrerer parallel arbeitender Bedieneinheiten (Busse), also eines Multi-Bus-Verbindungsnetzwerkes, in einem zunächst beliebigen System grundsätzlich untersucht. Es wird dabei der Verlauf der mittleren Wartezeit in den Sendespeichern der Stationen für n parallel arbeitende Bedieneinheiten diskutiert (n=1,2 oder 3).

Es wird eine symmetrische Belastung des Verbindungsnetzwer-

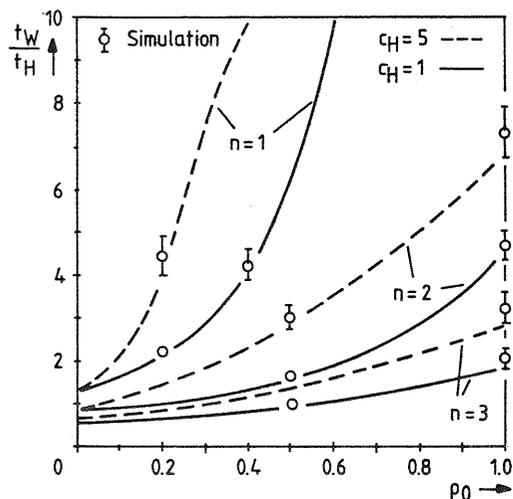


Bild 5.11: Einfluß der Busanzahl n auf die mittlere Wartezeit im Sendespeicher t_w

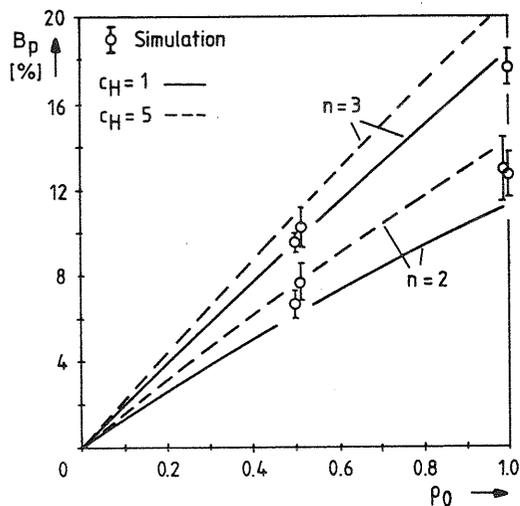


Bild 5.12: Einfluß der Busanzahl n auf die Sendeportblockierung im Sendespeicher B_p

kes angenommen. Das untersuchte System besteht aus $g=5$ identisch aufgebauten Stationen. Die maximale vorgegebene Belastung des Verbindungsnetzwerkes sei 1.0 Erlang, damit ergibt sich pro Station eine maximale Ankunftsrate von 0.2 Nachrichten pro mittlerer Busbelegungsdauer für die Übertragung einer Nachricht. Die Busverwaltungsdauer sei $t_0=0.5t_H$. Die Ergebnisse von Bild 5.11 zeigen, daß das Ein-Bus-Verbindungsnetzwerk durch den Verwaltungsaufwand, der durch die zyklische Buszuteilung entsteht, nur bis zu einem Angebot von 0.5 Erlang zufriedenstellende Wartezeiten in der Sendewarteschlange liefert. Die Reduktion der mittleren Wartezeit beim Einsatz von zwei Bussen kann durch die wesentlich kürzeren Besuchszeiten sowie durch die Lastteilung der Busse erklärt werden. Beim Einsatz von drei und mehr Bedieneinheiten nimmt die Belastung pro Bus und damit die mittlere Wartezeit einer Anforderung im Sendespeicher in geringerem Maße ab.

Bild 5.12 zeigt den Verlauf der Sendeportblockierung über dem Gesamtangebot. Die in Gleichung (5.20) angegebene Näherung zeigt für die untersuchten Parameterbereiche noch eine gute Übereinstimmung mit den durch Simulation bestimmten Werten.

5.5.4 Systeme mit unsymmetrischem Verkehrsangebot

Die Leistungsfähigkeit von Systemen mit zyklischer Zuteilung gemeinsam benutzter Betriebsmittel hängt auch stark von den Verkehrsbeziehungen zwischen den autonomen Stationen ab. Die einfache zyklische Buszuteilung behandelt alle Stationen gleich und nimmt deshalb keine Rücksicht auf ein unsymmetrisches Verkehrsaufkommen.

Die Anwendung der in diesem Abschnitt vorgestellten Ergebnisse liegt in der Leistungsuntersuchung von Informationssystemen in einer zukünftigen Büroumgebung. Das untersuchte System besteht aus einer zentralen Verarbeitungseinheit (z.B. File-Server) und vier Arbeitsplatzrechnern (Station 2,3,4,5).

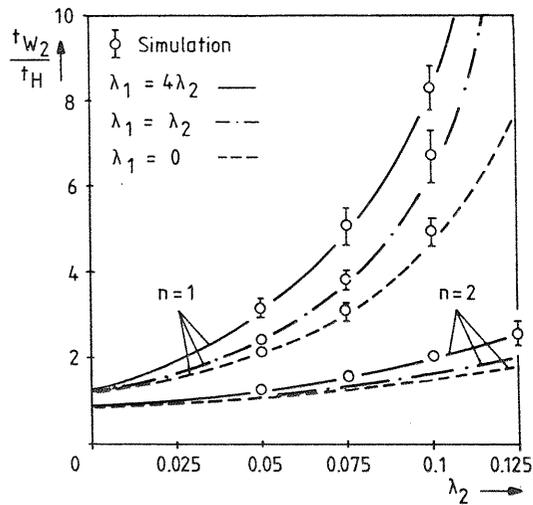


Bild 5.13: Einfluß unsymmetrischer Lastbedingungen auf die Wartezeit im Sendespeicher

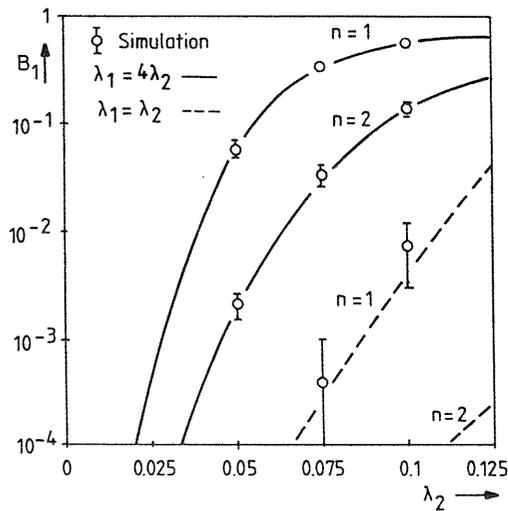


Bild 5.14: Einfluß unsymmetrischer Lastbedingungen auf die Blockierwahrscheinlichkeit des Sendespeichers

Alle Stationen kommunizieren über ein Uni- bzw. Multi-Bus-Verbindungsnetzwerk mit zyklischer Zuteilung. Es wird die mittlere Wartezeit im Sendespeicher eines Arbeitsplatzrechners (Bild 5.13) und die Blockierwahrscheinlichkeit des Sendespeichers im File-Server (Bild 5.14) für folgende drei unterschiedliche Verkehrsbeziehungen zwischen der Station 1 (File-Server) und den Stationen 2 bis 5 (Arbeitsplatzrechner) untersucht:

- a) Jede Nachrichtenübertragung eines Arbeitsplatzrechners erzeugt eine Nachrichtenübertragung des File-Servers;
- b) die Übertragungsanforderungen verteilen sich gleichmäßig auf alle Stationen (symmetrisches Verkehrsaufkommen);
- c) der File-Server erzeugt keine Übertragungsanforderungen, damit wird das Verbindungsnetzwerk durch ihn nicht belastet.

Die Verkehrsbelastung der Arbeitsplatzrechner (Stationen 2, 3, 4 und 5) ist identisch, im folgenden wird daher Station 1 mit Station 2 verglichen.

In Bild 5.13 ist der Verlauf der Wartezeit im Sendespeicher der Station 2 über der Ankunftsrate von Nachrichten, ebenfalls in der Station 2 dargestellt. Wiederum kann durch den Einsatz von zwei Bussen eine erhebliche Reduzierung der Wartezeiten erzielt werden. Darüber hinaus wirken sich unterschiedliche Verkehrsbeziehungen der Stationen weniger auf die Wartezeiten der Anforderungen im Sendespeicher aus. Für die Verkehrsbeziehung a) stellt Station 1 einen starken Engpaß des Systems dar. Dies zeigt sich vor allem durch die hohe Wahrscheinlichkeit für Blockierungen des Sendespeichers, siehe Bild 5.14.

6 ANALYSE VON MULTI-BUS-VERBINDUNGSNETZWERKEN UNTER EINSCHRÄNKUNG DER VERKEHRSLEISTUNG DURCH DEN EMPFÄNGER

Das folgende Kapitel behandelt die Einschränkung der Verkehrsleistung eines Multi-Bus-Verbindungsnetzwerkes durch verkehrsbedingte Empfangsspeicherblockierungen oder durch Bus-Interferenzen an den Empfangsports der Stationen. Die Analyse beruht auf der in Kapitel 5 vorgestellten Methode. Blockierungs- und Interferenzwahrscheinlichkeiten des Empfangsteils der Stationen stellen eine Rückwirkung auf die Charakteristik des Besuchszeitprozesses der Sende- speicher dar.

6.1 Modellierungsaspekte

Die Modellanalyse dieses Kapitels bezieht sich auf das bereits in Kapitel 3 vorgestellte Modell nach Bild 3.2, wobei nur symmetrische Systeme betrachtet werden, d.h. es wird angenommen, daß sich der Nachrichtenfluß mit der in Gleichung (3.1) angegebenen Verzweigungswahrscheinlichkeit gleich auf alle Stationen verteilt sowie alle Stationsparameter identisch sind. Der Stationsindex kann damit im folgenden entfallen.

Die Stationen besitzen jeweils einen Empfangsspeicher mit endlicher Kapazität r . Empfangene Nachrichten können im Empfangsspeicher bis zu ihrer Bedienung zwischengespeichert werden. Die Abfertigungsstrategie des Empfangsspeichers ist als FIFO angenommen. Nachrichten werden mit einer negativ-exponentiellen Bediendauer T_p bearbeitet. Das in Bild 3.2 dargestellte Modell ist zwischen Empfangsteil (Empfangswarteschlangen, Bedieneinheiten) und Sendeteil (Sendewarteschlangen, Busse) aller Stationen aufgetrennt. Beide Modellteile werden individuell untersucht, wobei in iterativer Weise die Analyse der Sendewarteschlangen durch die Ergeb-

nisse des Empfangsteils (Interferenzen, Blockierung) beeinflusst wird (siehe auch Anhang 2). Bei der Modellzerlegung gehen die Abhängigkeiten zwischen Sender und Empfänger verloren.

Im Falle der Einschränkung des Nachrichtentransfers (Interferenz, Speicherblockierung) am Empfänger sind die in Kapitel 3.3.3 beschriebenen Ablaufsteuerungen denkbar. Die Leistungsfähigkeit von Uni- bzw. Multi-Bus-Verbindungsnetzwerken mit den Ablaufsteuerungen (B) und (C) wurde in /47,48/ untersucht. Gegenstand dieses Kapitels ist die Leistungsfähigkeit eines Multi-Bus-Verbindungsnetzwerkes, welches bei Blockierung oder Interferenz am Empfänger nach der Ablaufsteuerung (A) arbeitet.

Bei Austausch einer Nachricht zwischen einer sendenden und einer empfangenden Station werden in der Analyse folgende Modellierungsaspekte berücksichtigt:

- Der Speicherplatz im Sender wird nur bei einer korrekten Übertragung der Nachricht freigegeben. Bei Blockierung oder bei Portinterferenz am Empfänger wird die Übertragung der Nachricht wiederholt, die Reihenfolge der Nachrichten einer Station bleibt dadurch erhalten.
- Zu Beginn der Nachrichtenübertragung wird festgestellt, ob eine Blockierung oder Interferenz am Empfänger vorliegt; damit ist die Verwendung von Speicherbereichen, die während eines Übertragungsvorganges im Empfangsspeicher freiwerden, nicht möglich.
- Erst nach der korrekten Übertragung beginnt die Nachricht im Empfangsspeicher zu warten bzw. wird bei freiem Prozessor sofort bedient. Damit wird eine eindeutige Trennung zwischen dem Übertragungsende der Nachricht im Sender und der Wartezeit bis zur Bearbeitung durch die Bedieneinheit im Empfänger erreicht.

6.2 Analyse des Empfängers einer Station

6.2.1 Wahrscheinlichkeitsverteilung der Systemzustände zu den Regenerationszeitpunkten

Der Empfangsteil einer Station wird im folgenden durch die diskrete Zufallsvariable X_n beschrieben. Der durch X_n definierte endliche Zustandsraum beschreibt die verschiedenen Belegungszustände des Empfangsspeichers unter Einschluß der Bedienungseinheit zu den Zeitpunkten der Nachrichtenankünfte ($t_n, n=0,1,2,\dots$). Befindet sich der Empfänger im Zustand $X_n = r+1$, dann wird eine übertragene Nachricht abgewiesen bzw. blockiert. Eine eintreffende Anforderung wird dagegen im Zustand 0, ohne zu warten, sofort im Empfänger bedient.

Die Verteilung der Systemzustände wird im folgenden mit der Methode der eingebetteten Markoff-Kette berechnet. Im Gegensatz zu der Analyse des Sendespeichers einer Station werden bei der Analyse des Empfängers beliebig verteilte Ankunftsabstände zwischen eintreffenden Nachrichten zugrundegelegt. Die Regenerationszeitpunkte der eingebetteten Markoff-Kette t_n^- entsprechen wie bei der Markoff-Kette des Sendespeichers den Zeitpunkten kurz vor dem Ende der gedächtnisbehafteten Phase, d.h. im vorliegenden Falle den Zeitpunkten der Nachrichtenankünfte im Empfänger. Die Komponenten des Wahrscheinlichkeitsvektors $\underline{p}^{(n)} = (p_0, \dots, p_k, \dots, p_{r+1})^{(n)}$ beschreiben die Wahrscheinlichkeit für k Anforderungen im Empfänger und sind in den Zeitpunkten t_n^- wie folgt definiert:

$$p_k^{(n)} = P(X_n^- = k), \quad k = 0, \dots, r+1. \quad (6.1)$$

Das Ende der Nachrichtenbearbeitung, sowie der Ankunftszeitpunkt einer Nachricht, erzeugen eine Zustandsänderung im Empfänger. Der Zustandsübergang des Empfängers bei der Ankunft einer Nachricht, also kurz nach dem Regenerationszeitpunkt (Zeitpunkt t_n^+), kann durch folgendes Gleichungssystem an-

gegeben werden (vgl. Abschnitt 5.3.1):

$$\begin{aligned} p_0^{(n)+} &= 0, \\ p_k^{(n)+} &= p_{k-1}^{(n)} \quad k = 1, 2, \dots, r, \\ p_{r+1}^{(n)+} &= p_r^{(n)} + p_{r+1}^{(n)}. \end{aligned} \quad (6.2)$$

Der Prozeßverlauf zwischen zwei Ankünften entspricht einem Sterbeprozess, wobei

$$b_i(t) = \frac{(t/t_p)^i}{i!} e^{-t/t_p}$$

der Wahrscheinlichkeit entspricht, daß in dem Zeitintervall (t_n, t_n+t) genau i Anforderungen durch den Empfänger bedient werden.

Das zeitliche Verhalten der Zustandswahrscheinlichkeiten nach den Regenerationszeitpunkten ist damit beschreibbar. Mit Hilfe der Gleichungen für die Zustandsübergänge in den Regenerationszeitpunkten (6.2) kann das Gleichungssystem für die zeitabhängigen Zustandswahrscheinlichkeiten der eingebetteten Markoff-Kette angegeben werden:

$$\begin{aligned} p_0^{(n)}(t) &= p_{r+1}^{(n)} \cdot \sum_{i=r+1}^{\infty} b_i(t) + \sum_{h=0}^r p_h^{(n)} \cdot \sum_{m=h+1}^{\infty} b_m(t) \\ p_k^{(n)}(t) &= p_{r+1}^{(n)} \cdot b_{r+1-k}(t) + \sum_{h=k-1}^r p_h^{(n)} \cdot b_{k-h+1}(t), \end{aligned} \quad (6.3)$$

$$k = 1, 2, \dots, r+1.$$

Entspricht die Zeitspanne t dem Ankunftsabstand der Nachrichten (T_A) mit der Verteilungsdichtefunktion $f_A(t)$, so ist es möglich, die Zeitabhängigkeit der Zustandswahrscheinlichkeiten zu eliminieren und das Übergangsverhalten zwischen den benachbarten Regenerationszeitpunkten (t_n) und (t_{n+1}) der

eingebetteten Markoff-Kette durch Bildung der totalen Wahrscheinlichkeit mit Hilfe der Verteilungsdichtefunktion des Ankunftsabstandes zu bestimmen. Damit entsteht mit

$$p_k^{(n+1)} = \int_{t=0}^{\infty} p_k^{(n)}(t) f_A(t) dt, \quad (6.4)$$

ein Gleichungssystem, welches die Berechnung der Zustandswahrscheinlichkeiten des Empfängers zu den eingebetteten Zeitpunkten ermöglicht:

$$p_0^{(n+1)} = p_{r+1}^{(n)} \cdot \sum_{i=r+1}^{\infty} q_i + \sum_{h=0}^r p_h^{(n)} \cdot \sum_{m=h+1}^{\infty} q_m, \quad (6.5)$$

$$p_k^{(n+1)} = p_{r+1}^{(n)} \cdot q_{r+1-k} + \sum_{h=k-1}^r p_h^{(n)} \cdot q_{k-h+1},$$

wobei $k = 1, 2, \dots, r+1$ und

$$q_i = \int_{t=0}^{\infty} b_i(t) \cdot f_A(t) dt, \quad i = 0, 1, 2, \dots, \quad (6.6)$$

der Wahrscheinlichkeit entspricht, daß zwischen zwei Regenerationszeitpunkten i Bedienungsdigungen im Empfänger auftreten.

Entsprechend den Ausführungen in Abschnitt 5.3.2 ist es möglich, auch hier die Zustandswahrscheinlichkeiten des Empfängers zu beliebigen Zeitpunkten zu berechnen.

6.2.2 Nachrichtenankunftsprozeß am Empfänger

Zur Berechnung der Zustandsverteilung nach Kapitel 6.2.1 werden beliebig verteilte Ankunftsabstände der Nachrichten vorausgesetzt. Im folgenden soll das erste und zweite Moment der Verteilungsdichtefunktion des Ankunftsabstandes T_A zweier aufeinanderfolgender Nachrichten am Empfänger einer

Station berechnet werden. Dabei werden auftretende Interferenzen oder Blockierungen der Übertragungswünsche an dem betrachteten Empfänger noch nicht beachtet.

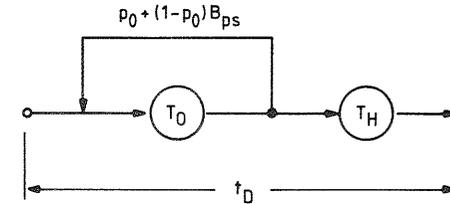


Bild 6.1: Phasenmodell des Ausgangsprozesses einer Bedieneinheit

In einem ersten Schritt zur Berechnung der Ankunftsverteilungsdichtefunktion wird die LST (Laplace-Stieltjes-Transformierte) des Ausgangsprozesses einer Bedieneinheit (T_D) mit Hilfe des in Bild 6.1 dargestellten Phasenmodells berechnet:

$$\Phi_D(s) = \frac{(1-p_0)(1-B_{ps})\Phi_O(s)\Phi_H(s)}{1 - (p_0 + (1-p_0)B_{ps})\Phi_O(s)}. \quad (6.7)$$

Unter Anwendung der Gleichung (4.8) ist es möglich, durch Differentiation der Gleichung (6.7) das erste und das zweite Moment des Ausgangsprozesses einer Bedieneinheit zu bestimmen. Die Interferenzen am Sendeport der Stationen sind in Gleichung (6.7) bereits berücksichtigt. Zur Unterscheidung von Nachrichtenblockierungen durch Sendeport- bzw. Empfangsportinterferenzen wird im folgenden die Wahrscheinlichkeit für eine Sendeportinterferenz mit B_{ps} bzw. die Wahrscheinlichkeit für eine Empfangsportinterferenz mit B_{pe} bezeichnet.

In einem weiteren Schritt wird der durch Gleichung (6.7) bestimmte Ausgangsprozeß einer Bedieneinheit (T_D) in g stationsindividuelle Eingangsprozesse (T_I) für die Empfänger

aller Stationen, entsprechend Gleichung (4.33) zerlegt. Danach erfolgt mit dem in Kapitel 4.3.6.2 beschriebenen Kompositionsverfahren eine Überlagerung der Eingangsprozesse (T_I) aller Bedieneinheiten an den Empfangsports der Empfänger. Der dadurch entstehende Nachrichtenankunftsprozeß an einem Empfänger einer Station wird durch die Zufallsvariable (T_A) beschrieben. Die Verteilungsdichtefunktion von T_A ist in den ersten beiden Momenten bekannt.

6.3 Wahrscheinlichkeit für Blockierung oder Interferenz

6.3.1 Empfangsspeicherblockierung

Begründet durch die eingangs beschriebene Speicherverwaltung des Empfängers einer Station, muß für eine blockierungsfreie Nachrichtenübertragung mindestens ein Speicherplatz im Empfänger zu Beginn des Übertragungsvorganges vorhanden sein. Die Empfangsspeicherblockierung muß deshalb zum Zeitpunkt des Übertragungsbeginns ermittelt werden.

Gemäß Ablaufsteuerung (A) wird die Nachricht unabhängig von Portinterferenz oder Speicherblockierung am Empfänger übertragen. Sie belegt damit das Übertragungsmedium auch bei einer Blockierung oder Interferenz am Empfänger.

Die Zufallsvariable T_A läßt sich als Summe von Bediendauer T_H und der Zeitspanne vom letzten Übertragungsende bis zum nächsten Übertragungsbeginn (T_A^*) interpretieren, d.h.

$$T_A = T_A^* + T_H \quad (6.8)$$

Da T_A und T_H in den ersten beiden Momenten bekannt sind, kann unter der Annahme der Unabhängigkeit von T_H und T_A^* , der Mittelwert und Variationskoeffizient von T_A^* in folgender Weise bestimmt werden:

$$t_A^* = t_A - t_H$$

$$c_A^{*2} = \frac{1}{t_A^{*2}} (c_A^2 t_A^2 - c_H^2 t_H^2) \quad (6.9)$$

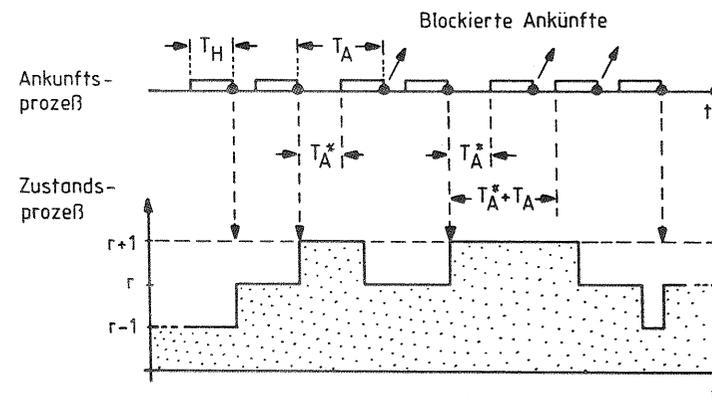


Bild 6.2: Nachrichtenblockierungen am Empfangsspeicher

Wenn sich der Empfänger kurz vor der letzten Nachrichtenankunft im Zustand r befand, wird die nächste eintreffende Nachricht blockiert, falls während der Zeitdauer T_A^* kein Bedienungsende im Empfänger auftritt und die momentane Belegung des Empfängers im darauffolgenden Ankunftsabstand endet, siehe Bild 6.2.

Die Wahrscheinlichkeit, daß ab dem Betrachtungszeitpunkt (Regenerationszeitpunkt in welchem der Zustandsprozeß in den Zustand r+1 übergeht) die Bedienung im Empfänger mindestens für die Dauer T_A^* besteht, entspricht der Wahrscheinlichkeit für die positive Differenz der Zufallsvariablen T_p und T_A^* :

$$P(T_P - T_A^* \geq 0) = \int_{t=0}^{\infty} e^{-t/t_P} \cdot f_A^*(t) dt. \quad (6.10)$$

Da die Bedienzeit im Empfänger negativ-exponentiell verteilt ist, läßt sich die Gleichung (6.10) auch als die Laplace-Transformierte der Verteilungsdichtefunktion von $f_A^*(t)$ an der Stelle $1/t_P$ interpretieren, d.h.

$$\Phi_A^*(1/t_P) = P(T_P - T_A^* \geq 0). \quad (6.11)$$

Wie in Bild 6.2 dargestellt, ist im Zustand r die Blockierung weiterer eintreffender Anforderungen möglich. Es werden genau x Anforderungen blockiert, falls in der Zeit

$$\underbrace{T_A^* + T_A^* + T_A^* + \dots + T_A^*}_{(x-1)\text{-mal}}$$

kein Bedienungsende auftritt und bis zur unmittelbar folgenden, $(x+1)$ -ten Nachrichtenanforderung die Bedienphase des Empfängers endigt. Die Wahrscheinlichkeit für die Blockierung von genau x Nachrichtenübertragungen am Empfänger errechnet sich dann zu:

$$P(T_P - T_A^* - T_A^* - \dots - T_A^* \geq 0) = \int_{t=0}^{\infty} e^{-t/t_P} \cdot f_A^*(t) \otimes f_A^*(t) \otimes \dots \otimes f_A^*(t) \cdot dt \quad (6.12)$$

Mit den Gleichungen (6.10 - 6.12) ist es möglich, die mittlere Anzahl blockierter Nachrichtenübertragungen N_b im Zustand r zu bestimmen:

$$N_b = \sum_{x=1}^{\infty} x \cdot \Phi_A^*(s) \cdot (\Phi_A(s))^{x-1} (1 - \Phi_A(s)), \quad s=1/t_P. \quad (6.13)$$

Die Blockierwahrscheinlichkeit des Empfängers entspricht dem Produkt von der Zustandswahrscheinlichkeit $p_r^{(n)}$ und der

mittleren Anzahl blockierter Anforderungen N_b im Zustand r , sie errechnet sich wie folgt:

$$B_m = \frac{\Phi_A^*(s) \cdot p_r^{(n)}}{1 - \Phi_A(s)}. \quad (6.14)$$

6.3.2 Empfangsportinterferenz

Durch die Koinzidenz von Nachrichtenübertragungen unterschiedlicher Bedieneinheiten am Empfangsport einer Station sind Interferenzen einzelner Übertragungsanforderungen möglich. Im Falle der Interferenz wird vorausgesetzt, daß die interferierende Übertragung einen momentan ablaufenden Übertragungsvorgang nicht stört.

Die Empfangsportinterferenz B_{pe} entspricht somit der Wahrscheinlichkeit, daß eine Anforderung zu Übertragungsbeginn den Empfangsport der Zielstation durch eine der anderen Bedieneinheiten belegt vorfindet, wobei die momentan ablaufende Nachrichtenübertragung erfolgreich ist, also zu Übertragungsbeginn keiner Portinterferenz oder Speicherblockierung unterlag. Erstes und zweites Moment des Ankunftsprozesses von erfolgreichen Übertragungen erhält man durch eine Zerlegung des Ankunftsprozesses mit der Speicherblockier- und Interferenzwahrscheinlichkeit des Empfängers.

Entsprechend den Ausführungen in Kapitel 5.4.4 zur Berechnung der Sendeportinterferenz kann die Empfangsportinterferenzwahrscheinlichkeit ebenfalls näherungsweise berechnet werden. Durch die Wahrscheinlichkeit für die positive Differenz von Vorwärtsrekurrenzzeit der Übertragungszeit und des Ankunftsabstandes erfolgreicher Nachrichtenübertragungen wird wie in Kapitel 5.4.4 nur eine wiederholte Empfangsportinterferenz berücksichtigt.

Die Wahrscheinlichkeit einer Empfangsportinterferenz (B_{pe})

kann wie folgt näherungsweise angegeben werden:

$$B_{pe} \cong \frac{(n-1)}{n} \frac{t_H}{t_A} (1 + P(T_H^V - T_A \geq 0)). \quad (6.15)$$

6.4 Der Besuchszeitprozeß bei Portinterferenzen und Speicherblockierung

Blockierungen an den Empfängern sowie Interferenzen an den Sende- und Empfangsports der Stationen des Multi-Bus-Verbindungsnetzwerkes beeinflussen die Leistungsfähigkeit des Gesamtsystems. Speicherblockierung und Portinterferenz wirkt sich insbesondere im Sendeteil der Stationen durch einen Anstieg der mittleren Wartezeit und damit durch eine höhere Blockierwahrscheinlichkeit der Sendewarteschlange aus.

Die Rückwirkung von Blockierungen und Interferenzen an den Empfängern auf die Sender wird in der Verkehrsanalyse durch eine Zerlegung des Zuteilungsprozesses (T_G) erreicht. Gemäß den Ausführungen in Kapitel 4.3.6 wird der Zuteilungsprozeß mit den Wahrscheinlichkeiten für Sende- und Empfangsportinterferenz und Empfangsspeicherblockierung in einen Komponentenprozeß zerlegt, der die interferenz- und blockierungsfreien Besuchszeitpunkte aller Bedieneinheiten an einer bestimmten Station beschreibt. Der resultierende Prozeß bewirkt dann Zustandsänderungen (erfolgreiche Nachrichtenübertragungen) im Sendespeicher einer Station.

Die Zerlegung erfolgt unter der Berücksichtigung, daß eine Interferenz oder Blockierung am Empfänger nur bei Buszuteilungen erfolgt, die keiner Sendeportinterferenz unterlagen. Mittelwert und Variationskoeffizient der Besuchszeitverteilungsfunktion für eine beliebige Station errechnen sich nach Gleichung (4.33) wie folgt:

$$t_V = \frac{t_G}{(1-B_{ps})(1-(B_m+B_{pe}))} \quad (6.16)$$

$$c_V^2 = (1-B_{ps})(1-(B_m+B_{pe}))c_G^2 + B_{ps} + (1-B_{ps})(B_m+B_{pe}).$$

6.5 Ergebnisse

6.5.1 Vergleich von Uni- und Doppel-Bus-System

Im folgenden werden die in Kapitel 5.5.3 behandelten Multi-Bus-Verbindungsnetzwerke betrachtet. Dabei soll die Leistungsfähigkeit der untersuchten Systeme, speziell unter dem Modellierungsaspekt der Speicherblockierung und Portinterferenz von Nachrichtenübertragungen am Empfänger, untersucht und miteinander verglichen werden.

Den in Bild 6.3 und 6.4 vorgestellten Ergebnissen liegt gemeinsam ein System mit fünf Stationen zugrunde, die mit einem Uni-Bus- bzw. mit einem Doppel-Bus-Verbindungsnetzwerk gekoppelt worden sind. Aus Übersichtlichkeitsgründen werden nur negativ-exponentiell verteilte Busbelegungszeiten betrachtet. Hyperexponentielle Busbelegungszeiten weisen ähnliche Effekte auf wie in den in Kapitel 5 gezeigten Ergebnisgraphen. Die mittlere Verwaltungsdauer einer Buszuteilung wird wie in Kapitel 5 mit der halben mittleren Busbelegungs-dauer angenommen. Alle zeitlichen charakteristischen Größen wurden auf die mittlere Busbelegungs-dauer (t_H) normiert. Die charakteristischen Größen sind über der Ankunftsrate der Nachrichten am Sendespeicher einer Station aufgetragen. Die mittlere Bearbeitungsdauer der Nachrichten im Empfänger einer Station, normiert auf die mittlere Busbelegungs-dauer, wurde dabei so gewählt, daß sie dem Kehrwert der maximalen Ankunftsrate einer Station entspricht ($t_p=5$).

Der Verlauf der Wartezeit im Sendespeicher ohne Einfluß von

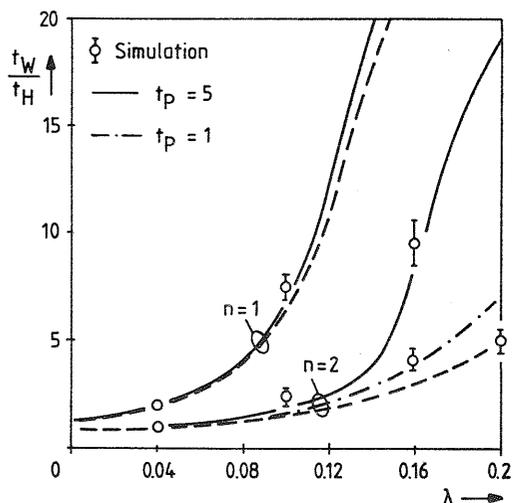


Bild 6.3: Einfluß von Interferenzen B_{pe} und Blockierungen B_m an den Empfängern auf die B_{pe} Wartezeiten in den Sendern t_w

Interferenzen oder Blockierungen am Empfänger ist durch die gestrichelten Graphen dargestellt, die den Ergebnissen nach Bild 5.11 entsprechen.

Vergleicht man die mittlere Wartezeit einer Nachricht im Sendespeicher der Stationen eines Uni-Bus-Systems mit einem Doppel-Bus-System (Bild 6.3), ist zunächst zu erkennen, daß das Uni-Bus-System einen Systemengpaß darstellt; auftretende Blockierungen und Interferenzen am Empfänger bewirken nur einen geringen Einfluß auf die Wartezeit im Sendespeicher der Station (durchgezogener Graph), da die Belastung des Empfängers durch den Systemengpaß begrenzt wird.

Der Systemengpaß wird durch den Einsatz eines Doppel-Bus-Systems beseitigt. Dadurch nimmt die durch das Verbindungsnetzwerk akzeptierte Last und damit auch die Verkehrsbelastung der Empfänger mit zunehmender Ankunftsrate ebenfalls zu. Aufgrund der untersuchten Ablaufsteuerung (A) entsteht

im Verbindungsnetzwerk durch speicherblockierte bzw. durch interferierende Übertragungen eine zusätzliche "innere" Last, welche sich durch einen starken Anstieg der mittleren Wartezeiten mit zunehmender Ankunftsrate immer mehr bemerkbar macht (Bild 6.3). Die Speicherblockierungen des Empfängers spielen dabei eine dominierende Rolle (Bild 6.4). Der strichpunktierter Graph in Bild 6.3 zeigt den Verlauf der Wartezeit, falls die Bearbeitungszeiten der Nachrichten im Empfänger so kurz sind ($t_p=1$), daß Speicherblockierungen vernachlässigbar sind und nur noch Portblockierungen im Verbindungsnetzwerk auftreten.

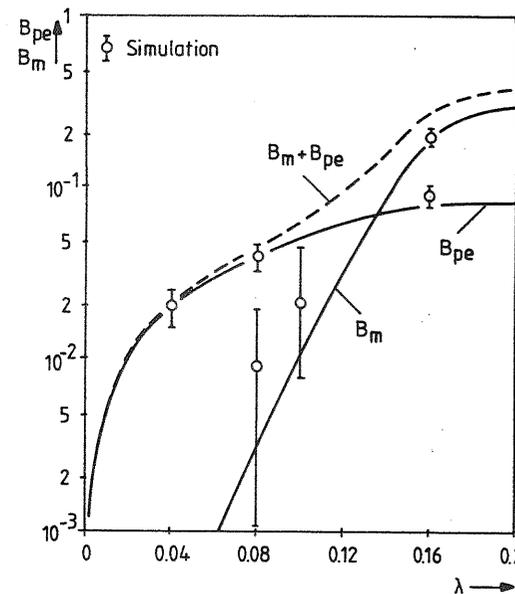


Bild 6.4: Interferenz- und Blockierwahrscheinlichkeit an den Empfängern

In Bild 6.4 ist der Verlauf der Empfangsportinterferenz (B_{pe}) und der Empfangsspeicherblockierung (B_m) für das Doppel-Bus-System dargestellt. Die Wahrscheinlichkeit einer Portinterferenz am Empfänger steigt proportional mit der Ankunfts-

rate der Nachrichten und kann deshalb bereits bei niedrigen Ankunftsrate festgestellt werden. Die Wahrscheinlichkeit einer Empfangsportinterferenz nimmt mit zunehmender Stationszahl ab. Wie bereits erwähnt, werden blockierte oder interferierende Übertragungen wiederholt. Das Verkehrsangebot an den Empfängern erhöht sich, so daß die Verkehrsbelastung der Empfänger dem durch das Verbindungsnetzwerk akzeptierten Nachrichtenverkehr entspricht. Die Empfangsspeicherblockierung steigt durch wiederholte Nachrichtenübertragungen mit zunehmender Ankunftsrate stark an (Bild 6.4). Der gestrichelte Graph in Bild 6.4 stellt die Summe von Portinterferenz und Speicherblockierung dar.

6.5.2 Leistungsverhalten des Paketvermittlungsknotens EWSP

6.5.2.1 Knotenarchitektur

Die zunehmende Akzeptanz paketvermittelnder Dienste in privaten und öffentlichen Nachrichtennetzen verlangt eine neue Generation von leistungsfähigen Knotenarchitekturen. Die Firma Siemens entwickelte, basierend auf dem Knoten EDX-P, den modular aufgebauten Paketvermittlungsknoten EWSP /64,65/. In Bild 6.5 ist die prinzipielle Architektur des Knotens EWSP dargestellt. Der Datenverkehr mehrerer Leitungen wird in einer Leitungseinheit zusammengefaßt und über ein Verbindungsnetzwerk, bestehend aus einem gedoppelten 16 bit breiten Daten-Bus bzw. -Ring, autonomen Vermittlungseinheiten zugeführt.

Das aus Sicherheitsgründen gedoppelte Verbindungsnetzwerk kann auch in Lastteilung arbeiten. Die Taktfrequenz einer Ring-Einheit beträgt 5MHz, so daß eine maximale Datenrate von 160 Mbit/s bei Parallelarbeit beider Ringe erzielt werden kann. Die Zugriffsregelung der Einheiten auf das Verbindungsnetzwerk erfolgt für ein Ring-Netzwerk individuell und nach

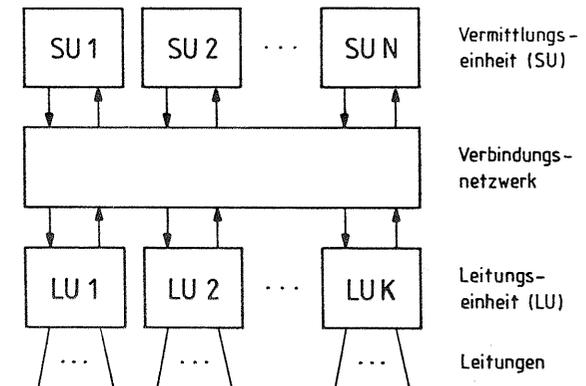


Bild 6.5: Architektur des Paketvermittlungsknotens EWSP

dem in Kapitel 2.3.2.2 beschriebenen zyklischen Zuteilungsverfahren.

Die maximal mögliche Ausbaustufe des Systems entspricht 255 Einheiten. Die Leitungs- und Vermittlungseinheiten sind, wie in Bild 1.2 dargestellt, durch endliche Sende- und Empfangsspeicher mit dem Ring-Netzwerk verbunden.

Die folgende analytische Studie untersucht das Leistungsverhalten des zentralen Teils der vorgestellten Architektur für ein System im Maximalausbau. Unter der Voraussetzung, daß in den Verarbeitungseinheiten kein Engpaß auftritt, werden die mittleren Zugriffszeiten der Einheiten auf das Verbindungsnetzwerk über der Ankunftsrate der Datenpakete angegeben.

Zur Vermittlung wird ein Datenpaket zweimal über das Kommunikationssystem transferiert (Eintransfer, Austransfer). Für die Verwaltung von Speicherbereichen in den Verarbeitungseinheiten ist pro Datenpakettransfer noch die Übertragung eines zusätzlichen Paketes mit Steuerinformation über das Verbindungsnetzwerk notwendig /66/.

Die Modellbildung des Kommunikationssystems (Verbindungsnetzwerk, Sende- und Empfangsspeicher) führt prinzipiell

auf das in Bild 3.2 dargestellte Warteschlangenmodell. Die Arbeitsweise des Ringprotokolls entspricht der in Kapitel 3.3.3 diskutierten und in diesem Kapitel behandelten Ablaufsteuerung (A).

6.5.2.2 Leistung des Verbindungsnetzwerkes

Die Leistungsuntersuchung wurde für ein symmetrisch belastetes System durchgeführt. Es sind 125 Vermittlungseinheiten und 125 Leitungseinheiten durch ein Verbindungsnetzwerk, bestehend aus einem Ring oder zwei unter Lastteilung arbeitenden Ringen, verbunden. Die mittlere Übertragungszeit eines Datenpaketes (t_H) errechnet sich in folgender Weise:

$$t_H = \frac{1}{2} \cdot (l_p + l_s) \cdot 200ns, \tag{6.17}$$

dabei entspricht l_p und l_s der Anzahl von 16 bit Worten eines Daten- bzw. Steuerpaketes. Die mittlere Verwaltungsdauer t_0 entspricht der Stationsverzögerung, d.h. dem Ringtakt (200ns). Sende- und Empfangsspeicher einer Einheit sind mit einer Kapazität von jeweils fünf Warteplätzen angenommen. Durch die Annahme von fünf Warteplätzen im Sender wird die zur Paketübertragung parallele Aufbereitung weiterer Pakete im Sender quantitativ mitberücksichtigt.

Bild 6.6 zeigt den Verlauf der mittleren Wartezeit eines Datenpaketes im Sendespeicher einer betrachteten Einheit über der Ankunftsrate von Datenpaketen. Die mittlere Länge der Datenpakete beträgt 80 byte (40 Worte), die der Steuerpakete 20 byte (10 Worte). Die mittleren Wartezeiten sind für ein Single-Ring-Netzwerk ($n=1$) und ein Doppel-Ring-Netzwerk ($n=2$) für unterschiedliche Variationskoeffizienten der Übertragungszeit dargestellt.

Für das Verbindungsnetzwerk mit einem Ring sind in einer Einheit bei der Ankunftsrate von 350P/s bereits Blockierungen in den Sendern festzustellen. Das Verbindungsnetzwerk des Kno-

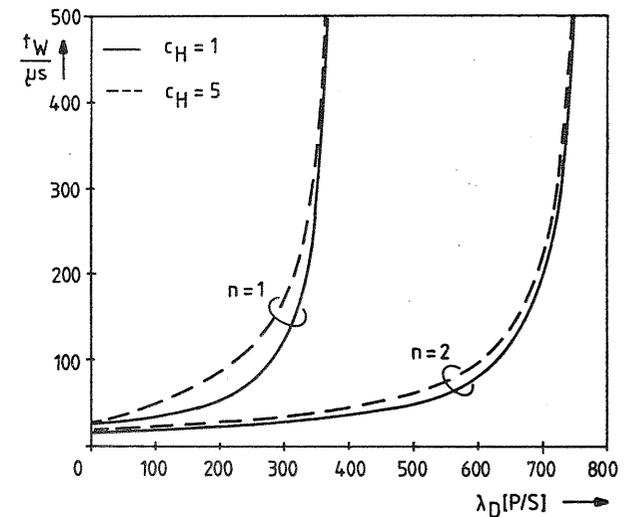


Bild 6.6: Einfluß des Variationskoeffizienten der Paketlängenverteilung c_H auf die Wartezeiten in den sendenden Einheiten t_W

tens ist dann in gesättigtem Zustand, die Wartezeit bis zum Ringzugriff steigt stark an. Das instabile Verhalten der Zugriffzeiten auf das Verbindungsnetzwerk begrenzt gleichzeitig den Durchsatz des Paketvermittlungsknotens. Eine Verdopplung des Durchsatzes ist, bei der an dieser Stelle diskutierten Parameterkonstellation, durch den Einsatz eines Doppel-Ring-Netzwerkes möglich (siehe Bild 6.6). Durch die große Anzahl der Stationen wirken sich hohe Varianzen der Paketübertragungszeit auf die Wartezeit im Sender kaum aus. Variiert man die mittlere Anzahl von Worten eines Datenpaketes (mittlere Datenpaketlänge), so verändert sich die Belastung des Verbindungsnetzwerkes und damit der maximal mögliche Durchsatz in Datenpaketen/Sekunde einer Einheit, siehe Bild 6.7. Für große mittlere Paketlängen erreicht das Verbindungsnetzwerk früher den gesättigten Zustand. Die endliche Kapazität des Empfangsspeichers ($r=5$) und die

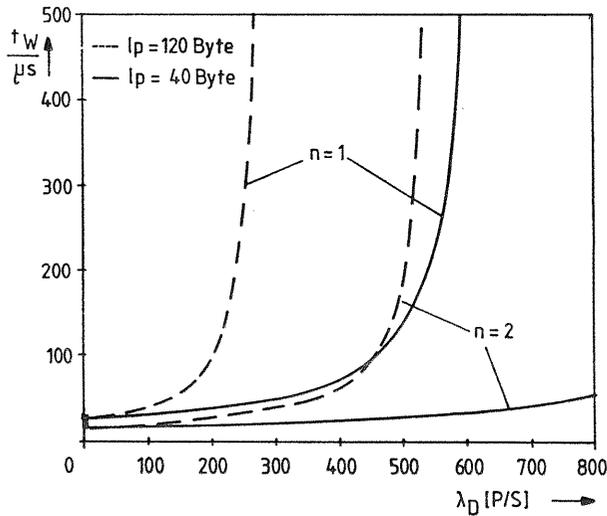


Bild 6.7: Einfluß der mittleren Paketlänge l_p auf die Wartezeiten in den sendenden Einheiten t_w

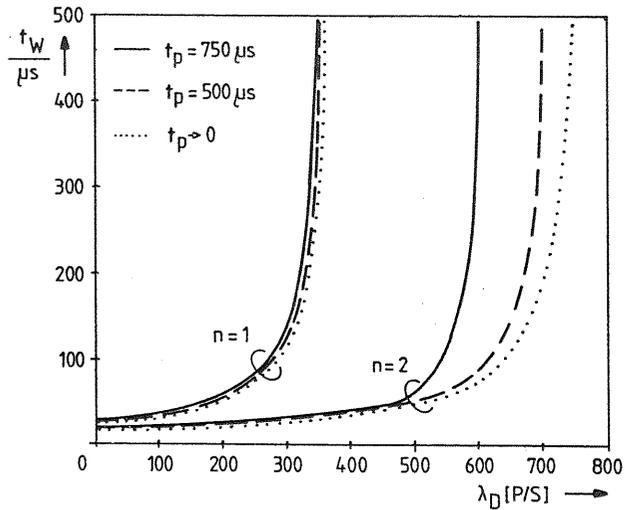


Bild 6.8: Einfluß der mittleren Bearbeitungsdauer t_p in den Empfängern auf die Wartezeiten t_w in den P_S Sendern

mittlere Bearbeitungsdauer eines Paketes im Empfänger (t_p) beeinflussen ebenfalls die Leistungsfähigkeit des Verbindungsnetzwerkes (Bild 6.8). Wie in Bild 6.3 zeigt sich, daß ein Single-Ring-Verbindungsnetzwerk bereits den Systemengpaß bildet und die durchgesetzte Paketrage die Empfänger nicht stark belastet. Blockierungen der Empfängerspeicher wirken sich deshalb kaum auf die Leistungsfähigkeit des Gesamtsystems aus.

Durch die Verdopplung der Ringkapazität (Doppel-Ring-Verbindungsnetzwerk) wird der Engpaß an Übertragungskapazität beseitigt. Da mit zunehmender Paketankunftsrate auch die Ankunftsrate und die Belastung des Empfängers steigt, kann durch die endliche Empfängerspeicherkapazität der Empfänger zum Systemengpaß werden. Bild 6.8 zeigt den Einfluß der mittleren Bearbeitungsdauer eines Paketes im Empfänger der Ziel-einheit auf die Wartezeit im Sendespeicher der Ursprungseinheit.

Aus Komplexitätsgründen wurde auf die Validierung der in diesem Abschnitt vorgestellten Ergebnisse mittels der zeit-treuen Simulation verzichtet. Erste Ansätze zur Simulation solcher komplexer Modelle wurden anhand des Beispiels EWSP in /66/ durchgeführt.

7 ZUSAMMENFASSUNG

Die vorliegende Arbeit untersuchte die Leistungsfähigkeit von Multi-Bus-Verbindungsnetzwerken, in welchen jeder Bus individuell in zyklischer Reihenfolge allen, mit Hilfe von Kommunikationsspeichern über das Verbindungsnetzwerk kommunizierenden, Einheiten zugeteilt wird.

Zur Leistungsbeurteilung wurden eine Modellierung des Kommunikationssystems (Multi-Bus-Verbindungsnetzwerk und Kommunikationsspeicher) durchgeführt und anschließend einige Anwendungsbeispiele des entwickelten Modells diskutiert.

Die Leistungsuntersuchung wurde mittels der wahrscheinlichkeitstheoretischen Modellanalyse durchgeführt, wobei Methoden wie eingebettete Markoff-Ketten, Zerlegung und Überlagerung von Punktprozessen sowie die Momentenapproximation von Verteilungsfunktionen angewendet wurden.

In der Verkehrsanalyse wurden erstes und zweites Moment der Busbelegungs- und Busverwaltungszeiten, die endliche Kapazität von Sende- und Empfangsspeicher der Einheiten sowie Konfliktsituationen, wie sie typischerweise beim Zugriff auf Ports und Empfangsspeicher der Einheiten auftreten, berücksichtigt und ihr Einfluß auf die Leistungsfähigkeit des Gesamtsystems untersucht.

Der Einsatz von Multi-Bus-Netzwerken zur Leistungssteigerung des Gesamtsystems, die Dimensionierung der Kommunikationsspeicher sowie der Einfluß von Speicherkapazität und Verarbeitungsgeschwindigkeit im Empfänger auf Durchsatz und Transferzeit von Nachrichten wurden an Hand einiger konkreter Fallstudien diskutiert und vorgestellt.

Die in dieser Arbeit untersuchte Problematik leistet einen Beitrag zur Leistungsuntersuchung der Architektur modularer Paketvermittlungsknoten, Token-gesteuerter Lokaler Netzwerke sowie Steuerungsstrukturen in Fernsprech-Vermittlungssystemen.

8 LITERATURVERZEICHNIS

- /1/ W.K. Giloi, "Rechnerarchitektur", Springer-Verlag, Berlin/Heidelberg/New York, 1981.
- /2/ L. Kleinrock, "Distributed Systems", Computer, Vol. 18, No. 11, (1985), pp. 90 - 103.
- /3/ P.P. Spies, "No-wait-send / Rendezvous", Informatik-Spektrum, Band 8, Heft 5, (1985), Seite 283-286.
- /4/ ISO 7498/CCITT Recommendation X.200: Reference Model of Open Systems Connection.
- /5/ IEEE Standard 802.3: CSMA/CD Access Method and Physical Layer Specifications.
- /6/ IEEE Standard 802.4: Token-Passing Bus Access Method and Physical Layer Specifications.
- /7/ W. Kessel, M. Klages, F. Hofmann, "Multi-Mikrocomputerkonzept mit Mailbox-Schnittstellen", Elektronik, Heft 21, (1984), Seite 229-232.
- /8/ M.A. Marsan, G. Balbo, G. Conte, "Comparative Performance Analysis of Single Bus Multiprocessor Architectures", IEEE Transactions on Computers, Vol. C-31, No. 12, (1982), pp. 1179-1191.

- /9/ M.A. Marsan, M. Gerla, "Markov Models for Multiple Bus Multiprocessor Systems", IEEE Transactions on Computers, Vol. C-31, No. 3, (1982), pp. 239-248.
- /10/ M.A. Marsan, G. Balbo, G. Conte, F. Gregoretti, "Modelling Bus Contention and Memory Interference in a Multiprocessor System", IEEE Transactions on Computers, Vol. C-32, No. 1, (1983), pp. 60-72.
- /11/ J. Kriz, "A Queueing Analysis of a Symmetric Multiprocessor with Shared Memories and Busses", IEE Proc., Vol. 130, Pt. E., No. 3, (1983), pp. 83-89.
- /12/ T.N. Mudge, H.B. Al-Sadoun, "A Semi-Markov Model for the Performance of Multiple-Bus Systems", IEEE Transactions on Computers, Vol. C-34, No. 10, (1985), pp. 934-942.
- /13/ C.R. Das, L.N. Bhuyan, "Bandwidth Availability of Multiple-Bus Multiprocessors", IEEE Transactions on Computers, Vol. C-34, No. 10, (1985), pp. 918-926.
- /14/ S. Hentschke, R. Meierl, P. Wildenauer, "Leistungsverhalten eines über einen gemeinsamen Bus kommunizierenden Mehrmikrorechner-Systems", Elektronische Rechenanlagen, 24. Jahrgang, Heft 6, (1982), Seite 255-262.
- /15/ I. Chlamtac, O. Ganz, "A Study of Communication Interfaces for Multiple Bus Networks", Computer Networks and ISDN Systems, Vol. 9, (1985), pp. 177-189.

- /16/ O. Ganz, I. Chlamtac, "Analysis of Multiple-Bus Synchronous and Asynchronous Communication Systems", Performance Evaluation, Vol. 5, No. 3, (1985), pp. 1-13.
- /17/ L.N. Bhuyan, "On the Performance of Loosely Coupled Multiprocessors", 11th Annual International Symposium on Computer Architecture, SIGARCH Newsletter, Vol. 12, Issue 3, (1984), pp. 256-262.
- /18/ M. A. Leibowitz, "An Approximate Method for Treating a Class of Multiqueue Problems", IBM Journal of Research and Development, Vol. 5, No. 3, (1961), pp. 204-209.
- /19/ M. A. Leibowitz, "A Note on Some Fundamental Parameters of Multiqueue Systems", IBM Journal of Research and Development, Vol. 6, No. 4, (1962), pp. 470-471.
- /20/ O. Hashida, "Analysis of Multiqueue", Review of the Electrical Communication Laboratories, Vol. 20, Nos. 3-4, (1972), pp. 189-199.
- /21/ O. Hashida, K. Ohara, "Line Accomodation Capacity of a Communication Control Unit", Review of the Electrical Communication Laboratories, Vol. 20, Nos. 3-4, (1972), pp. 231-239.
- /22/ W.W. Chu, A.G. Konheim, "On the Analysis and Modelling of a Class of Computer Communication Systems", IEEE Transactions on Communications, Vol. COM-20, No. 3, (1972), pp. 645-654.

- /23/ A.G. Konheim, B. Meister, "Waiting Lines and Times in a System with Polling", Journal of the Association for Computing Machinery, Vol. 21, No. 3, (1974), pp. 470-490.
- /24/ A.G. Konheim, "Chaining in a Loop System", IEEE Transactions on Communications, Vol. COM-24, No. 2, (1976), pp. 203-210.
- /25/ R.B. Cooper and G. Murray, "Queues served in Cyclic Order", Bell Systems Technical Journal, Vol. 48, No. 3, (1969), pp. 675-689.
- /26/ R.B. Cooper, "Queues served in Cyclic Order: Waiting Times", Bell Systems Technical Journal, Vol. 49, No. 3, (1970), pp. 399-413.
- /27/ M. Eisenberg, "Two Queues with Changeover Times", Operations Research, Vol. 19, No. 2, (1971), pp. 386-401.
- /28/ M. Eisenberg, "Queues with Periodic Service and Changeover Times", Operations Research, Vol. 20, No. 2, (1972), pp. 440-451.
- /29/ F.J. Kuehn, "Multiqueue Systems with Nonexhaustive Cyclic Service", Bell Systems Technical Journal, Vol. 58, No. 3, (1979), pp. 671-699.
- /30/ P.J. Kuehn, "Performance of ARQ-Protocols for HDX-Transmission in Hierarchical Polling Systems", Performance Evaluation, Vol. 1, No. 1, (1981), pp. 19-30.

- /31/ G.E. Mapp, D.R. Manfield, "Performance Analysis of Priority Polling Systems with Complex Cycles", Proc. of the 8th International Conference on Computer Communication, (ICCC), (1986), Munich, paper C5-2.
- /32/ D.R. Manfield, "Analysis of a Polling System for Two-Way Traffic", IEEE Transactions on Communications, Vol. COM-33, No. 9, (1985), pp. 1001-1006.
- /33/ D.R. Manfield, "Analysis of a Polling System with Priorities", IEEE Global Telecommunications Conference, San Diego, (1983), pp. 43.4.1-43.4.5.
- /34/ R.T. Carsten, E.E. Newhall, M.J.M. Posner, "A Simplified Analysis of Scan Times in an Asymmetrical Newhall Loop with Exhaustive Service", IEEE Transactions on Communications, Vol. COM-25, No. 9, (1977), pp. 951-957.
- /35/ W. Bux, H.L. Truong, "Mean-Delay Approximation for Cyclic-Service Queueing Systems", Performance Evaluation, Vol. 3, No. 3, (1981), pp. 187-196.
- /36/ W. Bux, "Local-Area Subnetworks: A Performance Comparison", IEEE Transactions on Communications, Vol. COM-29, No. 10, (1983), pp. 187-196.
- /37/ W. Bux, "Performance Issues in Local-Area Networks", IBM Systems Journal, Vol. 23, No. 4, (1984), pp. 351-374.

- /38/ M.J. Ferguson, Y. Aminetzah, "Exact Results for Non-Symmetric Token Ring Systems", IEEE Transactions on Communications Vol. COM-33, No. 3, (1985), pp. 223-231.
- /39/ M.J. Ferguson, "Mean Waiting Time for a Token Ring with Nodal Dependent Overheads", Proceedings of the 11th International Teletraffic Congress, Kyoto, Japan, (1985), pp. 4.2A-1.1 - 4.2A-1.7.
- /40/ P. Tran-Gia, T. Raith, "Multiqueue Systems with Finite Capacity and Nonexhaustive Cyclic Service", Proc. of the International Seminar on Computer Networking and Performance Evaluation, Tokyo, Japan, (1985), pp. 5.3.1 - 5.3.13
- /41/ P. Tran-Gia, T. Raith, "Approximation for Finite Capacity Multiqueue Systems", Tagungsband der GI/NTG-Fachtagung "Messung, Modellierung und Bewertung von Rechen-systemen", Dortmund, (1985), Seite 332-345.
- /42/ S.C. Agrawal, J.B. Buzen, A.K. Thareja, "A Unified Approach to Scan Time Analysis of Token Rings and Polling Networks", ACM Sigmetrics, Performance Eval. Review, Special Issue, Vol. 12, No. 3, (1984), pp. 176-185.
- /43/ I. Rubin, L.F.M. de Moraes, "Message Delay Analysis for Polling and Token Multiple-Access Schemes for Local Communication Networks", IEEE Journal on Selected Areas in Communications, Vol. SAC-1, No. 5, (1983), pp. 935-947.

- /44/ A.S. Sethi, T. Saydam, "Performance Analysis of Token Ring Local Area Networks", Computer Networks and ISDN Systems, Vol. 9, (1985), pp. 191-200.
- /45/ G.B. Swartz, "Polling in a Loop System", Journal of the Association for Computing Machinery, Vol. 27, No.1, (1980), pp. 42-59.
- /46/ P.J. Kühn, M. Langenbach-Belz, "Über die Wirksamkeit zyklischer Abfertigungsstrategien in Realzeitsystemen", Lecture Notes in Mathematics and Computer Science, No. 26, Springer-Verlag, (1974), Seite 463-472.
- /47/ T. Raith, P. Tran-Gia, "Performance Analysis of Polling Mechanisms with Receiver Blocking", Proceedings of the 8th International Conference on Computer Communication, (ICCC), (1986), Munich, pp. 577-582.
- /48/ T. Raith, "Performance Analysis of Multibus Interconnection Networks in Distributed Systems", Proceedings of the 11th International Teletraffic Congress, Kyoto, Japan, (1985), pp. 4.2A-5.1 - 4.2A-5.7.
- /49/ D.R. Cox, "A Use of Complex Probabilities in the Theory of Stochastic Processes", Proc. Cambr. Phil. Soc. 51 (1955), pp. 313-319.
- /50/ A. Papoulis, "Probability, Random Variables and Stochastic Processes", McGraw-Hill Book Company.

- /51/ E.W. Gnedenko, D. König, "Handbuch der Bedienungstheorie I", Akademie-Verlag Berlin 1983.
- /52/ P.J. Kuehn, "Approximate Analysis of General Queueing Networks by Decomposition", IEEE Transactions on Communications, Vol. COM-27, No. 2, (1979), pp. 113-126.
- /53/ P.J. Kuehn, "Analyse zufallsabhängiger Prozesse in Systemen der Nachrichtenvermittlung und Nachrichtenverarbeitung", Habilitationsschrift, Universität Stuttgart, (1981).
- /54/ J.D.C. Little, "A Proof of the Queueing Formula $L=\lambda W$ ", Operations Research 9, (1961), pp. 383-387.
- /55/ O. J. Boxma, "Two Symmetric Queues with Alternating Service and Switching Times", Proceedings of the 10th International Symposium on Computer Performance, Paris, pp. 409-431.
- /56/ K.S. Watson, "Performance Evaluation of Cyclic Service Strategies - A Survey", Proceedings of the 10th International Symposium on Computer Performance, Paris, pp. 521-533.
- /57/ H. Takagi, "Mean Message Waiting Time in a Symmetric Polling System", Proceedings of the 10th International Symposium on Computer Performance, Paris, pp. 293-302.

- /58/ H. Takagi, L. Kleinrock, "Analysis of Polling Systems", Japan Science Institute - Communication Networks, JSI Research Report, (1985).
- /59/ H. Takagi, M. Murata, "Queueing Analysis of Scan-type TDM and Polling Systems", Proceedings of the Intern. Seminar on Computer Networking and Performance Eval., Tokyo, Japan, (1985), pp. 5.2.1 - 5.2.13.
- /60/ H. Takagi, "On the Analysis of a Symmetric Polling System with Single-Message Buffers", Performance Evaluation, Vol. 5, No. 3, (1985), pp. 149-157.
- /61/ R.J.T. Morris, Y.-T. Wang, "Some Results for Multiqueue Systems with Multiple Cyclic Servers", Proceedings of the International Symposium on the Performance of Computer-Communication Systems, North-Holland, (1984), pp. 245-258.
- /62/ C. Mack, T. Murphy, N.L. Webb, "The Efficiency of N Machines Uni-Directionally Patrolled by One Operative when Walking Time and Repair Times are Constants", Journal of the Royal Statistical Society, Ser.B, Vol.19, No. 1, (1957), pp. 166-172.
- /63/ C. Mack, "The Efficiency of N Machines Uni-Directionally Patrolled by One Operative when Walking Time is Constant and Repair Times are Variable", Journal of the Royal Statistical Society, Ser.B, Vol.19, No. 1, (1957), pp. 173-178.

- /64/ J.F. Huber, A. v.Kienlin, "Functional and Performance Requirements in Future Packet Switching Systems", Proc. of the International Switching Symposium 1984, paper 43A2.
- /65/ E. Mair, H. Hausmann, R. Naeßl, "EWSP - A High Performance Packet Switching System", Proceedings of the 8th International Conference on Computer Communication, (ICCC), (1986), Munich, pp. 359-364.
- /66/ M.N. Huber, W.M. Kiesel, P.J. Kühn, T. Raith, G. Kerschner, "Performance Modelling of a Highly Modularized Packet Switching Node", Proceedings of the 8th International Conference on Computer Communication, (ICCC), (1986), Munich, pp. 600-605.

ANHÄNGE

Anhang 1: ALGORITHMUS FÜR DIE ANALYSE NACH KAPITEL 5

Die Berechnung der charakteristischen Größen des Multi-Bus-Verbindungsnetzwerkes ohne Berücksichtigung von Blockierungen oder Interferenzen am Empfänger einer Station erfolgt auf iterative Weise.

- A. Initialisierung der Zustandswahrscheinlichkeiten und der Zykluszeitsegmentdauer jeder Station.
- B. Iterations-Zyklus - Berechnungen für jede Station
 - a) Berechnung von Mittelwert und Variationskoeffizient der Zykluszeitsegmentdauer (Kapitel 5.4.1)
 - b) Berechnung der bedingten Zykluszeiten (Kapitel 5.4.6)
 - c) Überlagerung aller Zykluszeitprozesse zur Berechnung des Zuteilungsprozesses (Kapitel 5.4.3)
 - d) Berechnung der Sendeportinterferenz (Kapitel 5.4.4)
 - e) Zerlegung des Zuteilungsprozesses mit der Wahrscheinlichkeit für Sendeportinterferenz (Kapitel 5.4.5)
 - f) Approximation der Besuchszeitverteilungsfunktion (Kapitel 5.3.3)
 - g) Berechnung der Verteilung der Systemzustände im Sender (Kapitel 5.3.1)

Nachdem die Zustandswahrscheinlichkeiten für alle Stationen berechnet wurden, wird auf Konvergenz geprüft. Als Konvergenzkriterium wird die Differenz der Summe der mittleren Warteschlangenlängen aller Stationen aufeinanderfolgender Iterationszyklen verwendet, z.B. 10^{-6} .

- C. Konvergenz - Berechnung der charakteristischen Größen
 - a) Berechnung der zeitgemittelten Zustandswahrscheinlichkeiten des Senders (Kapitel 5.3.2)
 - b) Berechnung der stationsbezogenen charakteristischen Größen (Kapitel 4.5)

Anhang 2: ALGORITHMUS FÜR DIE ANALYSE NACH KAPITEL 6

Die Berechnung der charakteristischen Größen des Multi-Bus-Verbindungsnetzwerkes mit Berücksichtigung von Blockierungen oder Interferenzen am Empfänger einer Station erfolgt auf iterative Weise.

- A. Initialisierung der Zustandswahrscheinlichkeiten, der Zykluszeitsegmentdauer und der Blockierwahrscheinlichkeit des Empfängers für jede Station.
- B. Iterations-Zyklus - Berechnungen für jede Station
- a) Ausführung der Iterationsschritte a)-d) nach Anhang 1
 - b) Berechnung des Nachrichtenankunftsprozesses am Empfänger (Kapitel 6.2.2)
 - c) Berechnung der Verteilung der Systemzustände im Empfänger (Kapitel 6.2.1)
 - d) Berechnung der Empfangsspeicherblockierung (Kapitel 6.3.1)
 - e) Berechnung der Empfangsportinterferenz (Kapitel 6.3.2)
 - f) Zerlegung des Zuteilungsprozesses mit der Wahrscheinlichkeit für Sendeportinterferenz, Empfangsportinterferenz und Empfangsspeicherblockierung (Kapitel 6.4)
 - g) Approximation der Besuchszeitverteilungsfunktion (Kapitel 5.3.3)
 - h) Berechnung der Verteilung der Systemzustände im Sender (Kapitel 5.3.1)
- Prüfen auf Konvergenz (wie in Anhang 1)
- C. Konvergenz - Berechnung der charakteristischen Größen (wie in Anhang 1)

Lebenslauf

Name : Thomas Raith
Wohnort : 7016 Gerlingen, Pappelweg 5
Geboren : 24.10.1953 in Stuttgart Bad-Cannstatt
Familienstand : verheiratet, ein Kind

von April 1960 bis Aug. 1968 Grund- und Hauptschule in Stuttgart

von Sept. 1968 bis Aug. 1971 Berufsausbildung zum Fernmeldehandwerker bei der Deutschen Bundespost FA3, Stuttgart

von Sept. 1971 bis Aug. 1972 Berufsaufbauschule in Stuttgart mit dem Abschluß der Fachhochschulreife

von Sept. 1972 bis Juni 1974 Technische Oberschule in Stuttgart mit dem Abschluß der fachgebundenen Hochschulreife

von Juli 1974 bis Sept. 1975 Grundwehrdienst

von Okt. 1975 bis Dez. 1981 Studium der Elektrotechnik (Modell: Ing.-Informatik) an der Universität Stuttgart mit dem Abschluß Dipl.-Ing. der Elektrotechnik

von Jan. 1982 bis Dez. 1986 Wissenschaftlicher Mitarbeiter an der Universität Stuttgart am Institut für Nachrichtenvermittlung und Datenverarbeitung von Herrrn Prof. Dr.-Ing. P. Kühn

Die vorliegende Arbeit entstand während meiner Tätigkeit am Institut für Nachrichtenvermittlung und Datenverarbeitung der Universität Stuttgart.

Herrn Prof. Dr.-Ing. P. Kühn möchte ich an dieser Stelle für die Themenstellung der Arbeit sowie für die wertvollen Ratschläge und Anregungen bei der Durchführung der Arbeit herzlich danken.

Weiterhin möchte ich mich bei Dr.-Ing. P. Tran-Gia und allen Kollegen des Instituts für die wertvollen Diskussionen bedanken.

