Institut für Nachrichtenvermittlung und Datenverarbeitung

Universität Stuttgart

Prof. Dr.-Ing. A. Lotze

21. Bericht über verkehrstheoretische Arbeiten

Verkehrsfluß in Datennetzen

von

Ulrich Herzog

Institute of Switching and Data Technics

University of Stuttgart

Prof. Dr.-Ing. A. Lotze

21st Report on Studies in Congestion Theory

Traffic Flow in Data Networks

by

Ulrich Herzog

1973

TRAFFIC FLOW IN DATA NETWORKS

- AN OVERVIEW OF THE MONOGRAPH -

1. INTRODUCTION
2. THE GENERAL PROBLEM
3. IMPORTANCE OF THESE INVESTIGATIONS
4. CONTENT OF THIS MONOGRAPH

After a general introduction the basic objective of traffic theory for data networks is stated. Persuing statistics of data traffic and prognoses for the next decades the importance of these investigations becomes evident. Finally, a survey is given on the intent of this monograph and a brief summary of the individual chapters, as well.

1. INTRODUCTION

Tremendous developments in the area of computer science and -technology allow us to control many technical processes automatically. Time-sharing systems have been designed and implemented for the interactive use of computers. And computer networks make it possible to share computing power, program libraries and data bases, as well.
In many cases the objectives of the system design - performance and economy- can be reached only by a fast data switching and data transmission system between the data source and the processing unit.

For some decades there has existed a network for the transmission of digital messages, the TELEX data network. This worldwide network has reached a high degree of technical perfection and is most suitable for transmission of digital

data at a rate of 50 bit per second. The TELEX network, however, is often insufficient because transmission rates of 1200, 2400, 48000 bits per second or even more are required. Therefore different methods have been developed to allow data transmission of several thousand bits per second on telephone lines. Unfortunately, reduced quality of transmission, considerable delays before a connection is established and limited speed restrict the general applicability.
Data transmission on specially designed data networks is getting more and more popular: in many countries sophisticated data networks are planned, in other countries networks of considerable size already exist.
A typical example of an existing network is the S.I.T.A.-network. The "Société Internationale de Télécommunications Aéronautiques" operates a worldwide network which deals with the data traffic of more than 150 airline companies. Computerized data switching centers in New York, London, Frankfurt, Paris and many other places control traffic. "Host"-computers connected to these switching centers handle seat reservation, air traffic information and related topics. Fig. I.1. shows a portion of the network; some technical details can be found in section III. 6.

2. THE GENERAL PROBLEM

The manifold problems occuring in the design of such networks and treated in this monograph can be demonstrated most easily by means of an example:
Fig. I.2. shows schematically a data network with four nodes, linked by the so called "high level network" (HLN). Data sources (computers, telegraph machines, typewriter-

like data terminals,....) are connected with these nodes

- via individual lines (cf. node 1)
- via multiple lines (cf. node2)
- via loop lines (cf. node 3)
- via multiplexer or concentrators (▲) with or without satellite processors (▧) as shown for node 4.

This simple example already shows the three main components to be destinguished:

1) The HLN (high level network) connecting the nodes,
2) The LLN (low level network) with the data terminals,
3) The switching centers connecting both HLN and LLN as well as controlling the traffic in the network.

The scope of traffic theory is to investigate the traffic flow in such a network, i.e. to determine and value the influence of transmission, switching, message routing, etc. The goal of these investigations is to detect possible bottlenecks in subsystems as well as to find economical structures and operating modes for the total system.

Structure and operating mode of all three components - HLN, LLN and switching centers - influence the traffic flow in such a data network:

- Single lines between data terminals and switching centers are often poorly utilized. Then, special structures of the low level networks or concentrators may save considerable cost.
- A vital problem of networks is reliability and security. It is known, e. g. from investigations in the S.I.T.A. network that transmission faults (cable breakdown, transmission errors,...) occur ten times as much as errors in the computerized switching centers. Special topologies of

the network and sophisticated routing strategies reduce the influence of such faults and guarantee a satisfactory grade of service.
- Capability and speed of switching centers may be improved considerably using modern technologies. However they depend strongly on which functions are to be realized in software and which in hardware.

A variety of problems occur which can be solved satisfactorily only if at the same time technology, transmission- and switching technique, programming as well as traffic theory are taken into consideration.

3. IMPORTANCE OF THESE INVESTIGATIONS

The investigation of traffic flow as well as the optimal design of data networks are of considerable economic importance. This can be seen clearly when studying the recent traffic statistics and prognosis for data traffic in the next decade (cf. Hough /1/, Vollmeyer/2/). Two features characterize this development:

1) Data traffic is - similar to telephone traffic - growing exponentially and there are no indications of saturation.
2) The average yearly increase in data traffic amounts to 10 or 20 percent, sometimes even more.

If we do want to manage the total volume of data traffic being expected in the next decades we have to consider the stochastic traffic flow and find optimal economic structures.

4. CONTENT OF THIS MONOGRAPH

Organization and contend of this monograph are determined by several objectives. It has been tried to

- describe the basic intent of traffic theory and its methods,
- give a survey on the different structures and operating modes of data networks,
- discuss the various traffic problems and to survey the related literature,
- describe three yet unsolved problems, to outline a method of solution and to discuss numerical results, as well.

A rather extensive list of references facilitates further research in several directions: traffic theory, transmission- or switching technique, programming, etc.

Chapter II presents a description of traffic flow in data networks by means of mathematical methods. Traffic models and service strategies, most often used are presented. Furthermore, important critera for the grade of service are defined referring to the performance of the total network or partial systems.

Chapter III deals with various topologies and routing strategies for data networks. The structure and operating mode of typical switching centers is discussed and the function of concentrators and other switching arrangements is briefly described.
Then, a survey is given of different transmission facilities and terminals. Finally, some examples of existing and future data networks are presented.

Chapter IV surveys the various traffic problems in data networks and different solution methods.
According to the subsystems of complete data networks - HLN, LLN and switching centers - the individual problems

are summarized in three categories; publications dealing with these problems, if any, are cited. Finally, the standard methods for the treatment of traffic problems -simulation and analytical methods- are summarized.

All three chapters (II, III, and IV) are of basic, tutorial nature. They are intended to survey the structure of data networks and to indicate traffic problems and different methods of solutions, as well. The following chapters, however, deal with three yet unsolved problems in the area of the HLN, the LLN and the data switching centers.

In modern store- and forward networks data are not transmitted via a unique path. Flexible routing is possible taking into consideration the instantaneous traffic situation in the network. As a result, high security in case of local overload or breakdowns and high performance is achieved (cf. Boehm and Mobley /3/, Fultz and Kleinrock /4/). Usually, such networks are designed by means of methods originally introduced for networks without flexible routing. Hence, considerable error in design and performance estimation occur.

Chapter V presents an approximate method taking into consideration the special traffic characteristics in store- and - foreward networks with alternate routing. Beside the expected waiting- and response times interesting results for distribution functions are also presented.

Comparison between calculated and simulated values demonstrate the high accuracy of the method.
An english summary of this chapter has been published at the "Seventh International Teletraffic Congress", Stockholm, Sweden, June 1973, Doc. 415: Message-Switching Networks with Alternate Routing, cf. Appendix 1 .

In very large data networks so-called satellite processors are coming into use. The aim of these facilities is not only to concentrate the traffic but also to take over some of the control procedures of the switching center (line control, remote sensing of failures, measurement,...). Therefore, satellite processors help to utilize the LLN efficiently and release the switching centers from much routine work.

In chapter VI a special satellite processor is described for networks with packet-switching and priorities. The corresponding queuing model is presented and interesting performance values are exactly derived.

In order to improve the performance of data switching centers often reasonable combinations of preemptive and nonpreemptive priorities are used. Such strategies allow us to get minimal delays for the most urgent demands avoiding at the same time the disadvantages of pure preemptive priorities (number of interrupts very high, large overhead for interrupt handling or expensive hardware; cf. Martin/6/or Graef, Greiller and Hecht/7/).

In chapter VII it is shown how the variety of combinations of preemptive- and nonpreemptive (head-of-the-line) priorities can be uniformly described by means of the so called "Preemption-Distance" (it therefore seems reasonable to intruduce the unifying term "Preemption-Distance" Priorities).
Probabilities for waiting or interrupts, mean waiting times and other characteristic performance values are determined exactly for each class of demands. Arrival processes are assumed to be Poissonian, service times

are distributed according to a "General Erlangian" distribution, individually for each priority class. Therefore, any type of distribution function can be modelled with any required accuracy.
This feature is important because in time-sharing systems or data switching systems demands within different priority classes are often of different nature. They may be constant for the most urgent class 1 hyperexponential for class 2 and distributed according to a step function for class 3. All these types can be described by means of the General Erlangian distribution.
An english summary of this chapter has been published at the Second Texas Conference on Computing Systems, Austin, November 1973, Doc. 45: Efficient Priority Strategies for Switching Centers in Communication Networks, cf. Appendix 2 .

The intention of all three chapters (V, VI and VII) is to describe in detail the problem, to present the principle of the method of solution and to discuss these problems by means of numerical results. The mathematical derivations are brief, however many references are presented which make it possible to study the different derivations.

Finally chapter VIII presents a short summary and an outlook on problems to be solved in the near and more distant future.

INHALT

Seite

ABKÜRZUNGEN –

I. ÜBERSICHT 1

1. EINFÜHRUNG 1

2. PROBLEMSTELLUNG 3

3. BEDEUTUNG 5

4. INHALT 6

II. BESCHREIBUNG DES VERKEHRSFLUSSES IN DATENNETZEN 11

1. ALLGEMEINES 11

2. ANKUNFTS - UND ENDEPROZESSE 14

2.1. Anzahl der Verkehrsquellen 14
2.2. Wichtige Ankunfts-und Endeprozesse 15
2.3. Einige Eigenschaften von Prozessen mit negativ exponentieller Verteilungsfunktion 23

3. DER SYSTEMEINFLUSS 25
3.1. Abfertigung ohne Prioritäten 25
3.2. Abfertigung mit Prioritäten 27

4. CHARAKTERISTISCHE GRÖSSEN 29

III. STRUKTUR UND BETRIEBSART VON DATENNETZEN 33

1. ÜBERBLICK UND NOMENKLATUR 33

1.1. Unterscheidung nach der Art der Wegesuche 34
1.2. Unterscheidung nach Art der Vermittlung 35

2. NETZE 40

2.1. Ungeregelte Netze 40
2.1.1. Maschennetze 40
2.1.2. Sternnetz 41
2.1.3. Ringnetz 41
2.1.4. Linien - oder Reihennetz 42
2.2. Geregelte Netze 42
2.2.1. Prinzip 42
2.2.2. Verfahren der verkehrsabhängigen Wegesuche 44
2.2.2.1. Alternative Leitweglenkung 44
2.2.2.2. Adaptive Leitweglenkung 45
2.2.2.3. Zentrale Leitweglenkung 46

3. VERMITTLUNGSEINRICHTUNGEN 46

3.1. Datenvermittlungen 47

Seite

3.1.1. Durchschaltevermittlungen 47
3.1.2. Speichervermittlungen 49
3.1.3. Erhöhung der Leistungsfähigkeit 52
3.2. Vermittlungstechnische Einrichtungen für das Anschlußnetz 54
3.2.1. Überblick 54
3.2.2. Auswahl der geeigneten Struktur 55

4. ÜBERTRAGUNGSWEGE 56

4.1. Allgemeines 57
4.2. Übertragungswege 58

5. DATENSTATIONEN 60

6. EINIGE BEISPIELE REALER DATENNETZE 61
6.1. Bestehende Netze für die Übertragung von Daten 62
6.2. Projektierte Datennetze 63

IV. VERKEHRSTHEORETISCHE PROBLEME BEI DATENNETZEN UND IHRE BEHANDLUNG 65

1. AUFGABENSTELLUNG 65

2. DER WEG ZUR OPTIMIERUNG 66

3. VERKEHRSTHEORETISCHE PROBLEME BEI DER UNTERSUCHUNG VON TEILSYSTEMEN 68

3.1. Anschlußnetz 68
3.2. Datenvermittlungen 70
3.3. Vermittlungsnetz 72

4. METHODEN ZUR BEHANDLUNG VERKEHRSTHEORETISCHER PROBLEME 75

4.1. Simulationsverfahren 75
4.2. Berechnungsmethoden 77
4.2.1. Markoff'sche Prozesse 77
4.2.2. Nicht Markoff'sche Prozesse 79
4.2.3. Die Ergebnisse von Burke und Jackson 81
4.2.4. Die Unabhängigkeitsannahme von Kleinrock 82

V. DURCHLAUFZEITEN VON NACHRICHTEN IN NETZEN MIT SPEICHERVERMITTLUNGEN UND ALTERNATIVER LEITWEGLENKUNG 84

1. PROBLEMSTELLUNG 84

2. BESCHREIBUNG DES VERKEHRSABLAUFS 87

3. MODELLBILDUNG 88

3.1. Struktur und Betriebsart 88
3.2. Verkehr 89

Seite

4. PRINZIPIELLER LÖSUNGSWEG 90

4.1. Allgemeines 90
4.2. Verkehrsfluß innerhalb einer Datenvermittlung 91
4.2.1. Das Eingangssystem 92
4.2.2. Primärsysteme 92
4.2.3. Sekundärsysteme 93
4.3. Verkehrsfluß in einem Netz 94

5. BERECHNUNG DER CHARAKTERISTISCHEN GRÖSSEN 96

5.1. Das Eingangssystem 97
5.2. Primärsysteme 97
5.3. Sekundärsysteme 99
5.3.1. Berechnung der Zustandswahrscheinlichkeiten 100
5.3.2. Charakteristische Mittelwerte 102
5.3.3. Die Wartewahrscheinlichkeit 103
5.3.4. Die Wartezeitverteilungsfunktion 104
5.3.4.1. Wartezeitverteilungsfunktion für den Testruf 105
5.3.4.2. Wartezeitverteilungsfunktion für eine beliebige Anforderung 106

6. NUMERISCHE BEISPIELE, VERGLEICH MIT SIMULATIONS-ERGEBNISSEN 106

7. ZUSAMMENFASSUNG UND AUSBLICK 114

VI. EINE SATELLITENVERMITTLUNG FÜR "PACKET-SWITCHING" MIT PRIORITÄTEN 117

1. PROBLEMSTELLUNG 117

2. BESCHREIBUNG DES VERKEHRSABLAUFES 119

3. MODELLBI LDUNG 119

3.1. Struktur und Betriebsart 119
3.2. Verkehr 120

4. BERECHNUNG DER CHARAKTERISTISCHEN GRÖSSEN 121

4.1. Prinzip des Lösungsweges 121
4.2. Die Bearbeitungsphase 122
4.3. Die Übertragungsphase 123
4.3.1. Beschreibung der Gruppenankunft 123
4.3.2. Mittlere Wartezeiten und Wartebelastung 124
4.3.2.1. Zusammenhang zwischen den Wartezeiten von Anforderungen derselben Gruppe 126
4.3.2.2. Zusammenhang zwischen den Wartezeiten und der Wartebelastung 127

Seite

4.3.2.3. Rekursionsformel für die mittlere Wartezeit 128
4.3.2.4. Die mittlere Wartezeit einer Anforderung (Paket) 129
4.3.2.5. Die mittlere Durchlaufzeit einer Gruppe (Nachricht) 129
4.4. Die mittlere Durchlaufzeit einer Nachricht der Prioritätsklasse p durch die Satellitenvermittlung 129

5. NUMERISCHE BEISPIELE 130

5.1. Vergleich zwischen "packet" - und "message - switching" 130
5.2. Mittlere Wartezeit einzelner Pakete 131
5.3. Verzögerung einzelner Pakete einer Nachricht 131

6. ZUSAMMENFASSUNG UND AUSBLICK 135

VII. UNTERBRECHUNGS - DISTANZ PRIORITÄTEN 137

1. PROBLEMSTELLUNG 137

2. BESCHREIBUNG VON UD-PRIORITÄTEN 141

2.1. Gleiche Unterbrechungs-Distanz für alle Prioritätsklassen 141
2.2. Beliebige Unterbrechungs-Distanz für jede Prioritätsklasse 143

3. MODELLBILDUNG 145

3.1. Struktur und Betriebsart 145
3.2. Verkehr 146

4. BERECHNUNG DER CHARAKTERISTISCHEN GRÖSSEN 148

4.1. Mittlere Warte - und Durchlaufzeiten, Wartebelastung 148
4.1.1. Restbedienungszeiten 151
4.1.1.1. Gesamte Restbedienungszeit 151
4.1.1.2. Abzuwartende Restbedienungszeit 152
4.1.1.3. Nicht - unterbrochene Restbedienungszeit 157
4.1.2. Die Wartebelastung und die mittlere Wartezeit $w_I(p)$ 158
4.1.2.1. Die Wartebelastung 159
4.1.2.2. Die mittlere Wartezeit $w_I(p)$ 162
4.1.3. Die mittlere Wartezeit $w_{II}(p)$ 162
4.1.4. Die mittlere Wartezeit $w_{III}(p)$ 163
4.1.5. Rekursionsformeln für die mittleren Durchlauf-und Wartezeiten 164
4.1.6. Geschlossene Lösung 165
4.2. Die Unterbrechnugswahrscheinlichkeit und die mittlere Anzahl von Unterbrechungen 166

Seite

4.2.1. Die Unterbrechungswahrscheinlichkeit 166
4.2.2. Die mittlere Anzahl von Unterbrechungen 167
4.3. Die Wartewahrscheinlichkeit 168

5. NUMERISCHE BEISPIELE 169

5.1. Einheitliche Unterbrechungs-Distanz 169
5.2. Unterschiedliche Unterbrechungs-Distanz 170
5.3. Extreme Verteilungsfunktionen für die Bedienungszeiten 170

6. ZUSAMMENFASSUNG UND AUSBLICK 179

VIII. ZUSAMMENFASSUNG 181

SCHRIFTTUM 183

ABKÜRZUNGEN

1.) ALLGEMEIN EINGEFÜHRTE ABKÜRZUNGEN

n	Anzahl der Bedienungseinheiten oder Transportkanäle
s	Anzahl der Warteplätze in einem Wartespeicher
$P\{\ldots\}$	Wahrscheinlichkeit für das in Klammern angegebene Ereignis
$E[\ldots]$	Operator zur Bildung des Erwartungswertes
$Var[\ldots]$	Operator zur Bildung der Varianz σ^2
t	Länge eines Zeitintervalls oder Zeitpunkt
T_A	Ankunftsabstand (Zufallsvariable)
T_B	Bedienungs- oder Transportzeit (Zufallsvariable)
$A(\leq t)$	Verteilungsfunktion der Ankunftsabstände
$B(\leq t)$	Verteilungsfunktion der Bedienungs- oder Transportzeiten
$F(\leq t)$	Verteilungsfunktion, allgemein
a	mittlerer Ankunftsabstand
b	mittlere Bedienungs- oder Transportzeit
k	Ordnung der Erlang-k-Verteilung
l	Ordnung der Hyperexpoentiellen Verteilung
k,l	Ordnung einer "Allgemeinen Erlangverteilung"
α, λ	Anrufrate
ε, μ	Enderate
A	Verkehrsangebot (kurz Angebot genannt) in Erlang
Y	Verkehrsbelastung (Belastung) in Erlang
R	Mittelwert von Überlaufverkehr
D	Streuwert von Überlaufverkehr
Ω	Wartebelastung
W	Wartewahrscheinlichkeit
d	mittlere Durchlaufzeit, bezogen auf alle Anforderungen
w	mittlere Wartezeit, bezogen auf alle Anforderungen
t_w	mittlere Wartezeit der wartenden Anforderungen
x	Anzahl von Ereignissen (z.B. Ankünfte, Belegungen, etc.)

2.) ZUSÄTZLICHE ABKÜRZUNGEN ZU KAPITEL II

p_γ	Wahrscheinlichkeit, daß bei einer hyperexponentiellen Verteilungsfunktion die Phase Nr. γ bzw. bei einer "Allgemeinen Erlangverteilung" die Phasenkette Nr. γ durchlaufen wird ($\gamma = 1,2,\ldots,l$).
$P_x(t)$	Wahrscheinlichkeit, daß innerhalb eines Zeitintervalls der Länge t genau x Ereignisse eintreten
q	Anzahl von Verkehrsquellen
T_m	Mittelwert einer Verteilungsfunktion F ($\leq$ t)
T_γ	Mittlere Dauer der Phase Nr. γ bei einer hyperexponentiellen Verteilung bzw. mittlere Dauer der Phasenkette Nr. γ bei einer "Allgemeinen Erlangverteilung" ($\gamma = 1,2,\ldots,l$).

3.) ZUSÄTZLICHE ABKÜRZUNGEN ZU KAPITEL III

A	Auswerte-Einheit
DV	Datenvermittlung
H-Sp	Hintergrundspeicher-Einheit
Ki	Knoten Nr. i eines Netzes (i = 1,2,...)
L	Leitungen
LA	Leitungsanschluß-Einheit
Sp	Speicher-Einheit
St	Steuer-Einheit
SV	Satellitenvermittlung
Ti	Transitknoten Nr. i eines Netzes (i = 1,2,...)
V	Vermittlungs-Einheit
VN	Verbindungs-Netz
Z-Sp	Zwischenspeicher-Einheit

4.) ZUSÄTZLICHE ABKÜRZUNGEN ZU KAPITEL IV

$\emptyset$	Stochastischer Prozeß
$\emptyset_t$	Zustand, welchen der Prozeß $\emptyset$ zum Zeitpunkt t annimmt.

$p_t(x)$	Zustandswahrscheinlichkeit, d.h. Wahrscheinlichkeit, daß zum Zeitpunkt t der Zustand x angetroffen wird.
$\ddot{u}_{t_1,t_2}(x_1,x_2)$	Übergangswahrscheinlichkeit, d.h. Wahrscheinlichkeit, daß zum Zeitpunkt t_1 der Zustand x_1 und zum Zeitpunkt t_2 der Zustand x_2 angetroffen wird.

5.) ZUSÄTZLICHE ABKÜRZUNGEN ZU KAPITEL V

$p_{k,l}(u,z)$	Wahrscheinlichkeit, daß eine Nachricht vom Ursprungsknoten u zum Zielknoten z den Zwischenknoten k erreicht hat und auf der Leitung zum Knoten l übertragen wird.
Z	Anzahl der ankommenden und abgehenden Richtungsbündel in einer bestimmten Datenvermittlung (Anschlußnetz als eine Richtung gerechnet).
z	Richtung Nr. z (z = 1,2,.....Z; Richtung Nr. Z ist das Anschlußnetz der betrachteten Datenvermittlung).
η_{zj}	Aufteilungswahrscheinlichkeit, d.h. Wahrscheinlichkeit, mit welcher eine aus Richtung z eintreffende Nachricht in Richtung j übertragen werden soll (z,j = 1,2,....,Z).
$\{u_1,u_2\}$	Zustand, daß zu einem beliebigen Zeitpunkt (Stationarität!) genau u_1 Anforderungen im Primärsystem und u_2 Anforderungen im Sekundärsystem sind.
$p(u_1,u_2)$	Zustandswahrscheinlichkeit, d.h. Wahrscheinlichkeit, daß der Zustand $\{u_1,u_2\}$ existiert

6.) ZUSÄTZLICHE ABKÜRZUNGEN ZU KAPITEL VI

P	Anzahl der Prioritätsklassen
p	Prioritätsklassenindex (p = 1,2,..., P; p = 1 höchste, p = P niedrigste Prioritätsklasse).
m (p)	maximale Gruppengröße, d.h.: maximale Anzahl gleichzeitig eintreffender Anforderungen der Prioritätsklasse p.

$x(p)$	Größe einer Gruppe der Prioritätsklasse p ($x = 1,2,\ldots m(p)$)
$\bar{x}(p)$	mittlere Gruppengröße für Prioritätsklasse p
$q_x(p)$	Wahrscheinlichkeit, daß eine eintreffende Gruppe der Prioritätsklasse p aus x Anforderungen besteht.
$b_R(\mu)$	Mittlere Restbedienungszeit einer Anforderung der Prioritätsklasse p
$w_2(p,\gamma)$	mittlere Wartezeit einer Anforderung der Prioritätsklasse p, welche als γ-te Anforderung einer Gruppe eintrifft ($\gamma = 1,2,\ldots, m(p)$)
$v_2(p,\gamma)$	mittlere Verzögerung einer Anforderung der Prioritätsklasse p, welche als γ-te Anforderung einer Gruppe eintrifft gegenüber der ersten Anforderung dieser Gruppe ($\gamma = 2,3,\ldots m(p)$)
$w_{2P}(p)$	mittlere Wartezeit einer beliebigen Anforderung (Paket) der Prioritätsklasse p.
$d_{2N}(p)$	mittlere Durchlaufzeit einer ganzen Nachricht der Prioritätsklasse p (durch das Übertragungssystem).
$d_N(p)$	mittlere Durchlaufzeit einer ganzen Nachricht der Prioritätsklasse p durch die gesamte Satellitenvermittlung.

7.) ZUSÄTZLICHE ABKÜRZUNGEN ZU KAPITEL VII

UD	Unterbrechungs-Distanz
ξ	Wert der UD
P	Anzahl der Prioritätsklassen
p	Prioritätsklassenindex ($p = 1,2,\ldots, P$; $p = 1$ höchste, $p = P$ niedrigste Prioritätsklasse)
GE (p)	Bedienungszeiten für Anforderungen der Prioritätsklasse p gehorchen einer "Allgemeinen Erlangverteilung" (General Erlangian).
$q_\gamma(p)$	Wahrscheinlichkeit, daß bei der Verteilungsfunktion GE (p) die Phasenkette Nr. γ ausgewählt wird (vgl. Abschnitt II.2.2).

$f_p(t)$	Dichtefunktion zur Verteilungsfunktion GE (p)
$b_g(p)$	mittlere "gesamte Restbedienungszeit" einer Anforderung der Prioritätsklasse p, wenn diese Anforderung zu einem zufällig gewählten Zeitpunkt während der Bedienungsphase beobachtet wird.
$b_R(\leq p + \xi - 1)$	mittlere "abzuwartende Restbedienungszeit" einer Anforderung der Prioritätsklasse p. Diese Zeit entsteht dadurch, daß Anforderungen der Klassen 1 bis $p + \xi - 1$ nicht unterbrochen werden können.
$b_n(p)$	mittlere "nicht-unterbrochene Restbedienungszeit" einer Anforderung der Klasse p. Dauer der letzten Bedienungsphase einer Anforderung der Klasse p bis zum Verlassen des Systems.
$\lambda_u(p)$	Ankunftsrate von Anforderungen, welche eine Bedienungsphase der Prioritätsklasse p unterbrechen.
$\Omega_u(p)$	Wartebelastung durch Anforderungen der Prioritätsklasse p, welche während der Bedienungsphase unterbrochen und in die Warteschlange zurückgeschoben wurden.
$P_u(p)$	Unterbrechungswahrscheinlichkeit.
$P_e(p)$	Wahrscheinlichkeit, daß eine Anforderung beim Eintreffen sofort warten muß.
$u(p)$	mittlere Anzahl von Unterbrechungen je Anforderung der Prioritätsklasse p.
$d_e(p)$	Durchlaufzeit einer Anforderung der Klasse p ab dem Zeitpunkt der ersten Bedienung.
$v_e(p)$	mittlere Verweilzeit einer Anforderung der Klasse p im Wartespeicher ab dem Zeitpunkt der ersten Bedienung.

I. ÜBERSICHT

1. EINFÜHRUNG
2. PROBLEMSTELLUNG
3. BEDEUTUNG
4. INHALT DER ARBEIT

Nach einer allgemeinen Einführung wird die Problemstellung diskutiert, die dieser Arbeit zugrunde liegt. Die Bedeutung dieser Untersuchungen erkennt man sofort, wenn man die Verkehrsstatistiken und Prognosen der letzten Jahre verfolgt. Ein letzter Abschnitt gibt eine ausführliche Übersicht über den Inhalt der einzelnen Kapitel.

1. EINFÜHRUNG

Die stürmische Entwicklung auf dem Gebiet der Datenverarbeitung ermöglicht heute eine vollelektronische Steuerung und Regelung vieler technischer Prozesse. Teilnehmerrechensysteme erlauben dem Programmierer praktisch jederzeit einen direkten Zugriff zu einer Rechenanlage. Beim Verbundbetrieb mehrerer Großrechenanlagen können Programmbibliotheken und Dateien gemeinsam benutzt, ferner Überlastungen und Teilausfälle aufgefangen werden.
Häufig ist in diesen Fällen die schnelle Datenübertragung zwischen der Datenquelle und Datenverarbeitungsanlage eine Voraussetzung für eine wirtschaftliche Lösung.

Bereits seit Jahrzehnten gibt es ein Netz zur Übertragung von Nachrichten in digitaler Form, das Telex-Datennetz. Weltweit verzweigt, technisch auf sehr hohem Stand und weitgehend vollautomatisiert ist es für den Austausch digital verschlüsselter Nachrichten mit einer Übertragungsrate von 50 bit/s ideal geeignet. Oft reicht jedoch das Telexnetz nicht aus, da Übertragungsraten von 1200, 2400, 48000 bit/s und mehr für den Datenaustausch notwendig sind. Viele Industrieländer entwickelten deshalb in den letzten Jahren Verfahren, die eine Datenübertragung von einigen tausend bit pro Sekunde im

Fernsprechnetz erlauben. Eine für die Datenübertragung oft nicht ausreichende Übertragungsqualität, relativ große Verbindungsaufbauzeiten und begrenzte Übertragungsgeschwindigkeit schränken auch hier die Anwendbarkeit ein.

Datenübertragung in speziellen Datennetzen wird deshalb in Zukunft eine immer größere Bedeutung und Verbreitung gewinnen: aus vielen Ländern sind z.T. sehr detaillierte Pläne für das Einrichten von Datennetzen bekannt, teilweise existieren schon Netze von erheblichem Umfang.

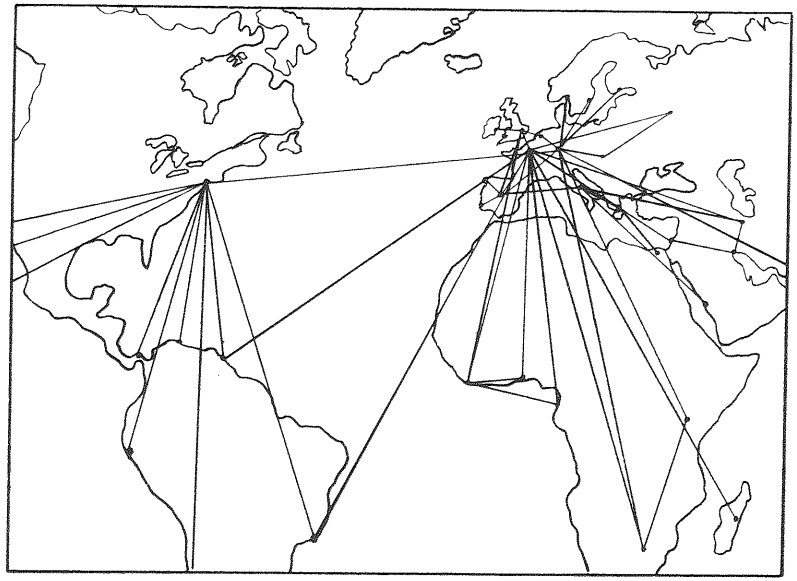


Bild I.1: Ausschnitt aus dem weltweiten Datennetz der S.I.T.A. (Société International de Télécommunications Aéronautiques)

Ein typisches Beispiel für ein bestehendes Datennetz ist das Netz der S.I.T.A. Die Société International de Télécommunications

Aéronautiques betreibt ein weltweites Datennetz, über welches der Datenverkehr von mehr als 150 Fluggesellschaften abgewikkelt wird. Vollelektronische Datenvermittlungen in New York, London, Frankfurt, Paris und vielen anderen Städten steuern den Verkehr. An diese Datenvermittlungen angeschlossene Rechner geben Auskunft über Platzreservierungen, Flugverkehr und vieles andere mehr. Bild I.1 zeigt einen Teil dieses Netzes, einige technische Daten sind in Abschnitt III.6 angegeben.

2. PROBLEMSTELLUNG

Die Vielzahl der Probleme, die bei der Dimensionierung derartiger Netze auftreten und in dieser Arbeit diskutiert bzw. untersucht werden, kann man am einfachsten an einem Beispiel erläutern:

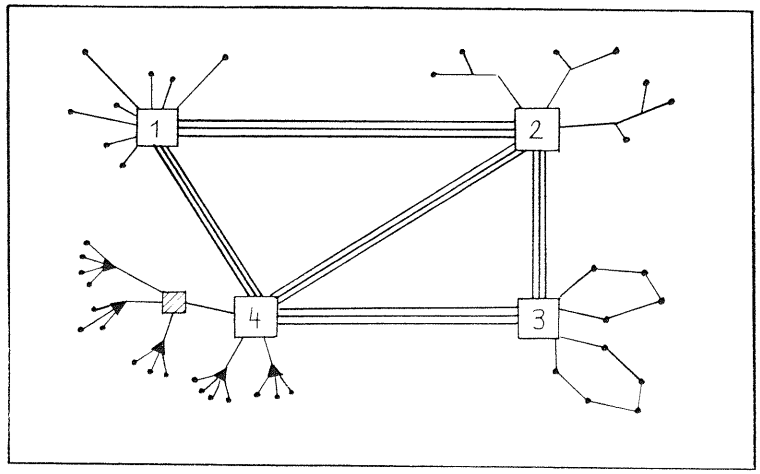


Bild I.2: Datennetz mit vier Datenvermittlungen
(• Datenstationen, ▲ Multiplexer oder Konzentratoren, ▨ Satellitenvermittlungen, □ Datenvermittlungen; eine Kombination verschiedener Strukturen des Anschlußnetzes an einer Datenvermittlung ist möglich)

Bild I.2 zeigt - schematisch dargestellt - ein Datennetz mit vier Netzknotenpunkten, die durch das sog. Vermittlungsnetz miteinander verbunden sind. Die Datenstationen der einzelnen Teilnehmer sind

- bei Knoten 1 direkt über Einzelleitungen,
- bei Knoten 2 über gemeinsame Sammelleitungen,
- bei Knoten 3 über Ringleitungen und
- bei Knoten 4 über Multiplexer und Konzentratoren, teilweise auch über eine Satellitenvermittlung

mit den rechnergesteuerten Zentren verbunden.

Bereits dieses einfache Beispiel zeigt die drei wesentlichen Komponenten, in die man ein Datennetz zerlegen kann:

1. Das Vermittlungsnetz, das die Knoten miteinander verbindet.
2. Das Anschlußnetz mit den Datenstationen. Das Anschlußnetz stellt die Verbindung zwischen den Datenstationen und dem Vermittlungsnetz her (Sonderfall eines Vermittlungsnetzes).
3. Die Datenvermittlungen mit Rechnern, die die Verbindung von Vermittlungsnetz und Anschlußnetz herstellen, sowie den Verkehrsfluß im Vermittlungsnetz steuern.

Aufgabe der Verkehrstheorie ist es, den Verkehrsfluß in einem Datennetz, d.h. die Übertragung, Vermittlung und Lenkung der Nachrichten und Steuerbefehle, zu untersuchen und zu beurteilen. Ziel dieser Untersuchung ist es, Engpässe in Teilsystemen zu entdecken und wirtschaftlich optimale Strukturen und Betriebsarten für das Gesamtsystem zu entwerfen.

Struktur und Betriebsart aller drei Komponenten - Anschlußnetz, Vermittlungsnetz und Datenvermittlungen - beeinflussen den Verkehrsfluß in einem Datennetz:

- Einzelleitungen zwischen Datenstationen und Datenvermittlung sind oft schlecht ausgelastet. Durch spezielle Strukturen des Anschlußnetzes und Verkehrskonzentration im Vorfeld können erhebliche Kosten eingespart werden.

- Ein zentrales Problem ist die Sicherung des Netzes gegen Störungen. Aus Untersuchungen beim Datennetz der S.I.T.A. ist beispielsweise bekannt, daß durch Übertragungsfehler (Störungen, Kabelausfälle, usw.) etwa zehnmal so viele Fehler auftreten als durch den Ausfall der Rechner. Verschiedene Netzformen und Verfahren der Leitweglenkung reduzieren diese Gefahr und garantieren auch im Falle der Überlastung eine ausreichende Verkehrsgüte.

- Die Leistungsfähigkeit der Datenvermittlungen kann durch den Einsatz moderner Technologien erheblich gesteigert werden. Entscheidend hängt sie jedoch stets davon ab, wo die Grenze zwischen "hardware" und "software" gezogen wird.

Überall treten dabei Probleme auf, die nur dann befriedigend gelöst werden können, wenn gleichzeitig neben der Technologie, Übertragungstechnik und Programmierung auch verkehrstheoretische Gesichtspunkte berücksichtigt werden.

3. BEDEUTUNG

Die wirtschaftliche Bedeutung der verkehrstheoretischen Untersuchung und Optimierung von Datennetzen erkennt man sofort, wenn man die Verkehrsstatistiken der letzten Jahre sowie die Prognosen für die nächsten Jahrzehnte verfolgt (s.z.B. Hough /1/, Vollmeyer /2/.).

Zwei Merkmale kennzeichnen die Entwicklung des Datenverkehrs:

1. Das Wachstum weist - wie das des Fernsprechverkehrs - überall Exponentialverhalten auf und es sind keinerlei Anzeichen für eine beginnende Sättigung vorhanden.

2. Die durchschnittlichen jährlichen Verkehrszuwachsraten betragen etwa 10 bis 20 Prozent, teilweise sogar mehr.

Will man in den nächsten Jahren die zu erwartenden Verkehrszunahmen bewältigen, so ist es deshalb unumgänglich, Netze verkehrstheoretisch zu untersuchen und wirtschaftlich optimale Strukturen zu finden.

4. INHALT DER ARBEIT

Vier wesentliche Gesichtspunkte bestimmen den Aufbau und Inhalt der vorliegenden Arbeit. Es wird versucht,

- eine Einführung in die verkehrstheoretischen Betrachtungsweise und ihre Methoden zu geben,
- die verschiedenen Möglichkeiten der Struktur und Betriebsart von Datennetzen kurz vorzustellen,
- Verkehrsprobleme zu diskutieren und einen Überblick über bisher durchgeführte Untersuchungen zu geben sowie
- drei bisher ungelöste Probleme zu beschreiben, Lösungswege zu skizzieren und Endergebnisse anzugeben.

Zahlreiche Literaturangaben ermöglichen eine Vertiefung in den verschiedenartigsten Richtungen: Verkehrstheorie, Übertragungs- und Vermittlungstechnik, Programmierung usw.

In Kapitel II wird gezeigt, wie mit Hilfe mathematischer Methoden der Verkehrsfluß in Datennetzen beschrieben werden kann. Die bekanntesten Verkehrsmodelle und Abfertigungsstrategien werden vorgestellt; wichtige Kriterien zur Beurteilung der Leistungsfähigkeit eines Gesamtnetzes, bzw. Teilsystems definiert.

In Kapitel III werden verschiedene Netzformen und Verfahren der Verkehrslenkung diskutiert. Die Struktur und der Aufbau typischer Datenvermittlungen werden kurz beschrieben, ferner die Wirkungsweise von Konzentratoren und anderen Vermittlungseinrichtungen des Anschlußnetzes erläutert.

Nach einem Überblick über die zur Verfügung stehenden Übertragungswege und Datenstationen werden noch einige Beispiele bestehender und zukünftiger Datennetze vorgestellt.

Kapitel IV gibt eine Übersicht über die verkehrstheoretischen Probleme in Datennetzen und die verschiedenen Lösungsmethoden.

Die Vielzahl der verschiedenen Einzelprobleme wird in drei großen Gruppen zusammengefaßt, aus der Literatur bekannte Lösungsalgorithmen werden - soweit vorhanden - zitiert. Eine Zusammenstellung der verschiedenen Methoden zur Behandlung der verkehrstheoretischen Probleme - Simulation und rechnerische Verfahren - beschließt dieses wichtige Kapitel.

Diese drei Kapitel (II bis IV) haben einen einführenden Charakter. Sie geben einen Überblick über den Aufbau von Datennetzen, zeigen Verkehrsprobleme auf und geben einen Überblick über die verschiedenen Lösungswege. In den folgenden Kapiteln werden drei bisher ungelöste Probleme mathematisch behandelt, je eines aus dem Bereich

der Vermittlungsnetze, der Anschlußnetze und der Datenvermittlungen.

In modernen Netzen mit Speichervermittlungen werden Nachrichten vom Sender zum Empfänger nicht auf einem fest vorgeschriebenen Weg übertragen. Eine verkehrsabhängige Wegesuche in den einzelnen Knotenpunkten ermöglicht eine flexible Verkehrslenkung und deshalb eine hohe Sicherheit im Falle von Überlastung und Ausfall von Teilen des Netzes (vgl. z.B. Boehm und Mobley /3/, Fultz und Kleinrock /4/). Derartige Netze werden heute noch nach denselben Verfahren dimensioniert, die für Netze ohne flexible Verkehrslenkung entwickelt wurden. Fehlplanungen und eine falsche Abschätzung der Leistungsfähigkeit sind die Folge.
In Kapitel V wird nun ein Näherungsverfahren vorgestellt, das die verkehrsgerechte Dimensionierung von Netzen mit Speichervermittlungen und alternativer Leitweglenkung erlaubt. Neben den Mittelwerten für Warte- und Durchlaufzeiten werden auch Verteilungsfunktionen angegeben. Der Vergleich von Ergebnissen aus Rechnung und Simulation zeigt die hohe Genauigkeit des Verfahrens.

Bei großen Datennetzen werden im Anschlußnetz Satellitenvermittlungen eingesetzt. Neben den Aufgaben der Verkehrskonzentration werden auch gewisse Teilaufgaben der Datenvermittlungen übernommen. Satellitenvermittlungen verbessern deshalb nicht nur die Auslastung der Anschlußleitungen, sondern befreien die Datenvermittlungen von zahlreichen Routineaufgaben (s.z.B. Martin /5/).
In Kapitel VI wird eine spezielle Satellitenvermittlung für ein Netz mit "packet - switching" und Prioritäten vorgestellt, das entsprechende Modell (mit Gruppenankunft) entwickelt und mathematisch exakt behandelt.

Zur Erhöhung der Leistungsfähigkeit von Datenvermittlungen benutzt man häufig sinnvolle Kombinationen sofort unterbrechender und nicht-unterbrechender Prioritäten. Dadurch können einerseits Wartezeiten für besonders dringliche Anforderungen auf ein Minimum beschränkt werden, auf der anderen Seite vermeidet man die zahlreichen Nachteile rein unterbrechender Prioritäten (unnötig große Anzahl von Unterbrechungen, sehr hoher Aufwand für Verwaltungszeiten im Prozessor; s. z.B. Martin /6/ sowie Graef, Greiller und Hecht /7/).

In Kapitel VII werden die verschiedenartigsten Kombinationen unterbrechender und nichtunterbrechender Prioritäten einheitlich beschrieben und unter dem Namen "Unterbrechungs-Distanz" Prioritäten zusammengefaßt. Warte- und Unterbrechungswahrscheinlichkeit, mittlere Wartezeit und verschiedene andere charakteristische Größen werden für jede Prioritätsklasse exakt bestimmt. Ankunftsabstände eintreffender Anforderungen gehorchen einer Poisson-Verteilung, für Bedienungszeiten werden pro Prioritätsklasse unterschiedliche "Allgemeine Erlangverteilungen" angenommen. Damit ist man in der Lage, jeden möglichen Verlauf einer Verteilungsfunktion von Bedienungszeiten individuell pro Prioritätsklasse, beliebig genau nachzubilden und zu untersuchen.
Beispielsweise sind bei Datenvermittlungen oder Time-Sharing Rechenanlagen die Bearbeitungszeiten für Anforderungen der verschiedenen Prioritätsklassen oft sehr unterschiedlich. Sie können bei der dringlichsten Prioritätsklasse (Klasse 1) konstant, bei Klasse 2 hyperexponentiell und bei Klasse 3 entsprechend einer Treppenfunktion verteilt sein. Alle diese Typen von Verteilungsfunktionen sind mit Hilfe der "Allgemeinen Erlangverteilung" beschreibbar.

In allen drei Kapiteln (V, VI und VII) wird versucht, die Problemstellung, den prinzipiellen Lösungsweg und die Ergebnisse ausführlich zu beschreiben. Mathematische Ableitungen werden relativ kurz gefaßt, zahlreiche Literaturhinweise ermöglichen jedoch ein spezielles Studium der verschiedenen Methoden.

Zum Schluß wird in Kapitel VIIIeine kurze Zusammenfassung sowie ein Ausblick auf jene Arbeiten gegeben, die in Zukunft durchzuführen sind.

II. BESCHREIBUNG DES VERKEHRSFLUSSES IN DATENNETZEN

1. ALLGEMEINES
2. ANKUNFTS-UND ENDEPROZESSE
3. SYSTEMEINFLUSS
4. CHARAKTER. GRÖSSEN

Der Verkehrsfluß in Datennetzen wird sowohl durch den Verkehrscharakter als auch durch den Systemeinfluß bestimmt.
In diesem Kapitel wird gezeigt, wie mit Hilfe mathematischer Methoden eine genaue Beschreibung des Verkehrscharakters möglich ist. Es wird ferner ein Überblick gegeben über die verschiedenen Komponenten, die notwendig sind, den Systemeinfluß zu beschreiben.
Die Verkehrsgüte eines Datennetzes (bzw. eines Teilsystems) kann mit Hilfe der sog. charakteristischen Größen beurteilt werden. Die wichtigsten charakteristischen Größen werden deshalb in einem letzten Abschnitt zusammengestellt und definiert.

1. ALLGEMEINES

Das Ablaufgeschehen in Datennetzen wird durch den Transport von Nachrichten auf den Übertragungswegen sowie deren Bearbeitung und Weiterleitung in den verschiedenen Vermittlungseinrichtungen (Datenvermittlungen, Multiplexer, etc.) bestimmt. Nicht ausreichende Übertragungsraten im Anschluß - oder Verbindungsnetz führen zu einem Rückstau der Nachrichten in den Datenvermittlungen. Ferner behindern hardwaremäßige Engpässe der Vermittlungseinrichtungen sowie die ungünstige Konzeption des Betriebssystems eine reibungslose Vermittlung zu übertragender Nachrichten.

Der zeitliche Ablauf derartiger Transport-, Warte- und Bearbeitungsphasen entspricht einem speziellen stochastischen Prozeß (Zufallsprozeß), welcher mit Methoden der Wahrscheinlichkeitstheorie beschrieben werden kann.

Vollständig beschrieben wird dieses zufällige, stochastische Verhalten durch folgende Kriterien:

- Ankunftsprozeß: Verteilung der Ankunftsabstände der Nachrichten bzw. Befehle. Die Verteilungsfunktion $A(\leq t)$ ist definiert als die Wahrscheinlichkeit, daß der Ankunftsabstand T_A höchstens gleich der beliebigen Zeit t ist:

$$A(\leq t) = P\{T_A \leq t\} \qquad \text{(II.1)}$$

Bezeichnet man den mittleren Ankunftsabstand mit a, so wird der Kehrwert als mittlere Ankunftsrate definiert:

$$\lambda = 1/a \qquad \text{(II.2)}$$

- Endeprozeß: Verteilung der Bedienungs - oder Transportzeiten T_b bzw. Verteilung der Länge T_b von Nachrichten oder Datenblöcken. Entsprechend Gl (II.1) und (II.2) gilt:

$$B(\leq t) = P\{T_B \leq t\} \qquad \text{(II.3)}$$

$$\mu = 1/b \qquad \text{(II.4)}$$

(Ankunfts- und Enderaten können auch Zeit- und/oder zustandsabhängig sein. In dieser Einführung wird jedoch auf eine derart allgemeine Darstellung aus Gründen der Übersichtlichkeit verzichtet).

- Systemeinfluß: Der Einfluß des Systems ist durch die Systemstruktur und die Art der Abfertigung ankommender und wartender

Nachrichten, Datenblöcke oder Befehle festgelegt.

Bild II.1 zeigt - schematisch dargestellt - die Struktur eines einfachen, aber typischen Bedienungssystems: von den Verkehrsquellen (z.B. Fernschreiber, Rechner, Übertragungskanäle) wird ein "Strom " von Bedienungswünschen erzeugt. Ist die Bedienungseinheit (z.B. Zentraler Prozessor einer

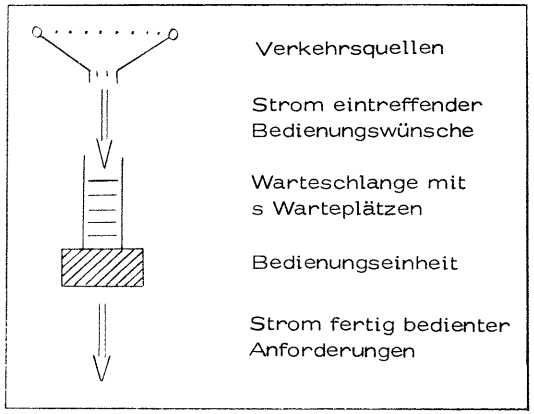


Bild II.1: Schematische Darstellung eines einfachen Bedienungssystems mit einer Warteschlange und einer Bedienungseinheit

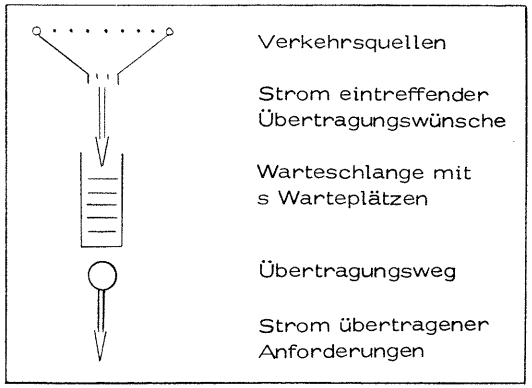


Bild II.2: Schematische Darstellung eines zu Bild II.1 analogen Übertragungssystems

Datenvermittlung, Steuerung einer Speichereinheit, etc.) momentan belegt, so wird in einer Warteschlange ein Indikator gesetzt, der die Ankunft des neuen Bedienungswunsches anzeigt. Die Warteschlange wird in einer bestimmten Reihenfolge (s. II.3) von der Bedienungseinheit abgearbeitet. Das entsprechende Beispiel eines einfachen Übertragungssystems zeigt Bild II.2.

2. ANKUNFTS- UND ENDEPROZESSE

Die Wahl eines geeigneten Verkehrsmodells ist von weittragender Bedeutung: je realistischer die gewählten Ankunfts- und Endeprozesse sind, desto besser werden die durch Analyse bzw. Synthese erzielten Ergebnisse der Wirklichkeit entsprechen.

In diesem Abschnitt wird der Einfluß einzelner Verkehrsquellen kurz diskutiert, ferner wird ein Überblick über die wichtigsten Ankunfts- und Endeprozesse gegeben.

2.1 Anzahl der Verkehrsquellen

In Wirklichkeit ist die Anzahl der Verkehrsquellen immer endlich (vgl. auch II.1); jede einzelne Verkehrsquelle hat ein individuelles Verkehrsbedürfnis. Der sog. Angebotsdruck, d.h. die erwartete Anzahl eintreffender Bedienungs- und Transportwünsche pro Zeiteinheit ist deshalb vom momentanen Belegungszustand abhängig: sind viele Verkehrsquellen momentan belegt (werden bedient oder warten), so ist der Angebotsdruck gering. Sind dagegen nur wenige Verkehrsquellen belegt, so ist der

Angebotsdruck groß.

Mathematisch exakt ist dieser Fall sehr schwierig zu behandeln, siehe z.B. Brandt /8/.

In vielen Fällen der Praxis ist die Anzahl der Verkehrsquellen wesentlich größer als die Anzahl der zur Verfügung stehenden Bedienungseinheiten oder Transportkanäle; der Anteil einer individuellen Verkehrsquelle am gesamten Verkehrsvolumen ist gering. In all diesen Fällen kann man ein idealisiertes Modell anwenden, das die mathematische Behandlung des Problems erheblich vereinfacht: die Anzahl der Verkehrsquellen q wächst über alle Grenzen und der Beitrag einer individuellen Quelle (Ankunftsrate α) strebt gegen null dergestalt, daß das Produkt $q.\alpha$ konstant ist. Bezeichnet man mit x die Anzahl momentan belegter Verkehrsquellen, so gilt deshalb für die gesamte Ankunftsrate

$$\lim_{\substack{q \to \infty \\ \alpha \to 0 \\ q \cdot \alpha = \text{konst.}}} \alpha \cdot (q - x) = \lambda \qquad \text{(II.5)}$$

In diesem indealisierten Verkehrsmodell ist also der Angebotsdruck - d.h. die Ankunftsrate λ - unabhängig vom Belegungszustand. Im folgenden Abschnitt werden wichtige Ankunfts- und Endeprozesse vorgestellt, die diese Forderung erfüllen.

2.2. Wichtige Ankunfts- und Endeprozesse

Bereits früher wurde auf die Bedeutung der gewählten Ankunfts-

und Endeprozesse hingewiesen. Die folgende Zusammenstellung gibt nun einen Überblick über die wichtigsten mathematischen Modelle, weitere Einzelheiten findet man z.B. in den Büchern von Feller /9/, Ferschl /10/, Gnedenko /11/, Morse /12/, Riordan /13/ und Syski /14/. Interessante Meßergebnisse von realen Systemen sowie deren Vergleich mit verschiedenen mathematischen Modellen werden in Arbeiten von Cole /15/, Jackson, Stubbs, Fuchs und Dudick /16, 17, 18/ vorgestellt, siehe ferner Hayward und Wilkinson /19,20/.

Negativ exponentielle Verteilung

Die bekannteste Verteilungsfunktion für Ankunfts- und Endeprozesse ist die negativ exponentielle Verteilung. Sie tritt nicht nur bei vielen Naturereignissen (Zerfall von Atomen, Brown'sche Bewegung, etc.) auf, sondern beschreibt häufig in guter Näherung die Verteilung von Ankunftsabständen sowie Bedienungs- und Transportzeiten in Daten- und Fernsprechnetzen.

Für die Verteilungsfunktion F ($\leq$ t) (Wahrscheinlichkeit, daß der Ankunftsabstand, die Bearbeitungs- oder die Transportzeit T höchstens gleich einer bestimmten Zeit t ist) gilt:

$$F(\leq t) = P\{T \leq t\} = 1 - e^{-\frac{t}{T_m}} \qquad \text{(II.6)}$$

mit T_m als Mittelwert der Verteilung von T. Bild II. 3 zeigt die entsprechende graphische Darstellung, einige wichtige Eigenschaften werden in Abschnitt II.2.3. zusammengestellt.

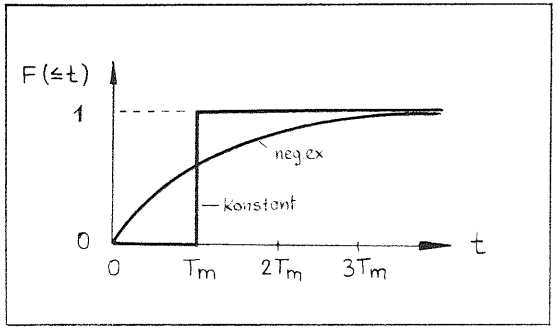


Bild II. 3: Verteilungsfunktion F (≤ t) einer negativ exponentiellen bzw. konstanten Verteilung

Die mathematische Behandlung derartiger Prozesse ist noch relativ einfach, da es sich in diesem Fall um sog. Markoff'sche Prozesse (s. IV.4.2) handelt.

Konstante Verteilung

Endeprozesse werden häufig besser durch eine konstante Verteilung beschrieben. Ein typisches Beispiel sind Datennetze mit "packet-switching" (s. III. 2.2). Für die Verteilungsfunktion gilt dann

$$F(\leq t) = P\{T \leq t\} = \begin{cases} 0 & \text{für} \quad 0 \leq t < T_m \\ 1 & \text{für} \quad t \geq T_m \end{cases} \qquad \text{(II.7)}$$

Die mathematische Behandlung derartiger Verkehre ist jedoch bedeutend schwieriger als im ersten Fall, da die sog. Markoff'sche Eigenschaft (Unabhängigkeit von der Vorgeschichte, s. IV. 4.2.1) verlorengeht. Der Verlauf der Verteilungsfunktion ist ebenfalls

in Bild II. 3 graphisch dargestellt.

Erlang - k - Verteilung:

Beide oben angegebenen Verteilungsfunktionen, die negativ exponentielle und die konstante, sind Grenzwerte der Erlangverteilung k - ter Ordnung (kurz Erlang - k - Verteilung):

$$F(\leq t) = P\{T \leq t\} = 1 - e^{-\frac{t}{T_m/k}} \sum_{\nu=0}^{k-1} \frac{\left(\frac{t}{T_m/k}\right)^{\nu}}{\nu!} \qquad \text{(II. 8)}$$

Diese Verteilung wird erfolgreich angewendet, um Ankunfts- und Endeprozesse zu beschreiben, deren Verteilungsfunktion zwischen konstanter und negativ exponentieller liegen. Bild II. 4 zeigt den Verlauf für verschiedene Ordnungsnummern k. Man erkennt sofort, daß sich für k = 1 die negativ exponentielle Verteilung ergibt.

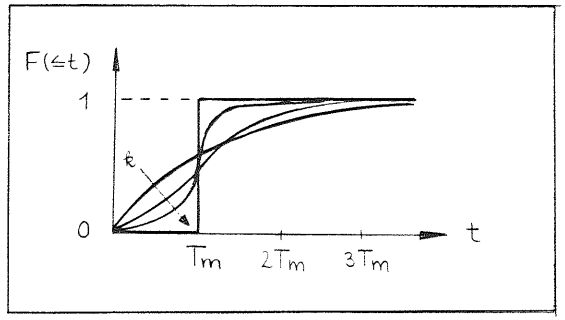


Bild II. 4: Verteilungsfunktion F (≤ t) einer Erlang - k - Verteilung

Durch einen Grenzübergang kann man ferner zeigen, daß für $k \to \infty$ die konstante und die Erlang - k - Verteilung identisch sind.

Zum Schluß sei noch eine physikalische Interpretation erwähnt, die den Charakter der Erlang - k - Verteilung sehr anschaulich erfaßt (vgl. Bild II.5). Eintreffende Anforderungen durchlaufen

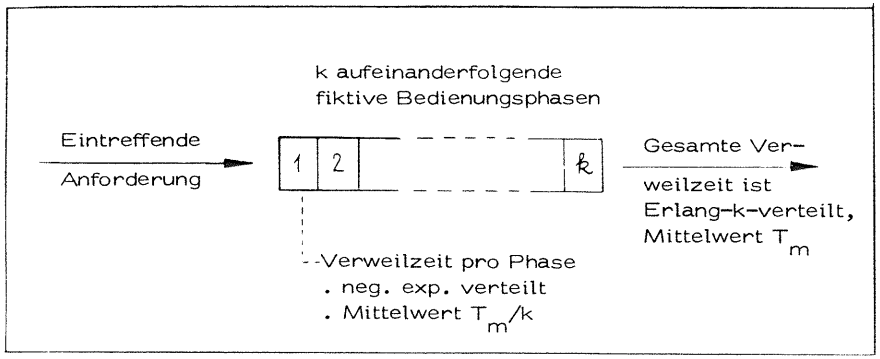


Bild II.5: Physikalische Interpretation zur Erlang - k - Verteilung

k aufeinanderfolgende fiktive Bedienungsphasen. Die Verweilzeit in jeder Phase ist negativ exponentiell verteilt mit dem Mittelwert Tm/k. Die gesamte Verweilzeit ist dann Erlang - k - verteilt mit dem Mittelwert Tm.

Hyperexponentielle Verteilung

Ankunfts-und Endeprozesse mit vielen sehr kurzen und vielen sehr langen, dagegen wenigen Ankunfts-oder Bedienungsintervallen mittlerer Länge können mit Hilfe einer hyperexponentiellen Verteilung l - ter Ordnung beschrieben werden:

$$\left.\begin{aligned} F(\leq t) = P\{T \leq t\} &= \sum_{\nu=1}^{l} p_\nu \left(1 - e^{-\frac{t}{T_\nu}}\right) \; ; \quad \sum_{\nu=1}^{l} p_\nu = 1 \\ E[T] = T_m &= \sum_{\nu=1}^{l} p_\nu \cdot T_\nu \end{aligned}\right\} \quad (II.9)$$

Der typische Verlauf hyperexponentieller Verteilungsfunktionen ist in Bild II.6, zusammen mit dem Grenzwert l = 1 (negativ exponentielle Verteilung), skizziert.

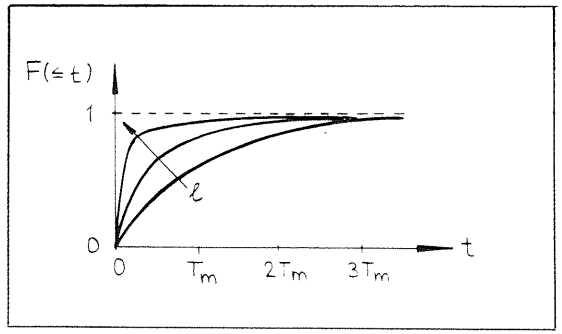


Bild II.6: Verteilungsfunktion F(≤t) einer hyperexponentiellen Verteilung l-ter Ordnung

Auch in diesem Fall ist eine physikalische Interpretation sehr leicht möglich (vgl. Bild II.7): eintreffende Anforderungen

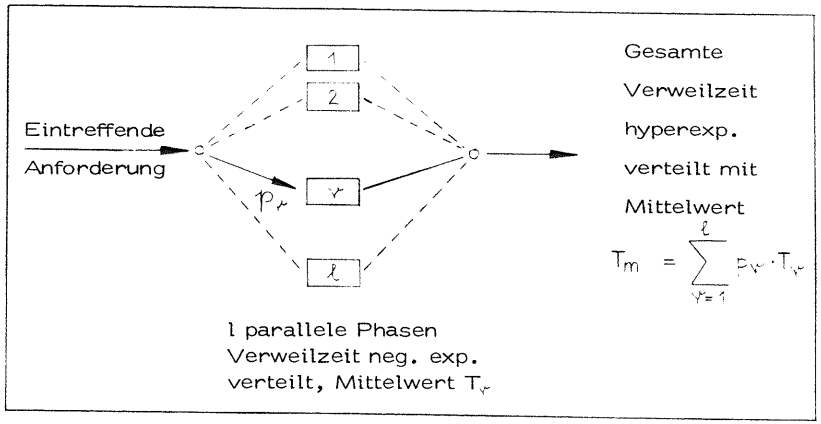


Bild II.7: Physikalische Interpretation zur hyperexponentiellen Verteilung

durchlaufen mit der Wahrscheinlichkeit p_ν eine negative exponentiell verteilte Bedienungsphase mit dem Mittelwert T_ν und verlassen anschließend das System. Die Verweilzeit aller Anforderungen entspricht deshalb einer Summe gewichteter Exponentialverteilungen, wie sie durch Gleichung (II.9) ausgedrückt wird.

Allgemeine Erlangverteilung

Eine Familie von Verteilungsfunktionen ohne Begrenzung ist die "Allgemeine (oder Gemischte) Erlangverteilung", bestehend aus einer Summe gewichteter Erlang - k Verteilungen.

$$F(=t) = P\{T \leq t\} = \sum_{\nu=1}^{\ell} p_\nu \left[1 - e^{-\frac{t}{T_\nu/k_\nu}} \sum_{\eta=0}^{k_\nu - 1} \frac{\left(\frac{t}{T_\nu/k_\nu}\right)^\eta}{\eta!} \right]$$

Normierungsbedingung $\sum_{\nu=1}^{\ell} p_\nu = 1$ (II.10)

Mittelwert $E[T] = T_m = \sum_{\nu=1}^{\ell} p_\nu \cdot T_\nu$

Auch in diesem Fall gibt die physikalische Interpretation in Bild II. 8 einen interessanten Aufschluß über den Typ der Verteilungsfunktion. Sofort erkennt man, daß alle bisher vorgestellten Verteilungsfunktionen sich als Sonderfall aus dieser Funktion ergeben.

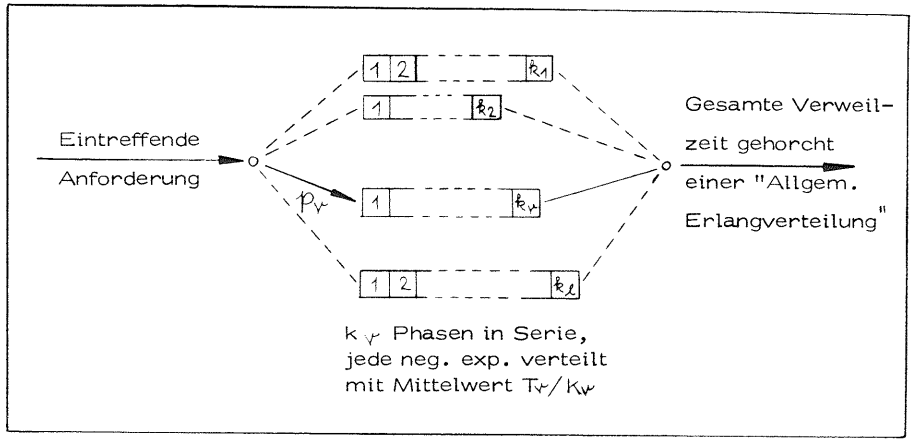


Bild II.8: Physikalische Interpretation zur "Allgemeinen Erlangverteilung"

Drei willkürlich herausgegriffene Beispiele in Bild II.9 veranschaulichen, daß mit Hilfe der "Allgemeinen Erlangverteilung" ein beliebig komplizierter Verlauf der Verteilungsfunktion von Ankunfts- und Endeprozessen beschrieben werden kann.

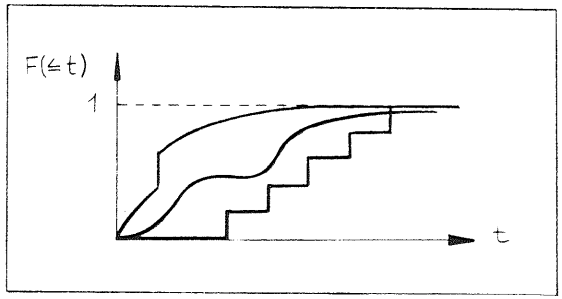


Bild II.9: Drei Beispiele für den Verlauf der Verteilungsfunktion der "Allgemeinen Erlangverteilung"

Erlang - k - Verteilung, hyperexponentielle Verteilung und die "Allgemeine Erlangverteilung" erlauben eine genaue Beschreibung vieler in der Praxis auftretenden Ankunfts- und Endeprozesse; die mathematische Behandlung derartiger Verkehrsmodelle ist jedoch häufig recht schwierig (vgl. auch Kapitel VII).

Gruppenankunft und Gruppenbedienung

Ein typisches Merkmal aller bisher behandelten Ankunfts- und Endeprozesse besteht darin, daß innerhalb eines infinitesimal kleinen Zeitelements dt höchstens ein Ankunfts- oder Endeereignis auftreten kann (Verteilungsfunktion beginnt immer im Ursprung $F(\leq 0) = 0$). Treffen an einer Bedienungsstelle gleichzeitig mehrere Bedienungswünsche ein, so spricht man von Prozessen mit Gruppenankunft.

Der Ankunftsabstand der einzelnen Gruppen ist häufig wie bei den Prozessen mit Einzelankunft verteilt (negativ exponentiell, konstant, etc.). Innerhalb einer Gruppe treffen mit Wahrscheinlichkeit p_x genau x Anforderungen ein. Analoge Überlegungen führen zu den Prozessen mit Gruppenabfertigung.

Ein typisches Beispiel für einen Prozess mit Gruppenankunft wird in Kapitel VI behandelt.

2.3. Einige Eigenschaften von Prozessen mit negativ exponentieller Verteilungsfunktion

. Die negativ exponentielle Verteilungsfunktion ist in Gleichung (II.6) angegeben, ihren Verlauf zeigt das Diagramm in Bild II.3.

. Mittelwert und Varianz können direkt aus der Verteilungsfunktion abgeleitet werden. Es gilt

$$\left.\begin{aligned} E[T] &= T_m \\ Var[T] &= \sigma^2 = T_m^2 \end{aligned}\right\} \quad (II.11)$$

. Die Wahrscheinlichkeit P_x (dt) daß innerhalb der infinitesimal kleinen Zeitspanne (t_1, t_1 + dt) genau x Ereignisse eintreten, ist unabhängig vom Zeitpunkt t_1 und unabhängig davon, ob im Zeitpunkt t_1 ein Ereignis eingetreten ist oder nicht, konstant.

Die Wahrscheinlichkeit, daß zwei oder mehr Ereignisse in der Zeitspanne dt eintreffen ist 0 (dt), also vernachlässigbar. Damit gilt:

$$\left.\begin{aligned} P_1(dt) &= \frac{1}{T_m} \cdot dt \\ P_{\geq 2}(dt) &= 0 \end{aligned}\right\} \quad (II.12)$$

. Betrachtet man ein Zeitintervall endlicher Länge to, so gehorcht die Wahrscheinlichkeit P_x (to), daß innerhalb dieses Intervalls genau x Ereignisse eintreten einer Poisson-Verteilung:

$$P_x(t_0) = \frac{\left(\frac{t_0}{T_m}\right)^x}{x!} \cdot e^{-\frac{t_0}{T_m}} \quad (II.13)$$

Ankunftsprozesse mit negativ exponentiell verteilten Ankunftsabständen werden deshalb häufig auch als Poisson-Prozeß bezeichnet.

. Der Mittelwert der Poisson-Verteilung, d.h. die mittlere Anzahl von Ereignissen in einem Zeitintervall der endlichen

Länge to, kann wieder leicht aus Gleichung (II.13) abgeleitet werden. Es ergibt sich:

$$E[x] = \frac{t_o}{T_m} \qquad \text{(II.14)}$$

3. DER SYSTEMEINFLUSS

Transport- und Bedienungsprozesse werden sowohl durch die Systemstruktur als auch durch die Art der Abfertigung eintreffender und wartender Anforderungen beeinflußt. Bereits im Abschnitt II.1 wurden zwei einfache Strukturmodelle vorgestellt (Bild II.1 und 2). Komplexe Strukturen für Netze und Datenvermittlungen werden in Kapitel III diskutiert, weitere Beispiele werden in Kapitel V bis VII vorgestellt und verkehrstheoretisch untersucht.

Von der Vielzahl verschiedenartiger Abfertigungsstrategien für eintreffende oder wartende Anforderungen können hier nur die wichtigsten behandelt werden. Sie werden im folgenden unterschieden nach der Abfertigung mit und ohne Prioritäten. Zahlreiche Literaturangaben siehe z.B. Herzog, Kühn und Zeh /21/ und Wagner /22/.

3.1 Abfertigung ohne Prioritäten

Die bekanntesten Abfertigungsstrategien für gleichberechtigte Anforderungen sind:

. FIFO — first-in, first-out, d.h. die zuerst eingetroffene Anforderung wird zuerst bedient. Im Gegensatz zu den später vorgestellten Strategien ist in diesem Fall stets gewährleistet, daß alle Anforderungen mit Sicherheit bedient werden. Nahezu alle Warteschlangen für Realzeitanforderungen werden deshalb nach der FIFO-Strategie organisiert und abgearbeitet.

. LIFO — last-in, first-out, d.h. die zuletzt eingetroffene Anforderung wird zuerst bedient. Der für die Organisation von LIFO-Warteschlangen erforderliche Zeitaufwand ist - im Gegensatz zu FIFO - sehr gering. Oft werden deshalb Warteschlangen für die Speicherplatzverwaltung nach LIFO organisiert.

. RANDOM — die wartenden Anforderungen werden zufällig ausgesucht und bedient. Derartige Abfertigungsdisziplinen treten oft auf, wenn die Auswahl der zu bedienenden Anforderung mit Hilfe einer Schaltung (hardware) erfolgt.

Sind mehrere Warteschlangen mit gleichberechtigen Anforderungen vorhanden, so kommen zu den oben vorgestellten Abfertigungsstrategien noch verschiedene Möglichkeiten zur Auswahl einer bestimmten Warteschlange hinzu (sog. interqueue - discipline, vgl. Kühn /23, 24/. Die taktmäßige Abfertigung mehrerer paralleler Warteschlangen wird von Langenbach - Belz /25/ behandelt.

3.2 Abfertigung mit Prioritäten

Bei Prioritäten unterscheidet man interne und externe Prioritäten. Interne Prioritäten werden bei an sich gleichrangigen Aufgaben gesetzt, um z.B. einen maximalen Durchsatz in einer Datenvermittlung zu erreichen. Externe Prioritäten geben Auskunft über die Wichtigkeit einer Anforderung (Nachricht, Befehl) im Vergleich mit anderen konkurrierenden Anforderungen. Sie sind oft fest vorgegeben für bestimmte Anforderungen, können aber auch wartezeitabhängig sein.

Sowohl interne als auch externe Prioritäten können rein unterbrechend, rein nichtunterbrechend oder eine Mischform zwischen beiden Extremen sein.

Unterbrechende Prioritäten

Rein unterbrechend bedeutet, daß die Bearbeitung einer Anforderung beim Eintreffen einer Anforderung höherer Priorität sofort unterbrochen wird. Die weniger dringliche Anforderung wird erst dann weiterbearbeitet, wenn die Anforderung höherer Priorität das System verlassen hat und keine neue in der Zwischenzeit eingetroffen ist.

Nichtunterbrechende Prioritäten

Bei den rein nichtunterbrechenden Strategien werden Anforderungen einer bestimmten Prioritätsklasse vor allen andern Wartenden

bedient, die weniger dringlich sind. Eine momentane Bearbeitungsperiode kann aber nicht unterbrochen werden.

Gemischte Prioritäten

Bei Echtzeit - Datenverarbeitung sind unterbrechende Prioritäten für spezielle unaufschiebbare Anforderungen notwendig, damit z.B. ein reibungsloser Verkehrsablauf in einem Datennetz garantiert werden kann. Andere Anforderungen sind oft weniger wichtig und rechtfertigen keine Unterbrechung. Teilweise sind sogar unterbrechende Prioritäten unsinnig, z.B. wenn eine unterbrechende Nachricht zwar wichtiger ist als die unterbrochene, der Zeitaufwand für die Unterbrechung jedoch größer ist als die notwendige restliche Bedienungszeit.

Praktisch in allen Echtzeit - Datenverarbeitungssystemen benutzt man eine Kombination aus unterbrechenden und nichtunterbrechenden Prioritäten, sog. gemischte Prioritäten. Bei den gemischten Prioritäten kann man zwei große Gruppen unterscheiden, zeitabhängige Prioritäten und Unterbrechungs - Distanz Prioritäten.

Zeitabhängige Prioritäten

Derartige Abfertigungsstrategien treten hauptsächlich bei Teilnehmer-Rechensystemen auf (s.z.B. Herzog, Kühn und Zeh /21/). Sie erfordern einen relativ hohen Aufwand im Betriebssystem für die Warteschlangen - Organisation und werden deshalb vorzugsweise bei Systemen mit wenigen Prioritätsklassen und großer zulässiger Antwortzeit benutzt.

Unterbrechungs - Distanz Prioritäten

Für Prozeß - Steuerungen, die Vermittlung von Daten in einem Datennetz oder für die Behandlung rechnerinterner Prioritäten benützt man sinnvolle Kombinationen sofort unterbrechender und nichtunterbrechender Prioritäten. Die verschiedenartigsten Kombinationen werden vom Verfasser /26/ erstmals einheitlich beschrieben und unter dem Begriff Unterbrechungs - Distanz (UD) Prioritäten zusammengefaßt. Mittlere Wartezeiten wurden für jede Prioritätsklasse bestimmt, wenn die Bedienungszeiten Erlang - k - verteilt sind. In Kapitel VII der vorliegenden Arbeit werden UD-Prioritäten ausführlich beschrieben. Es wird ferner gezeigt, daß auch Systeme mit Bedienungszeiten, die einer "Allgemeinen Erlangverteilung" (s. II. 2.2) gehorchen, mathematisch behandelt werden können.

4. CHARAKTERISTISCHE GRÖSSEN

Ziel der verkehrstheoretischen Untersuchungen ist es, mit Hilfe der Angaben über Ankunfts- und Endeprozesse, Systemstruktur und Abfertigungsstrategie typische Gütemerkmale abzuleiten. Mit Hilfe dieser charakteristischen Größen ist man dann in der Lage, die Effektivität eines Datennetzes oder Teilsystems zu beurteilen bzw. zu verbessern (vgl. Kapitel IV). Die wichtigsten charakteristischen Größen sind im folgenden zusammengestellt.

Durchlaufzeit: Die Durchlaufzeit einer bestimmten Anforderung ist gleich der Summe aller Warte-, Bedienungs- und Transportzeiten, beginnend beim Eintritt der Anforderung in ein System

bis zum Verlassen. Teilweise benutzt man für die Durchlaufzeit auch die Begriffe Antwort- oder Reaktionszeit.

Angaben über die minimale, mittlere und maximale Durchlaufzeit ermöglichen eine gute Abschätzung der Leistungsfähigkeit eines Systems.

Durchsatz: Der mittlere oder maximale Durchsatz ist gleich der mittleren oder maximalen Anzahl von Anforderungen, welche pro Zeiteinheit in einem System transportiert oder bearbeitet wird.

Belastung: Die mittlere Belastung, kurz Belastung genannt, ist die mittlere Anzahl gleichzeitig belegter Bedienungseinheiten oder Transportkanäle.

Wartewahrscheinlichkeit: Wahrscheinlichkeit, daß eine eintreffende Anforderung überhaupt warten muß.

Mittlere Wartezeit: Für die mittlere Zeit die eine eintreffende Anforderung warten muß, werden nahezu immer zwei Werte, angegeben: die mittlere Wartezeit bezogen auf alle Anforderungen oder nur auf jene, die auch tatsächlich warten müssen.

Wartebelastung: Die Wartebelastung, häufig auch als mittlere Warteschlangenlänge bezeichnet, wird definiert als die mittlere Anzahl gleichzeitig auf Bedienung oder Transport wartender Anforderungen.

Unterbrechungswahrscheinlichkeit: Bei Systemen mit Prioritäten ist die Wahrscheinlichkeit, daß eine Anforderung mindestens einmal während der Bedienungs- oder Transportphase unterbrochen wird, von Interesse.

Mittlere Anzahl von Unterbrechungen: Die mittlere Anzahl von Unterbrechungen einer bestimmten Anforderung ist eine wichtige charakteristische Größe. Sie gibt z.B. Auskunft über die zusätzliche Belastung eines Prozessors durch die Unterbrechungsbehandlung.

Blockierungswahrscheinlichkeit: Bei Systemen mit begrenztem Speicherraum (sog. Warte-Verlust-Systeme) oder Systemen ohne Wartemöglichkeit ist die sog. Blockierungswahrscheinlichkeit von Interesse. Sie wird definiert als Wahrscheinlichkeit, daß alle zur Verfügung stehenden Bedienungseinheiten (oder Kanäle) und (wenn vorhanden) alle Warteplätze belegt sind.

Verlust- und Überlaufwahrscheinlichkeit: Wahrscheinlichkeit, daß eine eintreffende Anforderung weder bedient noch gespeichert werden kann, da das System momentan blockiert ist. Im Falle der Verlustwahrscheinlichkeit wird die eintreffende Anforderung vom System abgewiesen, im Falle der Überlaufwahrscheinlichkeit einem anderen System zur Weiterbearbeitung angeboten.

Verteilungsfunktionen: Charakteristische Mittelwerte und Wahrscheinlichkeiten geben zwar eine sehr wesentliche Auskunft über das Verkehrsverhalten eines Systems. In vielen Fällen der Praxis sind diese Angaben ausreichend. Eine allumfassende Auskunft über

die Schwankungen der Größen um die Mittelwerte liefern aber erst die Verteilungsfunktionen. Typische Beispiele sind die Wartezeitverteilungsfunktion, die Verteilungsfunktion der in Speicher wartenden Anforderungen, die Verteilungsfunktion der Blockierungszeiten, etc.

III. STRUKTUR UND BETRIEBSART VON DATENNETZEN

1. ÜBERBLICK UND NOMENKLATUR
2. NETZE
3. VERMITTLUNGSEINRICHTUNGEN
4. ÜBERTRAGUNGSWEGE
5. DATENSTATIONEN
6. EINIGE BEISPIELE REALER DATENNETZE

In diesem Kapitel wird eine Übersicht gegeben über die verschiedenen Netzformen und die verschiedenen Möglichkeiten der Verkehrslenkung. Es wird weiter die Struktur und der Aufbau typischer Datenvermittlungen beschrieben, ferner die Wirkungsweise von Konzentratoren, Multiplexern und anderen Vermittlungseinrichtungen kurz erläutert(entsprechende verkehrstheoretische Probleme s. Kapitel IV). Nach einem Überblick über die heute für den Aufbau von Datennetzen zur Verfügung stehenden Übertragungswege und Datenstationen werden noch einige Beispiele realer Datennetze vorgestellt.

1. ÜBERBLICK UND NOMENKLATUR

In der Literatur gibt es eine Vielzahl von Schlagwörtern und Begriffen, welche dazu benützt werden, die Funktion von Datennetzen zu beschreiben.

Da der Schwerpunkt der hier vorliegenden Arbeit auf der Beschreibung des Verkehrsflusses liegt, werden hier Datennetze nach folgenden zwei Kriterien unterschieden:

. nach der Art der Wegesuche innerhalb des Netzes,
. nach der Art der Vermittlung innerhalb der Netzknoten.

1.1. Unterscheidung nach der Art der Wegesuche

Ungeregelte Netze. Bei den ungeregelten Netzen ist der Weg, den die zu übertragende Nachrichten vom Ausgangsknoten zum Bestimmungsort zurücklegen sollen, fest vorgeschrieben (feste Leitweglenkung). Die Verkehrssituation im Netz wird nicht berücksichtigt: ist der vorgeschriebene Weg momentan blockiert, so müssen die zu übertragenden Nachrichten warten, bis dieser wieder frei ist. Die Vor- und Nachteile ungeregelter Netze werden in Abschnitt III. 2 diskutiert.

Eine Variante ungeregelter Netze sind Netze mit zufälliger Leitweglenkung, d.h. der einzuschlagende Weg wird mit einer bestimmten Wahrscheinlichkeit - unabhängig von der momentanen Verkehrssituation ausgewählt. Eine derart zufällige Leitweglenkung wird in verschiedenen verkehrstheoretischen Arbeiten angenommen, da sich in diesem Fall, ähnlich wie bei fester Leitweglenkung, erhebliche Vereinfachungen bei der mathematischen Untersuchung ergeben.

Geregelte Netze. Hier wird der Weg der zu übertragenden Nachricht nicht fest vorgeschrieben. Je nach momentaner Verkehrssituation entscheidet man, welcher Weg gewählt werden soll. Verschiedene Verfahren der Wegesuche werden in Abschnitt III. 2.2. vorgestellt.

1.2 Unterscheidung nach der Art der Vermittlung

Je nach der Art der Vermittlung der zu übertragenden Nachrichten innerhalb der Datenvermittlung unterscheidet man:

Netze mit Durchschaltevermittlungen. Zunächst wird die gesamte Verbindung von der Datenquelle zur Datensenke über sämtliche Teilstrecken aufgebaut, d.h. zwischen Absender und Empfänger besteht eine direkte Verbindung. Erst dann werden die Daten übertragen.

Netze mit Speichervermittlungen (Teilstreckenvermittlung). Bei Netzen mit Speichervermittlungen erfolgt die Durchschaltung der Leitung und die Übertragung der Nachrichten abschnittsweise von Knoten zu Knoten. Ist der vorgesehene Weg zwischen zwei Knoten blockiert, so werden die Daten solange gespeichert, bis diese Teilstrecke wieder frei ist.

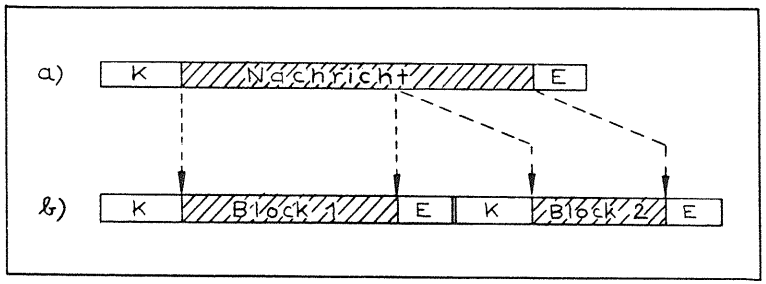


Bild III.1.: Datenformate für "message - switching" (a) und "packet-switching" (b). Im Datenkopf (K) werden Adreß- und Prioritätsangaben sowie Angaben zur Zeichenbehandlung gegeben. Der Datenendeteil (E) zeigt das Ende der Nachricht (bzw. des Blocks) an und enthält Kontrollinformation zur Fehlererkennung bzw. Fehlerkorrektur (s.Neubauer und Polakowski /27/, White /28/, Dell /29/ sowie Chu /30/)

Neben dem seit vielen Jahren bekannten "message-switching" - Verfahren (die gesamte zwischen Quelle und Senke zu übertragende Nachricht wird als ein Block von Daten zusammengefaßt und übertragen) setzt sich in neuerer Zeit immer mehr das "packet-switching" - Verfahren durch. Hier werden die zu übertragenden Nachrichten (z.B. in der Datenvermittlung) in einzelne "Pakete" mit einer maximalen Länge (z.B. 1000 bit) aufgeteilt, getrennt übertragen, und dann in Zielknoten wieder gesammelt (vgl. Bild III.1).

Der wesentliche Vorteil des "packet-switching" gegenüber dem herkömmlichen "message-switching" liegt in der einheitlich vorgeschriebenen Länge der zu übertragenden Datenformate. Die "Datenbehandlung" innerhalb der Datenvermittlung ist deshalb geschickt durchzuführen. Bei "packet-switching" entfällt weiter der Nachteil, daß dringende Nachrichten undefiniert lange warten müssen, wenn gerade eine sehr lange Nachricht geringer Priorität übertragen wird. Die Wartezeit ist durch das feste Format der "Pakete" begrenzt.

Durchschalte - contra Teilstreckenvermittlung. Heftige Diskussionen werden schon seit Jahren geführt über die Vorzüge und Nachteile von Netzen mit Durchschalte - bzw. Teilstreckenvermittlung. In einem CCITT - Vorschlag /31/, von Davies /32/, Gauthier /33/, Holloway und Hart /34/, Martin /5/, Pierce /35/ und vielen anderen Autoren werden Für und Wider beider Prinzipien gegenübergestellt; verkehrstheoretische Betrachtungen siehe auch Closs /36/. Die wesentlichen Argumente wurden in Bild III.2 und III.3 zusammengefaßt.

Betrachtet man die bestehenden großen Datennetze, so erkennt man folgende Tendenz: private und öffentliche Netze, in welchen große Entfernungen zu überbrücken sind, benutzen vorzugsweise das Teilstreckenprinzip; für öffentliche Netze in Europa mit einer großen Teilnehmerdichte und relativ geringen Entfernungen wird das Durchschalteprinzip bevorzugt.

Fast in allen geplanten großen Datennetzen der (nahen und ferneren) Zukunft liegt zwar der Schwerpunkt auf einem Vermittlungsprinzip, die andere Art der Durchschaltung wird jedoch ebenfalls vorgesehen.

AUFWAND FÜR	DURCHSCHALTEVERMITTLUNG	TEILSTRECKENVERMITTLUNG
1. ÜBERTRAGUNGSWEGE	Betrieb als Verlustsystem, deshalb rel. geringe Auslastung möglich (hohe Kosten bei langen Übertragungswegen)	Betrieb als Wartesystem mit hoher Leitungsausnutzung
2. DATENVERMITTLUNGEN	Sende-, Empfangs- und Vermittlungseinrichtungen	Prinzipiell wie Durchschaltevermittlungen, jedoch zusätzlich hoher Aufwand an Speicherkapazität für Zwischenspeicherung
3. VERBINDUNGSAUFBAU	Aufbau der gesamten Verbindung mit Hilfe eines Signalisierungssystems, das nach dem Teilstreckenprinzip arbeitet, anschließend Übertragung	Übertragung der Daten von Knoten zu Knoten
Gesamte Übertragungszeit für eine bestimmte Verbindung	Setzt sich aus einer (nahezu konstanten) Verbindungsaufbauzeit(während der der Übertragungsweg nicht belegt ist!) und der Zeit zur Übertragung der Nachricht zusammen	Setzt sich bei n Knoten aus n zufälligen Wartezeiten (während denen der Übertragungsweg voll benutzt wird!) und n mal der Übertragungszeit der Nachricht zusammen.
kurze Nachrichten	Anteil der Verbindungsaufbauzeit zu groß; schlechte Ausnutzung der Übertragungswege	günstig
lange Nachrichten	günstig	Mehrmaliges Zwischenspeichern von großen Datenmengen ungünstig
Verkehrsspitzen	empfindlich, da Verlustsystem	unempfindlich gegen kurzzeitige Verkehrsspitzen

Bild III.2: Vergleich einiger wesentlicher Argumente, die für das Prinzip der Durchschalte- bzw. Teilstreckenvermittlung sprechen (vgl. Abschnitt III. 1.2.).

AUFWAND FÜR	DURCHSCHALTEVERMITTLUNG	TEILSTRECKENVERMITTLUNG
4. ZUSÄTZLICHE LEISTUNGSMERKMALE		
Möglichkeit für Dialog	automatisch erfüllt	nur näherungsweise erfüllt in Netzen mit sehr kurzer Durchlaufzeit
Prioritäten	ungünstig, da Unterbrechung einer bestehenden Verbindung erforderlich	nahezu beliebig
Mehrfachverteilung von Nachrichten	nur als Konferenzschaltung möglich	beliebig
Nachrichtenformat	beliebig	vorgeschrieben, aber versch. Formate möglich
Verkehr zwischen DEE unterschiedlichen Nachrichtenformats	praktisch nicht möglich	einfach realisierbar
Verkehr zwischen DEE unterschiedlicher Datenrate	praktisch nicht möglich	leicht möglich

Bild III.3: Vergleich einiger Argumente, die für das Prinzip der Durchschalte- bzw. Teilstreckenvermittlung sprechen (vgl. Abschnitt III.1.2)

2. NETZE

2.1. Ungeregelte Netze

Als ungeregelte Netze werden alle Netze definiert, bei welchen der Weg für eine zu übertragende Nachricht fest vorgeschrieben ist (vgl. III.1). Im folgenden werden die vier Grundstrukturen ungeregelter Netze vorgestellt. Vor- und Nachteile werden kurz erläutert.

2.1.1 Maschennetz (Bild III.4):

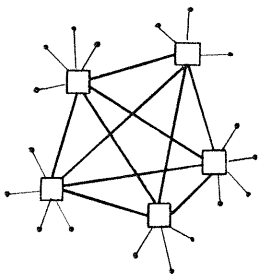
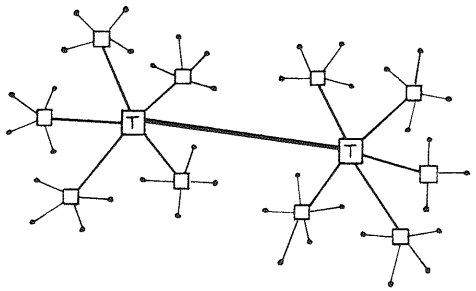


Bild III. 4: Maschennetz

Jeder Knoten ist mit jedem anderen durch einen Übertragungsweg direkt verbunden, d.h. es findet keinerlei Konzentration des Verkehrs im Netz statt. Die Nachrichten werden deshalb auf dem kürzest möglichen Weg übertragen. Ein weiterer Vorteil ist die Unabhängigkeit der einzelnen Richtungen bei Störungen, d.h. der Ausfall eines Knotens bzw. Übertragungsweges beeinflußt nicht den Verkehr im übrigen Netz.

Nachteilig ist, daß die Verkehre zwischen zwei Knoten relativ klein sind. Die Ausnutzung der Übertragungswege ist deshalb gering und die Kosten pro Verkehrseinheit hoch.

2.1.2 Sternnetz (Bild III.5):



Alle Verkehre eines Einzugsbereichs werden in einem sog. Transitknoten T zusammengefaßt und über einen Weiterverkehrsweg zum nächsten Transitknoten übertragen. Erst dort werden die Verkehre zu den verschiedenen Bestimmungsknoten aufgeteilt.

Bild III.5.: Sternnetz

Vorteilhaft ist die hohe Ausnutzung der Übertragungswege und die Ersparnis von Leitungen und Kosten. Ein großer Nachteil ist, daß der Ausfall eines Transitknotens oder eines Weiterverkehrsweges katastrophale Folgen hat, da der Verkehrsfluß eines großen Bereichs blockiert wird. In Sternnetzen sind ferner Umwege bei der Verkehrsführung nicht zu vermeiden.

2.1.3 Ringnetz (Bild III.6):

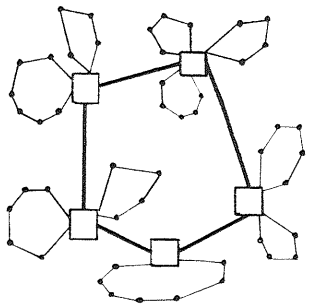


Bild III.6.: Ringnetz

Alle Knoten eines Netzes (und/oder einer Netzebene) sind durch eine Ringleitung verbunden. Die Steuerung des Verkehrsflusses erfolgt nach verschiedenen Strategien (nach Bedarf auf fester oder beliebiger Zeitlage oder nach Aufforderungen durch den "Steuerknoten", vgl. Pierce /35/ und Martin /5/.

Vor- und Nachteile von Netzen mit Ringstruktur werden z.Zt. in der Literatur intensiv diskutiert (s.z.B. Pierce /35/). Die Hauptvorteile sind ein einfacher Steuerungsalgorithmus und deshalb geringe Vermittlungskosten pro Knoten. Ferner ist das Netz leicht erweiterbar durch das Aufschneiden einer Verbindungsleitung und das Einfügen eines neuen Knotens. Nachteilig ist vor allem die mangelhafte Zuverlässigkeit im Falle von Kabelstörungen sowie die relativ schlechte Auslastung der Ringleitung (werden auch Teilnehmer mit sehr geringem Datenvolumen in einen Ring eingeschlossen, so sind die Kosten für die Datenendeinrichtung ziemlich hoch).

2.1.4 Linien - oder Reihennetz (Bild III.7):

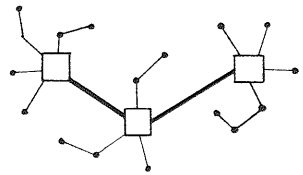


Bild III.7: Linien- oder Reihennetz

Alle Knoten eines Netzes (oder einer Netzebene) sind längs einer Linie angeordnet. Die Vor- und Nachteile entsprechen weitgehend jenen des Ringnetzes (vgl. III.2.1.3). Liniennetze eignen sich deshalb besonders als Sammelleitungen im Anschlußnetz, weniger jedoch für große Netzstrukturen.

2.2. Geregelte Netze

2.2.1 Prinzip

Große Datennetze sind in ihrer Grundstruktur im wesentlichen

Sternnetze. Um die Nachteile des Sternnetzes (vgl. III.2.1.2) zu vermeiden werden zusätzliche Querwege mit Überlaufmöglichkeit auf das Sternnetz eingerichtet. Bild III.8 zeigt ein Beispiel eines derartigen Verbundnetzes: der Hauptteil des Verkehrs (70-90%) von Knoten 1 nach Knoten 3 wird - wie beim Maschensystem - auf einem direkten Weg übertragen. Der restliche Teil wird - wie beim Sternnetz - gemeinsam mit anderen Verkehren auf einem hochbelasteten Weitverkehrsbündel übertragen.

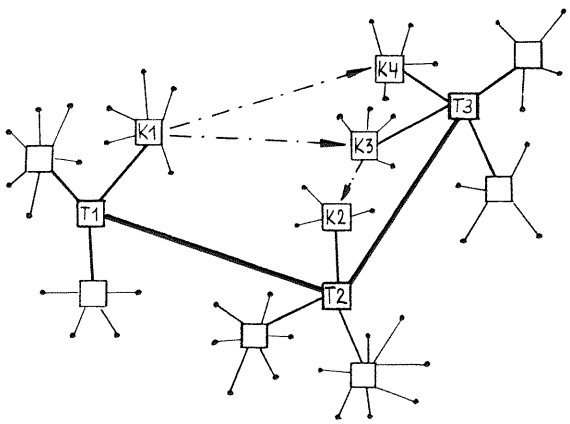


Bild III. 8: Verbundnetz

Zu übertragende Nachrichten von Knoten 1 nach Knoten 3 werden deshalb zuerst dem Direktweg (K1 - K3) angeboten. Bei Blockierung dieses Weges durch eine Warteschlange bestimmter Länge wird der Verkehr eventuell einem zweiten Querweg (K1 - K4 -T3- K3) und dann dem sog. Kennzahlweg (K1-T1 - T2- T3 - K3) angeboten.

Ein Verbundnetz verbindet deshalb die Vorteile der beiden Grundstrukturen Stern- und Maschennetz (hohe Verkehrsgüte und Sicherheit) und vermeidet deren Nachteile weitgehend. Der zusätzliche Steuerungsaufwand im Vergleich zu einem ungeregelten Netz ist

abhängig von der gewählten Art der Leitweglenkung (s. nächster Abschnitt)

2.2.2 Verfahren der verkehrsabhängigen Wegesuche

Es gibt eine Vielzahl mehr oder weniger unterschiedlicher Vorschläge für die verkehrsabhängige Wegesuche (Zusammenstellungen s. Boehm und Mobley /3/, Fultz und Kleinrock /4/, Silk /37/, Prosser /38/, ferner Butrimenko /39/, Brandt und Chrétien /40/, Beeforth, Grimsdahl, Halsall und Woollons /41/, Davies /42/, Frank und Chou /43/, Kahn und Crowther /44/, Lotze /45/, Pedersen und Fu /46/).

Um einen Überblick über die wichtigsten Prinzipien zu geben, werden die Verfahren in drei Klassen eingeteilt: alternative, adaptive und zentrale Leitweglenkung.

2.2.2.1 Alternative Leitweglenkung

Der Verkehr kann alternativ auf mehreren Wegen zum Zielknoten geleitet werden. Nur die Verkehrssituation im betrachteten Knoten wird berücksichtigt.

Bild III.9 zeigt als Beispiel die Leitwegtafel, welche im Vermittlungsrechner des Knotens K 1 (Bild III.8) gespeichert ist und für die Wegesuche jeder zu übertragenden Nachricht benutzt wird: ist der erste Weg blockiert, so wird die Übertragung der Nachrichten auf dem zweiten Weg versucht, usw.

Ist momentan kein geeigneter Weg frei, so wird der "Übertragungswunsch" im Falle von Durchschaltevermittlungen abgelehnt

ZIEL ↓	1. Weg führt über	2. Weg führt über	3. Weg führt über
K2	K3	K4	T1
K3	K3	K4	T1
K4	K4	K3	T1
≡			

Bild III. 9: Beispiel für eine Leitwegtafel bei alternativer Leitweglenkung

(Besetztzeichen), bei Speichervermittlungen in eine Warteschlange eingereiht (vgl. Abschnitt III.3 und Kapitel V).

Der Vorteil alternativer Leitweglenkungsverfahren liegt in dem relativ geringen zusätzlichen Steuerungsaufwand im Vergleich zu ungeregelten Netzen. Der Nachteil ist eine geringe (automatische) Flexibilität bei Störungen.

2.2.2.2 Adaptive Leitweglenkung:

ZIEL ↓	Abgehende Wege in Richtung T1	K3	K4
K2	8 s	20 s	16 s
K3	10 s	18 s	14 s
K4	17 s	27 s	13 s
≡			

Bild III. 10: Beispiel für eine Leitwegtafel bei adaptiver Leitweglenkung

In diesem Fall wird zur Steuerung des Verkehrs auch die Verkehrssituation im Gesamtnetz berücksichtigt: zu bestimmten Zeiten wird die Verkehrsverteilung aller abgehenden Richtungen eines Knotens zu allen Nachbarknoten übertragen. Diese zusätzliche

Information erlaubt eine flexiblere Wegesuche als bei alternativer Leitweglenkung. Das Verfahren ist deshalb verkehrsmäßig noch günstiger als alternative Leitweglenkung, der zusätzliche Steuerungsaufwand jedoch beträchtlich. Bild III.10 zeigt als Beispiel die Leitwegtafel, mit welcher im Knoten K 1 (Bild III.8) die Verkehrslenkung erfolgt: gespeichert sind Schätzwerte für die momentan zu erwartenden Wartezeiten auf den verschiedenen abgehenden Wegen für bestimmte Zielknoten.

2.2.2.3 Zentrale Leitweglenkung:

Eine logische Fortsetzung dieser Wegesuchstrategien wäre ein zentral gesteuertes Gesamtnetz, bei dem für jede Nachricht der absolut optimale Weg bestimmt wird. Der Steuerungsaufwand wird jedoch in diesem Fall derartig groß, daß diese Lösung für große nationale, internationale und interkontinentale Netze in den nächsten Jahren sicher nicht in Frage kommt.

3. VERMITTLUNGSEINRICHTUNGEN

In diesem Abschnitt wird die Struktur und der Aufbau typischer Datenvermittlungen skizziert, ferner die Wirkungsweise von wichtigen Komponenten des Anschlußnetzes erläutert. Detaillierte Beschreibungen findet man beispielsweise in den Veröffentlichungen über Projekte (s. Abschnitt III.6, s. ferner auch Schomburg /47/.). Verkehrstheoretische Probleme werden im nächsten Kapitel (Abschnitt IV. 3.1 und IV. 3.2) diskutiert.
Die zunehmende Zentralisierung und Spezialisierung einzelner Systemeinheiten bei vollelektronischen Vermittlungseinrichtungen

erhöht die Leistungsfähigkeit und Flexibilität. Auf der anderen Seite hat aber der Ausfall einer zentralen Einheit katastrophale Folgen. Notwendige Gegenmaßnahmen (hardware- und software-Kontrollen, Duplizieren von wichtigen Systemeinheiten, etc.) können hier nicht diskutiert werden, auf ihre große Bedeutung sei jedoch hingewiesen (s.z.B. Martin /6,48/, Stimler /49/).

3.1. Datenvermittlungen

3.1.1 Durchschaltevermittlungen

Drei wesentliche Komponenten charakterisieren den Aufbau einer Durchschaltevermittlung (vgl. Bild III. 11):

- die Leitungsanschluß-Einheit (LA),
- die Vermittlungs-Einheit (V), und
- die Steuer-Einheit (St)

Aufgabe der LA-Einheit ist der Empfang und das Senden von Codezeichen, Serien/Parallel bzw. Parallel/Serien-Wandlung sowie das Puffern von Zeichen (Die Begriffe "Senden" und "Empfang" fassen zahlreiche Einzelaufgaben zusammen, so z.B. das Abtasten der Leitungen oder die Anzeige von Polaritätswechseln, das Regenerieren von Impulsen, Synchronisation usw.). Moderne Datennetze erlauben nicht nur die Übertragung von Nachrichten mit einheitlicher Geschwindigkeit, sondern ermöglichen - den Bedürfnissen der verschiedenartigen Benutzer entsprechend - die Auswahl einer geeigneten Übertragungsgeschwindigkeit (s. z.B. Atkin /50/, Gabler und Staudinger /51/.). In diesem Fall

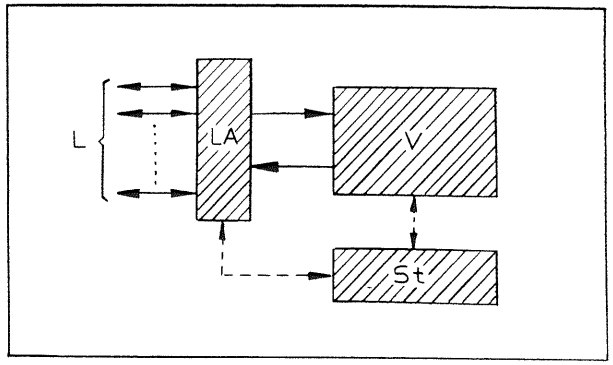


Bild III.11: Struktur einer Durchschaltevermittlung (Die Abkürzungen bedeuten: L=Leitungen, LA=Leitungsanschluß-Einheit, V=Vermittlungs-Einheit, St=Steuer-Einheit, Wege zur Übertragung von Nachrichten, --- Wege zur Übertragung von Steuerinformationen)

übernimmt die LA-Einheit oft noch das Umsetzen auf eine einheitliche Übertragungsgeschwindigkeit zum Vermitteln innerhalb der Durchschaltevermittlung (Blackburn und Muench /52/).

Aufgabe der Vermittlungs-Einheit (V) ist die Vermittlung eintreffender Nachrichten in eine gewünschte Richtung. Diese erfolgt entweder im Raummultiplex (s.z.B. Melis und Verschueren /53/), im Zeitmultiplex (Adam und Orbell /54/), im Adress-Code-Multiplex (Schaffer /55/) oder einer Kombination dieser Möglichkeiten (Fick und Kammerl /56/, Schaffer /55/). Der Verbindungs- Auf- bzw. Abbau (Auswertung der Wählinformationen, Leitweglenkung, Prüfen der Berechtigung, usw) wird durch die Steuer-Einheit kontrolliert. Während früher festprogrammierte Steuer-Einheiten benutzt wurden setzen sich heute immer mehr die flexibleren speicherprogrammierten Steuerungen durch.

Alle handelsüblichen Durchschaltevermittlungen lassen sich durch die hier beschriebene Struktur darstellen. Erhebliche Unterschiede treten jedoch auf bei der Ausführung der einzelnen Funktionseinheiten durch "hardware" und "software" (s. auch Abschnitt III.3.1.3, ferner die Literaturangaben zu Abschnitt III.6).

3.1.2 Speichervermittlungen

Halbelektronische Speichervermittlungen

Bei modernen halbelektronischen Speichervermittlungen findet man eine den Durchschaltevermittlungen sehr verwandte Struktur (vgl. Bild III.12, ferner Austin und Huck /57/, Störmer /58/): Die eintreffenden Nachrichten werden von der Leitungsanschluß-Einheit (LA) empfangen und dann in einer Speichereinheit (z.B. mehrere kleine Kernspeicher, vgl. Austin und Huck /57/) gespeichert. Ist die gesamte Nachricht eingetroffen,

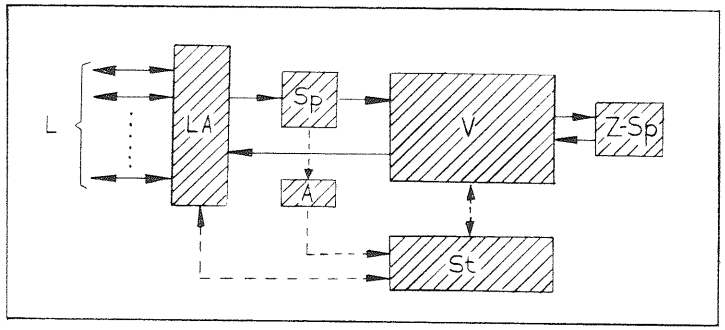


Bild III.12: Struktur einer halbelektronischen Speichervermittlung (L=Leitungen, LA=Leitungsanschluß-Einheit, Sp=Speicher-Einheit, V=Vermittlungs-Einheit, Z-Sp= Zwischenspeicher-Einheit, A=Auswerte-Einheit, St=Steuer-Einheit)

so erkennt eine spezielle Schaltung der Speicher-Einheit das "Nachrichtenende" und veranlaßt das Abschalten der Leitung. Eine Auswerte-Einheit (A) übernimmt die Auswertung des Nachrichtenkopfes (Aufbau einer Nachricht vgl. Abschnitt III.1) und gibt die entsprechenden Informationen an die Steuer-Einheit (St). Die anschließende Übertragung dieser Nachricht von der Speicher-Einheit zu einer bestimmten abgehenden Leitung erfolgt über die von der Steuer-Einheit überwachte Vermittlungs-Einheit (Raummultiplex mit Wählern, Schutzgas-Kontakten, etc). Ist die abgehende Leitung momentan belegt, so erfolgt eine Zwischenspeicherung in einer Zwischenspeicher-Einheit (Band- oder Trommelspeicher).

Vollelektronische Speichervermittlungen

In Verbindung mit den enormen Fortschritten auf dem Gebiet elektronischer Rechenanlagen hat sich auch bei Speichervermittlungen die vollelektronische Technik weitgehend durchgesetzt. Bild III.13 zeigt die Struktur einer derartigen Vermittlung,

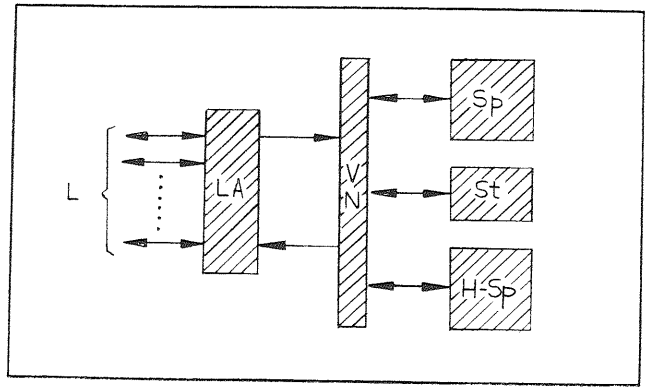


Bild III.13: Struktur einer vollelektronischen Speichervermittlung (L=Leitungen, LA=Leitungsanschluß-Einheit, VN=Verbindungs-Netz, Sp=Speicher-Einheit, St=Steuer-Einheit, H-Sp= Hintergrundspeicher)

welche - von der Grundkonzeption her - sehr weitgehend einer modernen Multiprogramm - Rechenanlage entspricht (vgl. z.B. Martin /6/, Stimler /49/, Swoboda /59/):
Wieder empfängt oder sendet, puffert und wandelt eine Leitungsanschluß-Einheit (LA) die zu übertragenden Nachrichten. Wichtige Aufgaben sind auch die redundante Codierung beim Senden sowie die Fehlerprüfung und Quittungsgabe beim Empfang von Zeichen (bei Durchschaltevermittlungen übernehmen normalerweise Nachrichten-Sender und - Empfänger diese Aufgaben).

Verbindungsnetz, Speicherhierarchie

Die Übernahme der Nachrichten von der Leitungsanschluß-Einheit (LA) in den Speicherbereich (Sp, H-Sp) erfolgt über das Verbindungsnetz (VN), ein sog. "Bus-System", das den Datenaustausch zwischen allen System-Einheiten steuert (s. z.B. Becker und Aupperle /60/). Der Speicherbereich umfaßt dabei neben dem modular aufgebauten Hauptspeicher (Sp) auch noch eine Anzahl von Hintergrund-Speichern (H-Sp) (Speicherhierarchie aus Kern-, Trommel-, Platten- und Bandspeichern, vgl. Martin /6/, Gunderson /61/).

Ist die zu übertragende Nachricht im Speicherbereich gespeichert, so wird in einer Eingangs-Warteschlange ein Indikator gesetzt. Dieser Indikator zeigt der Steuer-Einheit (St) an, daß in einem bestimmten Bereich des Speichers eine eingetroffene Nachricht auf die Übertragung zur nächsten Datenvermittlung (bzw. zu einer Datenstation) wartet.

Steuer-Einheit, Systemprogramme

Drei Arten von Systemprogrammen kennzeichnen die vielfältigen Aufgaben der Steuer-Einheit:

- das Hauptprogramm
- Programme zur Nachrichtenbehandlung, sowie
- Steuer - und Kontrollprogramme.

Das Hauptprogramm steuert den gesamten Betriebsablauf, d.h. es bearbeitet Unterbrechungsanforderungen und ruft die einzelnen Programme auf. Typische Aufgaben für Programme zur Nachrichtenbehandlung sind das Bearbeiten von Eingangswarteschlangen, die Leitweglenkung, das Setzen von Indikatoren in Ausgangswarteschlangen, aber auch die Zuordnung besonderer Leistungsmerkmale und Kostenabrechnungen. Die dritte Gruppe (Steuer- und Kontrollprogramme) dient zur Steuerung und Überwachung der verschiedenen Systemeinheiten.

Ausführliche Beschreibung von Systemprogrammen siehe z.B. Martin /48/ sowie Jilek und Jokay /62/.

3.1.3 Erhöhung der Leistungsfähigkeit

Die Leistungsfähigkeit, Flexibilität und Wirtschaftlichkeit einer Datenvermittlung (Durchschalte - oder Speichervermittlung) ist sehr weitgehend davon abhängig, wo die Schnittstelle zwischen "hardware" und "software" liegt. Frühe Entwürfe vollelektroni-

scher Datenvermittlungen zeigten die Tendenz, möglichst viele Funktionen durch Programme abzuwickeln. Eine Folge davon war die Überlastung der Steuereinheit und deshalb eine Überlastung des gesamten Systems.

Die Entlastung der Steuereinheit ist durch folgende Maßnahmen möglich:

a) Übernahme verschiedener Aufgaben durch die Leitungsanschluß-Einheit (LA), indem

- viele verhältnismäßig einfachen und immer wiederkehrenden Aufgaben der Zeichenbehandlung durch "hardware" ausgeführt werden

- bei sehr großen Datenvermittlungen die Funktion der Leitungsanschluß-Einheit durch einen Prozeßrechner mit speziellen LA-Komponenten übernommen wird. In diesem Fall können deshalb bereits auf dieser Ebene "software"-Aufgaben durchgeführt werden (vgl. z.B. Davies /63/, Martin /5/).

b) System-Einheiten möglichst weitgehend autonom, so daß beispielsweise der Datentransfer zwischen Leitungsanschluß-Einheit und Speichereinheit ohne Überwachung durch die Steuereinheit erfolgen kann (direkter Speicherzugriff, vgl. z.B. Becker und Aupperle /60/, Duke und Mc Intyre /64/).

c) Mehrere Systemeinheiten desselben Typs arbeiten parallel nach einem "load sharing"-Verfahren (Martin /6,48/).

3.2. Vermittlungstechnische Einrichtungen für das Anschlußnetz

3.2.1. Überblick

Bild III. 14 zeigt einige typische Formen von Anschlußnetzen: bei den Beispielen 1 bis 3 sind die Teilnehmerstationen entweder über Einzel-, Sammel- oder Ringleitungen mit der Datenvermittlung (Durchschalte- oder Speicherverm.) verbunden. In allen drei Fällen sind für die Verkehrslenkung mehr oder weniger aufwendige Schaltungen direkt in der Teilnehmerstation notwendig (s. z.B. Martin /5/, Pierce /35/).

Um hohe Leitungskosten zu vermeiden werden häufig auch Multiplexer oder Konzentratoren - wie in Beispiel 4 eingesetzt: bereits im Vorfeld findet eine starke Verkehrskonzentration statt, die

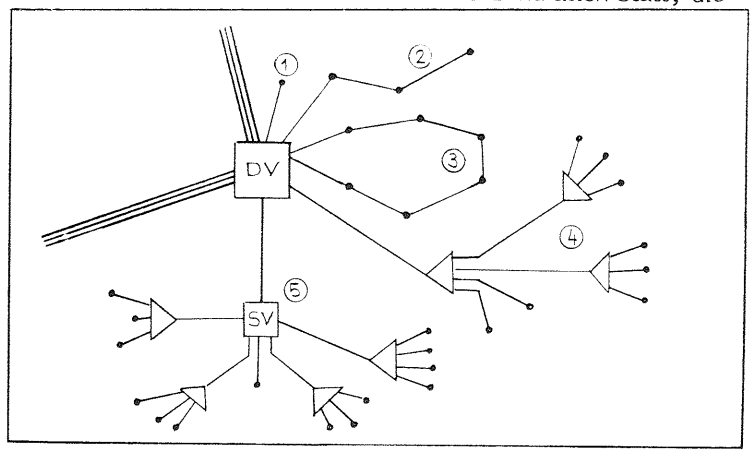


Bild III.14.: Fünf verschiedene Beispiele für den Anschluß der Teilnehmerstationen an eine Datenvermittlung (• = Teilnehmerstationen, △ Multiplexer oder Konzentratoren, SV = Satellitenvermittlung, DV = Datenvermittlung, vgl. Text)

Anschlußleitungen werden deshalb gut ausgenutzt. Ein Multiplexer gruppiert mehrere Datenströme geringer Geschwindigkeit zu einem gemeinsamen Strom hoher Geschwindigkeit zusammen ohne Änderung der Datenstruktur. Ein Konzentrator dagegen empfängt Daten, kann u.U. auch größere Datenmengen zwischenspeichern, umcodieren und umstrukturieren sowie Aufgaben der Fehlerkontrolle übernehmen. Im Gegensatz zum Multiplexer ist in diesem Falle also auch eine Datenreduktion möglich.

Beispiel 5 zeigt eine weitere Möglichkeit der Verkehrskonzentration im Anschlußnetz: zur besseren Auslastung der Anschlußleitungen, vor allem aber auch zur Entlastung der zentralen Datenvermittlung wird der Datenstrom nicht nur über Multiplexer und Konzentratoren geführt, sondern zusätzlich über eine Satelliten-Vermittlung. Diese Anschlußvermittlung besteht üblicherweise aus einer kleinen oder mittleren Prozeßrechenanlage. Neben den Aufgaben des Konzentrators werden auch gewisse Teilaufgaben der Datenvermittlung (insbesondere der Leitungsanschluß-Einheit!) übernommen. In verkehrsarmen Zeiten kann die Auslastung der Rechenanlage durch sog. Hintergrundarbeit erhöht werden (s.z.B. Duke und Mc Intyre /64/, Martin /5/).

3.2.2. Auswahl der geeigneten Struktur

Die im obigen Abschnitt vorgestellten Beispiele geben einen knappen Überblick über die zahlreichen vermittlungstechnischen Einrichtungen des Anschlußnetzes. Die Auswahl der günstigsten Konfiguration des Anschlußnetzes hängt von zahlreichen Komponenten

ab, vor allem aber von

- den Anforderungen der Teilnehmer,
- dem Verkehrsvolumen,
- der geographischen Lage der Teilnehmer sowie
- den Kosten für Datenstationen, Vermittlungseinrichtungen und Übertragungswege.

(vgl. Abschnitt IV.4: verkehrstheoretische Probleme bei der Untersuchung von Teilsystemen).

4. ÜBERTRAGUNGSWEGE

Dieser Abschnitt gibt einen kurzen Überblick über die in der BRD zur Verfügung stehenden Übertragungswege und ihre wesentlichen Eigenschaften.

Da nicht auf übertragungstechnische Gesichtspunkte eingegangen werden kann, sei insbesondere auf folgende einführende Literatur hingewiesen:

Tietz /65/ gibt eine kurze Zusammenfassung und Definition der wichtigsten Betriebsweisen und Verfahren der Datenübertragung. Die übertragungstechnischen Grundlagen werden von Kaiser /66/ behandelt. Kaiser, Lang und Ohnsorge /67/, Voss /68/ und Haas /69/ geben einen Überblick über die technischen Möglichkeiten schneller Datenübertragung.

Zur Mehrfachausnutzung von Übertragungswegen wird heute

Frequenzmultiplex bevorzugt, in Zukunft werden jedoch Zeitmultiplexverfahren eine größere Bedeutung erhalten (s.z.B. Irmer /70/, Kaiser /71/, Kwan /72/, Poschenrieder /73/ und Setzer /74/.

4.1. Allgemeines

Nach dem deutschen Fernmelderecht muß man für alle Datenübertragungen die über ein privates Grundstück hinausgehen, die "Allgemeinen Übertragungswege" der Deutschen Bundespost (DBP) benutzen.

Für den Aufbau eines Datenübertragungssystems gibt es deshalb in derBRD folgende zwei Möglichkeiten:

a) Integration in ein bestehendes öffentliches Netz (Telex-, Datex-, Fernsprechnetz),

b) Mieten sog. überlassener Leitungen bei der DBP (Telegrafenleitungen, Fernsprechstromwege, Breitbandübertragungswege).

Die Entscheidung über die Auswahl der günstigsten Möglichkeit ist abhängig von den

- Anforderungen an die Übertragungsgeschwindigkeit, von den
- Anforderungen an die Übertragungssicherheit, von der
- zulässigen Systemantwortzeit (Verbindungs- auf und abbauzeiten spielen hier eine entscheidende Rolle), und von den
- Kosten für den Übertragungsweg (reicht das zu übertragende Datenvolumen aus, um eine Leitung zu mieten?).

Im nächsten Abschnitt werden die zur Verfügung stehenden Übertragungswege kurz miteinander verglichen, Angaben über die kostengünstige Auswahl eines Übertragungsweges findet man beispielsweise bei Tietz /75/ und Vetter /76/ (vgl. auch Abschnitt IV.4.).

4.2. Übertragungswege

Bild III.15 gibt eine kurze tabellarische Zusammenstellung der in der BRD zur Verfügung stehenden Übertragungswege:

Mit dem Telex-Netz steht uns heute ein weltweites Netz (ca. 400.000 Teilnehmer) zur Verfügung, das es erlaubt, Daten in alle Welt zu übertragen. Die Übertragungsgeschwindigkeit von 50 bit/s sowie die geringe Bitfehlerwahrscheinlichkeit sind für viele Anwendungsfälle ausreichend (Reisebüros, Handel, Banken etc.).

Für Verbindungen innerhalb Deutschlands sowie für die Datenübertragung nach Frankreich erlaubt das Datex-Netz einen Datenaustausch mit einer Geschwindigkeit von 200 bit/s.

Beide Wählnetze haben den Nachteil einer relativ geringen Übertragungsgeschwindigkeit und eines sehr langsamen Verbindungsaufbaus (Wählervermittlungen mit elektromechanischen Wählern). Die DBP plant deshalb den Aufbau eines "Integrierten digitalen Datenübertragungsnetzes" mit verschiedenen Übertragungsgeschwindigkeiten (50, 200, 2400, 9600, 48000 bit/s) und kurzer Verbindungsaufbauzeit ($\lesssim$ 0,1 Sekunden; elektronisches Datenvermittlungssystem EDS). Einen umfassenden Überblick über den technischen Aufbau des Netzes und der Vermi ttlungen sowie zahlreiche weiterführende Literaturangaben findet man bei Gabler und Staudinger /51/ sowie im

Literaturverzeichnis zu Abschnitt III.6.

Ab 1973 sollen jährlich 3 bis 5 EDS - Vermittlungen in Betrieb genommen werden und die bestehenden Vermittlungen des Datex - Netzes ersetzen, ferner soll das bestehende Telex - Netz ebenfalls in das EDS - Netz integriert werden. Etwa 1980 soll der Ausbau des Netzes abgeschlossen sein. Es wird daran gedacht, einen Teil

Art des Übertragungsweges	Übertragungsgeschwindigkeit bit/s	Bitfehlerwahrscheinlichkeit (mittlere)
1) Telex-Netz	50	$5...10 \cdot 10^{-6}$
2) Datex-Netz	200	$2... 8 \cdot 10^{-6}$
3) Integriertes digitales Datenübertragungsnetz	50, 200, 2400 9600, 48000	?
4) Öffentliches Fernsprechnetz	600, 1200 (evtl. 2400)	$1... 10 \cdot 10^{-5}$
5) Überlassene		
Telex-	50	$2... 20 \cdot 10^{-7}$
Datex-	200	$1....10 \cdot 10^{-7}$
Fernsprech-	$\leqslant$4800	$1.. 10 \cdot 10^{-6}$
Breitbandwege	$\geqslant$48000	$\approx 1 \cdot 10^{-6}$

Bild III.15.: Tabellarische Zusammenstellung der Übertragungswege. Das unter Punkt 3 aufgeführte Netz wird ab 1973 aufgebaut

des stark anwachsenden "schnellen" Datenverkehrs (2400-48000 bit/s) in der Zwischenzeit über ein (zunächst handvermitteltes) öffentliches Breitband-Datennetz abzuwickeln (s. Gabler /77/ und Tietz /75/).

Bereits heute können über das öffentliche Fernsprechnetz Daten mit einer Geschwindigkeit von 600 und 1200 (teilweise 2400) bit/s übertragen werden. Erhebliche Verbindungsaufbauzeiten und relativ grosse Bitfehlerwahrscheinlichkeiten sind jedoch - vor allem auch bei internationalen Verbindungen - nicht zu vermeiden. Durch die hohe Teilnehmerdichte und kostengünstige Lösungsmöglichkeiten für Datenstationen mit sehr geringem Datenvolumen wird das Fernsprechnetz auch in Zukunft für die Datenübertragung von grossem Interesse sein (s. z.B. Bocker /79,80/, Goslau in /80,81/, Kaiser /71/, Oden /82/, Tietz /75/ und Vollmeyer /2/).

Überlassene Leitungen (Telex-, Datex - und Fernsprechleitungen sowie Breitbandübertragungswege) werden heute vor allem aus folgenden zwei Gründen eingesetzt:

- hohes Datenvolumen, so daß die Kosten geringer sind als in einem öffentlichen Netz (s.z.B. Tietz /75/ und Vetter /76/),

- kurze Verbindungsaufbauzeiten (fest geschaltete Einzelleitungen oder Netze, in welchen der Verkehr durch private Datenvermittlungen gelenkt wird; vgl. Schreiber /83/)

Überlassene Leitungen haben zwar prinzipiell denselben Aufbau wie Verbindungen der Wählnetze; da die einzelnen Leitungsabschnitte jedoch nicht über Wählvermittlungen des öffentlichen Netzes verbunden sind, ist die Fehlerhäufigkeit wesentlich geringer als bei öffentlichen Netzen.

5. DATENSTATIONEN

Über die Vielzahl der verschiedenartigen Datenstationen (Daten-

endeeinrichtungen und Datenübertragungseinrichtungen) sei nur ein kurzer Überblick gegeben. Datenstationen können aus einer einfachen Eingabetastatur, einem Fernschreiber oder einem Datensichtgerät bestehen; kleinere Prozessrechner sind oft über ein Netz mit Großrechenanlagen verbunden, verschiedene Großrechenanlagen tauschen Programme und Daten untereinander aus.

Die Auswahl einer geeigneten Datenstation hängt nicht nur von Art und Umfang der vorgesehenen Anwendung ab. Einen entscheidenden Einfluß hat auch die Struktur und Betriebsart des Netzes, mit welchem die Datenstation verbunden werden soll.

Aufgabe und Funktionsweise verschiedener Datenstationen werden u.a. von Letsche /84/ und Tietz /75/ beschrieben. Ein Ansatz zur Klassifizierung von Datenstationen wird von Schreiber /83/ gemacht, s. ferner auch Tietz /85/. Wichtige Entwicklungstendenzen auf dem Gebiet der "Endgeräte an Datennetzen" wurden bei einer NTG - Diskussionstagung /86/ aufgezeigt.

6. EINIGE BEISPIELE REALER DATENNETZE

In den vorhergehenden Abschnitten wurde der Schwerpunkt auf die prinzipielle und systematische Beschreibung der Struktur und Betriebsart von Datennetzen gelegt. Einige Beispiele bestehender und geplanter Datennetze sollen nun einen Überblick geben über den derzeitigen Stand sowie die zu erwartende Entwicklung in den nächsten Jahren.

Die zitierten Literaturstellen geben einen Überblick über den Gesamtaufbau der Netze. Weitere zusammenfassende Aufsätze über Erfahrungen mit bestehenden Netzen sowie über Pläne für die

Zukunft findet man in dem "state of the art report: computer networks" /87/. Detaillierte Angaben über Planung und Aufbau einzelner Netze sind in den zusätzlichen Veröffentlichungen des Schrifttumsverzeichnisses zu finden.

6.1. BESTEHENDE NETZE FÜR DIE ÜBERTRAGUNG VON DATEN

Die bestehenden öffentlichen Netze (Telex-,Datex-,Fernsprechnetz; vgl. Abschnitt III.4) erlauben eine vollautomatische Datenübertragung mit nur geringer Übertragungsgeschwindigkeit. In einigen Ländern stehen außerdem handvermittelte Breitbandwege zur Verfügung (s. z.B. Becker und Aupperle /60/, Gibson und Millard /88/).

Netze mit großer Ausdehnung und hoher Übertragungsgeschwindigkeit werden heute vor allem von privaten Gesellschaften eingesetzt. Zwei Beispiele sollen stellvertretend für die Vielzahl privater Datennetze genannt werden, das ARPA - und das S.I.-T.A.-Netz.

Das ARPA - Netz

Das ARPA-Netz (Advanced Research Projects Agency) verbindet (zur Zeit) 40 Universitäten der U.S.A. miteinander und ermöglicht den Austausch von Programmen und Daten zwischen den verschiedenen Universitätsrechenanlagen. Die Übertragung der Daten erfolgt in Paketen (vgl. Abschnitt III.1.2) über ein 50 k Baud-Netz mit Speichervermittlungen (Interface Message Processor). Roberts und Wessler /89/ sowie Heart in /87/.

Das SITA - Netz

Die S.I.T.A. (Société Internationale de Télécommunications Aéronautiques) betreibt ein weltweites Datennetz, über welches der Datenverkehr von mehr als 150 Fluggesellschaften abgewickelt wird.

Alle Datenstationen (Flugbuchungspulte , Fernschreiber etc.) werden über weitverzweigte Anschlußnetze mit Multiplexern und Satelliten - Vermittlungen zusammengefaßt und mit elf großen Datenvermittlungen verbunden. Diese elf Speichervermittlungen in New York, Paris, Frankfurt, Hongkong und anderen Städten lenken den Verkehr in einem 2,4(4,8) kBaud-Vermittlungsnetz. Vollelektronische Satelliten - und Datenvermittlungen ermöglichen auch in diesem weltweiten Datennetz die für Platzbuchungsdienste geforderten kurzen Antwortzeiten. SITA-Report /90/ sowie Brandt Chrétien /40/.

6.2. Projektierte Datennetze

Pläne für schnelle Datenübertragungsnetze existieren in zahlreichen Ländern. Einige Projekte, die in relativ kurzer Zeit schon betriebsfähig sein sollen, sind in Bild III.16 zusammengestellt. Bei den Zeitangaben ist zu berücksichtigen, daß verschiedene Netze zunächst handvermittelt und in erst späteren Jahren vollautomatisch betrieben werden. Fast alle Netze werden zunächst entweder als reine Durchschalte - oder reine Speichervermittlungsnetze betrieben, sehen aber für einen späteren Zeitpunkt gemischten Betrieb vor.

Projekt	Eigentümer	Erster Einsatz ab	Angebotene Geschwindigkeits-Klassen kbit/s	Literatur, Bemerkungen
DATRAN	Data Transmission Company, U.S.A.	1974	0,15/4,8/9,6/14,4	Fisher und Sligh /91/
Bell System's Digital Data System	AT & T, U.S.A.	1974	2,4/4,8/9,6/56	Blackburn und Muench /52/
Datennetz mit EDS-Datenvermittlungen	Western Union, U.S.A.	?	s. unter EDS	Gabler u. Staudinger /51/
TCTS-Digital Data Network	Trans-Canada Telephone Systems, Can.	1972	2,4/4,8/9,6	Bender u. Riml /92/
Synchronous Data Network	Britische Postverwaltung	Versuch ab 1974	0,6/2,4/9,6/48	Dell /29/
CADUCÉE	Französische Postverwaltung	1972	0,6/1,2/4,8	Jousset und Profit /93/ Übergangslösung bis zum Einsatz des Datennetzes HERMES (Despres /94/)
Public Data Network	Schwedische Postverwaltung	?	≥ 0,3/0,6/1,2/2,4/4,8/9,6	Lindberg /95/
Datennetz mit dem elektronischen Datenvermittlungssystem (EDS)	Deutsche Bundespost	1973	0,05/0,2/2,4/9,6/(48)	Gabler und Staudinger /51/

Bild III.16: Einige Beispiele für große öffentliche Datennetze, die in diesem Jahrzent noch zum Einsatz kommen sollen (vgl. auch Abschnitt III.6.2)

IV. VERKEHRSTHEORETISCHE PROBLEME BEI DATENNETZEN UND IHRE BEHANDLUNG

1. AUFGABENSTELLUNG
2. DER WEG ZUR OPTIMIERUNG
3. VERKEHRSTHEORETISCHE PROBL. BEI DER UNTERS. VON TEILSYSTEMEN
4. METHODEN ZUR BEHANDLUNG VERKEHRSTHEOR. PROBLEME

In einem kurzen Abschnitt wird das Hauptproblem aller verkehrstheoretischen Untersuchungen in Datennetzen definiert, ein zweiter Abschnitt zeigt den prinzipiellen Lösungsweg. Die Vielzahl der verschiedenartigen Einzelprobleme wird - entsprechend der Aufteilung eines Datennetzes - in drei großen Gruppen zusammengefaßt und diskutiert. Hinweise auf Lösungsansätze werden - wenn vorhanden - gegeben. Ein letzter Abschnitt gibt einen Überblick über die verschiedenen Methoden zur Behandlung verkehrstheoretischer Probleme. Es werden ferner einige grundlegende Annahmen und Ergebnisse zusammengefaßt, die häufig angewendet werden.

1. AUFGABENSTELLUNG

Hauptaufgabe der Verkehrstheorie ist die Untersuchung des Verkehrsflusses in Datennetzen, die Definition charakteristischer Gütemerkmale sowie das Aufdecken von systeminternen Engpässen. Ziel sämtlicher Untersuchungen ist es, wirtschaftlich optimale Netzformen und Betriebsarten zu finden

- für eine bestimmte vorgeschriebene Verkehrsgüte,
- sowie vorgeschriebene Sicherheit im Falle von Überlastung oder Ausfall von Netzknoten und Bündeln.

(in einigen Arbeiten findet man eine etwas abgewandelte Zielsetzung: optimale Verkehrsgüte bei vorgeschriebenen Kosten).

2. DER WEG ZUR OPTIMIERUNG

Bild IV.1 gibt einen Überblick über die fünf wesentlichen Schritte,

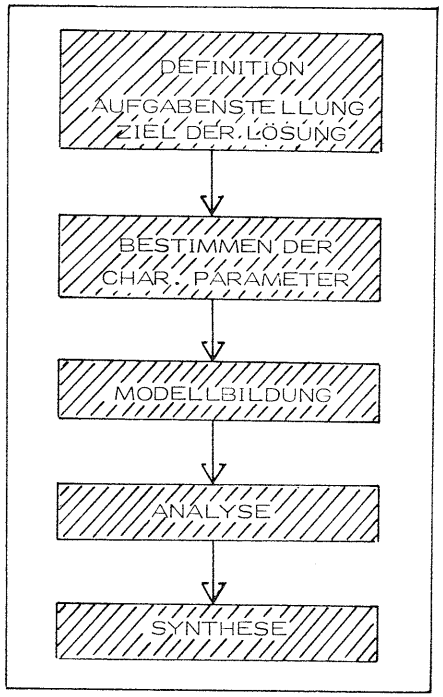


Bild IV.1.: Der Weg zur Optimierung

die zur Lösung eines Optimierungsproblems notwendig sind. An erster Stelle ist die Aufgabenstellung sowie das Ziel der Problem-

lösung zu präzisieren, d.h. es muß festgelegt werden, welche Anforderungen das System in bezug auf Verkehrsgüte und Sicherheit erfüllen soll. Bereits hier sind gewisse Vorentscheidungen über die Betriebsart des Netzes (ungeregelt oder geregelt, alternative oder adaptive Leitweglenkung, usw.) zu treffen. In einem zweiten Schritt sind die charakteristischen Parameter der Verkehre, Vermittlungseinrichtungen und Übertragungswege festzustellen.

Mit Hilfe dieser Daten ist man dann in der Lage, Struktur und Verkehrsmodelle des realen Systems zu entwerfen.

Die bei der Analyse erzielten Ergebnisse werden entscheidend von den gewählten Verkehrs - und Strukturmodellen beeinflußt: je genauer die Modelle den tatsächlichen Verhältnissen entsprechen, desto besser sind die Ergebnisse, verglichen mit der Wirklichkeit.Eine sorgfältige Modellbildung ist deshalb besonders wichtig. Bei der Analyse komplexer Systeme wird man zunächst Simulationsmethoden anwenden und dann anschließend (oder parallel) rechnerische Verfahren entwickeln.

Schließlich erfolgt in einem letzten Schritt die Synthese optimaler Strukturen, die die Grundlage für den wirtschaftlichen Einsatz von Datennetzen liefert.

Die Optimierung großer Datennetze ist mit den heute zur Verfügung stehenden mathematischen Hilfsmitteln in einem Schritt nicht möglich. Deshalb untersucht und optimiert man zuerst Teilsysteme und bestimmt das Gesamtoptimum durch Iteration (vgl. Herzog /130/).

3. VERKEHRSTHEORETISCHE PROBLEME BEI DER UNTERSUCHUNG VON TEILSYSTEMEN

Bei der Beschreibung von Datennetzen haben wir drei große Teilbereiche, das Anschlußnetz, die Datenvermittlungen und das Vermittlungsnetz unterschieden. Ein analoges Vorgehen bietet sich bei der Klassifizierung verkehrstheoretischer Probleme an. Aufgabenstellungen, die durch zwei oder alle drei Teilbereiche beeinflußt werden, werden entsprechend dem Schwerpunkt des Problems eingeordnet.

3.1. Anschlußnetz

Um die Frage nach der optimalen Struktur eines Anschlußnetzes bearbeiten zu können, sind zunächst zahlreiche Einzelfragen zu beantworten. Typische Beispiele sind:

- Leitungsausnutzung und Antwortzeiten bei verschiedenen Netzstrukturen (Einzel-, Sammel-, Ringleitungen etc.)
- Einfluß von Prioritäten auf die Wartezeiten für Anforderungen verschiedener Dringlichkeit.
- Beeinflussung der Verkehrsgüte durch Verkehrsquellen (Datenstationen) unterschiedlicher Intensität und Verkehrsart.
- Wirksamkeit von Multiplexern, Konzentratoren und Satellitenvermittlungen.

. Individuelle Zwischenspeicher in den Datenstationen oder gemeinsame Pufferspeicher, z.B. im Konzentrator.

. Geographische Lage von Knotenpunkten, Leitungsführung sowie Auswahl geeigneter Übertragungswege.

Sehr eingehend wurde in den letzten Jahren die verkehrstheoretische Leistungsfähigkeit von Netzen mit Ringstruktur untersucht, siehe z.B. Avi-Itzhak /131/, Chu und Konheim /132/, Gall und Müller /133/, Hayes und Sherman /134/, Konheim und Meister /135/ sowie Spragins /136/. Einsatzmöglichkeiten von Multiplexern und Konzentratoren werden von Botsch /137/, Chu /138-140/ sowie Rudin /141-143/ diskutiert und mathematisch behandelt. Abschätzung der Leitungsausnutzung und Antwortzeiten im Anschlußnetz, siehe z.B. Martin /6/, Stimler /49/ und Weingarten /144/.

In Kapitel VI der vorliegenden Arbeit wird gezeigt, wie man die Leistungsfähigkeit einer Satellitenvermittlung für ein Datennetz mit "packet-switching" und Prioritäten exakt bestimmen kann.

Ansätze zur optimalen Auswahl von sog. zentralisierten Datennetzen lassen sich auch bei der Planung des Anschlußnetzes anwenden, da in diesem Fall die Datenvermittlung als Zentrale betrachtet werden kann. Doll /145/ und Gerich /146/ geben einen Überblick über verschiedene Ansätze, siehe ferner auch Chandy und Russel /147/, Esau und Williams /148/ sowie Frank, Frisch, Van Slyke und Chou /149/. Bahl und Tang /150/ geben einen Algorithmus an zur Bestimmung der optimalen Lage von Konzentratoren. Andere Lösungsansätze zur Planung des Anschlußnetzes findet man bei Hyvärinen /151/ und Randazzo /152/.

3.2. Datenvermittlungen

Das Hauptproblem verkehrstheoretischer Fragestellungen bei Datenvermittlungen ist die Berechnung der Gesamtleistungsfähigkeit (Durchsatz, Durchlaufzeit etc.) sowie das Ermitteln maschineninterner Engpässe. Aus der Vielzahl der Einzelfragen seien nur einige besonders wichtige herausgegriffen (siehe auch Margopoulos und Williams /153/):

. Vergleich verschiedener Struktur- und Betriebsarten bezüglich der Gesamtverzögerung von Nachrichten.

. Untersuchung der Wirksamkeit von Einzelspeichern (Puffer-, Haupt-, und Hintergrundspeicher) und Speicherhierarchien.

. Dimensionierung der Vermittlungseinheit bei Durchschaltevermittlungen und halbelektronischen Speichervermittlungen (vgl. Abschnitt III.3.1).

. Erhöhung der Leistungsfähigkeit durch Prioritätsmechanismen.

. Einfluß von "Hintergrundarbeit" auf die Reaktionszeit und den Durchsatz des Systems.

Zahlreiche Fragestellungen sind identisch oder mindestens eng verwandt mit jenen, welche bei modernen Großrechnersystemen auftreten. Eine Zusammenfassung und Klassifizierung bereits vorliegender Untersuchungen sowie über 70 weiterführende Literaturstellen findet man in der Arbeit von Herzog, Kühn und Zeh /21/. Einige zusätzliche Hinweise auf vorliegende Untersuchungen gibt die folgende Zusammenstellung:

Zahlreiche Hinweise für die schnelle Abschätzung der Gesamtleistungsfähigkeit geben Chang /154/, IBM /155/, Martin /6/ und Stimler /49/. Spragins /156/ untersucht mit Hilfe der Simulation eine Datenvermittlung bei extremer Belastung und zeigt anhand von Kurven Grenzen der maximalen Belastungsfähigkeit auf. Hauth /157/ untersucht analytisch das Modell einer Datenvermittlung mit internen (unterbrechenden) Prioritäten und prüft seine Ergebnisse mit Hilfe der Simulation. Den Einfluß schneller Datenkanäle auf die Gesamtleistungsfähigkeit eines Systems bestimmt Gay /158/.

Ein wichtiges Problem bei Datenvermittlungen ist die Speicherorganisation. Chang /159/ gibt eine Übersicht über die verschiedenen Techniken der Zwischenspeicherung und vergleicht sie bezüglich der Leistungsfähigkeit miteinander. Bricault und Delgavis /160/ geben einen Algorithmus an zur Abschätzung des Speicherbedarfs im Hauptspeicher, welcher zur Zwischenspeicherung von Nachrichten notwendig ist. Speicherhierarchien werden u.a. von Seaman, Lind und Wilson /161/, Iazeolla /162/, Bovet und De Julio /163/ analytisch untersucht.

In /26/ wurde vom Verfasser gezeigt, wie mit Hilfe sog. Unterbrechungsdistanz-Prioritäten (vgl. Abschnitt II. 3.2) die Leistungsfähigkeit einer Speichereinheit gesteigert werden kann. In Kapitel VII der vorliegenden Arbeit wird das Verkehrsmodell stark erweitert, ferner werden neben der mittleren Wartezeit nun auch andere wichtige charakteristische Größen bestimmt.

Eine große Gruppe von Arbeiten befaßt sich mit der Dimensionierung von "mehrstufigen Koppelanordnungen" für die Vermitt-

lungseinheit (V) bei Durchschaltevermittlungen oder halbelektronische Speichervermittlungen (vgl. Kapitel III. 3.1). Einen ausführlichen Überblick gibt Kümmerle /164, 165/, siehe ferner Bazlen /166/, Herzog /167/, Lörcher /168/ und Lotze /169/ für Durchschaltevermittlungen sowie Hieber /170/ für halbelektronische Speichervermittlungen mit "Raummultiplex - Durchschaltung". Huber /171/ zeigte erstmals, daß Zeitmultiplex-Koppelanordnungen abgebildet werden können auf äquivalente Raummultiplex-Systeme. Die für Raummultiplex entwickelten Berechnungsmethoden können deshalb auch für die Untersuchung vieler Zeitmultiplex-Systeme angewendet werden. Spezielle nichtblockierende Zeitmultiplex-Koppelanordnungen werden von Tanaka /172/ untersucht, s. ferner auch Cantor /173/. Ein Verfahren zur Optimierung mehrstufiger Koppelanordnungen wurde erstmals von Lotze /174/ vorgeschlagen.

3.3. Vermittlungsnetz

Die in diesem Fall zu untersuchenden Probleme sind jenen, die bei der Planung des Anschlußnetzes auftreten, teilweise sehr ähnlich, normalerweise aber komplexer. Fragen der Sicherheit und Überlastbarkeit im Falle verschiedener Struktur- und Betriebsarten spielen hier eine besondere Rolle. Einige typische Problemstellungen seien hier kurz erwähnt (siehe auch Kroneberg /175/ und Schwartz /176/):

- Leitungsausnutzung und Antwortzeiten bei verschiedenen Netzkonfigurationen (Stern-, Maschen-, Verbundnetz etc.) und Durchschalteprinzipien (Speicher- oder Durchschalte-

vermittlung).

- Leitungsausnutzung und Antwortzeiten bei verschiedenen Verfahren der Leitweglenkung (feste, alternative, adaptive oder zentrale Leitweglenkung).

- Sicherheit bei Überlastung oder Bündel- und Netzknotenausfällen.

- Empfindlichkeit gegen Verkehrsschwankungen und Änderungen des Verkehrscharakters.

- Erweiterungsmöglichkeit im Falle wachsenden Verkehrs der angeschlossenen Datenstationen bzw. größerer räumlichen Ausdehnung des Netzes.

Grundlage vieler Untersuchungen bei Netzen mit Speichervermittlungen sind die wichtigen Arbeiten von Burke /177/ und Jackson /178/ sowie die Unabhängigkeitsannahme von Kleinrock /179/ (siehe Abschnitt IV. 4.2). Kriterien zur Optimierung von Datennetzen mit fester Leitweglenkung werden ebenfalls von Kleinrock erstmals angegeben, später dann in Arbeiten von Meister, Müller und Rudin /180,181/ sowie Fratta und Gerla /182/ verfeinert und erweitert. Meister, Müller und Rudin /181/ berücksichtigen dabei erstmals den Einfluß von Bearbeitungszeiten in den Datennetzen. Einen Überblick über die verkehrstheoretischen Untersuchungen, die bei der Planung des ARPA-Netzes durchgeführt wurden, geben Frank, Kahn und Kleinrock /183/.

Die bisher in diesem Abschnitt diskutierten Untersuchungen beziehen

sich alle auf Netze mit Speichervermittlungen; Wartezeiten bei gemischtem Speicher- und Durchschaltevermittlungsbetrieb werden von Ottermann /184/ und Thierer /185/ untersucht. Closs /36/ untersucht in einer umfangreichen Arbeit die Antwortzeit in Netzen mit Speicher- bzw. Durchschaltevermittlungen mit Berücksichtigung der Signalisierung. In einer Arbeit von Silk /186/ wird der Einfluß konstanter Nachrichtenlänge (packetswitching!) auf die Gesamtverzögerungszeit von Nachrichten in einem Netz mit Speichervermittlungen näherungsweise untersucht.

Alle bisher zitierten Arbeiten untersuchen Netze mit fester oder zufälliger Leitweglenkung (Definition siehe Abschnitt III.1.1). Große Datennetze besitzen jedoch flexible, verkehrsabhängige Verfahren der Leitweglenkung, d.h. je nach momentaner Verkehrssituation wird der Verkehrsfluß gesteuert. Lotze /187/ gibt einen Überblick über die Dimensionierungsverfahren für derartige Netze mit Durchschaltevermittlungen. Prioritätsstrategien für diese Netze behandelt Katzschner /188/. Verschiedene Strategien der Leitweglenkung für Systeme mit Wartemöglichkeit werden von Herzog und Kühn /189/ diskutiert und miteinander verglichen. Erstmals wird dabei eine Berechnungsmethode vorgestellt, die die Bemessung von Netzen mit alternativer Leitweglenkung ermöglicht. Dieses Verfahren wird in Kapitel V der vorliegenden Arbeit erweitert. Das Modell der Datenvermittlung wird verfeinert. Neben den Mittelwerten für Warte- und Durchlaufzeiten werden jetzt auch wichtige Ergebnisse für Wartezeitverteilungsfunktionen angegeben.

4. METHODEN ZUR BEHANDLUNG VERKEHRSTHEORETISCHER PROBLEME

Für die analytische Untersuchung des Verkehrsflusses in Datennetzen gibt es zwei Möglichkeiten, die Simulation und die rein rechnerischen Verfahren. Beide Methoden werden in den folgenden Abschnitten kurz beschrieben, Literaturhinweise ermöglichen ein eingehendes Studium.

Für die Synthese optimaler Strukturen werden neben den klassischen Optimierungsverfahren häufig auch graphentheoretische Methoden benützt. Eine immer größere Bedeutung finden auch Methoden der linearen und dynamischen Programmierung. Bücher von Zurmühl /190/, Busacker und Saaty /191/, Churchman, Ackhoff und Arnoff /192/, Elmaghraby /193/, Hu /194/, Nemhauser /195/ sowie Zimmermann und Zielinski /196/ geben eine verständliche und umfassende Einführung in die verschiedenartigen Methoden.

4.1. Simulationsverfahren

Bei der Simulation von Datenflüssen auf Digitalrechnern bildet man das zu untersuchende System und den Verkehrsfluß mit Hilfe eines Programms im Rechner nach. Man entwirft ein Modell des realen Systems und der tatsächlichen Verkehrsabläufe und kann damit - zeitlich gerafft gegenüber der Wirklichkeit - die Leistungsfähigkeit des Systems untersuchen.

Der Hauptvorteil der Simulation gegenüber den rein rechnerischen

Verfahren liegt darin, daß man neue Probleme, deren Lösung noch nicht bekannt ist, in relativ kurzen und gut abschätzbaren Zeiträumen untersuchen kann, während ein Erfolg bei den rein rechnerischen Verfahren nicht immer gewährleistet ist. Simulation wird man deshalb immer dort als ersten Schritt anwenden, wo die Probleme komplex werden, also auch bei Datennetzen (Prinzip der Simulation siehe z.B. Gordon /197/, Huber und Wagner /198/ sowie Kampe /199/; die Simulation von Systemen mit kleiner Quellenzahl behandelt Weissschuh /200/; Simulation von Netzen siehe Boehm und Mobley /201/, Caplan und Yonda /202/, Price /203/, Seaman /204/ und Slight /205/).

Heute sind wir bei der Simulation nicht mehr auf die für diesen Zweck recht unhandlichen allgemeinen Programmiersprachen angewiesen wie z.B. ALGOL oder FORTRAN.

Spezielle Programmiersprachen erlauben eine relativ einfache und schnelle Beschreibung eines Systems. Die Simulationssprachen sind meistens sehr flexibel und sind zur Simulation nahezu aller Systemkonfigurationen geeignet. Ein in einer Simulationssprache programmiertes Grundmodell kann stetig erweitert werden, bis es sehr genau dem Verkehrsverhalten des realen Systems entspricht. Heute gibt es schon zahlreiche Simulationssprachen, von denen vor allem GPSS, SIMULA, SIMPAC und SIMSCRIPT sehr häufig angewendet werden (siehe z.B. Gordon /197/ und Kampe /199/).

Aufgabe eines Systemplaners ist nicht nur die Analyse, sondern vor allem die Synthese, also der Entwurf wirtschaftlich optimaler Netze. Optimierung ist aber mit Hilfe der Simulation schlecht

möglich. Durch systematische Variation der Struktur- und Verkehrsparameter kann man zwar wirtschaftliche Lösungen finden, aber hier stößt man wegen der enormen Rechenzeiten selbst bei den größten Rechnern bald auf eine Grenze.

Echte Optimierung ist mit sinnvollem Aufwand nur bei den rein rechnerischen Verfahren möglich.

4.2. Berechnungsmethoden

Den Verkehrsfluß in Datennetzen kann man - wie bereits in Kapitel II ausführlich erläutert - als stochastischen Prozeß beschreiben und mit Methoden der Wahrscheinlichkeitsrechnung untersuchen. Als Ergebnis erhält man charakteristische Größen, wie z.B. Warte- oder Antwortzeiten, Belastungen für Speicher oder Bedienungseinheiten usw. Entsprechend den verschiedenartigen Berechnungsmethoden wollen wir zwei große Gruppen von Prozessen unterscheiden, sog. Markoff'sche und Nichtmarkoff'sche Prozesse.

4.2.1. Markoff'sche Prozesse

Ein Prozess $\emptyset$ werde zu einem beliebigen Zeitpunkt t_n beobachtet. Ist der weitere Verlauf dieses Prozesses nur vom Zustand x_n zur Zeit t_n und nicht von der Vorgeschichte $(t_{n-1}, t_{n-2}, \ldots t_0)$ abhängig, so nennt man den Prozeß einen Markoff-Prozeß (nach dem russischen Mathematiker A.A. Markoff).

Formelmäßig erfaßt wird diese Unabhängigkeit von der Vorgeschichte durch folgende Beziehung für die bedingten Zustandswahrscheinlichkeiten:

$$\left.\begin{array}{l} P\{\Phi_{t_{n+1}}=j \mid \Phi_{t_0}=x_0, \Phi_{t_1}=x_1, \ldots, \Phi_{t_n}=x_n\} = P\{\Phi_{t_{n+1}}=j \mid \Phi_{t_n}=x_n\} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad t_0 < t_1 < t_2 < \ldots \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad n = 0, 1, 2, \ldots \end{array}\right\} \quad \text{(IV.1)}$$

Führt man für die Zustandswahrscheinlichkeiten $P\{\emptyset_{t_n} = x_n\}$ (Prozeß $\emptyset$ zum Zeitpunkt t_n im Zustand x_n) die Abkürzung $p_{t_n}(x_n)$ und sinngemäß für die Übergangswahrscheinlichkeiten $\ddot{u}_{t_n,t_{n+1}}(x_n,x_{n+1})$ ein, so gilt ferner:

$$p_{t_{n+1}}(x_{n+1}) = \sum_{x_n} p_{t_n}(x_n) \cdot \ddot{u}_{t_n,t_{n+1}}(x_n, x_{n+1}) \qquad \text{(IV.2)}$$

d.h. die Zustandswahrscheinlichkeiten zu einem Zeitpunkt t_{n+1} sind nur von den Zustandswahrscheinlichkeiten des vorhergehenden Zeitpunktes t_n und den Übergangswahrscheinlichkeiten im betrachteten Zeitintervall abhängig. Gleichung (IV.2) gilt für jeden beliebigen Zustand und wird als Chapman - Kolmogoroff'sches Gleichungssystem bezeichnet.

Markoff'sche Prozesse haben den großen Vorteil, daß man stets denselben Lösungsweg einschlagen kann:

- Aufstellen des Chapman-Kolmogoroff'schen Gleichungssystems.
- Bestimmen und Einsetzen der Übergangswahrscheinlichkeiten.
- Auflösen des Gleichungssystems.

. Bestimmen der charakteristischen Größen mit Hilfe der Zustandswahrscheinlichkeiten.

Schwierigkeiten - teilweise allerdings erhebliche - treten vor allem erst bei der Auflösung bzw. Auswertung des linearen Gleichgewichtssystems auf.

Durch das Einführen von faktoriellen Momenten und erzeugenden Funktionen (siehe z.B. Kapitel V) ist es jedoch oft möglich, ohne explizit die Zustandswahrscheinlichkeiten zu berechnen, wichtige charakteristische Größen direkt anzugeben (z.B. Mittelwerte und Streuung). Markoff'sche Prozesse werden in der einschlägigen Literatur sehr ausführlich behandelt, siehe z.B. Cox und Smith /206/, Ferschl /10/, Lahres /207/, Morse /12/, Saaty /208/, Syski /14/, Takács /209/. Ein typisches Beispiel zweidimensionaler Markoff'scher Prozesse wird in Kapitel V beschrieben und analytisch behandelt.

4.2.2 Nicht-Markoff'sche Prozesse

Nicht-Markoff'sche Prozesse sind alle jene Prozesse, bei welchen der Prozeßverlauf - mindestens für eine gewisse Zeit - von der Vorgeschichte abhängig ist. Ein einheitlicher Lösungsansatz wie bei den Markoff'schen Systemen ist nicht mehr möglich.

Dennoch versucht man durch einen Trick, sich die Vorteile, die sich bei der Behandlung Markoff'scher Systeme ergeben, zunutze zu machen: man ersetzt das zeitlich abhängige Verhalten eines eindimensionalen Nicht-Markoff'schen Prozesses durch

das Verweilen in fiktiven exponentiellen Phasen, d.h. durch einen mehrdimensionalen Markoff-Prozeß (Phasenmethode, siehe Ferschl /10/, Morse /12/ usw.).

Eine zweite Möglichkeit ist das Verfahren der Eingebetteten Markoff-Kette: man betrachtet einen beliebigen stochastischen Prozeß nur zu solchen diskreten Zeitpunkten, bei denen die Zustände die Markoff-Eigenschaft besitzen (siehe Ferschl /10/, Morse /12/ usw.).

Beide Methoden sind dazu geeignet, zeitabhängige Prozesse zu beschreiben, also auch z.B. das Einschwingverhalten eines Systems zu untersuchen. Diese allgemeine Anwendbarkeit bringt zwangsläufig den Nachteil relativ umfangreicher Gleichungssysteme mit sich. Oft sind deshalb zwar prinzipielle Lösungsansätze möglich, eine numerische Auswertung scheitert aber selbst auf den größten Datenverarbeitungsanlagen.

Bei Prozessen, die bereits zu Beginn der Beobachtung im eingeschwungenen Zustand sind (sog. stationäre Prozesse) erlaubt oft die Momentenmethode mit relativ geringem Aufwand die Bestimmung des ersten Moments (z.B. der Wartezeitverteilungsfunktion), teilweise auch höherer Momente. Man verfolgt hierbei das Schicksal einer individuellen Anforderung (sog. Testruf) vom Eintreffen in das System bis zum Verlassen mit allen Möglichkeiten für Unterbrechung, Zurückschieben im Speicher usw. Durch das Bilden von Erwartungswerten ist dann oft eine Bestimmung der Momente möglich. Zwei Beispiele für die Anwendung der Momentenmethode findet man in den Kapiteln VI und VII.

4.2.3. Die Ergebnisse von Burke und Jackson

Die Annahme negativ exponentiell verteilter Ankunftsabstände entspricht in vielen Fällen sehr gut den Verhältnissen in der Wirklichkeit (vgl. Abschnitt II.2.2). In diesem Fall ist aber der Ankunftsprozeß ein spezieller Markoff-Prozeß (sog. Poisson-Prozeß), der die analytische Behandlung sehr vereinfacht.

Burke /177/ gibt Bedingungen an, unter welchen die Abstandsverteilung der Anforderungen, die ein System verlassen (und eventuell einem zweiten System angeboten werden), ebenfalls negativ exponentiell verteilt sind. Diese Bedingungen sind in Bild IV.2 zusammengefaßt (Verallgemeinerung siehe Boes /210/, Burke /211/ und Shanbhag /212/).

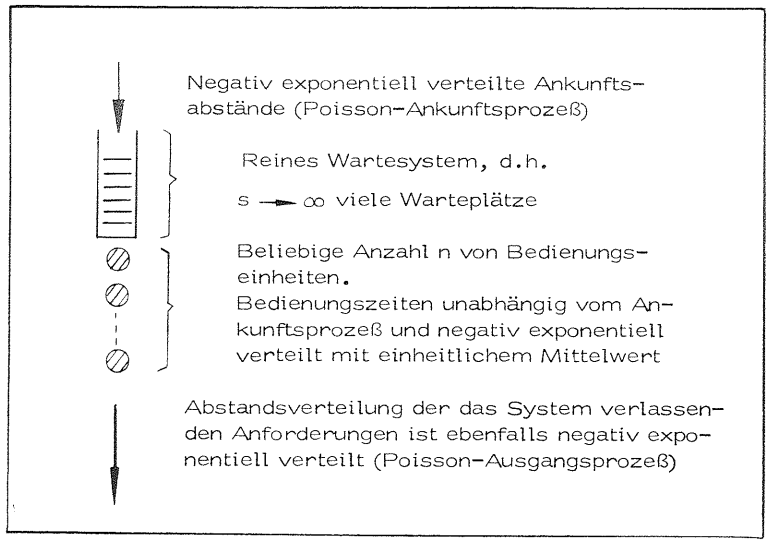


Bild IV.2: Das System von Burke /177/

Jackson /178/ untersucht ganze Netze von "Burke'schen Systemen" und zeigt, daß die Zustandswahrscheinlichkeiten der einzelnen Teilsysteme in einem beliebigen Zeitpunkt unabhängig voneinander sind. Die (mehrdimensionalen) Zustandwahrscheinlichkeiten des Gesamtsystems ergeben sich deshalb als Produkt der Zustandswahrscheinlichkeiten aller Teilsysteme.

4.2.4. Die Unabhängigkeitsannahme von Kleinrock

In einem ungeregelten Netz mit Speichervermittlungen (Definitionen siehe Abschnitt III.1) werden Nachrichten übertragen, deren Länge - insgesamt gesehen - einer negativ exponentiellen Verteilung gehorchen sollen. Die Übertragungszeit einer bestimmten Nachricht ist jedoch auf jeder Teilstrecke proportional zu ihrer festen Länge.

Betrachtet man als einfachstes Beispiel zwei aufeinanderfolgende Teilstrecken, so erkennt man, daß bei der zweiten Teilstrekke eine gewisse Abhängigkeit zwischen dem Ankunftsabstand und den Übertragungszeiten besteht: eine Nachricht Nr. j kann frühestens im Knoten vor der zweiten Teilstrecke eintreffen, wenn die Übertragungsdauer der Nachricht Nr. j-1 zu Ende ist. Damit sind aber die Voraussetzungen zur Anwendung der Ergebnisse von Burke und Jackson nicht mehr exakt erfüllt.

Eine exakte Bestimmung derartiger Abhängigkeiten ist mit erträglichem Aufwand bei realen Systemen nur noch selten möglich (siehe z.B. Krämer /213/). Man nimmt deshalb näherungsweise an, daß die Unabhängigkeit der Übertragungszeiten auf den ein-

zelnen Teilstrecken erfüllt ist. Kleinrock /179/ zeigte mit Hilfe der Simulation in einer Reihe von Beispielen, daß diese Annahme erlaubt ist. Ihre Auswirkung auf die charakteristischen Größen ist umso geringer, jemehr Verkehrsflüsse aus unterschiedlichen Richtungen einem System angeboten werden. Dies ist aber gerade bei Knotenpunkten von Datennetzen der Fall.

V. DURCHLAUFZEIT VON NACHRICHTEN IN NETZEN MIT SPEICHERVERMITTLUNGEN UND ALTERNATIVER LEITWEGLENKUNG

1. PROBLEMSTELLUNG
2. BESCHREIBUNG DES VERKEHRSABLAUFS
3. MODELLBILDUNG
4. PRINZIPIELLER LÖSUNGSWEG
5. BERECHNUNG DER CHARAKT. GRÖSSEN
6. NUM. BEISPIELE, VERGLEICH MIT SIMULATIONSERGEBNISSEN
7. ZUSAMMENFASSUNG UND AUSBLICK

Der Nachrichtenfluß in modernen Netzen mit verkehrsabhängiger Leitweglenkung entspricht speziellen stochastischen Prozessen. Sollen Fehlplanungen vermieden werden, so muß dieser spezielle Charakter bei der Dimensionierung von Netzen berücksichtigt werden. In diesem Kapitel wird ein Näherungsverfahren vorgestellt, welches mit hoher Genauigkeit die Berechnung von Netzen mit Speichervermittlungen und alternativer Leitweglenkung erlaubt. Neben charakteristischen Mittelwerten und Einzelwahrscheinlichkeiten werden auch wichtige Angaben über Wartezeitverteilungsfunktionen vorgestellt.

1. PROBLEMSTELLUNG

Aus Gründen der Wirtschaftlichkeit, ferner bei Überlastung und Teilausfällen sind Netze mit verkehrsabhängiger Leitweglenkung Netzen ohne diese Möglichkeit überlegen (vgl. Abschnitt III.2.2). In diesem Kapitel werden derartige Netze mit Speichervermittlungen und alternativer Leitweglenkung untersucht: der Verkehr

in eine bestimmte Richtung wird zunächst einem "Primärweg" angeboten; ist dieser direkte Weg momentan "überlastet", so wird der Verkehr einem zweiten Weg (Sekundärweg), eventuell auch weiteren Wegen angeboten.

Derartige ein- oder mehrmals abgewiesene Teilverkehre (sog. Überlaufverkehre) haben völlig andere stochastische Eigenschaften als die den Primärwegen angebotenen Gesamtverkehre (Zufallsverkehr).

Den Einfluß dieser speziellen Eigenschaften auf die charakteristischen Größen veranschaulicht das Beispiel in Bild V.1: aufgetragen sind Wartewahrscheinlichkeit und Überlaufverkehr eines Sekundärsystems, dem ebenfalls Überlaufverkehr von einem Primärsystem angeboten wird. Die durchgezogenen Linien zeigen exakt berechnete Werte, die gestrichelt eingezeichneten Kurven entsprechen angebotenem Zufallsverkehr gleichen Mittelwerts. Wartewahrscheinlichkeit und Überlaufverkehr sind bei Zufallsverkehr erheblich kleiner; für andere charakteristische Größen ergibt sich dieselbe Tendenz! Dieses typische Verhalten ist für Fernsprechnetze wohl bekannt (siehe z.B. Bretschneider /214/, Lotze /215, 45/, Wilkinson /216/). Moderne Dimensionierungsvorschriften, wie z.B. die Richtlinien der Deutschen Bundespost für die Dimensionierung des gesamten Orts- und Landesfernwahlnetzes, berücksichtigen deshalb den speziellen Charakter von Überlaufverkehr.

Will man bei Datennetzen Fehlplanungen vermeiden, so ist es unumgänglich, auch dort für die Zukunft entsprechende Dimensionierungsverfahren und -vorschriften zu entwickeln.

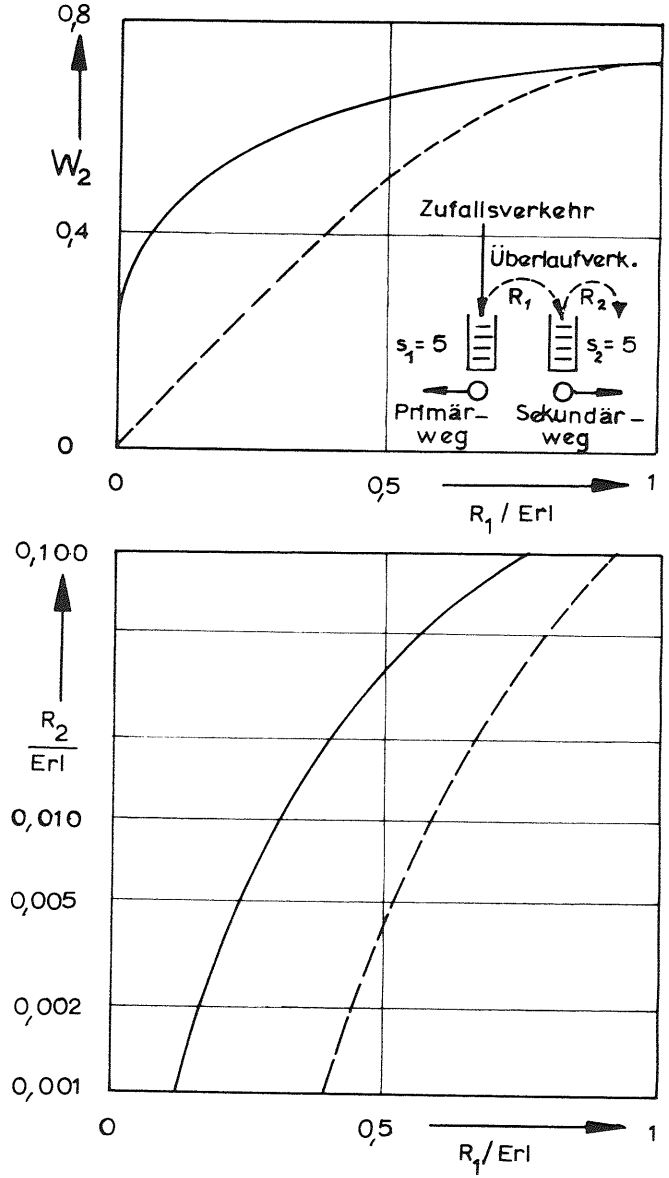


Bild V.1: Wartewahrscheinlichkeit W_2 und Überlaufverkehr R_2 für ein Sekundärsystem, dem selbst Überlaufverkehr (Mittelwert R_1) angeboten wird. Der Sekundärweg wird (in diesem Beispiel) erst dann benutzt, wenn am Primärweg bereits $S_1 = 5$ Nachrichten warten. Gestrichelt eingetragen sind jene Werte, die sich bei angebotenem Zufallsverkehr gleichen Mittelwerts ergeben (vgl. Text)

Verschiedene Strategien der Leitweglenkung für Systeme mit Wartemöglichkeit werden von Herzog und Kühn /189/ diskutiert. Erstmals wird dabei das Grundprinzip einer Berechnungsmethode vorgestellt, welche die Bemessung von Netzen mit alternativer Leitweglenkung ermöglicht. Diese Methode wird im folgenden erweitert, insbesondere wird

- der Verkehrsfluß innerhalb einer einzelnen Datenvermittlung mit Hilfe eines verfeinerten Strukturmodells untersucht und

- die Gesamtdurchlaufzeit von Nachrichten in einem Netz mit mehreren Datenvermittlungen (Knoten) bestimmt.

Neben den charakteristischen Mittelwerten und Einzelwahrscheinlichkeiten werden jetzt auch interessante Ergebnisse über Wartezeitverteilungsfunktionen vorgestellt.

2. BESCHREIBUNG DES VERKEHRSABLAUFS

In Abschnitt III. 3.1.2 wurde der Aufbau typischer Speichervermittlungen beschrieben: die vom Vermittlungs- und Anschlußnetz eintreffenden Nachrichten werden in der Speichereinheit gespeichert. Danach wird in einer Eingangswarteschlange ein Indikator gesetzt. Diese Warteschlange ist entweder individuell oder gemeinsam für alle Richtungen vorhanden und wird von der zentralen Steuereinheit (Prozessor) bearbeitet.

Der Prozessor wertet die Information des "Nachrichtenkopfes"

aus und entscheidet mit Hilfe der "Leitwegtafel", in welcher Richtung die Nachricht übertragen wird: ist der direkte Weg überlastet, d.h. wartet bereits eine bestimmte (beliebig vorgebbare) Anzahl von Nachrichten, so wird in der Ausgabeliste des Sekundärweges ein Indikator gesetzt (einfachster Fall; mehrfacher Überlauf ist möglich).

Die Ausgabeliste wird von der Ausgabesteuerung bearbeitet, d.h. die Nachrichten werden - ohne weitere Beanspruchung des Prozessors - zum betreffenden Nachbarknoten bzw. zu den Datenendgeräten des Anschlußnetzes übertragen.

3. MODELLBILDUNG

3.1. Struktur und Betriebsart

Bild V.2 zeigt das Strukturmodell einer Datenvermittlung, welches dem oben beschriebenen Verkehrsablauf entspricht. Die von

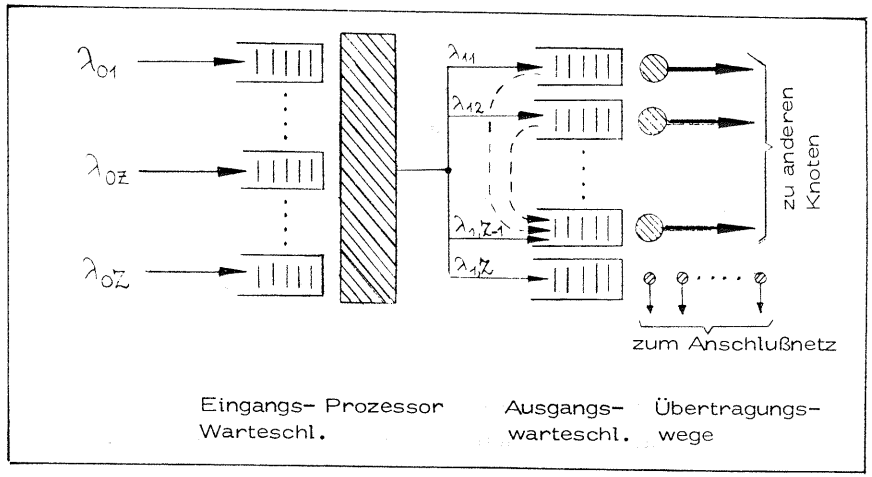


Bild V.2: Strukturmodell einer Datenvermittlung. Die gestrichelten Pfeile zeigen den Überlaufverkehr an (vgl. Text)

den (Z-1) Nachbarknoten und dem Anschlußnetz eintreffenden Anforderungen (Nachrichten) werden in den Eingangswarteschlangen gespeichert. Alle Eingangswarteschlangen werden vom Prozessor (z.B. in zyklischer Reihenfolge) bearbeitet. Nach dieser Bedienungsphase werden die Anforderungen entsprechend der alternativen Leitwegstrategie in eine bestimmte Ausgangswarteschlange eingereiht und warten auf Übertragung in die betreffende Richtung. Alle Eingangswarteschlangen sind unbegrenzt, d.h. keine Anforderung wird abgewiesen (dynamische Speicherplatzzuweisung). Die wartenden Anforderungen aller Eingangs- und Ausgangswarteschlangen werden in der Reihenfolge ihres Eintreffens abgearbeitet (FIFO). Ausführliche Begründung beider Annahmen, siehe Abschnitt II. 3.1 und VII. 3.1.

3.2. Verkehr

Die aus einer Richtung eintreffenden Anforderungen haben negativ exponentiell verteilte Ankunftsabstände, d.h. die Ankunftsprozesse haben die Verteilungsfunktion

$$A_{oz}(\leq t) = 1 - e^{-\frac{t}{a_{oz}}} = 1 - e^{-\lambda_{oz} \cdot t} \qquad (V.1)$$

mit dem mittleren Ankunftsabstand a_{oz} und der Ankunftsrate λ_{oz} (z = 1 .... Z).

Die Prozessor-Bedienungszeiten sind für alle Anforderungen einheitlich negativ exponentiell verteilt (andere Modelle siehe Zusammenfassung und Ausblick):

$$B_{oz}(\leq t) = 1 - e^{-\frac{t}{b_o}} \qquad (V.2)$$

mit dem Mittelwert b_o .

Die Aufteilungswahrscheinlichkeit η_{zj} gibt an, mit welcher Wahrscheinlichkeit eine aus Richtung z eintreffende Nachricht in der abgehenden Richtung j (Primärweg!) übertragen werden soll. Der einer Richtung j angebotene "Direktverkehr" hat deshalb die Ankunftsrate

$$\lambda_{1j} = \sum_{z=1}^{Z} \lambda_{oz} \cdot \eta_{zj} \qquad (V.3)$$

Die Übertragungsgeschwindigkeit kann individuell pro abgehende Richtung verschieden sein. Bei negativ exponentiell verteilten Nachrichtenlängen gilt deshalb für die Verteilungsfunktion der Übertragungszeiten:

$$B_{1j}(\leq t) = 1 - e^{-\frac{t}{b_{1j}}} \qquad (V.4)$$

mit der mittleren Übertragungszeit b_{1j} pro Nachricht.

4. PRINZIPIELLER LÖSUNGSWEG

4.1. Allgemeines

Die exakte Untersuchung des Verkehrsflusses in einem Datennetz ist prinzipiell möglich (vgl. Abschnitt IV. 4.2): vieldimensionale Zustandswahrscheinlichkeiten werden eingeführt, das Chapman-Kolmogoroff'sche Gleichungssystem wird aufgestellt und die Auswertung erfolgt mit Hilfe eines Relaxationsverfahrens

(siehe z.B. Schehrer /217/). Für Netzstrukturen, die von praktischem Interesse sind, versagt jedoch dieses Verfahren sofort, da die Auflösung der Gleichungssysteme, selbst auf den größten Datenverarbeitungsanlagen, nicht mehr möglich ist.

Im folgenden wird deshalb ein Näherungsverfahren vorgestellt, das die Berechnung des Verkehrsflusses auch in großen Netzen ermöglicht. Zunächst wird der Verkehrsfluß in einer Datenvermittlung, in einem weiteren Abschnitt innerhalb eines Netzes untersucht.

4.2. Verkehrsfluß innerhalb einer Datenvermittlung

Bei der mathematischen Behandlung kann man folgende vier Teilsysteme unterscheiden (vgl. Bild V. 2):

- das Eingangssystem mit den Eingangswarteschlangen und dem Prozessor als Bedienungseinheit,
- Primärsysteme, d.h. Ausgangswarteschlangen und Übertragungswege, denen ausschließlich Direktverkehr angeboten wird,
- Sekundärsysteme mit angebotenem Überlaufverkehr, häufig auch mit Direktverkehr sowie
- das Teilnehmersystem, das die Datenvermittlung mit dem Anschlußnetz und den Datenstationen verbindet.

Das Teilnehmersystem wird im folgenden nicht mehr betrachtet, verwiesen sei jedoch auf Abschnitt IV. 3.1.

4.2.1. Das Eingangssystem

Das Eingangssystem besteht aus einem reinen Wartesystem mit einer oder mehreren parallelen Warteschlangen und einer Bedienungseinheit. Derartige Systeme wurden schon in zahlreichen Arbeiten untersucht und mathematisch behandelt. Formeln für alle charakteristischen Größen können direkt übernommen werden.

Nach dem Theorem von Burke (ausführliche Beschreibung siehe Abschnitt IV. 4.2.3) ist der Ausgangsprozeß dieses Systems ein Poisson-Prozeß, d.h. die Ankunftsabstände der Nachrichten, die den Primär- und Sekundärsystemen oder dem Teilnehmersystem angeboten werden, gehorchen wieder einer negativ exponentiellen Verteilung.

4.2.2. Primärsysteme

Primärsystemen wird ausschließlich Direktverkehr angeboten. Formeln für viele charakteristischen Größen, wie z.B. die mittlere Warte- und Durchlaufzeit, die Wartewahrscheinlichkeit oder die Wartezeitverteilungsfunktion, sind deshalb bekannt (vgl. Abschnitt IV. 4.2).

Die von Primärsystemen abgewiesenen Überlaufverkehre werden nicht nur durch das erste Moment (Mittelwert R_1), sondern

auch durch das zweite Moment (Varianz σ_1^2 bzw. Streuwert $D_1 = \sigma_1^2 - R_1$) charakterisiert. Damit kann der spezielle stochastische Charakter von Überlaufverkehr erfaßt werden.

4.2.3. Sekundärsysteme

Im allgemeinsten Fall werden einem Sekundärsystem sowohl Direktverkehre(A_{1i}, $D_{1i} = 0$) als auch Überlaufverkehre (R_{1j}, D_{1j}) angeboten. Der gesamte, einem Sekundärsystem angebotene Verkehr (R_1, D_1) kann deshalb durch die Summe aller Mittelwerte und die Summe aller Streuwerte beschrieben werden:

$$\left.\begin{aligned} \overline{R_1} &= \sum_i A_{1i} + \sum_j R_{1j} \\ \overline{D_1} &= \sum_j D_{1j} \end{aligned}\right\} \quad (V.5)$$

Die Berechnung der charakteristischen Größen des Sekundärsystems erfolgt mit Hilfe eines "Ersatzsystems", bestehend aus einer Bedienungseinheit und S^* Warteplätzen. Das "erzeugende Angebot" A^* wird dergestalt bestimmt, daß ein Überlaufverkehr mit Mittelwert $\overline{R_1}$ und Streuwert $\overline{D_1}$ erzeugt wird (vgl. Bild V.3). Mit anderen Worten: alle tatsächlichen Verkehre, entweder direkt oder als Überlaufverkehr von verschiedenen Primärsystemen angeboten, werden näherungsweise durch einen Ersatzverkehr gleichen Mittelwerts $\overline{R_1}$ und gleichen Streuwerts $\overline{D_1}$ ersetzt.

Den Verkehrsfluß in dieser, aus Ersatzsystem und tatsächlichem

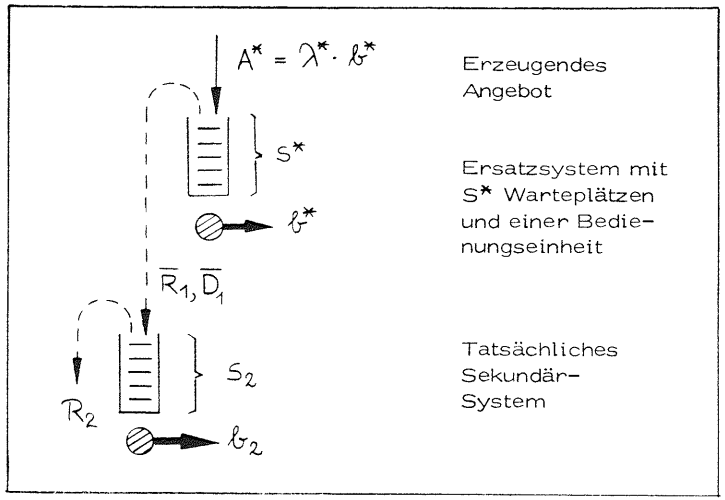


Bild V. 3: Ersatzanordnung zur Berechnung der charakteristischen Größen eines Sekundärsystems

Sekundärsystems bestehenden, Anordnung kann man durch einen zweidimensionalen Markoff'schen Prozeß beschreiben. Charakteristische Mittelwerte, Einzelwahrscheinlichkeiten und die Wartezeitverteilungsfunktion werden sowohl für die Gesamtanordnung, als auch für das zu untersuchende Sekundärsystem bestimmt.

4.3. Verkehrsfluß in einem Netz

Mit Hilfe des Verfahrens, das im vorhergehenden Abschnitt skizziert wurde, ist es möglich, den Verkehrsfluß innerhalb einer bestimmten Vermittlung bei vorgegebenen Verkehren zu

untersuchen. In einem Netz mit vielen Knoten sind die Verkehre voneinander abhängig, d.h. der auf den verschiedenen Wegen fließende Verkehr muß iterativ bestimmt werden. Individuell für jede Verbindung zwischen zwei Knoten können damit alle charakteristischen Größen angegeben werden.

Die Gesamtdurchlaufzeit für Nachrichten vom Ursprungs- zum Zielknoten setzt sich aus der Summe der Durchlaufzeiten für Nachrichten durch die einzelnen Knoten, gewichtet mit der Wahrscheinlichkeit, daß der betreffende Weg eingeschlagen wird, zusammen.

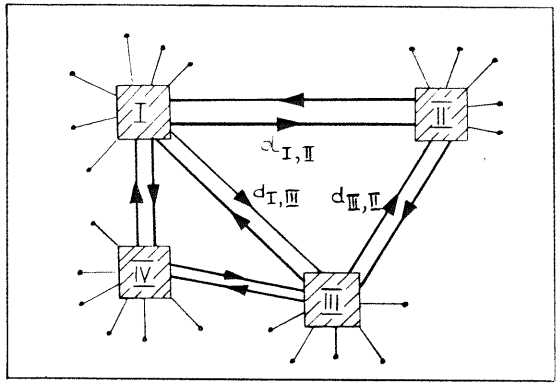


Bild V.4: Beispiel zur Berechnung der Gesamtdurchlaufzeit von Nachrichten vom Ursprungs- zum Zielknoten (vgl. Text)

Bild V.4. zeigt dazu ein einfaches Beispiel mit vier Knoten. Nachrichten vom Ursprungsknoten I zum Zielknoten II benutzen entweder den Primärweg I - II oder den Sekundärweg I - III - II.

Bezeichnet man mit

$d_{k,l}$ die mittlere Durchlaufzeit aller Nachrichten, die auf der Leitung vom Knoten k zum Knoten l übertragen werden, und mit

$p_{k,l}(u,z)$ die Wahrscheinlichkeit, daß eine Nachricht vom Ursprungsknoten u zum Zielknoten z den Zwischenknoten k erreicht hat und auf der Leitung zum Knoten l übertragen wird,

so gilt in unserem Beispiel für die gesamte mittlere Durchlaufzeit aller Nachrichten vom Ursprungsknoten I zum Zielknoten II:

$$d(I,II) = p_{I,II}(I,II) \cdot d_{I,II} + p_{I,III}(I,II) \cdot \left\{ d_{I,III} + 1 \cdot d_{III,II} \right\}$$

Entsprechende Beziehungen ergeben sich für die anderen Verkehre.

5. BERECHNUNG DER CHARAKTERISTISCHEN GRÖSSEN

In diesem Abschnitt werden bekannte Formeln angegeben und neue Ergebnisse abgeleitet. Damit ist man in der Lage, wichtige charakteristische Größen einzelner Teilsysteme, einzelner Datenvermittlungen und ganzer Netze mit Speichervermittlungen und alternativer Leitweglenkung anzugeben.

Entsprechend Abschnitt V. 4.2. wird bei der mathematischen Behandlung unterschieden zwischen dem Eingangssystem, Primärsystemen und Sekundärsystemen.

5.1. Das Eingangssystem

Das Eingangssystem entspricht einem rein Markoff'schen Wartesystem mit einer gemeinsamen oder Z parallelen Warteschlangen (vgl. Bild V.2). Formeln für alle charakteristischen Größen können deshalb aus der Literatur übernommen werden. Nach der bekannten Formel von Erlang (siehe z.B. Brockmeyer, Halstrøm und Jensen /218/) gilt beispielsweise für einen einfachen Fall – entweder eine gemeinsame Warteschlange und beliebige Ankunftsraten λ_{oz} (z = 1 ....Z) oder Z zyklisch abgefertigte Warteschlangen und gleiche Ankunftsraten – für die mittlere Wartezeit aller Anforderungen

$$\boxed{w_o = \frac{1}{\lambda_o} \cdot \frac{A_o^2}{1 - A_o}} \qquad (V.6)$$

wo
$$\lambda_o = \sum_{z=1}^{Z} \lambda_{oz}$$
$$A_o = \lambda_o \, b_o$$

Andere Abfertigungsstrategien bei ungleichen Ankunftsraten siehe z.B. Kühn /23, 24/.

5.2. Primärsysteme

Primärsysteme entsprechen ebenfalls rein Markoff'schen Systemen. Ist der Speicherraum begrenzt (s_1 Speicherplätze), so wird ein Teil des angebotenen Direktverkehrs abgewiesen. Dieser Überlaufverkehr wird durch seinen Mittelwert R_1 und

den Streuwert D_1 charakterisiert (vgl. Bild V.5)

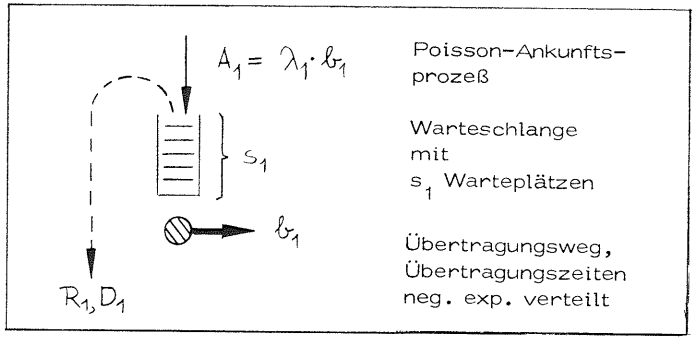


Bild V. 5: Primärsystem mit angebotenem Direktverkehr

Die Berechnung beliebiger Momente des Überlaufverkehrs wurde bereits in /189/ durchgeführt: zweidimensionale Zustandswahrscheinlichkeiten ermöglichen die Beschreibung des Überlaufverkehrs. Durch das Aufstellen des Chapman-Kolmogoroff'schen Gleichungssystems sowie das Einführen "faktorieller Momente" und "erzeugender Funktionen" wurde ein Lösungsalgorithmus gefunden. Bei einheitlicher Übertragungsgeschwindigkeit für Primär- und Sekundärweg kann außerdem eine geschlossene Lösung angegeben werden. In diesem Fall gilt für den Mittelwert R_1 und den Streuwert D_1 des Überlaufverkehrs:

$$\boxed{\begin{aligned} R_1 &= \frac{A_1^{s_1+2}}{1 + A_1 \cdot \dfrac{1 - A_1^{s_1+1}}{1 - A_1}} \\ D_1 &= R_1^2 \cdot \left\{ \frac{c_1 \cdot A_1}{c_2 \cdot R_1} - 1 \right\} \end{aligned}} \qquad (V.7)$$

wo

$$C_1 = \frac{1}{2k_2}\left\{\left[1+A_1-(k_1-k_2)\right]\cdot(k_1+k_2)^{s_1+1} - \left[1+A_1-(k_1+k_2)\right](k_1-k_2)^{s_1+1}\right\}$$

$$C_2 = 1+\frac{1}{2k_2}\left\{\left[1+A_1-(k_1-k_2)\right]\sum_{\nu=0}^{s_1}(k_1+k_2)^{\nu+1} - \left[1+A_1-(k_1+k_2)\right]\sum_{\nu=0}^{s_1}(k_1-k_2)^{\nu+1}\right\}$$

$$k_1 = 1+\frac{A_1}{2}$$

$$k_2 = \sqrt{1-\left(\frac{A_1}{2}\right)^2}$$

Formeln, die den Warteprozeß charakterisieren, wie z.B. die Wartezeitverteilungsfunktion, Wartewahrscheinlichkeit oder Wartebelastung, findet man in der einschlägigen Literatur (siehe Abschnitt IV. 4.2). Speziell für die mittlere Wartezeit aller Anforderungen, die nicht abgewiesen werden, gilt

$$\boxed{w_\gamma = \begin{cases} b_1\cdot A_1\cdot\dfrac{1-A_1^{s_1}[s_1+1-s_1\cdot A_1]}{1-A_1}\cdot\dfrac{1-A_1^{s_1+2}}{(1-A_1^{s_1+1})^2} & A_1 \gtrless 1 \\ b_1\cdot\dfrac{s_1+2}{2} & A_1 = 1 \end{cases}} \quad (V.8)$$

5.3. Sekundärsysteme

Der einem Sekundärsystem angebotene Verkehr hat andere Eigenschaften als Poisson-Verkehr. Im folgenden wird gezeigt, wie mit Hilfe eines zweidimensionalen Markoff-Prozesses auch in diesem Fall, alle charakteristischen Größen - einschließlich der Wartezeitverteilungsfunktion - bestimmt werden können.

Untersucht wird jene "Ersatzanordnung", auf welche alle möglichen tatsächlichen Anordnungen zurückgeführt werden können: ein Primär- und ein Sekundärsystem (vgl. Bild V.3).

Aus Gründen der Übersichtlichkeit werden alle Größen, die sich auf das Primärsystem beziehen, mit dem Index "1", jene des Sekundärsystems mit dem Index "2" versehen.

5.3.1. Berechnung der Zustandwahrscheinlichkeiten

Bezeichnet man mit $\{u_1, u_2\}$ den Zustand, daß zu einem beliebigen Zeitpunkt genau u_1 Anforderungen im (Ersatz-) Primärsystem und u_2 Anforderungen im (interessierenden) Sekundärsystem sind, so gilt nach dem Chapman-Kolmogoroff'schen Gleichungssystem für die Zustandswahrscheinlichkeiten:

$u_1 < 1+s_1$, $u_2 = 1+s_2$:

$$\varepsilon_1\cdot p(1,1+s_2) - (\varepsilon_2+\lambda_1)\cdot p(0,1+s_2) = 0 \quad (V.9)$$

$$\varepsilon_1\cdot p(u_1+2,1+s_2) - (\varepsilon_1+\varepsilon_2+\lambda_1)\cdot p(u_1+1,1+s_2) + \lambda_1 p(u_1,1+s_2) = 0$$

Gleichung (V.9) stellt ein System von (s_1+1) linearen, homogenen Differenzengleichungen zweiter Ordnung mit konstanten Koeffizienten dar. Mit Hilfe der Z-Transformation (Doetsch /219/) ist es deshalb möglich, alle Zustandswahrscheinlichkeiten $p(u_1, 1+s_2)$ als Funktion von $p(0, 1+s_2)$ auszudrücken:

$$p(u_1,1+s_2) = p(0,1+s_2)\cdot\left\{\left[f+A_1\right]\beta_{u_1} - A_1\cdot\beta_{u_1-1}\right\} \quad (V.10)$$

mit $\beta_{\mu_1} = \dfrac{\alpha_1^{\mu_1} - \alpha_2^{\mu_1}}{\alpha_1 - \alpha_2}$; $f = \dfrac{\varepsilon_2}{\varepsilon_1}$; $A_1 = \dfrac{\lambda_1}{\varepsilon_1}$

$$\alpha_1, \alpha_2 = \frac{(1 + f + A_1) \pm \sqrt{(1 + f + A_1)^2 - 4 A_1}}{2}$$

Völlig analog ergibt sich für die anderen Bereiche

$\mu_1 \leq s_1$, $0 < \mu_2 \leq s_2$:

$$p(\mu_1, \mu_2) = p(0, \mu_2) \cdot \left\{ [f + A_1] \cdot \beta_{\mu_1} - A_1 \cdot \beta_{\mu_1 - 1} \right\} - \sum_{\xi = 0}^{\mu_1 - 1} p(\xi, \mu_2 + 1) \cdot \beta_{\mu_1 - \xi} \qquad (V.11)$$

und

$\mu_1 \leq s_1$, $\mu_2 = 0$:

$$p(\mu_1, 0) = p(0, 0) \cdot \left\{ A_1 \cdot \beta'_{\mu_1} - A_1 \beta'_{\mu_1 - 1} \right\} - \sum_{\xi = 0}^{\mu_1 - 1} p(\xi, 1) \cdot \beta'_{\mu_1 - \xi} \qquad (V.12)$$

mit $\beta'_{\mu_1} = \dfrac{(\alpha'_1)^{\mu_1} - (\alpha'_2)^{\mu_1}}{\alpha'_1 - \alpha'_2}$; $\alpha'_1, \alpha'_2 = \dfrac{(f + A_1) \pm \sqrt{(f + A_1)^2 - 4 A_1}}{2}$

$\mu_1 = 1 + s_1$, $0 \leq \mu_2 \leq s_2$:

$$f \cdot p(1 + s_1, \mu_2 + 2) - (1 + f + A_1) \cdot p(1 + s_1, \mu_2 + 1) + A_1 \cdot p(1 + s_1, \mu_2) = - A_1 \cdot p(s_1, \mu_2 + 1) \qquad (V.13)$$

Mit Hilfe der Gl. (V.10) bis (V.12) kann man alle Zustandswahrscheinlichkeiten $p(u_1, u_2)$ als Funktion des "Grundzustandes" $p(0, u_2)$ und "höherer" Werte $p(\eta, u_2 + 1)$ $(\eta = 0, 1, \ldots u_1 - 1)$ darstellen. Beginnt man deshalb mit dem Eckwert $u_2 = s_2 + 1$ und ersetzt systematisch die unbekannten Wahrscheinlichkeiten, so ist es möglich, alle Zustandswahrscheinlichkeiten $p(u_1, u_2)$ durch die Grenzwerte $p(0, u_2), \ldots, p(0, s_2 + 1)$ zu ersetzen. Die zweidimensionalen Zustandsbeziehungen können also auf ein eindimensionales Gleichungssystem zurückgeführt

werden (Schehrer /217/). Zusammen mit Gl. (V.13) und der Normierungsbedingung ergibt sich schließlich die gesuchte explizite Lösung für die Zustandswahrscheinlichkeiten.

$$\boxed{p(\mu_1, \mu_2) = \frac{A_1^{\mu_1}}{1 + A_1 \cdot \dfrac{1 - A_1^{s_1 + 1}}{1 - A_1}} \; \frac{\sum_{\varrho = \mu_2}^{1 + s_2} c_{\mu_1, \mu_2, \varrho} \cdot b_\varrho}{\sum_{\gamma = 0}^{1 + s_2} \sum_{\varrho = \gamma}^{1 + s_2} c_{\mu_1, \gamma, \varrho} \cdot b_\varrho}} \qquad (V.14)$$

$$c_{\mu_1, \mu_2, \varrho} = - \sum_{\xi = 0}^{\mu_1 - 1} c_{\xi, \mu_2 + 1, \varrho} \cdot \beta_{\mu_1 - \xi} \qquad \text{für } \mu_2 \neq \varrho ;\ \mu_2 \neq 0 ;\ \varrho \geq \mu_2 ;$$

$$c_{\mu_1, 0, \varrho} = - \sum_{\xi = 0}^{\mu_1 - 1} c_{\xi, 1, \varrho} \cdot \beta'_{\mu_1 - \xi} \qquad \mu_2 \neq \varrho ;\ \mu_2 = 0 ;$$

$$c_{\mu_1, \mu_2, \varrho} = (f + A_1) \cdot \beta_{\mu_1} - A_1 \cdot \beta_{\mu_1 - 1} \qquad \mu_2 = \varrho ;\ \mu_2 \neq 0 ;$$

$$c_{\mu_1, \mu_2, \varrho} = A_1 \cdot \beta'_{\mu_1} - A_1 \cdot \beta'_{\mu_1 - 1} \qquad \mu_2 = \varrho ;\ \mu_2 = 0 ;$$

$$c_{\mu_1, \mu_2, \varrho} = 0 \qquad \mu_2 > \varrho ;$$

$$b_{\mu_2} = \sum_{z_1 = \mu_2 + 1}^{1 + s_2} a_{\mu_2, z_1} \cdot \sum_{z_2 = z_1 + 1}^{1 + s_2} a_{z_1, z_2} \cdot \sum_{z_3 = z_2 + 1}^{1 + s_2} a_{z_2, z_3} \cdot \ldots \sum_{z_w = z_{w-1} + 1}^{1 + s_2} a_{z_{w-1}, z_w} ; \quad w = 1 + s_2 - \mu_2$$

$$a_{\mu_2, \varrho} = \frac{-A_1 \cdot c_{1 + s_1, \mu_2, \varrho} + (1 + f + v \cdot A_1) \cdot c_{1 + s_1, \mu_2 + 1, \varrho} - v \cdot c_{1 + s_1, \mu_2 + 1, \varrho} - A_1 c_{s_1, \mu_2 + 1, \varrho}}{A_1 \cdot c_{1 + s_1, \mu_2, \mu_2}}$$

$$v = \begin{cases} 1 & \text{für } 0 \leq \mu_2 < s_2 \\ 0 & \text{sonst} \end{cases}$$

5.3.2. Charakteristische Mittelwerte für Sekundärsysteme

Belastung:

$$Y_2 = \sum_{\mu_1 = 0}^{1 + s_1} \sum_{\mu_2 = 1}^{1 + s_2} p(\mu_1, \mu_2) = \frac{\lambda_1}{\varepsilon_2} \cdot \left[p(\mu_1 = s_1 + 1) - p(s_1 + 1, s_2 + 1) \right] \qquad (V.15)$$

Abgewiesener Verkehr:

$$R_2 = \frac{\lambda_1}{\varepsilon_2} \cdot p(1+s_1, 1+s_2) \qquad (V.16)$$

Wartebelastung

$$\Omega_2 = \sum_{u_1=0}^{1+s_1} \sum_{u_2=1}^{1+s_2} (u_2-1) \cdot p(u_1, u_2) \qquad (V.17)$$

Mittlere Wartezeit aller nichtabgewiesenen Anforderungen

$$w_{\gamma_2} = \frac{\Omega_2}{\gamma_2 \cdot \varepsilon_2} \qquad (V.18)$$

Mittlere Wartezeit aller Anforderungen, die im Sekundärsystem warten müssen

$$t_{w_2} = \frac{\Omega_2}{\lambda_1 \cdot \sum_{u_2=1}^{s_2} p(1+s_1, u_2)} \qquad (V.19)$$

5.3.3. Die Wartewahrscheinlichkeit für Sekundärsysteme

Als Wartewahrscheinlichkeit W_2 für das Sekundärsystem wird definiert:

W_2 ist die Wahrscheinlichkeit, daß eine eintreffende Anforderung warten muß, unter der Bedingung, daß sie dem Sekundärsystem angeboten wird. Damit gilt:

$$W_2 = \frac{\sum_{u_2=1}^{s_2} p(1+s_1, u_2)}{\sum_{u_2=0}^{s_2+1} p(1+s_1, u_2)} \qquad (V.20)$$

5.3.4. Die Wartezeitverteilungsfunktion für Sekundärsysteme

Als Wartezeitverteilungsfunktion $F_2(\leq t)$ für das Sekundärsystem wird definiert:

$F_2(\leq t)$ ist gleich der Wahrscheinlichkeit, daß eine beliebige Anforderung im Sekundärsystem höchstens die Zeit t warten muß, unter der Bedingung, daß sie dem Sekundärsystem überhaupt angeboten wird.

Zur Bestimmung der Verteilungsfunktion betrachtet man zunächst einen Testruf, der dem Sekundärsystem angeboten wird, wenn dort gerade u_2 Anforderungen auf Bedienung warten ($u_2 = 0, 1, \ldots, s_2+1$). Die Wahrscheinlichkeit, daß dieser Testruf höchstens die Zeit t warten muß, wird mit $F_{u2}(\leq t)$ bezeichnet.

Für die gesuchte Wartezeitverteilungsfunktion $F_2(\leq t)$ gilt damit:

$$F_2(\leq t) = \frac{\sum_{u_2=0}^{s_2+1} \lambda_1 \cdot dt \cdot p(s_1+1, u_2)\, F_{u_2}(\leq t)}{\sum_{u_2=0}^{s_2+1} \lambda_1 \cdot dt \cdot p(s_1+1, u_2)}$$

$$F_2(\leq t) = \frac{1}{p(u_1 = s_1+1)} \cdot \sum_{u_2=0}^{s_2+1} p(s_1+1, u_2) \cdot F_{u_2}(\leq t) \qquad (V.21)$$

5.3.4.1. Wartezeitverteilungsfunktion für den Testruf

$F_{u2}(\leq t)$ ist die Wahrscheinlichkeit, daß ein Testruf, der u_2 Anforderungen im Sekundärsystem antrifft, höchstens die Zeit t warten muß.

$F_{u2}(> t)$ ist deshalb die Wahrscheinlichkeit, daß der Testruf länger als die Zeit t warten muß. Drei Fälle werden bei der Berechnung dieser Wahrscheinlichkeit unterschieden:

$u_2 = 0$: der Testruf wird sofort bedient, deshalb gilt:

$$F_0(> t) = 0 \qquad (V.22)$$

$0 < u_2 \leq s_1$: der Testruf muß warten. Länger als die Zeit t muß der Testruf warten, wenn in dieser Zeit entweder keine, eine, oder höchstens (u_2-1) Anforderungen bedient werden und das System verlassen. Im letzten Fall sitzt der betrachtete Testruf zur Zeit t auf dem ersten Warteplatz, kann also frühestens zum Zeitpunkt (t + dt) bedient werden.

$$F_{u_2}(>t) = \sum_{\nu=0}^{u_2-1} \frac{(\varepsilon_2 t)^\nu}{\nu!} \cdot e^{-\varepsilon_2 t} \qquad (V.23)$$

(vgl. Abschnitt II.2.3: einige Eigenschaften von Prozessen mit negativ exponentiell verteilten Bedienungszeiten).

$u_2 = s_2 + 1$: der Testruf wird zwar dem Sekundärsystem angeboten, kann aber nicht gespeichert werden, da alle Warteplätze belegt sind. Deshalb gilt

$$F_{s_2+1}(>t) = 0 \qquad (V.24)$$

Die gesuchte Wahrscheinlichkeit $F_{u2}(\leq t)$ ergibt sich jeweils als Komplement zur Wahrscheinlichkeit, daß der Testruf länger als die Zeit t warten muß.

5.3.4.2. Wartezeitverteilungsfunktion für eine beliebige Anforderung

Aus den Gl. (V.21) bis (V.24) ergibt sich direkt die gesuchte Wartezeitverteilungsfunktion:

$$\boxed{F_2(\leq t) = 1 - \frac{e^{-\varepsilon_2 t}}{p(u_1 = s_1 + 1)} \cdot \sum_{u_2=1}^{s_2} p(s_1+1, u_2) \sum_{\nu=0}^{u_2-1} \frac{(\varepsilon_2 t)^\nu}{\nu!}} \qquad (V.25)$$

6. NUMERISCHE BEISPIELE, VERGLEICH MIT SIMULATIONSERGEBNISSEN

Im vorhergehenden Abschnitt wurde gezeigt, daß alle charakteristischen Größen, sowohl für das Eingangssystem, als auch für alle Primärsysteme exakt bestimmt werden können.
Alle angegebenen charakteristischen Werte für Sekundärsysteme sind Näherungswerte, da die tatsächlich angebotenen Über-

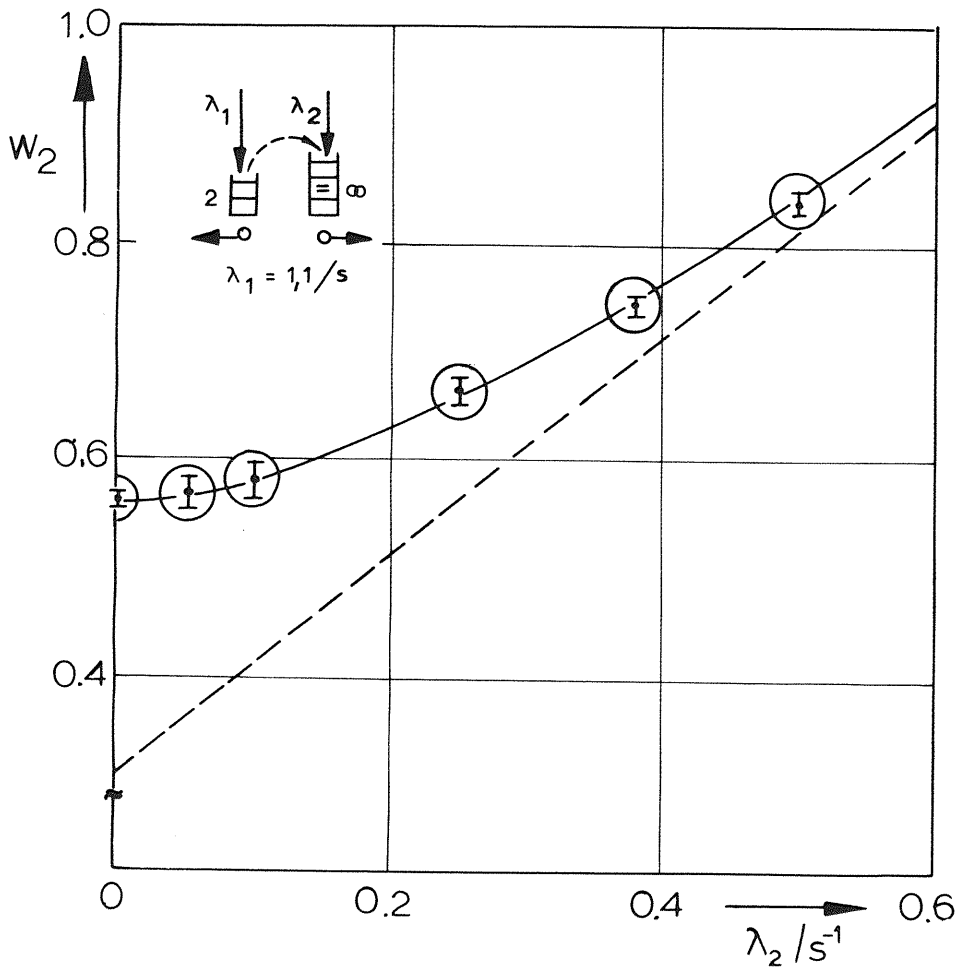


Bild V.6.: Wartewahrscheinlichkeit W_2 für ein Sekundärsystem, dem außer Überlaufverkehr auch Direktverkehr mit der Ankunftsrate λ_2 angeboten wird. Die Übertragungsrate ist für beide Kanäle gleich (ε = 1/s). Vergleich zwischen Rechnung (——) und Simulation (I ; statistische Sicherheit S = 95 %). Vernachlässigt man den speziellen Charakter des angebotenen Überlaufverkehrs, so ergibt sich die gestrichelte Linie (- - -).

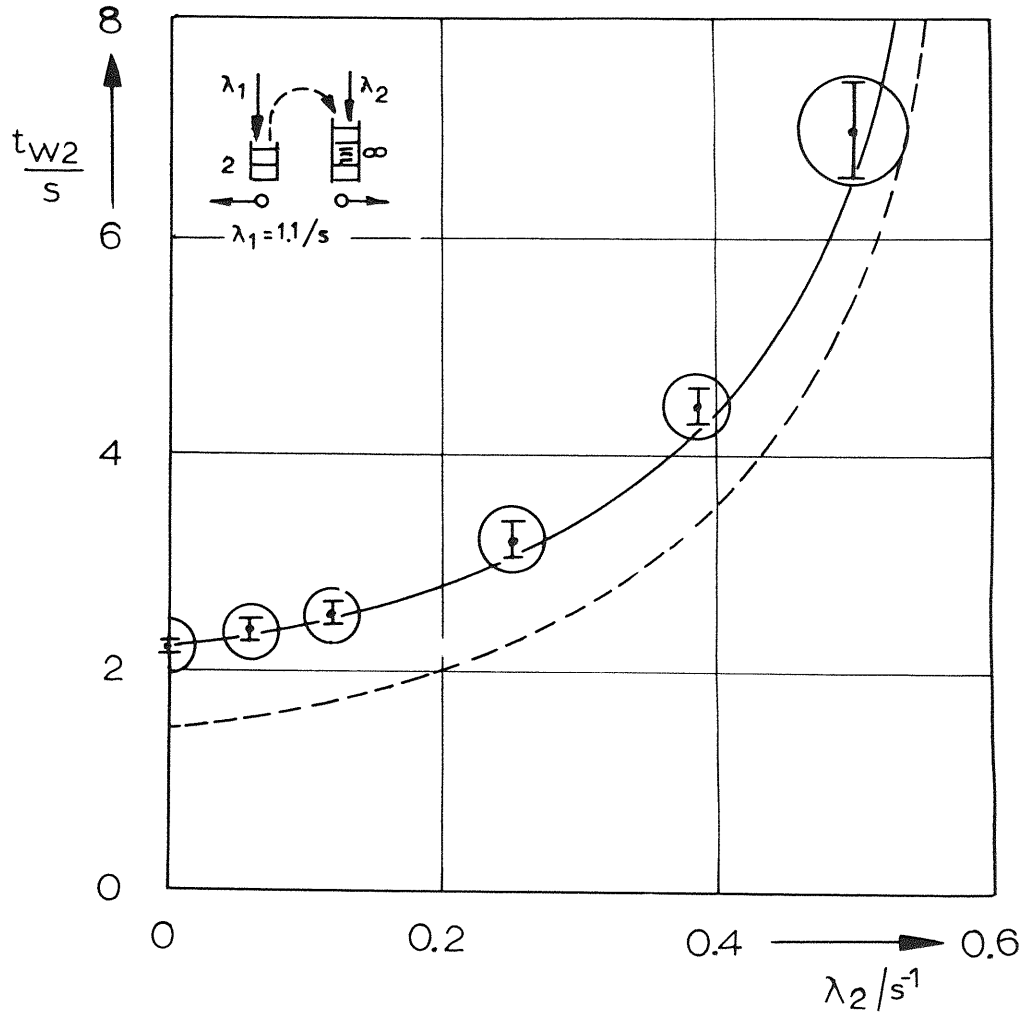


Bild V.7.: Mittlere Wartezeit für Anforderungen, die im Sekundärsystem warten müssen. Vergleich zwischen Rechnung (——) und Simulation (I ; statistische Sicherheit S = 95 %). Vernachlässigt man den speziellen Charakter des angebotenen Überlaufverkehrs, so ergibt sich die gestrichelte Kurve (- - -).

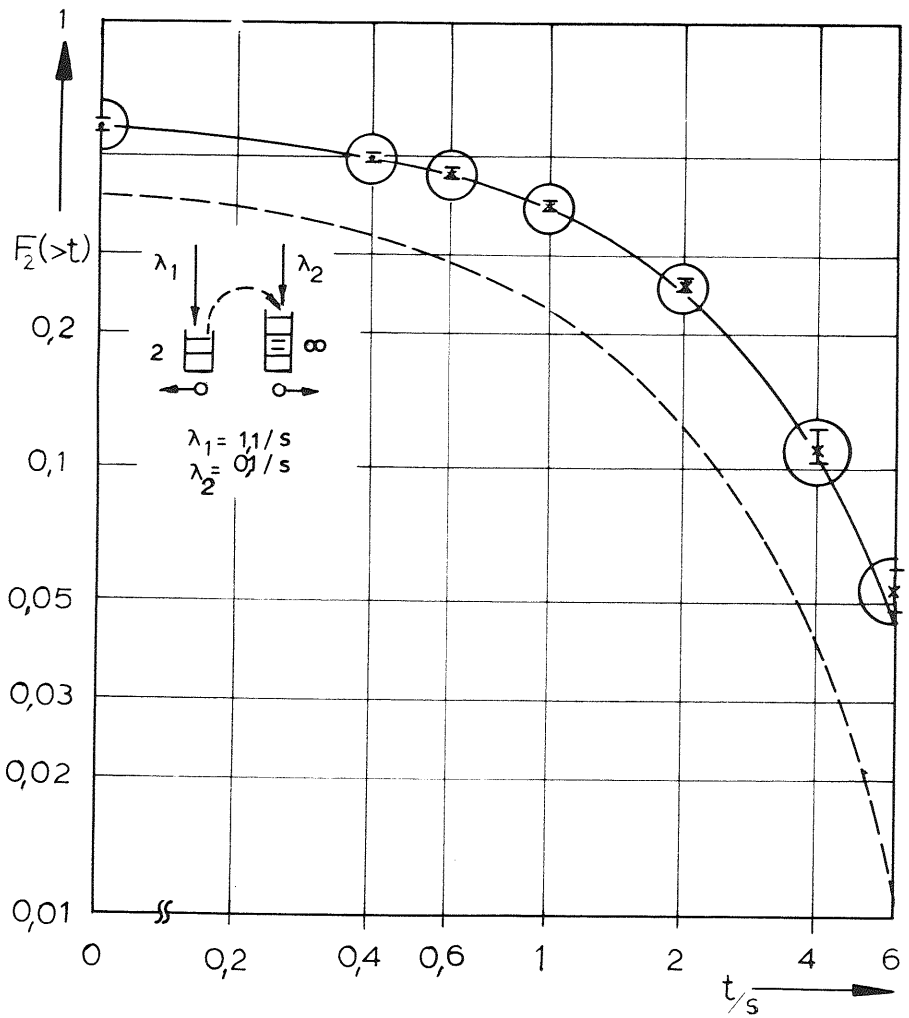


Bild V.8.: Komplement $F_2(>t)$ zur Wartezeitverteilungsfunktion aller Anforderungen, die dem Sekundärsystem angeboten werden. Vergleich zwischen Rechnung (——) und Simulation (I; statistische Sicherheit S = 95 %). Vernachlässigt man den speziellen Charakter des angebotenen Überlaufverkehrs, so ergibt sich die gestrichelte Kurve (- - -).

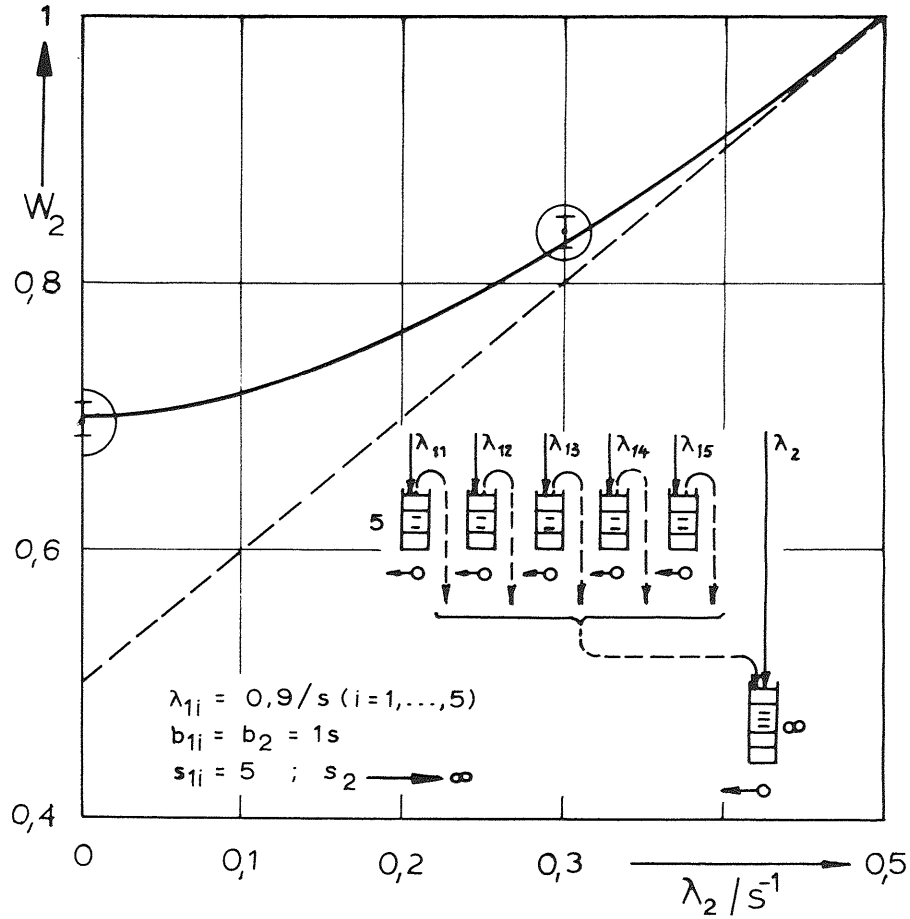


Bild V.9.: Wartewahrscheinlichkeit W_2 für ein Sekundärsystem, dem Überlaufverkehr von fünf Primärsystemen sowie Direktverkehr angeboten wird. Vergleich zwischen Rechnung (——) und Simulation (I; statistische Sicherheit S = 95 %). Vernachlässigt man den speziellen Charakter der angebotenen Überlaufverkehre, so ergibt sich die gestrichelte Kurve.

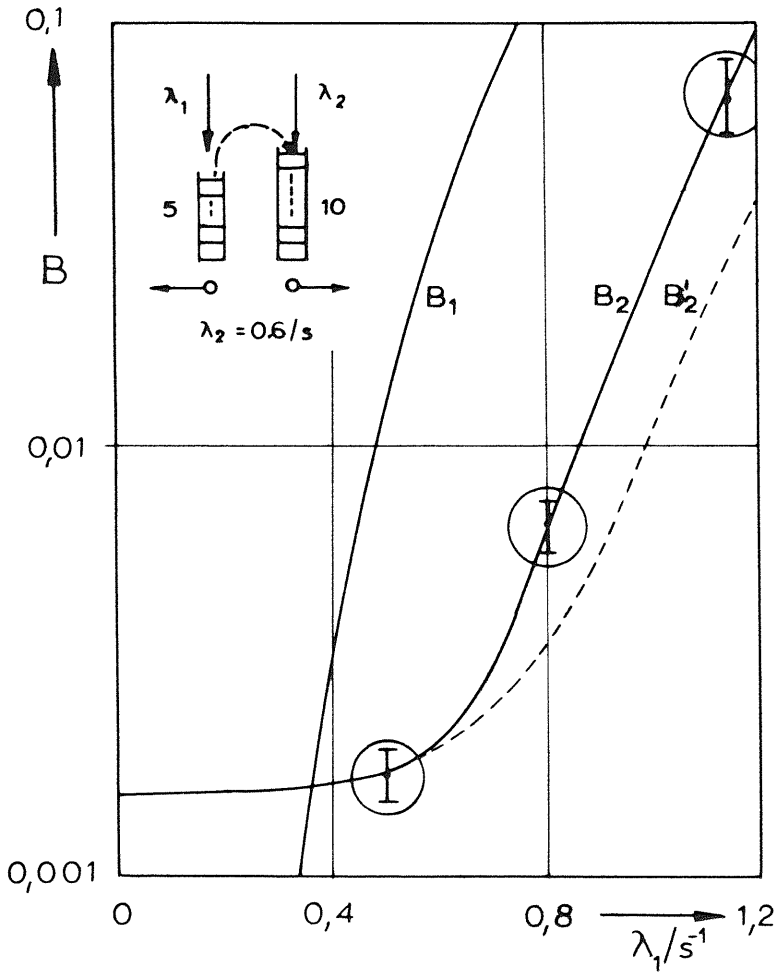


Bild V.10.: Verlustwahrscheinlichkeit für Anforderungen, die dem Primärsystem (B_1) bzw. dem Sekundärsystem (B_2) angeboten werden. Vergleiche zwischen Rechnung (——) und Simulation (I; statistische Sicherheit S = 95 %). Vernachlässigt man den speziellen Charakter des Überlaufverkehrs, so ergibt sich die gestrichelte eingezeichnete Verlustwahrscheinlichkeit B_2'.

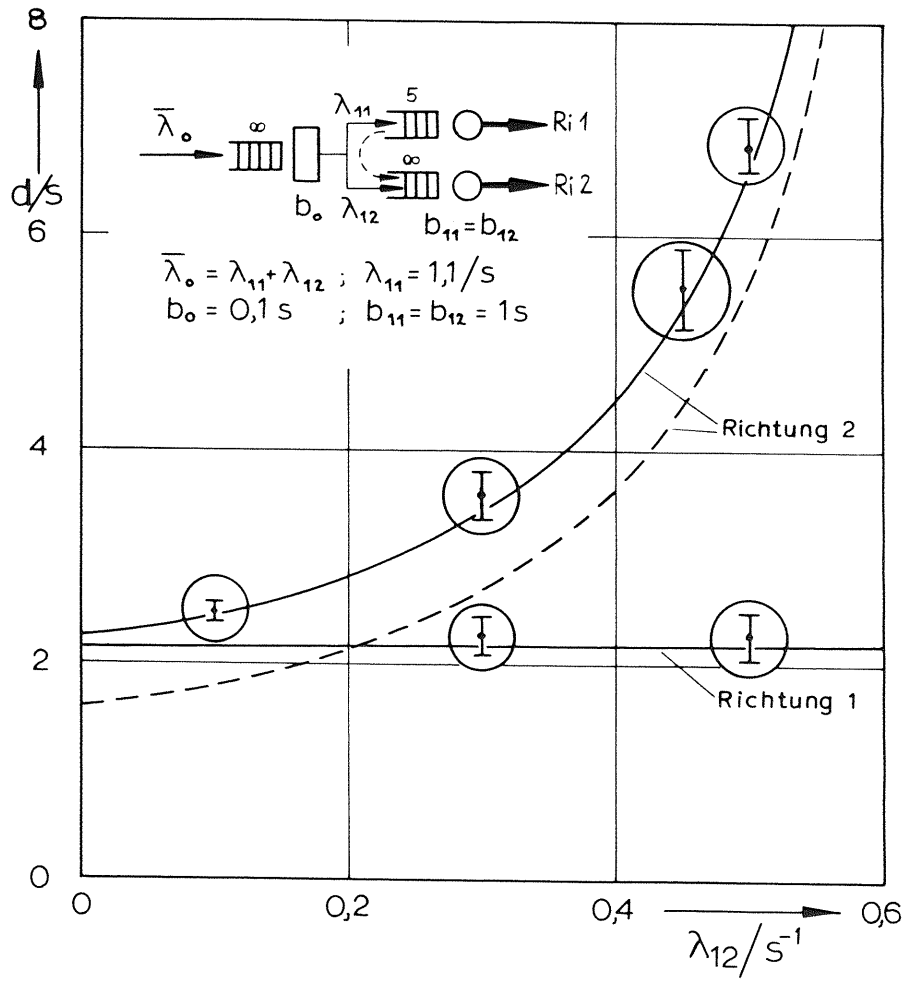
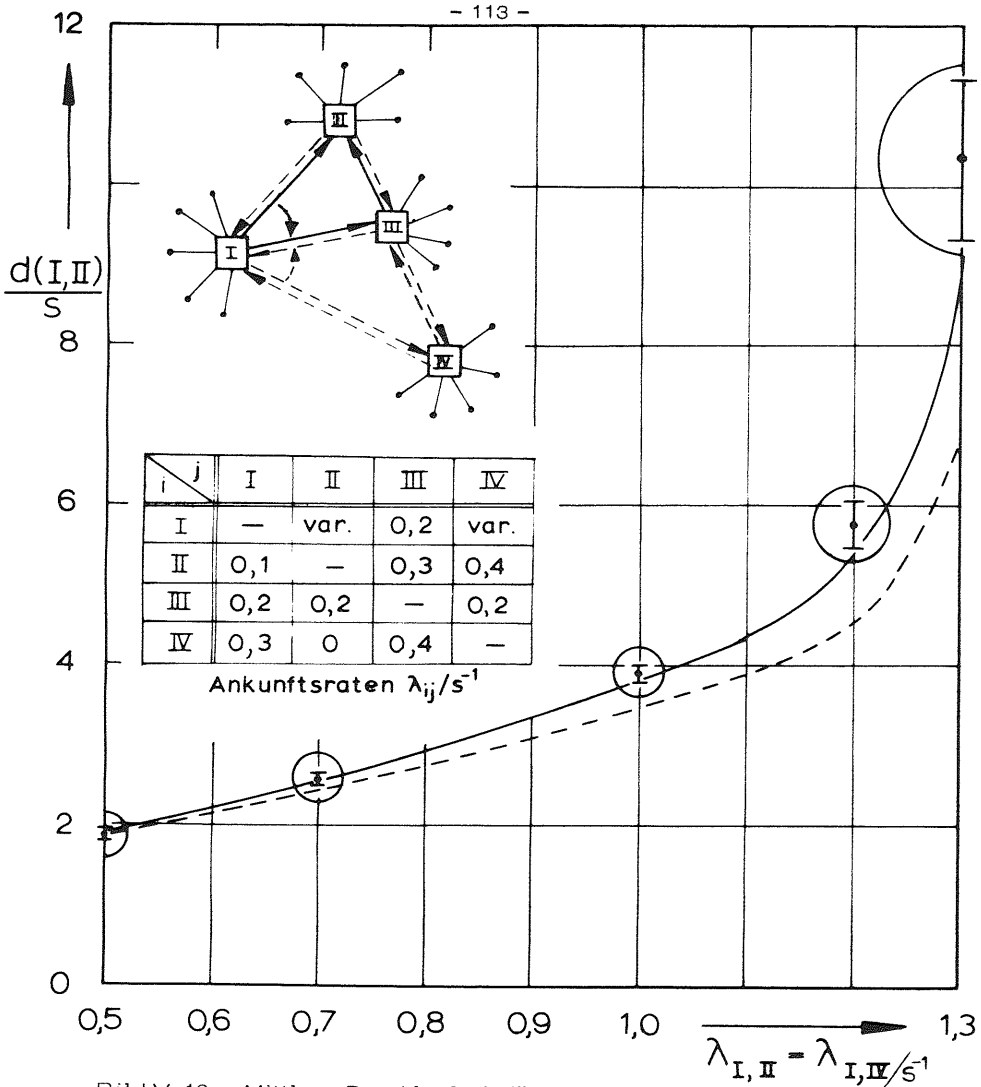


Bild V.11.: Gesamte mittlere Durchlaufzeiten von Nachrichten, die in Richtung 1 bzw. 2 übertragen werden. Vergleich zwischen Rechnung (——) und Simulation (I; statistische Sicherheit S = 95 %). Vernachlässigt man den speziellen Charakter des angebotenen Überlaufverkehrs, so ergibt sich die gestrichelte Kurve (– – –).



i \ j	I	II	III	IV
I	—	var.	0,2	var.
II	0,1	—	0,3	0,4
III	0,2	0,2	—	0,2
IV	0,3	0	0,4	—

Bild V.12.: Mittlere Durchlaufzeit für Nachrichten, die von Ursprungsknoten I zum Zielknoten II übertragen werden. Beispiel eines Netzes mit vier Knoten. Vergleich zwischen Rechnung (——) und Simulation (I; statistische Sicherheit S = 95 %). Vernachlässigt man den speziellen Charakter des Überlaufverkehrs, so ergibt sich die gestrichelte Kurve (- - -). Enderaten $\varepsilon_{ij} = 1/s^{-1}$.

laufverkehre durch einen fiktiven Verkehr ersetzt werden, der nur in den beiden wichtigsten Momenten exakt stimmt.

Mit Hilfe zweier Simulationsprogramme wurde deshalb eine große Anzahl von Systemen untersucht. Der Vergleich zwischen simulierten und berechneten Werten zeigt die Genauigkeit des angegebenen Näherungsverfahrens.

Einige typische Beispiele zeigen die Bilder V.6 bis V.12. Neben den berechneten und simulierten Ergebnissen sind zusätzlich gestrichelte Kurven eingetragen. Diese Kurven ergeben sich, wenn man bei der Berechnung der Sekundärsysteme den speziellen Charakter der Überlaufverkehre nicht berücksichtigt. Alle Diagramme zeigen, daß das tatsächliche Verhalten teilweise erheblich unterschätzt wird.

7. ZUSAMMENFASSUNG UND AUSBLICK

Netze mit verkehrsabhängiger Leitweglenkung sind Netzen ohne diese Möglichkeit aus Gründen der Wirtschaftlichkeit und Sicherheit überlegen.

Die dabei auftretenden Überlaufverkehre haben einen völlig anderen stochastischen Charakter als Direktverkehre. Sollen Fehlplanungen vermieden werden, so ist dieser spezielle Charakter bei der Dimensionierung der Netze zu berücksichtigen.

Die exakte Bestimmung des Verkehrsflusses in derartigen Netzen ist mit den heute zur Verfügung stehenden mathemati-

schen Hilfsmitteln nicht möglich.

In diesem Kapitel wurde nun ein Näherungsverfahren vorgestellt, das es erlaubt, wichtige charakteristische Größen einzelner Datenvermittlungen und ganzer Netze mit alternativer Leitweglenkung zu bestimmen. Zahlreiche Vergleiche zwischen berechneten und simulierten Ergebnissen zeigen die Genauigkeit des angegebenen Verfahrens.

Die vorliegenden Untersuchungen bezogen sich auf Netze mit einmaligem Überlauf. Bei großen Netzen könnte es jedoch sinnvoll sein, einen mehrfachen oder wechselseitigen Überlauf von Verkehren zuzulassen. Untersuchungen, das neue Verfahren in dieser Richtung zu erweitern, sind vorgesehen.

Neben den herkömmlichen Netzen mit "message-switching" bekommen Netze mit "packet-switching" eine immer größere Bedeutung. Für derartige Netze, bei denen pro Paket eine konstante Übertragungszeit notwendig ist, scheint eine Lösung - nach dem vorgestellten Prinzip - ebenfalls möglich zu sein.

Die vorgestellte Strategie der verkehrsabhängigen Leitweglenkung ist nur eine, mit geringem Aufwand realisierbare, Möglichkeit. Zahlreiche andere Strategien wurden in der Literatur vorgeschlagen und können nur mit Hilfe aufwendiger Simulationsprogramme untersucht werden. Gelänge es, auch in diesen Fällen genaue Näherungsverfahren zu finden, so könnten in Zukunft verschiedene Strategien in großem Umfang systematisch miteinander verglichen werden.

Ziel sämtlicher verkehrstheoretischen Untersuchungen ist nicht nur die Analyse des Verkehrsflusses, sondern vor allem die Synthese optimaler Strukturen. Durch das Einführen von Kostenfunktionen ist es zwar möglich, den Verkehrsfluß in kleinen Netzen optimal zu steuern, bei der Synthese großer Netze mit verkehrsabhängiger Leitweglenkung ergeben sich jedoch zahlreiche, bisher ungelöste Probleme.

VI. EINE SATELLITENVERMITTLUNG FÜR "PACKET - SWITCHING" MIT PRIORITÄTEN

1. PROBLEMSTELLUNG
2. BESCHREIBUNG DES VERKEHRSABLAUFS
3. MODELLBILDUNG
4. BERECHNUNG DER CHARAKT. GRÖSSEN
5. NUMERISCHE BEISPIELE
6. ZUSAMMENFASSUNG UND AUSBLICK

Satellitenvermittlungen bekommen in weitverzweigten Anschlußnetzen eine immer größere Bedeutung. Vorgestellt wird in diesem Kapitel eine Satellitenvermittlung für ein Netz mit "packet-switching" und Prioritäten. Berechnet werden mittlere Warte- und Durchlaufzeiten, individuell für jede Prioritätsklasse. Numerische Beispiele zeigen die verkehrstheoretischen Vorteile gegenüber dem "message-switching". Es wird ferner gezeigt, daß die Abschätzung der Verkehrsleistung mit bekannten Formeln für Prozesse mit Einzelankunft sehr ungenau ist.

1. PROBLEMSTELLUNG

In weitverzweigten Anschlußnetzen werden neben Multiplexern und Konzentratoren immer häufiger sog. Satellitenvermittlungen eingesetzt. Satellitenvermittlungen haben nicht nur die Aufgabe der Verkehrskonzentration, sondern übernehmen auch gewisse Teilaufgaben zur Entlastung der Datenvermittlungen (vgl. Abschnitt III.3.2, vermittlungstechnische Einrichtungen für das Anschlußnetz).

Neben dem bekannten "message-switching"-Verfahren scheint sich in neuerer Zeit immer mehr das "packet-switching" durchzusetzen: die einzelnen Nachrichten werden nicht mehr in einem Block beliebiger Länge, sondern in mehreren "Paketen" einheitlicher Länge übertragen (vgl. Abschnitt III.1.2: Netze mit Speichervermittlungen). Das Zerlegen der Nachrichten in einzelne Pakete kann durch die Datenendeinrichtungen selbst oder durch vermittlungstechnische Einrichtungen des Netzes erfolgen.

In diesem Kapitel wird eine Satellitenvermittlung betrachtet, die diese Aufgabe der Zerlegung übernimmt.

Zuerst wird der Verkehrsablauf beschrieben, d.h. es wird der Weg einer Nachricht vom Eintreffen bis zum Verlassen der Satellitenvermittlung verfolgt.

Je nach Dringlichkeit erhalten die einzelnen Nachrichten (und deshalb auch die Pakete) eine bestimmte Priorität. Die einzelnen Pakete haben eine Länge, die derart festgelegt wird, daß die technisch aufwendige Unterbrechnung von Paketen geringer Priorität nicht mehr sinnvoll ist (vgl. Abschnitt III.1.2).

Zur verkehrstheoretischen Beschreibung des Ablaufgeschehens genügt in diesem Fall nicht mehr die Annahme eines Prozesses mit Einzelankunft. Das Zerlegen einer Nachricht in einzelne Pakete entspricht vielmehr einem Prozeß mit Gruppenankunft (vgl. Abschnitt II.2.2: wichtige Ankunfts- und Endeprozesse).

Mittlere Warte- und Durchlaufzeiten werden, individuell für jede Prioritätsklasse bestimmt.

2. BESCHREIBUNG DES VERKEHRSABLAUFS

Von den Teilnehmern eintreffende Übertragungswünsche werden in einem Pufferspeicher gespeichert. Ist eine bestimmte Nachricht vollständig eingetroffen, so wird in der Eingangswarteschlange ein Indikator (Marke) gesetzt.
Diese Eingangswarteschlange wird vom Prozessor bearbeitet: die Nachrichten werden - sofern dies nicht schon früher geschehen ist - auf Übertragungsfehler geprüft und in x einzelne Pakete zerlegt. Je nach Dringlichkeit der Nachricht werden in einer der P parallelen Ausgangswarteschlangen x neue Indikatoren gesetzt. Die Anzahl dieser Marken ist gleich der Anzahl der Pakete, in welche die gesamte Nachricht zerlegt wurde.
Alle auf Übertragung wartenden Pakete werden in der Reihenfolge ihrer Dringlichkeit in Richtung Datenvermittlung abgesendet.

3. MODELLBILDUNG

3.1. Struktur und Betriebsart

Bild VI.1 zeigt ein Strukturmodell, welches dem im vorhergehenden Abschnitt beschriebenen Verkehrsablauf entspricht.

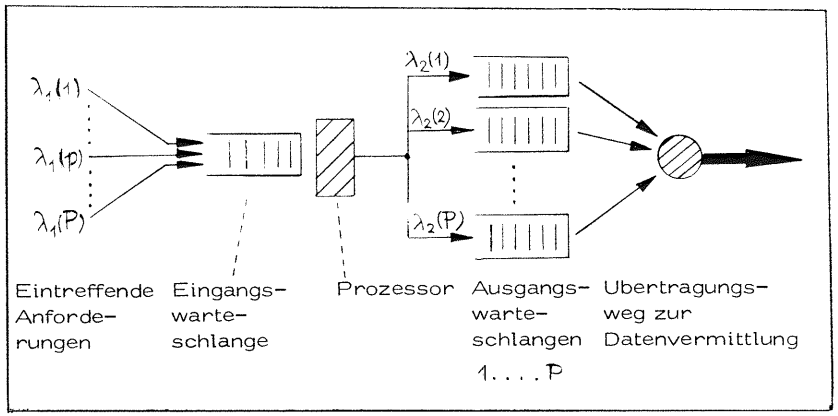


Bild VI.1: Strukturmodell der untersuchten Satellitenvermittlung

Die eintreffenden Anforderungen aller Prioritätsklassen (Nachrichten verschiedener Dringlichkeit) werden in einer gemeinsamen Eingangswarteschlange aufgenommen. Alle Anforderungen werden - ohne Berücksichtigung der Priorität - in der Reihenfolge ihres Eintreffens vom Prozessor bearbeitet. Nach dieser Bearbeitungsphase wird - entsprechend der Anzahl der Pakete, in die eine Nachricht zerlegt wurde - eine bestimmte Anzahl von Anforderungen in die Ausgangswarteschlange der betreffenden Priorität gesetzt.

Die Übertragung der in den verschiedenen Ausgangswarteschlangen wartenden Anforderungen erfolgt nach höchster Dringlichkeit mit nichtunterbrechender Priorität.

Alle Warteschlangen sind unbegrenzt, d.h. keine Anforderungen werden abgewiesen (dynamische Speicherplatzzuweisung!). Die wartenden Anforderungen einer bestimmten Warteschlange werden in der Reihenfolge ihres Eintreffens abgearbeitet (FIFO). Ausführliche Begründung beider Annahmen, siehe Abschnitt II.3.1 und VII. 3.1.

3.2. Verkehr

Eintreffende Anforderungen (Nachrichten) der Prioritätsklasse p haben negativ exponentiell verteilte Ankunftsabstände, d.h. der Ankunftsprozeß entspricht einem Poisson-Prozeß mit der Verteilungsfunktion:

$$A_{1p}(\leq t) = 1 - e^{-\frac{t}{a_1(p)}} = 1 - e^{-\lambda_1(p)\cdot t} \qquad \text{(VI.1)}$$

mit dem mittleren Ankunftsabstand a_1 (p) bzw. der Ankunftsrate λ_1(p).

Die Prozessor-Bedienungszeiten seien für alle Prioritätsklassen einheitlich negativ exponentiell verteilt (andere Modelle siehe Zusammenfassung und Ausblick):

$$B_{1p}(\leq t) = 1 - e^{-\frac{t}{b_1}} \qquad \text{(VI.2)}$$

mit dem Mittelwert b_1.

Das Zerlegen der Nachrichten in eine Anzahl von Paketen gleicher Länge wird durch die Wahrscheinlichkeit q_{fx} beschrieben: q_{fx} (p) ist die Wahrscheinlichkeit, daß eine eintreffende Nachricht der Priorität p in x Pakete aufgeteilt wird (x = 1 .... m (p); m (p) ist die maximale Anzahl von Paketen pro Nachricht). Die Übertragungszeit für jedes Paket ist konstant. Für die Verteilungsfunktion gilt deshalb

$$B_{2p}(\leq t) = \begin{cases} 0 & \text{für } 0 \leq t < b_2 \\ 1 & t \geq b_2 \end{cases} \qquad \text{(VI.3)}$$

4. BERECHNUNG DER CHARAKTERISTISCHEN GRÖSSEN

4.1. Prinzip des Lösungsweges

Das Modell in Bild VI.1 zeigt, daß man bei der Behandlung einer eintreffenden Nachricht folgende zwei Phasen unterscheiden kann:

- Bearbeitungsphase, d.h. das Warten in der Eingangswarteschlange und die Bearbeitung im Prozessor,

- Übertragungsphase, d.h. das Warten in einer der P parallelen Ausgangswarteschlangen und die Übertragung zur Datenvermittlung.

Die Berechnung der Wartezeit in der Eingangswarteschlange ist mit Hilfe Erlangs bekannter Wartezeitformel möglich. Der Ausgangsprozeß dieses ersten Systems ist nach dem Theorem von Burke (ausführliche Beschreibung siehe Abschnitt IV. 4.2.3) ebenfalls ein Poisson-Prozeß. Die Ankunftsabstände der einzelnen "Paketgruppen", die im Übertragungssystem eintreffen, sind deshalb negativ exponentiell verteilt.

Das Übertragungssystem entspricht einem Wartesystem mit Gruppenankunft, konstanten Bedienungszeiten und nichtunterbrechenden Prioritäten. Es wird gezeigt, wie mit Hilfe der Momentenmethode auch in diesem Fall mittlere Wartezeiten individuell für jede Prioritätsklasse und für jedes Paket einer Gruppe berechnet werden können.

Charakteristische Größen für die Satellitenvermittlung, wie z.B. die mittlere Wartezeit pro Paket oder die Durchlaufzeit einer Nachricht können damit bestimmt werden.

4.2. Die Bearbeitungsphase

Nach der bekannten Formel von Erlang (siehe z.B. Brockmeyer, Halstrøm und Jensen /218/ gilt für die mittlere Wartezeit

einer beliebigen Nachricht:

$$\boxed{w_1(p) = \frac{1}{\lambda_1} \frac{A_1^2}{1 - A_1}} \qquad \text{(VI.4)}$$

wo

$$\lambda_1 = \sum_{p=1}^{P} \lambda_1(p)$$

$$A_1 = \lambda_1 \cdot b_1$$

4.3. Die Übertragungsphase

4.3.1. Beschreibung der Gruppenankunft

Der Ankunftsabstand der einzelnen Gruppen entspricht einem Poisson-Prozeß mit der Verteilungsfunktion

$$A_{2p}(\leq t) = A_{1p}(\leq t) = 1 - e^{-\lambda_1(p)t} \qquad \text{(VI.5)}$$

Die eintreffende Gruppe hat mit der beliebig vorgebbaren Wahrscheinlichkeit $q_x(p)$ die Größe x (x = 1.... m (p)). Bezeichnet man mit m (p) die maximale Gruppengröße so gilt für die mittlere Gruppengröße der Priorität p:

$$\bar{x}(p) = \sum_{x=1}^{m(p)} x \cdot q_x(p) \qquad \text{(VI.6)}$$

und deshalb für die mittlere Anzahl pro Zeiteinheit eintreffender Anforderungen (Pakete)

$$\lambda_2(p) = \lambda_1(p) \cdot \bar{x}(p) \qquad \text{(VI.7)}$$

Nach Definition gilt ferner für das Angebot pro Prioritätsklasse

$$A_2(p) = \lambda_2(p) \cdot b_2 \qquad \text{(VI.8)}$$

4.3.2. Mittlere Wartezeiten und Wartebelastung

Zum Zeitpunkt der Beobachtung treffe eine Gruppe der Größe x und der Priorität p ein. Untersucht wird das Schicksal der Υ-ten Anforderung der Gruppe, der sog. Testruf ($\Upsilon = 1,\ldots x$).

Die mittlere Wartezeit $w_2(p, \Upsilon)$ des Testrufs setzt sich aus folgenden vier Anteilen zusammen (vgl. Bild VI.2).

- Erwartungswert der Restbedienungszeit b_R. Mit Wahrscheinlichkeit $A_2(i)$ trifft die Gruppe (und deshalb auch der Testruf auf ein Paket der Klasse i, das augenblicklich übertragen wird. Da das Eintreffen des Testrufs bei einem Poisson-Prozeß zu jedem beliebigen Zeitpunkt gleichwahrscheinlich ist, ist die Restbedienungszeit $b_2 / 2$ (ausführliche Behandlung von Restbedienungszeizeiten siehe Abschnitt VII. 4.1.1). Insgesamt gilt deshalb für die mittlere Restbedienungszeit:

$$b_R(p) = b_R = \sum_{i=1}^{P} \frac{b_2}{2} \cdot A_2(i) \qquad \text{(VI.9)}$$

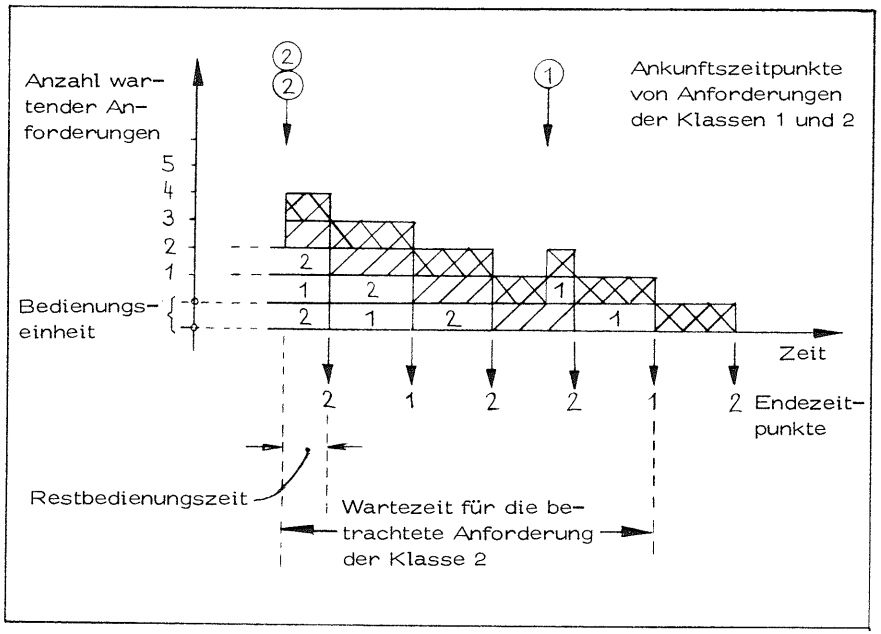


Bild VI.2: Einfaches Beispiel für das Schicksal und die Wartezeit einer Anforderung bei Gruppenankunft

. Wartezeit w_I (p). Pro Prioritätsklasse warten im Mittel Ω (i) Anforderungen, jede benötigt die Übertragungszeit b_2. Die mittlere Wartezeit für den Testruf durch derartige Anforderungen ist damit:

$$w_I(p) = \sum_{i=1}^{p} \Omega(i) \cdot b_2 \qquad \text{(VI.10)}$$

. Wartezeit w_{II} (p). Diese Wartezeit entsteht dadurch, daß der betrachtete Testruf das ν-te Paket der eintreffenden Gruppe ist:

$$w_{II}(p,\nu) = w_{II}(\nu) = (\nu - 1) \cdot b_2 \qquad \text{(VI.11)}$$

. Wartezeit w_{III} (p). Während der gesamten Wartezeit w_2 (p, ν) des Testrufs treffen Anforderungen aller Prioritätsklassen mit der Rate λ_2 (i) ein. Alle Anforderungen der Klassen 1 bis (p-1) werden vor dem betrachteten Testruf bedient. Die zusätzliche Wartezeit ist deshalb

$$w_{III}(p,\nu) = \sum_{i=1}^{p-1} \lambda_2(i) \cdot w_2(p,\nu) \cdot b_2 \qquad \text{(VI.12)}$$

4.3.2.1. Zusammenhang zwischen den Wartezeiten von Anforderungen derselben Gruppe

Mit den Gl. (VI.9 bis VI.12) gilt für die mittlere Wartezeit eines Testrufs, der am Anfang einer Gruppe eintrifft:

$$w_2(p,1) = \frac{b_2}{2} \cdot \sum_{i=1}^{p} A_2(i) + b_2 \cdot \sum_{i=1}^{p} \Omega(i) + w_2(p,1) \cdot \sum_{i=1}^{p-1} A_2(i) \qquad \text{(VI.13)}$$

und nach einer Umformung:

$$w_2(p,1) = \frac{b_2}{2} \cdot \frac{\sum_{i=1}^{p} A_2(i) + 2 \cdot \sum_{i=1}^{p} \Omega(i)}{1 - \sum_{i=1}^{p-1} A_2(i)}$$

Entsprechend ergibt sich für (ν = 2):

$$w_2(p,2) = \frac{b_2}{2} \cdot \frac{\sum_{i=1}^{p} A_2(i) + 2\sum_{i=1}^{p} \Omega(i) + 2}{1 - \sum_{i=1}^{p-1} A_2(i)}$$

und deshalb

$$w_2(p,2) = w_2(p,1) + \frac{b_2}{1 - \sum_{i=1}^{p-1} A_2(i)}$$

Durch vollständige Induktion ergibt sich schließlich der Zusammenhang zwischen der Wartezeit der ersten und ν-ten Anforderung eines Blocks:

$$\boxed{w_2(p,\nu) = w_2(p,1) + \frac{(\nu - 1) \cdot b_2}{1 - \sum_{i=1}^{p-1} A_2(i)}} \qquad \text{(VI.14)}$$

Die Wartezeiten aller Anforderungen eines Blocks können direkt berechnet werden, wenn es gelingt, die Wartezeit der ersten Anforderung zu bestimmen.

4.3.2.2. Zusammenhang zwischen den Wartezeiten und der Wartebelastung

Pro Zeiteinheit treffen $\lambda_1(p)$ Gruppen im System ein. Mit Wahrscheinlichkeit $q_x(p)$ ist die Gruppengröße x, d.h. die gesamte mittlere Wartezeit der x Anforderungen ist mit Gl. (VI.14):

$$\sum_{\nu=1}^{x} w_2(p,\nu) = x \cdot w_2(p,1) + \frac{b_2}{1 - \sum_{i=1}^{p-1} A_2(i)} \cdot \binom{x}{2}$$

Die Wartebelastung, d.h. die mittlere Anzahl wartender Anforderung der Prioritätsklasse p ist deshalb:

$$\Omega(p) = \sum_{x=1}^{m(p)} \lambda_1(p) \left[x \cdot w_2(p,1) + \frac{b_2}{1 - \sum_{i=1}^{p-1} A_2(i)} \binom{x}{2} \right] q_x(p)$$

Nach einer Vereinfachung des Ausdrucks ergibt sich schließlich

$$\boxed{\Omega(p) = w_2(p,1) \cdot \lambda_2(p) + \frac{\lambda_1(p) \cdot b_2}{1 - \sum_{i=1}^{p-1} A_2(i)} \sum_{x=1}^{m(p)} \binom{x}{2} q_x(p)} \qquad \text{(VI.15)}$$

4.3.2.3. Rekursionsformel für die mittlere Wartezeit

Setzt man Gl. (VI.15) in Gl. (VI.13), so ergibt sich die mittlere Wartezeit der ersten Anforderung einer Gruppe

$$\boxed{w_2(p,1) = b_2 \cdot \frac{\frac{1}{2}\sum_{i=1}^{p} A_2(i) + \sum_{i=1}^{p-1} \Omega(i) + \frac{\lambda_1(p) \cdot b_2}{1 - \sum_{i=1}^{p-1} A_2(i)} \sum_{x=1}^{m(p)} \binom{x}{2} q_x(p)}{1 - \sum_{i=1}^{p} A_2(i)}} \qquad \text{(VI.16)}$$

Diese Gleichung ist eine Rekursionsformel, aus welcher im Zusammenhang mit der Formel für die Wartebelastung (VI.15) - beginnend bei der höchsten Prioritätsklasse- die mittlere Wartezeit aller ersten Anforderungen berechnet werden kann.

Mit Hilfe von Gl. (VI.14) kann deshalb auch die mittlere Wartezeit jeder beliebigen Anforderung einer Gruppe bestimmt werden.

4.3.2.4. Die mittlere Wartezeit einer Anforderung (Paket)

Pro Zeiteinheit treffen im Mittel $\lambda_2(p)$ Anforderungen (Pakete) der Prioritätsklasse p ein. Die mittlere Wartezeit einer Anforderung dieser Klasse beträgt deshalb

$$w_{2P}(p) = \frac{\Omega(p)}{\lambda_2(p)} \qquad \text{(VI.17)}$$

Berechnung der Wartebelastung siehe Gl. (VI.15)und (VI.16).

4.3.2.5. Die mittlere Durchlaufzeit einer Gruppe (Nachricht)

Die Durchlaufzeit einer bestimmten Gruppe, d.h. einer bestimmten Nachricht, ist gleich der Zeitspanne zwischen dem Eintreffen der Gruppe in einer Ausgangswarteschlange und dem Zeitpunkt, zu welchem die letzte Anforderung (Paket) die Satellitenvermittlung verlassen hat. Für die mittlere Durchlaufzeit pro Prioritätsklasse p gilt deshalb:

$$\boxed{d_{2N}(p) = \sum_{x=1}^{m(p)} \left[w_2(p,x) + b_2 \right] \cdot q_x(p)} \qquad \text{(VI.18)}$$

Berechnung der Wartezeit w_2 (p,x) siehe Gl. (VI.14) und (VI.16).

4.4. Die mittlere Durchlaufzeit einer Nachricht der Prioritätsklasse p durch die Satellitenvermittlung

Die mittlere Durchlaufzeit d_N (p) durch die gesamte Satellitenvermittlung setzt sich aus der Warte- und Bedienungszeit der "Bearbeitungsphase" und der Durchlaufzeit der "Übertragungsphase" zusammen (vgl. Abschnitt VI. 4.1: Prinzip des Lösungsweges). Deshalb gilt

$$\boxed{d_N(p) = w_1(p) + b_1 + d_{2N}(p)} \qquad \text{(VI.19)}$$

Die Wartezeit in der Eingangswarteschlange w_1 (p) ist durch Gl. (VI.4) gegeben, die Durchlaufzeit d_{2N} (p) aus der Gl. (VI.18) bestimmbar.

5. NUMERISCHE BEISPIELE

In den Bildern VI. 3 bis VI.5 sind einige interessante Ergebnisse zusammengestellt. Aus Gründen der Übersichtlichkeit wurden bei allen Beispielen zwei Prioritätsklassen angenommen.

5.1. Vergleich zwischen "packet- " und "message-switching"

Bild VI.3 zeigt die mittlere Durchlaufzeit d_N (1) und Wartezeit w_2 (1) für Nachrichten der Prioritätsklasse 1 in einer Satellitenvermittlung. Der Vergleich der Kurven zeigt deutlich, daß Warte- und Durchlaufzeiten bei "packet-switching" erheblich geringer sein können, selbst wenn man annimmt, daß bei "message-switching" keine Bearbeitungszeiten notwendig sind ($b_1 = 0$).

5.2. Mittlere Wartezeit einzelner Pakete

Bild VI.4 zeigt ein Beispiel für die mittlere Wartezeit $w_2(2,\nu)$ eines Pakets Nr. ν einer bestimmten Nachricht der Prioritätsklasse 2.

Gestrichelt eingezeichnet sind Ergebnisse für die mittlere Wartezeit eines Pakets, wenn man fälschlicherweise annimmt, daß die Anforderungen entsprechend einem Prozeß mit Einzelankunft eintreffen. In beiden Fällen ist die mittlere Ankunftsrate beider Prioritätsklassen gleich.

Das Diagramm zeigt, daß mit Hilfe der bekannten Formeln für Prozesse mit Einzelankunft die Abschätzung der Wartezeiten sehr ungenau wird. Besonders ungenau wird die Abschätzung der Warte- und Durchlaufzeiten einzelner Nachrichten, da diese nur von der Wartezeit des letzten Pakets abhängen.

5.3. Verzögerung einzelner Pakete einer Nachricht

Bild VI.5 zeigt die Verzögerung $v_2(\nu)$ der Pakete Nr. ν = 2,3 und 4 einer Nachricht der Prioritätsklasse 2 im Vergleich zum ersten Paket. Erwartungsgemäß ist diese Verzögerung stark abhängig von der Anzahl eintreffender Anforderungen der Prioritätsklasse 1 (Ankunftsrate $\lambda_2(1)$).

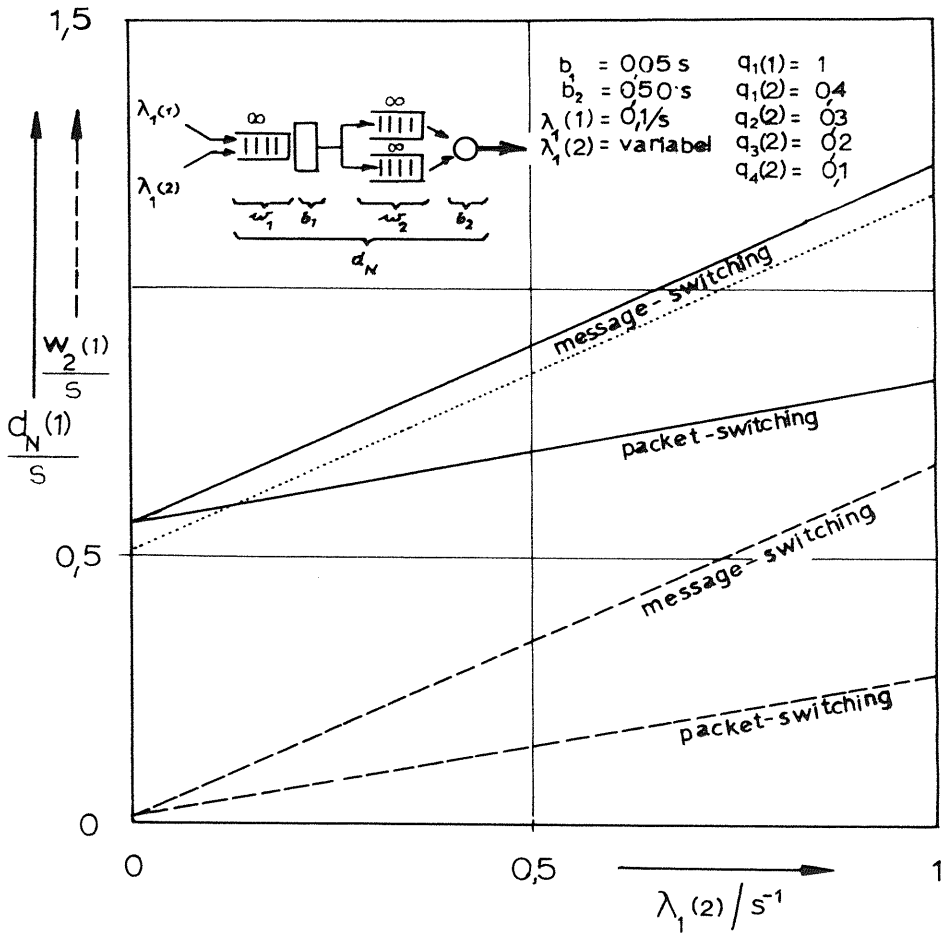


Bild VI.3.: Gesamte mittlere Durchlaufzeit $d_N(1)$ und mittlere Wartezeit $w_2(1)$ für Nachrichten der Prioritätsklasse 1. Vergleich zwischen "packet-" und "message-switching". Die punktierte Linie zeigt die Durchlaufzeit $d_N(1)$ bei "message-switching", wenn die Bearbeitungszeit b_1 vernachlässigbar klein ist.

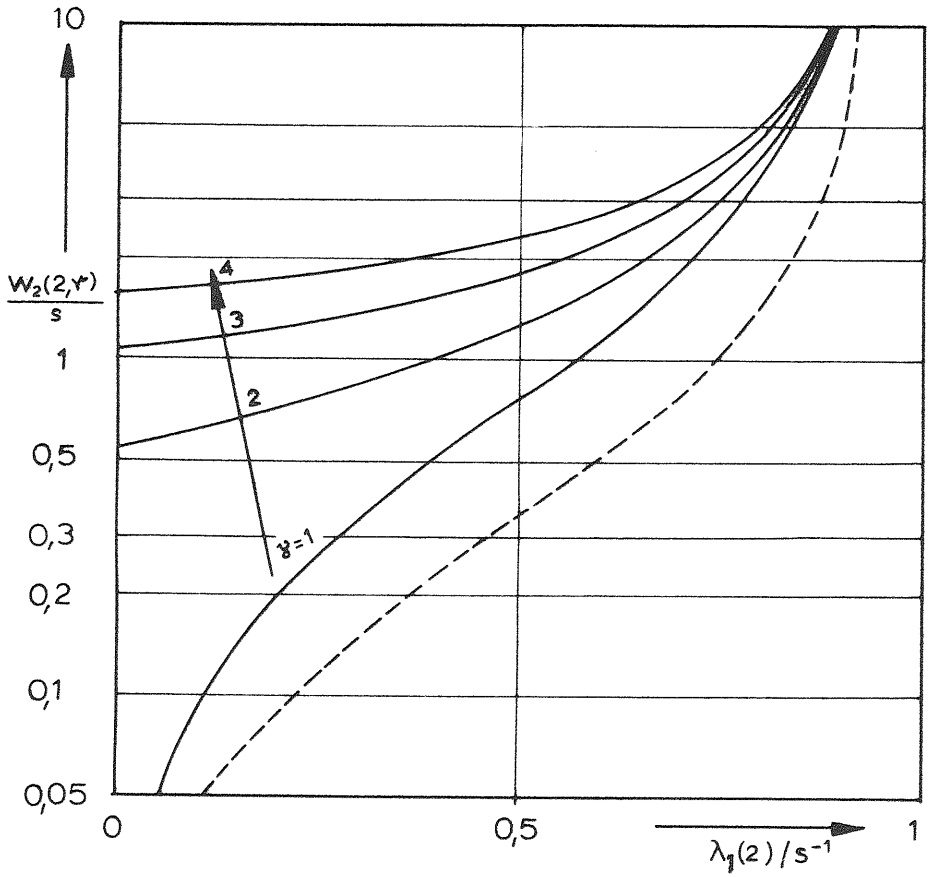


Bild VI.4.: Mittlere Wartezeit $w_2(2,\nu)$ von Paket Nr. ν ($\nu = 1,2,3,4$) einer bestimmten Nachricht der Prioritätsklasse 2. Die gestrichelte Kurve entspricht einem Prozeß mit Einzelankunft bei gleicher Ankunftsrate. Struktur und Verkehrsparameter wie im vorhergehenden Bild.

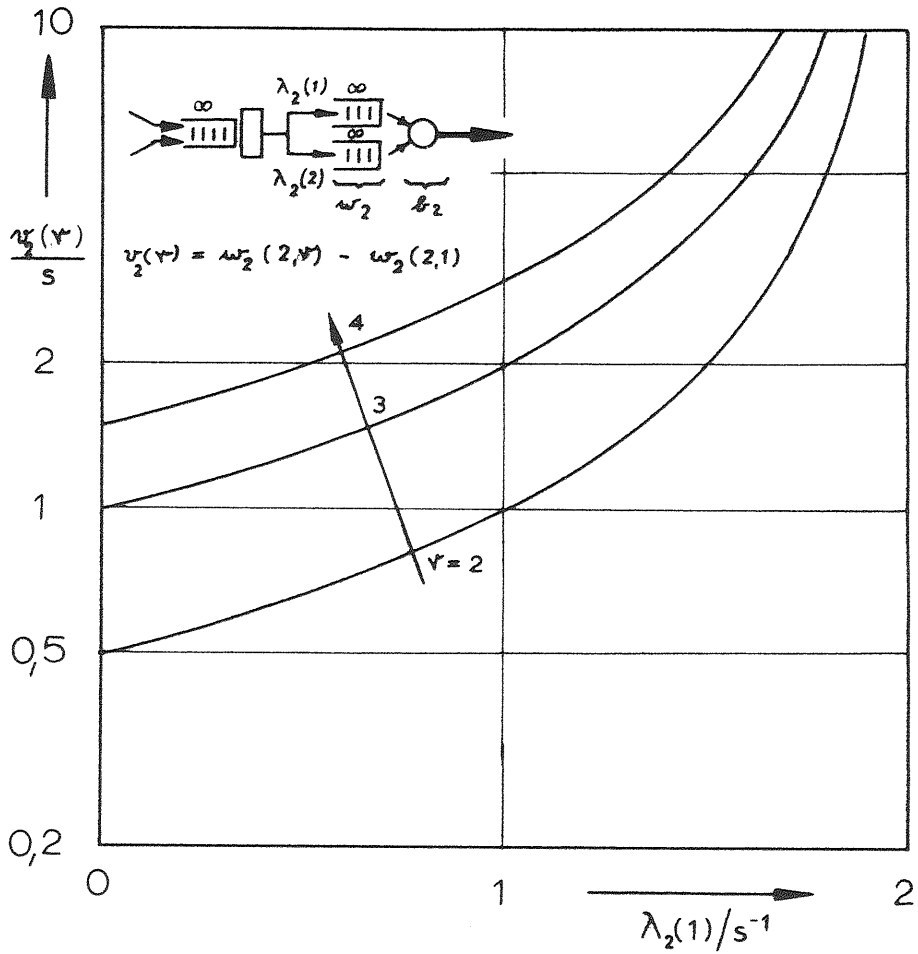


Bild VI.5.: Verzögerung $v_2(\nu)$ von Paket Nr. ν im Vergleich zum ersten Paket einer Nachricht der Prioritätsklasse 2. Die Verzögerung ist unabhängig von der Ankunftsrate $\lambda_2(2)$ der betrachteten Prioritätsklasse 2; die Übertragungszeit b_2 beträgt pro Paket 0,5 Sekunden

6. ZUSAMMENFASSUNG UND AUSBLICK

In diesem Kapitel wurde der Verkehrsfluß in einer Satellitenvermittlung für "packet-switching" mit Prioritäten untersucht. Verteilung und mittlere Anzahl von Paketen pro eintreffende Nachricht sind für jede Prioritätsklasse beliebig vergebbar.

Individuell pro Prioritätsklasse wurden Formeln abgeleitet für die Wartebelastung sowie mittlere Warte- und Durchlaufzeiten für einzelne "Pakete" oder ganze Nachrichten. Numerische Beispiele zeigen deutlich, daß die mittleren Durchlaufzeiten für dringliche Nachrichten kürzer sind als beim herkömmlichen "message-switching".

Für die Verteilungsfunktion der Prozessor- und Bearbeitungszeit wurde der mathematisch am einfachsten zu behandelnde Fall einer negativ exponentiellen Verteilungsfunktion angenommen. Andere Typen von Verteilungsfunktionen sind denkbar und sollen in Zukunft untersucht werden.

Bei der Berechnung wurde ferner angenommen, daß alle Pakete eine einheitliche Länge haben. In diesem Fall wird das letzte Paket einer bestimmten Nachricht, wenn erforderlich, mit "Leerbits" aufgefüllt. Sieht man für dieses letzte Paket eine variable Länge vor, so scheint - nach demselben Lösungsprinzip - ebenfalls die Berechnung der charakteristischen Größen möglich zu sein. Ganz sicher erlauben jedoch die vorgestellten Formeln auch bei variabler Länge eine gute Abschätzung der Durchlaufzeiten.

Bei einem anderen Typ von Satellitenvermittlungen werden die

eintreffenden Nachrichten nicht zuerst vollständig gespeichert und dann erst behandelt, sondern während der gesamten Empfangszeit überwacht und stückweise bearbeitet. Die mittleren Durchlaufzeiten von Nachrichten werden deshalb geringer, der "technische" Aufwand jedoch größer. Auch in diesem Fall scheint eine Berechnung der charakteristischen Größen möglich zu sein und soll in einer neuen Arbeit untersucht werden.

VII UNTERBRECHUNGS-DISTANZ PRIORITÄTEN

1. PROBLEMSTELLUNG
2. BESCHREIBUNG VON UD-PRIORITÄTEN
3. MODELLBILDUNG
4. BERECHNUNG CHARAKT. GRÖSSEN
5. NUMERISCHE BEISPIELE
6. ZUSAMMENFASSUNG UND AUSBLICK

Alle modernen Echtzeit-Datenverarbeitungsanlagen benützen sinnvolle Kombinationen rein unterbrechender und nichtunterbrechender Prioritäten. Die Vielzahl der möglichen Kombinationen wird einheitlich beschrieben (UD-Prioritäten) und einheitlich mathematisch behandelt. Numerische Beispiele für Wartezeiten sowie für die Anzahl der Unterbrechungen verdeutlichen die Vorteile von UD-Prioritäten.

1. PROBLEMSTELLUNG

Die wichtigste Forderung bei Echtzeit-Datenverarbeitung ist die schnelle Reaktion des Systems auf besonders dringliche Anforderungen. Dies ist möglich mit Hilfe rein unterbrechender Prioritäten (vgl. Abschnitt II. 3.2).

Jede Unterbrechung erfordert jedoch einen gewissen Aufwand an Verwaltungsarbeit:

- die Unterbrechungsursache muß festgestellt werden,
- die Registerinhalte müssen "gerettet" werden,
- die Warteschlangen werden neu organisiert usw.

Damit nun diese zusätzliche Systembelastung möglichst gering bleibt, benützen alle modernen Echtzeit-Datenverarbeitungsanlagen sinnvolle Kombinationen sofort unterbrechender und nichtunterbrechender Prioritäten.

Bild VII.1 zeigt ein kleines, jedoch typisches Beispiel für Prioritäten bei Prozeßrechnern:

Für spezielle unaufschiebbare Anforderungen, wie z.B. Alarme (Klasse 1) und Signale zur Prozeß-Steuerung (Klasse 2) benötigt man sofort unterbrechende Prioritäten. Daneben gibt

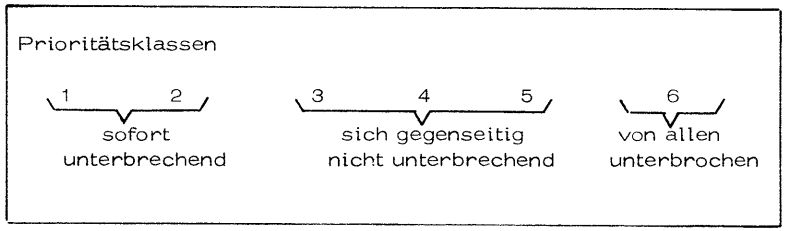


Bild VII.1: Typisches Beispiel für Prioritäten zur Steuerung des Ablaufgeschehens bei Prozeßrechnern.

es eine große Anzahl von Auswerteprogrammen, die zwar unterschiedliche Priorität haben, eine gegenseitige Unterbrechung aber nicht rechtfertigen (Klasse 3,4 und 5). Schließlich gibt es noch die Routineprogramme, die immer sofort unterbrochen werden, wenn eine Anforderung höherer Priorität eintrifft.

Je nach Prioritätsklasse haben wir also einen bestimmten

Abstand zur nächsten Prioritätsklasse, die unterbrochen wird, die sog. Unterbrechungs-Distanz (vgl. Bild VII.2).

Neue Anforderung der Klasse	unterbricht nicht Klasse	unterbricht Klasse	Unterbrechungs-Distanz
1	1	2,3,4,5,6	1
2	1,2	3,4,5,6	1
3	1,2,3,4,5	6	3
4	1,2,3,4,5	6	2
5	1,2,3,4,5	6	1
6	1,2,3,4,5,6	-	-

Bild VII.2: Unterbrechungs-Distanz zum Prioritätsbeispiel im vorhergehenden Bild VII.1.

Ein weiteres interessantes Beispiel ist die Prioritätsteuerung bei der Speichereinheit des elektronischen Datenvermittlungssystems EDS (siehe Gabler /100/): die Prioritäten können so gesteuert werden, daß z.B. eine Priorität mit dem Rang 21 erst von einer Priorität mit dem Rang 16 und dringlicher unterbrochen werden kann (Priorität 1 hat höchste Dringlichkeit). Die Unterbrechungs-Distanz ist also in diesem Fall $\xi = 5$.

Im folgenden wird gezeigt, wie die Vielzahl der verschiedenartigen Kombinationen sofort unterbrechender und nichtunterbrechender Prioritäten einheitlich beschrieben werden kann mit Hilfe der Unterbrechungs-Distanz ξ.

Die Unterbrechungs-Distanz kann für alle Prioritätsklassen gleich oder beliebig unterschiedlich sein. Zunächst wird deshalb der - leichter überschaubare - Fall gleicher Unterbrechungs-Distanz behandelt. Durch das Einführen "leerer" Prioritätsklassen (Klassen mit der Ankunftsrate null) können alle interessanten Möglichkeiten beliebiger Unterbrechungs-Distanz für jdede Prioritätsklasse auf diesen Fall zurückgeführt werden.

In einem weiteren Abschnitt wird gezeigt, wie für derartige Systeme, individuell für jede Prioritätsklasse:

- die Wartewahrscheinlichkeit,
- die mittlere Warte- und Durchlaufzeit,
- die Wartebelastung
- die Unterbrechungswahrscheinlichkeit sowie
- die mittlere Anzahl von Unterbrechungen

bestimmt werden können. Als Grenzfälle ergeben sich die bekannten Ergebnisse für Systeme mit rein unterbrechenden bzw. rein nichtunterbrechenden Prioritäten (vgl. z.B. Martin /6/, Jaiswal /237/). Der einzige in der Literatur mathematisch behandelte Fall einer Kombination der beiden Prioritätsstrategien - Chang /238/: einige Klassen unterbrechen alle Anforderungen geringerer Priorität, die restlichen Klassen können sich gegenseitig nicht unterbrechen - ist als Sonderfall in der angegebenen Lösung enthalten.

2. BESCHREIBUNG VON UD-PRIORITÄTEN

2.1. Gleiche Unterbrechungs-Distanz für alle Prioritätsklassen

Bereits im ersten Abschnitt wurde die Unterbrechungs-Distanz eingeführt: sie gibt den Abstand an zwischen einer beliebigen, aber bestimmten Prioritätsklasse und der nächsten Prioritätsklasse, die von ihr unterbrochen wird.

Anforderungen der Prioritätsklasse p ($p = 1,2,\ldots. P$; Klasse 1 ist am dringlichsten; vgl. Bild VII.3) unterbrechen deshalb nur Anforderungen

Prioritätsklassen

$$\underbrace{1,2,\ldots,p-\xi,}_{\text{unterbrechen Klasse p}}\quad \underbrace{p-\xi+1,\ldots,p-1,}_{\text{unterbrechen Klasse p nicht}}\ \boxed{p},\ \underbrace{p+1,\ldots,p+\xi-1,}_{\text{nicht unterbroch. von Klasse p}}\quad \underbrace{p+\xi,\ldots,P}_{\text{unterbrochen von Klasse p}}$$

Bild VII.3: Unterbrechungs-Distanz Prioritäten mit einheitlicher Unterbrechungs-Distanz ξ.

der Klassen ($p + \xi$) bis P, nicht aber jene der dazwischen liegenden Klassen ($p + 1$) bis($p + \xi - 1$). Auf der anderen Seite werden Anforderungen der betrachteten Klasse p nur durch Klasse 1 bis ($p - \xi$), nicht aber durch die Klassen ($p - \xi + 1$) bis ($p - 1$) unterbrochen.

Die bereits bekannten Sonderfälle von Unterbrechungs-Distanz

Prioritäten ergeben sich für

$\xi = 1$: rein unterbrechende Prioritäten, und
$\xi = P$: rein nichtunterbrechende Prioritäten.

Einige Beispiele sind in Bild VII.4 tabellarisch zusammengestellt.

Unterbrechungs-Distanz ξ	Neue Anforderung der Klasse p	unterbricht nicht Klasse	unterbricht Klasse
1	1	1,	2,3,4,
1	2	1,2,	3,4,
1	3	1,2,3,	4,
1	4	1,2,3,4	-
2	1	1,2,	3,4,
2	2	1,2,3,	4,
2	3	1,2,3,4,	-
2	4	1,2,3,4,	-
4	1	1,2,3,4	-
4	2	1,2,3,4	-
4	3	1,2,3,4	-
4	4	1,2,3,4	-

Bild VII.4: Drei Beispiele von UD-Prioritäten mit einheitlicher Unterbrechungs-Distanz (Anzahl der Prioritätsklassen P = 4; für ξ =1 ergibt sich der Sonderfall rein unterbrechender, für ξ = P=4 der Fall nichtunterbrechender Prioritäten).

2.2. Beliebige Unterbrechungs-Distanz für jede Prioritätsklasse

Führt man individuell für jede Prioritätsklasse p die Unterbrechungs-Distanz $\xi(p)$ ein und definiert sie analog zu Abschnitt VII.2.1, so sind beliebige Kombinationen rein unterbrechender und nichtunterbrechender Prioritäten möglich. Die einheitliche Darstellung und einheitliche mathematische Behandlung ist zwar möglich, wird jedoch ziemlich unübersichtlich.

Eine viel elegantere Lösung ergibt sich, wenn man eine einheitliche Unterbrechungs-Distanz ξ wählt und "leere" Prioritätsklassen einführt: zwischen die tatsächlich vorhandenen Prioritätsklassen schiebt man fiktive Prioritätsklassen mit der Ankunftsrate null.

Durch diesen Trick gelingt es, alle für die praktische Anwendung wichtigen Strategien übersichtlich und mit relativ geringem Aufwand mathematisch zu behandeln.

Bild VII.5 zeigt, wie sich durch das Einfügen leerer Prioritätsklassen das bereits in Bild VII.1 vorgestellte Beispiel für die Steuerung des Ablaufgeschehens bei Prozeßrechnern ergibt.

Vorgegebene Unterbrechungs-Distanz ξ	Neue Anforderung der Klasse p	unterbricht Klasse	tatsächliche Unterbrechungs-Distanz
3	1	4 bis 12	3
3	2	5 bis 12	3
3	3	6 bis 12	3
3	4	7 bis 12	3
3	5	8 bis 12	3
3	6	9 bis 12	3
3	7	10 bis 12	3
3	8	11 bis 12	3
3	9	12	3
3	10	-	3
3	11	-	3
3	12	-	3
Einfügen "leerer" Prioritätsklassen 2,3,5,6,10 und 11:			
3	1	4,7,8,9,12	1
3	4	7,8,9,12	1
3	7	12	3
3	8	12	2
3	9	12	1
3	12	-	-

Bild VII.5: Unterbrechungs-Distanz Prioritäten mit unterschiedlicher Unterbrechungs-Distanz (6 tatsächliche und 6 leere Prioritätsklassen; vgl. auch Text zu Abschnitt VII.2.2).

3. MODELLBILDUNG

3.1. Struktur und Betriebsart

Bild VII.6 zeigt schematisch dargestellt, das untersuchte Modell: Die Anforderungen werden, entsprechend ihrer Priorität in P parallele Warteschlangen eingeordnet. Alle Warteschlangen sind unbegrenzt, d.h. jede eintreffende Anforderung kann gespeichert und bedient werden. Diese Annahme ist auch in Wirklichkeit fast immer erfüllt, vor allem bei modernen Anlagen mit dynamischer Speicherplatzzuweisung.

Als Abfertigungsstrategie wird eine beliebige Art von UD-Prioritäten vorausgesetzt, die wir im letzten Abschnitt kennengelernt haben.

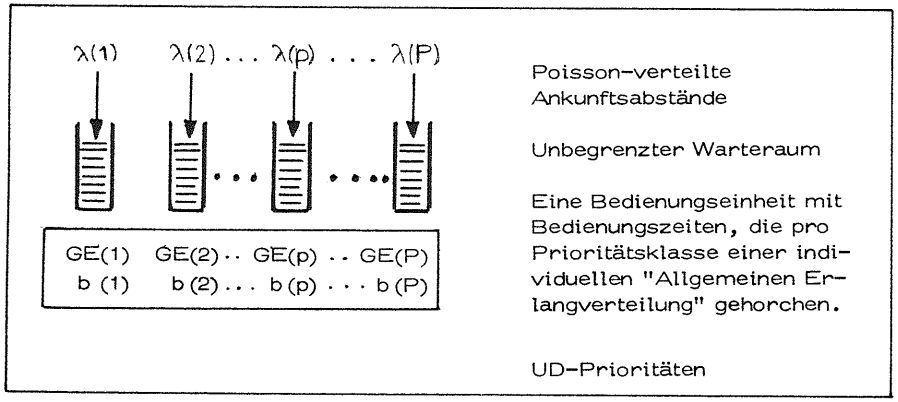


Bild VII.6: Das untersuchte Modell (P Prioritätsklassen, λ Ankunftsrate, b mittlere Bedienungszeit, GE Abkürzung für die "Allgemeine Erlangverteilung" (General Erlangian).

Innerhalb einer Warteschlange, d.h. innerhalb einer Prioritätsklasse werden die Anforderungen in der Reihenfolge ihres Eintreffens bedient (FIFO). Diese Annahme entspricht ebenfalls der Wirklichkeit und den besonderen Anforderungen bei Echtzeit-Datenverarbeitungssystemen (vgl. Abschnitt II.3.1).

3.2. Verkehr

Eintreffende Anforderungen einer Prioritätsklasse p (p = 1 .... P) haben negativ exponentiell verteile Ankunftsabstände, d.h. wir haben individuell pro Prioritätsklasse einen Poisson-Ankunftsprozeß (vgl. Kapitel II):

$$A_p(\leq t) = 1 - e^{-\frac{t}{a(p)}} = 1 - e^{-\lambda(p)\cdot t} \qquad \text{(VII.1)}$$

mit dem mittleren Ankunftsabstand a (p) bzw. der Ankunftsrate λ (p).

Die Bedienungszeiten gehorchen pro Prioritätsklasse p einer "Allgemeinen Erlang-Verteilung":

$$B_p(\leq t) = \sum_{\nu=1}^{\ell(p)} q_\nu(p)\left(1 - e^{-\frac{t}{b_\nu(p)/k_\nu(p)}} \cdot \sum_{\eta=0}^{k_\nu(p)-1} \frac{\left(\frac{t}{b_\nu(p)/k_\nu(p)}\right)^\eta}{\eta!}\right)$$

Dichte

$$f_p(t) = \frac{d\,B_p(\leq t)}{dt} = \sum_{\nu=1}^{\ell(p)} \frac{e^{-\frac{t}{b_\nu(p)/k_\nu(p)}}}{b_\nu(p)/k_\nu(p)} \cdot \frac{\left(\frac{t}{b_\nu(p)/k_\nu(p)}\right)^{k_\nu(p)-1}}{(k_\nu(p)-1)!} \cdot q_\nu(p) \qquad \text{(VII.2)}$$

Normierungsbedingung:

$$1 = \sum_{\nu=1}^{\ell(p)} q_\nu(p)$$

Mittelwert:

$$b(p) = \sum_{r=1}^{\ell(p)} q_r(p) \cdot b_r(p)$$

Varianz:

$$\sigma^2(p) = \sum_{r=1}^{\ell(p)} \frac{k_r(p)+1}{k_r(p)} \cdot b_r(p)^2 \cdot q_r(p) - b(p)^2$$

(vgl. Kapitel II, insbesondere Gl. (II.10); da p üblicherweise auch zur Bezeichnung einer Prioritätsklasse benutzt wird, wird hier die Aufteilungswahrscheinlichkeit mit q_r bezeichnet).

Beispielsweise können deshalb die Bedienungszeiten für Klasse 1 konstant, für Klasse 2 hyperexponentiell, für Klasse 3 beliebig Erlang - k - verteilt und bei den restlichen Prioritätsklassen nach einer Treppenfunktion verteilt sein (vgl. auch Bild II.9).

Für die Berechnung wird Stationarität des Verkehrs vorausgesetzt, d.h. es wird angenommen, daß das System bereits zu Beginn der Beobachtung im eingeschwungenen Zustand ist.

Vernachlässigt wird die Zeit zur Behandlung einer Unterbrechung. Diese Vernachlässigung ist häufig erlaubt. Vor allem bei großen Echtheit - Datenverarbeitungssystemen sind alle Registersätze doppelt oder mehrfach vorhanden und können innerhalb kürzester Zeit ($\leq 1\ \mu s$) umgeschaltet werden. In großen Systemen erfolgt außerdem die Analyse der Unterbrechungsursache weitgehend "hardware"-mäßig, so daß das

Betriebssystem weiter entlastet wird.

Für kleine Anlagen können diese Zeiten nicht oder nur in erster Näherung vernachlässigt werden. Intensive Untersuchungen, auch diese Verwaltungszeiten bei der Berechnung zu berücksichtigen, sind z.Zt. im Gange (vgl. Zusammenfassung und Ausblick, Abschnitt VII.6).

4. BERECHNUNG DER CHARAKTERISTISCHEN GRÖSSEN

4.1. Mittlere Warte- und Durchlaufzeiten, Wartebelastung

Die mittleren Warte- und Durchlaufzeiten pro Prioritätsklasse werden mit Hilfe der Momentenmethode bestimmt (Bild VII.7. zeigt zur Veranschaulichung den einfachsten, nichttrivialen Fall, siehe ferner auch Kapitel IV und VI).

Die mittlere Durchlaufzeit d (p) einer Anforderung der Prioritätsklasse p (p = 1,...P) setzt sich aus folgenden fünf Anteilen zusammen:

- Erwartungswert b_R ($\leq p + \xi - 1$) der abzuwartenden Restbedienungszeit für Anforderungen der Prioritätsklassen 1 bis ($p + \xi - 1$). Diese Anforderungen können von der betrachteten Prioritätsklasse p nicht unterbrochen werden;

- mittlere Wartezeit w_I (p), die notwendig ist, alle Anforderungen der Klassen 1 bis p zu bedienen, die bereits

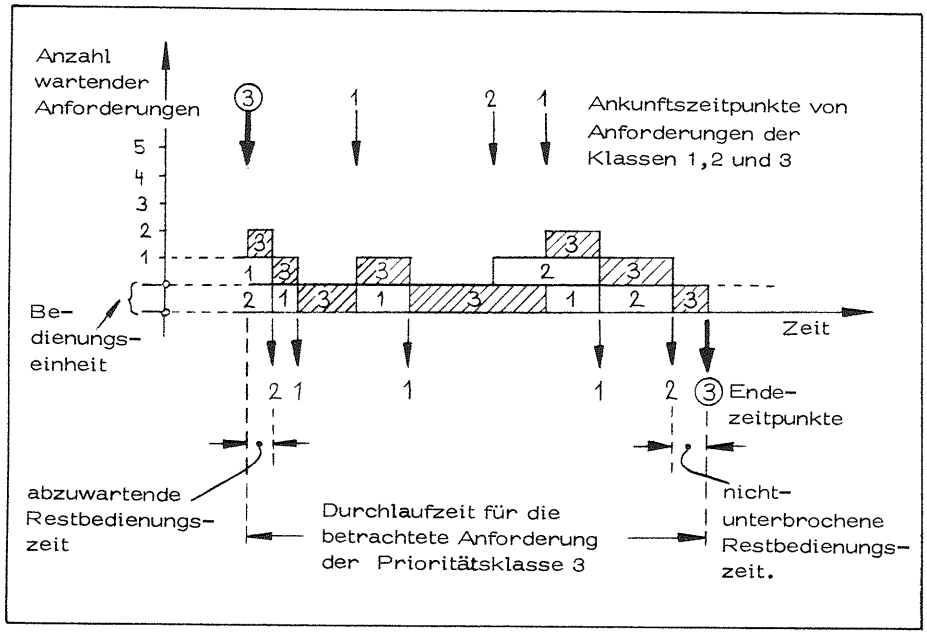


Bild VII.7: Einfachstes Beispiel eines Wartesystems mit UD-Prioritäten:
drei Prioritätsklassen (P = 3) mit der Unterbrechungs-Distanz zwei (ξ = 2).
Schicksal und Durchlaufzeit einer Anforderung der niedrigsten Prioritätsklasse.

zum Zeitpunkt des Eintreffens im System warten;

- mittlere Bedienungszeit b (p) der betrachteten Prioritätsklasse;

- mittlere Wartezeit $w_{II}(p)$. Diese Zeit ist notwendig, um alle während der Durchlaufzeit d (p) eintreffenden und unterbrechenden Anforderungen der Klassen 1 bis (p - ξ) zu bedienen;

- mittlere Wartezeit $w_{III}(p)$. Alle Anforderungen nicht-unterbrechender Prioritätsklassen (p - ξ +1) bis (p-1), die während der Durchlaufzeit d (p), jedoch vor der letzten unterbrechenden Anforderung eintreffen, werden vor der betrachteten Anforderung bedient.

Für die mittlere Durchlaufzeit gilt deshalb

$$d(p) = b_R(\leq p+\xi-1) + w_I(p) + b(p) + w_{II}(p) + w_{III}(p) \qquad \text{(VII.3)}$$

und wegen

$$d(p) = w(p) + b(p) \qquad \text{(VII.4)}$$

folgt weiter für die mittlere Wartezeit

$$w(p) = b_R(\leq p+\xi-1) + w_I(p) + w_{II}(p) + w_{III}(p) \qquad \text{(VII.5)}$$

Die vier unbekannten Anteile werden in den folgenden Abschnitten bestimmt.

4.1.1. Restbedienungszeiten

4.1.1.1. Gesamte Restbedienungszeit

Ein "Testruf" der Klasse p trifft auf eine Anforderung der Klasse v im System (vgl. Bild VII.8). Unabhängig davon, ob der Testruf unterbricht oder nicht unterbricht wird die "gesamte Restbedienungszeit" definiert als restliche notwendige Bedienungszeit für die Anforderung der Klasse v.

Da das Ergebnis bereits aus der berühmten Formel von Pollaczek -Khintchine bekannt ist, wird die Ableitung knapp gefaßt (vgl. z.B. Ferschl /10/).

Ein Charakteristikum des Poisson-Prozesses ist die Unabhängigkeit von der Zeit: die Ankunft einer Anforderung ist zu jedem beliebigen Zeitpunkt gleichwahrscheinlich (Grundlagen siehe Abschnitt II.2.3). Bei einer zufälligen Länge $T_b(v)$ der

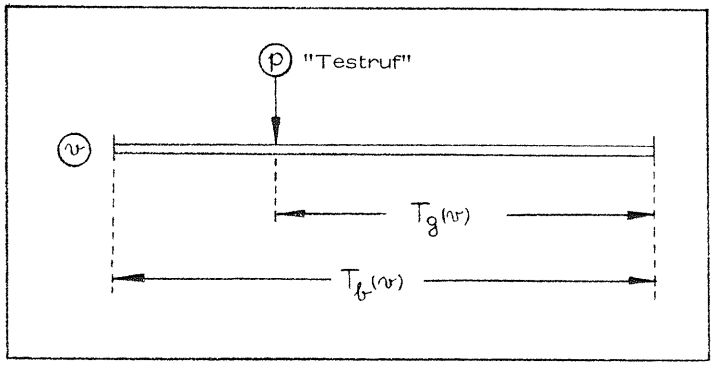


Bild VII.8: Gesamte Restbedienungszeit $T_g(v)$ (vgl. Text).

Bedienungszeit ist deshalb die mittlere Restbedienungszeit

$$E\left[T_g(v)\right]_{T_b(v)} = \frac{1}{2} \cdot T_b(v) \qquad \text{(VII.6)}$$

Da während der gesamten Bedienungsperiode der Länge $T_b(v)$ im Mittel

$$S_1(T_b(v)) = \lambda(p) \cdot T_b(v)$$

Anforderungen der Prioritätsklasse p eintreffen ist die Summe der Restbedienungszeiten

$$S_2(T_b(v)) = \lambda(p) \cdot T_b(v) \cdot \frac{1}{2} T_b(v)$$

$$S_2(T_b(v)) = \frac{1}{2} \lambda(p) \cdot T_b(v)^2$$

Für beliebig verteilte Bedienungszeiten ergibt sich deshalb die mittlere gesamte Restbedienungszeit $b_g(v)$ zu:

$$b_g(v) = \frac{E\left[S_2(T_b(v))\right]}{E\left[S_1(T_b(v))\right]} = \frac{b(v)}{2} \left\{ 1 + \frac{\sigma^2(v)}{b(v)^2} \right\} \qquad \text{(VII.7)}$$

Aus Gl. (VII.7) folgt weiter mit Gl. (VII.2) für Bedienungszeiten, die einer "Allgemeinen Erlangverteilung" gehorchen:

$$\boxed{b_g(v) = \frac{1}{2 \cdot b(v)} \sum_{\mu=1}^{\ell(v)} \frac{k_\mu(v)+1}{k_\mu(v)} \cdot b_\mu(v)^2 \cdot q_\mu(v)} \qquad \text{(VII.8)}$$

4.1.1.2. Abzuwartende Restbedienungszeit

Trifft ein Testruf der Klasse p auf eine Anforderung der Klasse v, so ist die abzuwartende Restbedienungszeit abhängig von

den betreffenden Prioritätsklassen und abhängig von der Unterbrechungs-Distanz ξ .

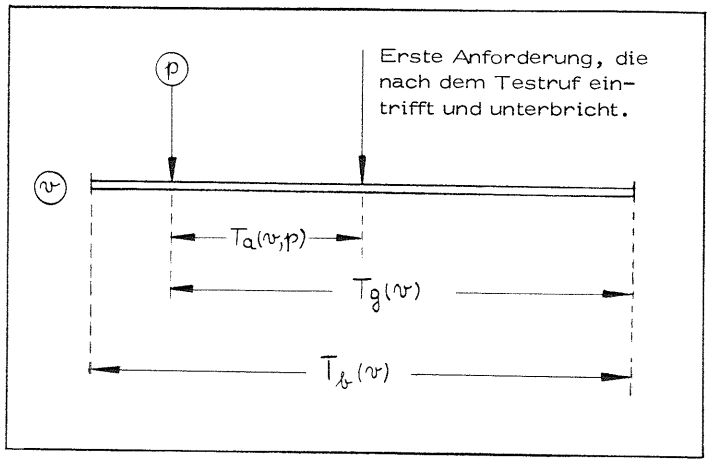
<u>$v = (p + \xi)$ bis P:</u> offensichtlich ist in diesem Fall die abzuwärtende Restbedienungszeit null, da die Anforderung der Klasse v sofort unterbrochen wird (vgl. Bild VII.3).

<u>$v = 1$ bis p:</u> hier ist die abzuwartende Restbedienungszeit gleich der gesamten Restbedienungszeit, da die angetroffene Anforderung mindestens die gleiche Priorität hat wie der Testruf.

<u>$v = (p + 1)$ bis $p + \xi - 1$):</u> in diesem Fall kann der Testruf die angetroffene Anforderung nicht unterbrechen, obwohl er einer wichtigeren Prioritätsklasse angehört. Abzuwarten ist jedoch nur die Zeit bis die erste unterbrechende Anforderung eintrifft. Die abzuwartende Restbedienungszeit ist also kleiner als die gesamte Restbedienungszeit.

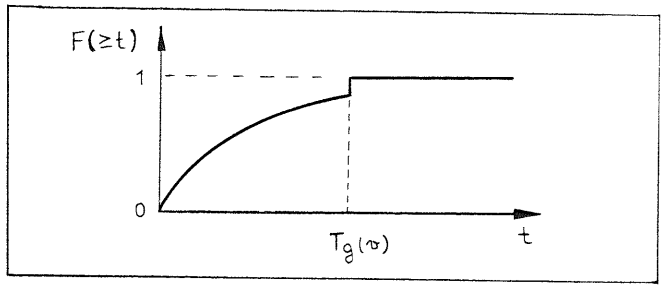
Zur Berechnung betrachte man - analog zum vorhergehenden Abschnitt - eine Anforderung der Klasse v mit der zufälligen Bedienungszeit T_b (v) (vgl. Bild VII.9). Die gesamte Restbedienungszeit für einen eintreffenden Testruf ist zwar T_g (v), abzuwarten ist jedoch nur die Zeitspanne T_a (v, p). Da die Ankunft der unterbrechenden Anforderungen ebenfalls einem Poisson-Prozeß entspricht, ist dieses Intervall T_a (v, p) negativ exponentiell verteilt mit der Ankunftsrate λ_u (v):

$$\lambda_u(v) = \sum_{i=1}^{v-\xi} \lambda(i) , \qquad \text{(VII.9)}$$



<u>Bild VII.9:</u> Die abzuwartende Restbedienungszeit T_a (v,p).

kann jedoch höchstens die Länge T_g (v) erreichen. Die Verteilungsfunktion hat deshalb den in Bild VII.10 skizzierten Verlauf. Der



<u>Bild VII.10:</u> Verteilungsfunktion der abzuwartenden Restbedienungszeit bei festen T_g (v).

Mittelwert dieser Verteilungsfunktion ergibt sich durch Integration:

$$E\left[T_a(v,p)\right]_{T_g(v)} = \int_{t=0}^{\infty} t \cdot dF(\geq t)$$

$$= \int_{t=0}^{T_g(v)} t \cdot \lambda_u(v) e^{-\lambda_u(v) \cdot t} dt + T_g(v) \cdot e^{-\lambda_u(v) T_g(v)} \qquad \text{(VII.10)}$$

$$= \frac{1}{\lambda_u(v)} \left(1 - e^{-\lambda_u(v) T_g(v)} \right)$$

Für eine zufällige Bedienungszeit der Länge $T_b(v)$ schwankt die gesamte Restbedienungszeit gleichverteilt zwischen 0 und T_b (v) (vgl. Abschnitt VII.4.1.1.1). Deshalb gilt:

$$E\left[T_a(v,p)\right]_{T_b(v)} = E\left[E\left[T_a(v,p)\right]_{T_g(v)}\right]$$

$$= \int_{T_g(v)=0}^{T_b(v)} \frac{1}{\lambda_u(v)} \left(1 - e^{-\lambda_u(v) T_g(v)}\right) \frac{1}{T_b(v)} \, d T_g(v)$$

$$= \frac{1}{\lambda_u(v)} \left\{ 1 - \frac{1}{\lambda_u(v) T_b(v)} \left(1 - e^{-\lambda_u(v) T_b(v)}\right) \right\} \qquad \text{(VII.11)}$$

Damit ist die mittlere abzuwartende Restbedienungszeit für eine Testanforderung der Klasse p bestimmt, wenn sie auf eine Anforderung der Klasse v mit der bestimmten Bedieungszeit $T_b(v)$ trifft.

Analog zum vorhergehenden Abschnitt kann nun die Formel für Bedienungszeiten, die einer "Allgemeinen Erlangverteilung" gehorchen, bestimmt werden:

$$E\left[T_a(v,p)\right] = \frac{E\left[S_2(T_b(v))\right]}{E\left[S_1(T_b(v))\right]}$$

wo

$$E\left[S_1(T_b(v))\right] = \lambda(p) \cdot b(p)$$

$$E\left[S_2(T_b(v))\right] = \int_{T_b(v)=0}^{\infty} \lambda(p) \cdot T_b(v) \cdot E\left[T_a(v,p)\right]_{T_b(v)} \cdot f_v(T_b(v)) \, d T_b(v)$$

Durch Einsetzen, Integration und Umformung ergibt sich mit Hilfe der Gleichungen (VII.2) und (VII.11) die mittlere abzuwartende Restbedienungszeit beim Antreffen einer beliebigen Anforderung der Prioritätsklasse v:

$$\boxed{E\left[T_a(v,p)\right] = \frac{1}{\lambda_u(v)^2 \cdot b(v)} \cdot \left\{ \lambda_u(v) \cdot b(v) - 1 + \sum_{r=1}^{\ell(v)} \frac{q_r(v)}{\left(\lambda_u(v) \cdot b_r(v)/k_r(v) + 1\right)^{k_r(v)}} \right\}} \qquad \text{(VII.12)}$$

Damit sind für alle drei Fälle, die bei der Berechnung zu unterscheiden waren, die mittleren abzuwartenden Restbedienungszeiten bekannt.

Ein eintreffender Testruf der Klasse p trifft mit der Wahrscheinlichkeit A (v) in der Bedienungseinheit auf eine Anforderung der Klasse v. Insgesamt ergibt sich deshalb für die gesuchte mittlere abzuwartende Restbedienungszeit

$$\boxed{\begin{aligned} b_R(\leq p+\xi-1) = & \sum_{v=1}^{p} \frac{A(v)}{2\,b(v)} \sum_{r=1}^{\ell(v)} \frac{k_r(v)+1}{k_r(v)} \cdot b_r(v)^2 \cdot q_r(v) \\ & + \sum_{v=p+1}^{p+\xi-1} \frac{A(v)}{\lambda_u(v)^2 \cdot b(v)} \cdot \left\{ \lambda_u(v) \cdot b(v) - 1 + \right. \\ & \left. + \sum_{r=1}^{\ell(v)} \frac{q_r(v)}{\left(\lambda_u(v) \cdot b_r(v)/k_r(v) + 1\right)^{k_r(v)}} \right\} \end{aligned}} \qquad \text{(VII.13)}$$

4.1.1.3. Nicht-unterbrochene Restbedienungszeit

Zur Bestimmung der mittleren Wartezeiten w_I (p) und w_{III}(p) muß die nicht - unterbrochene Restbedienungszeit berechnet werden. Dies ist jener Anteil der Bedienungszeit b (p) eines Testrufs der Klasse p, welcher nicht mehr von Anforderungen höherer Priorität (1.... p - ξ) unterbrochen wird (vgl. Bild VII.11 und das einfache Beispiel in Bild VII.7).

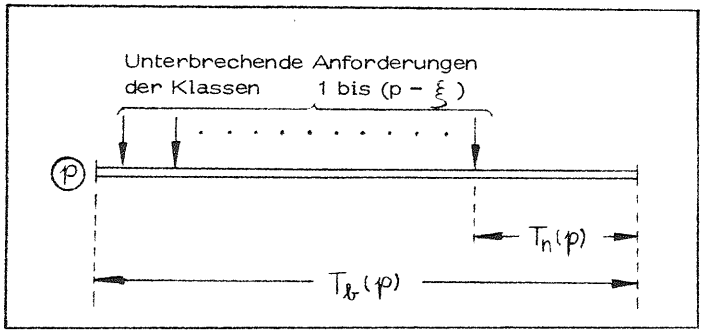


Bild VII.11: Zur Definition der nicht - unterbrochenen Restbedienungszeit T_n (p) eines Testrufs der Klasse p.

Die Berechnung erfolgt völlig analog zum vorhergehenden Abschnitt. Zusätzlich wird vom sog."reversiblen Verhalten" Poisson'scher Prozesse Gebrauch gemacht: der Prozeßverlauf kann auch von rückwärts betrachtet werden, da das Eintreffen einer Anforderung zu jedem Zeitpunkt gleichwahrscheinlich ist (siehe z.B. Reich /239/).

Die nicht - unterbrochene Restbedienungszeit T_n (p) ist deshalb

ebenfalls negativ exponentiell verteilt, kann jedoch maximal die Länge T_b (p) (gesamte Bedienungszeit) andauern.

Damit gilt

$$E[T_n(p)]_{T_b(p)} = \int_{t=0}^{T_b(p)} t \cdot \lambda_u(p) \cdot e^{-\lambda_u(p) \cdot t} dt + T_b(p) \cdot e^{-\lambda_u(p) T_b(p)}$$

$$= \frac{1}{\lambda_u(p)} \cdot \left\{ 1 - e^{-\lambda_u(p) \cdot T_b(p)} \right\}$$

und

$$b_n(p) = E\left[E[T_n(p)]_{T_b(p)}\right]$$

$$\boxed{b_n(p) = \frac{1}{\lambda_u(p)} \left\{ 1 - \sum_{r=1}^{\ell(p)} \frac{q_r(p)}{(\lambda_u(p) \cdot b_r(p)/k_r(p) + 1)^{k_r(p)}} \right\}} \quad \text{(VII.14)}$$

4.1.2. Die Wartebelastung und die mittlere Wartezeit w_I(p)

Beim Eintreffen des Testrufs warten im Mittel pro Prioritätsklasse Ω (i) Anforderungen (i = 1.... P). Alle wartenden Anforderungen, die den Klassen 1 bis p angehören, werden vor dem Testruf bedient.

Ein Teil der wartenden Anforderungen der Klasse i wurde bereits teilweise bedient, dann jedoch unterbrochen. Diese Anforderungen benötigen nur noch eine Restbedienungszeit. Die restlichen Anforderungen warten auf ihre gesamte Bedienungszeit.

4.1.2.1. Die Wartebelastung

Die mittlere Anzahl $\Omega(i)$ wartender Anforderungen der Klasse i ist gleich der mittleren Anzahl pro Zeiteinheit eintreffender Anforderungen $\lambda(i)$, multipliziert mit deren mittlerer Wartezeit w (i):

$$\Omega(i) = \lambda(i) \cdot w(i)$$

Ersetzt man die Wartezeit durch die Durchlaufzeit d (i), so gilt ferner

$$\Omega(i) = \lambda(i) \left\{ d(i) - b(i) \right\} \qquad \text{(VII.15)}$$

Die unbekannte Durchlaufzeit d (i) kann in der Gesamtformel eliminiert werden (Abschnitt VII.4.1.5.).

Die mittlere Anzahl wartender Anforderungen, die teilweise schon bedient wurden, läßt sich ebenfalls mit Hilfe der Momentenmethode bestimmen: man verfolgt das Schicksal eines Testrufs ab dem Zeitpunkt der ersten Bedienung bis zum Verlassen des Systems. Die Differenz zwischen dieser Zeitspanne und der tatsächlichen Bedienungszeit ist die Verweilzeit im Wartespeicher nach der ersten Unterbrechung.

Bild VII.12 zeigt ein typisches Beispiel. Die Durchlaufzeit ab dem Zeitpunkt der ersten Bedienung T_{de} (i) setzt sich aus folgenden Anteilen zusammen:

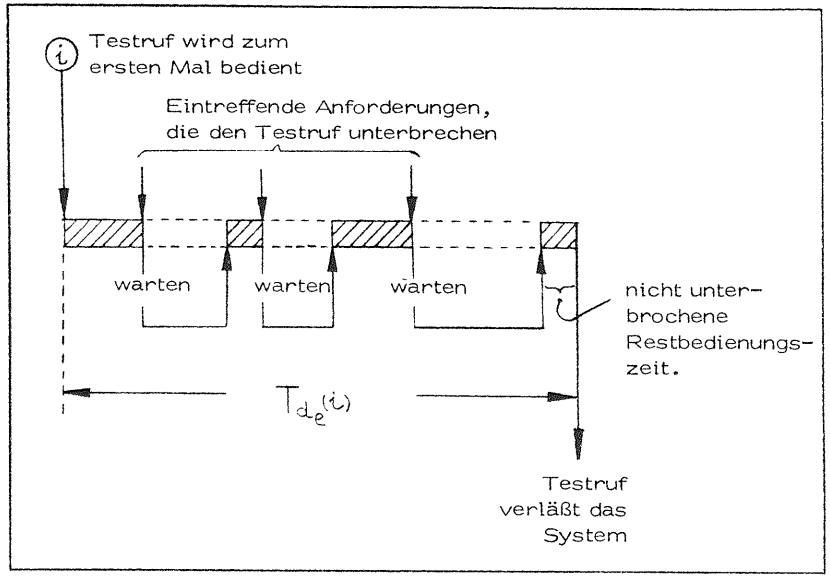


Bild VII.12 Das Schicksal eines Testrufs ab dem Zeitpunkt der ersten Bedienung bis zum Verlassen des Systems.

- der Bedienungsdauer des betrachteten Testrufs;
- der im Mittel während dieser Zeit eintreffenden Anforderungen der Prioritätsklassen 1 bis $(i-\xi)$;
- der im Mittel während der Durchlaufzeit T_{de} (i), jedoch vor der letzten Unterbrechung eintreffenden Anforderungen der nichtunterbrechenden Prioritätsklassen $(i - \xi + 1)$ bis $(i-1)$.

Analog zu den vorhergehenden Abschnitten ergibt sich deshalb für die Erwartungswerte die folgende Beziehung:

$$d_e(i) = b(i) + \sum_{j=1}^{i-\xi} \lambda(j)\, d_e(i) \cdot b(j) + \sum_{j=i-\xi+1}^{i-1} \lambda(j) \left\{ d_e(i) - b_n(i) \right\} \cdot b(j)$$

Nach einer Umformung ergibt sich für die gesuchte mittlere Durchlaufzeit ab dem Zeitpunkt der ersten Bedienung:

$$d_e(i) = \frac{b(i) - b_n(i) \sum_{j=i-\xi+1}^{i-1} A(j)}{1 - \sum_{j=1}^{i-1} A(j)}$$

Die mittlere Verweilzeit im Wartespeicher ab dem Zeitpunkt der ersten Bedienung ist deshalb:

$$v_e(i) = d_e(i) - b(i)$$

$$v_e(i) = \frac{b(i) \sum_{j=1}^{i-1} A(j) - b_n(i) \sum_{j=i-\xi+1}^{i-1} A(j)}{1 - \sum_{j=1}^{i-1} A(j)}$$

Insgesamt treffen pro Zeiteinheit λ (i) Anforderungen der betrachteten Prioritätsklasse ein. Die mittlere Anzahl wartender Anforderungen, die teilweise schon bedient, dann jedoch unterbrochen wurden, ist deshalb:

$$\Omega_u(i) = \lambda(i) \cdot v_e(i)$$

$$\boxed{\Omega_u(i) = \lambda(i) \cdot \frac{b(i) \sum_{j=1}^{i-1} A(j) - b_n(i) \sum_{j=i-\xi+1}^{i-1} A(j)}{1 - \sum_{j=1}^{i-1} A(j)}} \qquad \text{(VII.16)}$$

Nicht-unterbrochene Restbedienungszeit b_n(i) siehe Gl. (VII.14).

4.1.2.2. Die mittlere Wartezeit w_I(p)

Alle Anforderungen der Klassen 1 bis p, die beim Eintreffen des Testrufs warten, werden vor diesem bedient. Die dadurch entstehende mittlere Wartezeit ist

$$w_I(p) = \sum_{i=1}^{p} \left(\Omega(i) - \Omega_u(i) \right) \cdot b(i) + \sum_{i=1}^{p} \Omega_u(i) \cdot b_g(i)$$

und wegen der Beziehung (VII.15)

$$w_I(p) = \sum_{i=1}^{p} \left[A(i) \cdot \left\{ d(i) - b(i) \right\} - \Omega_u(i) \left\{ b(i) - b_g(i) \right\} \right] \qquad \text{(VII.17)}$$

Wartebelastung Ω_u(i) und Restbedienungszeit b_g (i) siehe Gl. (VII.16 bzw. (VII.8). Die unbekannte Durchlaufzeiten d (i) können in der Gesamtformel rekursiv bestimmt werden (Abschnitt VII.4.1.5).

4.1.3. Die mittlere Wartezeit w_{II} (p)

Während der Durchlaufzeit d (p) treffen im Mittel λ_u(p). d (p) Anforderungen der unterbrechenden Prioritätsklassen im System ein. Alle werden vor dem betrachteten Testruf bedient. Die zusätzliche Wartezeit ist deshalb

$$w_{II}(p) = \sum_{i=1}^{p-\xi} \lambda(i) \cdot d(p) \cdot b(i)$$

$$\boxed{w_{II}(p) = d(p) \cdot \sum_{i=1}^{p-\xi} A(i)} \qquad \text{(VII.18)}$$

Die unbekannte Durchlaufzeit d (p) kann in der Gesamtformel eliminiert werden (Abschnitt VII. 4.1.5).

4.1.4. Die mittlere Wartezeit $w_{III}(p)$

Anforderungen der Klassen ($p - \xi + 1$) bis ($p-1$) können den betrachteten Testruf vom Rang p nicht unterbrechen, obwohl sie eine höhere Priorität haben. Wird jedoch der Testruf unterbrochen, so schieben sich alle wartenden Anforderungen dieser Klassen vor den unterbrochenen Testruf.

Die mittlere Wartezeit durch derartige Anforderungen, die während der Durchlaufzeit, jedoch vor der letzten Unterbrechung eintreffen, ist

$$w_{III}(p) = \sum_{i=p-\xi+1}^{p-1} \lambda(i) \left\{ d(p) - b_n(p) \right\} \cdot b(i)$$

$$\boxed{w_{III}(p) = \left\{ d(p) - b_n(p) \right\} \cdot \sum_{i=p-\xi+1}^{p-1} A(i)} \qquad \text{(VII.19)}$$

Nicht - unterbrochene Restbedienungszeit b_n (p) siehe Gl. (VII.14). Die unbekannte Durchlaufzeit d (p) kann in der Gesamtformel eliminiert werden (nächster Abschnitt).

4.1.5. Rekursionsformeln für die mittleren Durchlauf- und Wartezeiten

Die mittlere Durchlaufzeit besteht aus fünf verschiedenen Anteilen, vgl. dazu Gl. (VII.3). Alle Anteile konnten in den vorhergehenden Abschnitten entweder explizit angegeben werden oder selbst als Funktion der Durchlaufzeit, der betrachteten und höherer Prioritätsklassen, bestimmt werden.

Aus Gl. (VII.3) folgt deshalb mit Gln. (VII.17), (VII.18) und (VII.19):

$$d(p) = b_R(\leq p+\xi-1) + \sum_{i=1}^{p} \left[A(i)\left\{ d(i) - b(i) \right\} - \Omega_U(i)\left\{ b(i) - b_g(i) \right\} \right] + b(p) + d(p) \cdot \sum_{i=1}^{p-\xi} A(i) + \left\{ d(p) - b_n(p) \right\} \cdot \sum_{i=p-\xi+1}^{p-1} A(i)$$

und nach einer einfachen Umformung:

$$\boxed{d(p) = \left\{ \sum_{i=1}^{p-1} A(i) \cdot d(i) + b(p) + b_R(\leq p+\xi-1) - \sum_{i=1}^{p} A(i) \cdot b(i) - \sum_{i=1}^{p} \Omega_U(i)\left\{ b(i) - b_g(i) \right\} - b_n(p) \sum_{i=p-\xi+1}^{p-1} A(i) \right\} \cdot \left\{ 1 - \sum_{i=1}^{p} A(i) \right\}^{-1}} \qquad \text{(VII.20)}$$

Gl. (VII.20) ist eine Rekursionsformel, aus welcher - beginnend bei der höchsten Prioritätsklasse p = 1 - die mittleren Durchlaufzeiten für jede Klasse bestimmt werden können.

Speziell für die wichtigste Prioritätsklasse gilt:

$$\boxed{d(1) = \frac{b_R(\leq \xi)}{1 - A(1)} + b(1)} \qquad \text{(VII.21)}$$

Die mittlere Wartezeit pro Prioritätsklasse ist gleich der Differenz zwischen der Durchlaufzeit und der Bedienungszeit. Nach einer Umformung folgt deshalb aus Gleichung (VII.20) für die mittlere Wartezeit aller Anforderungen einer beliebigen Prioritätsklasse p:

$$\boxed{\begin{aligned} w(p) = \Bigg\{ & \sum_{i=1}^{p-1} A(i)\cdot w(i) + b_R(\leq p+\xi-1) + b(p)\sum_{i=1}^{p-1} A(i) - \\ & - \sum_{i=1}^{p} \Omega_u(i)\left\{b(i)-b_g(i)\right\} - b_n(p)\sum_{i=p-\xi+1}^{p-1} A(i)\Bigg\}\cdot\left\{1-\sum_{i=1}^{p} A(i)\right\}^{-1} \end{aligned}} \quad \text{(VII.22)}$$

4.1.6. Geschlossene Lösung

Eine geschlossene Lösung für die mittleren Durchlauf- bzw. Wartezeiten ist ebenfalls möglich. Zunächst wird aus den Gleichungen (VII.20) bzw. (VII.22) eine lineare inhomogene Differenzengleichung erster Ordnung mit variablen Koeffizienten abgeleitet. Die Lösung des homogenen Gleichungssystems ist einfach. Wendet man anschließend das Verfahren von Lagrange (siehe z.B. Meschkowski /240/) zur Bestimmung einer "partikulären" Lösung des inhomogenen Gleichungssystems an, so ergibt sich – im Zusammenhang mit der homogenen Lösung – schließlich die gesuchte Gesamtlösung.

Die in den Gln. (VII.20) und (VII.22) angegebenen rekursiven Lösungen sind für die Auswertung auf einer Datenverarbeitungsanlage wesentlich besser geeignet als die geschlossenen Lösungen. Da außerdem einige Seiten für mathematische Lösungsansätze und Umformungen notwendig sind, soll die geschlossene Lösung an anderer Stelle veröffentlicht werden.

4.2. Die Unterbrechungswahrscheinlichkeit und die mittlere Anzahl von Unterbrechungen

4.2.1. Die Unterbrechungswahrscheinlichkeit

Die Unterbrechungswahrscheinlichkeit $P_u(p)$ ist die Wahrscheinlichkeit, daß eine Anforderung der Klasse p mindestens einmal während der Bedienungszeit unterbrochen wird.

Bild VII. 13 zeigt einen Testruf der betrachteten Prioritätsklasse p. Die zufällige Bedienungszeit des Testrufs sei $T_b(p)$. Mindestens eine unterbrechende Anforderung trifft ein mit der Wahrscheinlichkeit (vgl. Gl. (II.13)):

$$\begin{aligned} P_u(p)\big|_{T_b(p)} &= \sum_{x=1}^{\infty} \frac{(\lambda_u(p)\cdot T_b(p))^x}{x!}\cdot e^{-\lambda_u(p)T_b(p)} \\ &= 1 - e^{-\lambda_u(p)\cdot T_b(p)} \end{aligned} \quad \text{(VII.23)}$$

wo

$$\lambda_u(p) = \sum_{i=1}^{p-\xi} \lambda(i)$$

Gehorchen die Bedienungszeiten der betrachteten Prioritätsklasse p einer "Allgemeinen Erlangverteilung", so ergibt sich für die gesuchte Unterbrechungswahrscheinlichkeit:

$$P_u(p) = E\left[P_u(p)\big|_{T_b(p)}\right] = \int_{T_b(p)=0}^{\infty} P_u(p)\big|_{T_b(p)} \cdot f_p(T_b(p)\cdot dT_b(p)$$

$$P_u(p) = 1 - \sum_{\nu=1}^{\ell(p)} \frac{q_\nu(p)}{\left(\lambda_u(p)\cdot b_\nu(p)/k_\nu(p)+1\right)^{k_\nu(p)}} \qquad \text{(VII.24)}$$

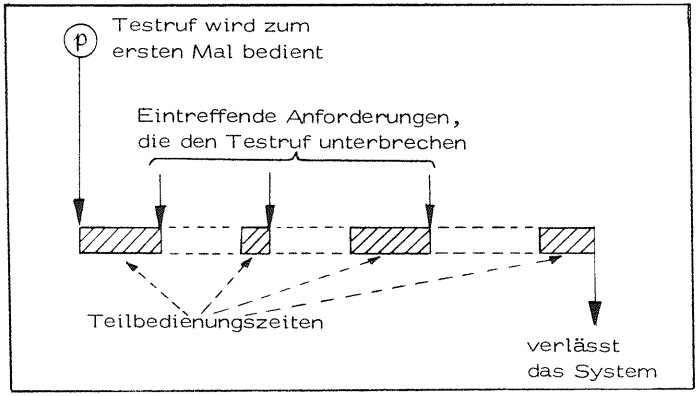


Bild VII.13: Zur Berechnung der Unterbrechungswahrscheinlichkeit, Abschnitt VII.4.2.1

4.2.2. Die mittlere Anzahl von Unterbrechungen

Die mittlere Anzahl von Unterbrechungen, denen ein Testruf der Klasse p mit der zufälligen Bedienungszeit T_b (p) unterworfen ist, ergibt sich zu:

$$u(p)\big|_{T_b(p)} = \lambda_u(p)\cdot T_b(p)$$

Entsprechend ergibt sich für die mittlere Anzahl von Unterbrechungen je Anforderung bei Bedienungszeiten, die einer

"Allgemeinen Erlangverteilung" gehorchen:

$$u(p) = E\left[u(p)\big|_{T_b(p)}\right] = \lambda_u(p)\cdot b(p) \qquad \text{(VII.25)}$$

4.3. Die Wartewahrscheinlichkeit

Als Wartewahrscheinlichkeit W (p) wird definiert die Wahrscheinlichkeit, daß eine Anforderung der Klasse p mindestens einmal warten muß.

Der Testruf muß sofort beim Eintreffen warten, wenn gerade eine Anforderung der nichtunterbrechbaren Prioritätsklassen 1 bis $(p+\xi-1)$ bedient wird. Dies ist der Fall mit der Wahrscheinlichkeit

$$P_e(p) = \sum_{i=1}^{p+\xi-1} A(i)$$

Trifft der Testruf auf ein leeres System oder nur auf Anforderungen der unterbrechbaren Prioritätsklassen $(p+\xi)$ bis P, so wird er sofort bedient, wird jedoch mit der Unterbrechungswahrscheinlichkeit P_u (p) in den Wartespeicher zurückgedrängt.

Insgesamt gilt deshalb für die Wartewahrscheinlichkeit:

$$W(p) = \sum_{i=1}^{p+\xi-1} A(i) + \left\{1 - \sum_{i=1}^{p+\xi-1} A(i)\right\}\cdot P_u(p) \qquad \text{(VII.26)}$$

Die Unterbrechungswahrscheinlichkeit P_u (p) ist aus Gl. (VII.24) bekannt.

5. NUMERISCHE BEISPIELE

Drei Beispiele sollen zeigen, welche verschiedenartigen Abfertigungsstrategien und Verkehrsmodelle jetzt mathematisch behandelt werden können und welche Vorteile UD-Prioritäten gegenüber rein unterbrechenden bzw. nichtunterbrechenden Strategien besitzen.

5.1. Einheitliche Unterbrechungs-Distanz

Bild VII.14 zeigt den Einfluß der Unterbrechungs-Distanz ξ auf die mittlere Wartezeit aller Anforderungen der einzelnen Prioritätsklassen. Verkehrsangebot und Verkehrscharakter werden dabei konstant gehalten.

Für die dringlichen Prioritätsklassen sinkt die mittlere Wartezeit mit kleiner werdender Unterbrechungs-Distanz ξ erheblich. Bei der wichtigsten Prioritätsklasse liegt z.B. zwischen dem Maximalwert ($\xi = 6$, d.h. rein nichtunterbrechend) und dem Minimalwert ($\xi = 1$, rein unterbrechend) der Faktor 250!

Den Preis, den man für die niedrigen Wartezeiten bezahlen muß, zeigt das nächste Bild VII.15: mit kleiner werdender Unterbrechungs-Distanz steigt die Anzahl der notwendigen Unterbrechungen stark an.

5.2. Unterschiedliche Unterbrechungs-Distanz

Bild VII. 16 zeigt die mittleren Wartezeiten für das schon mehrfach benützte Beispiel einer sinnvollen UD-Strategie für Prozeßrechner (vgl. auch Bild VII. 1,2 und 4).

Die beiden dringlichen Klassen (Alarme und Prozeß-Steuerung) haben die minimale Wartezeit eines rein unterbrechenden Systems. Die Klassen 3 bis 5 haben Wartezeiten, die alle in der gleichen Größenordnung liegen. Schließlich haben die Routineprogramme die höchst mögliche Wartezeit.

Obwohl in diesem Beispiel für die wichtigen Prioritätsklassen minimale Wartezeiten erzielt wurden, ist die Anzahl der notwendigen Unterbrechungen im Vergleich mit rein unterbrechenden Prioritäten stark gesunken (vgl. Bild VII.17).

Wartewahrscheinlichkeiten, Wartewahrscheinlichkeiten beim Eintreffen und Unterbrechungswahrscheinlichkeiten für alle Prioritätsklassen zeigt das Bild VII.18.

5.3. Extreme Verteilungsfunktionen für die Bedienungszeiten

Die verschiedenartigen Verteilungsfunktionen, die durch die "Allgemeine Erlangverteilung" erfaßt werden, soll das nächste Beispiel verdeutlichen.

Bild VII. 19 zeigt die (gemischte)Unterbrechungsstrategie und den Typ der gewählten Verteilungsfunktionen für die Bedienungszeiten,

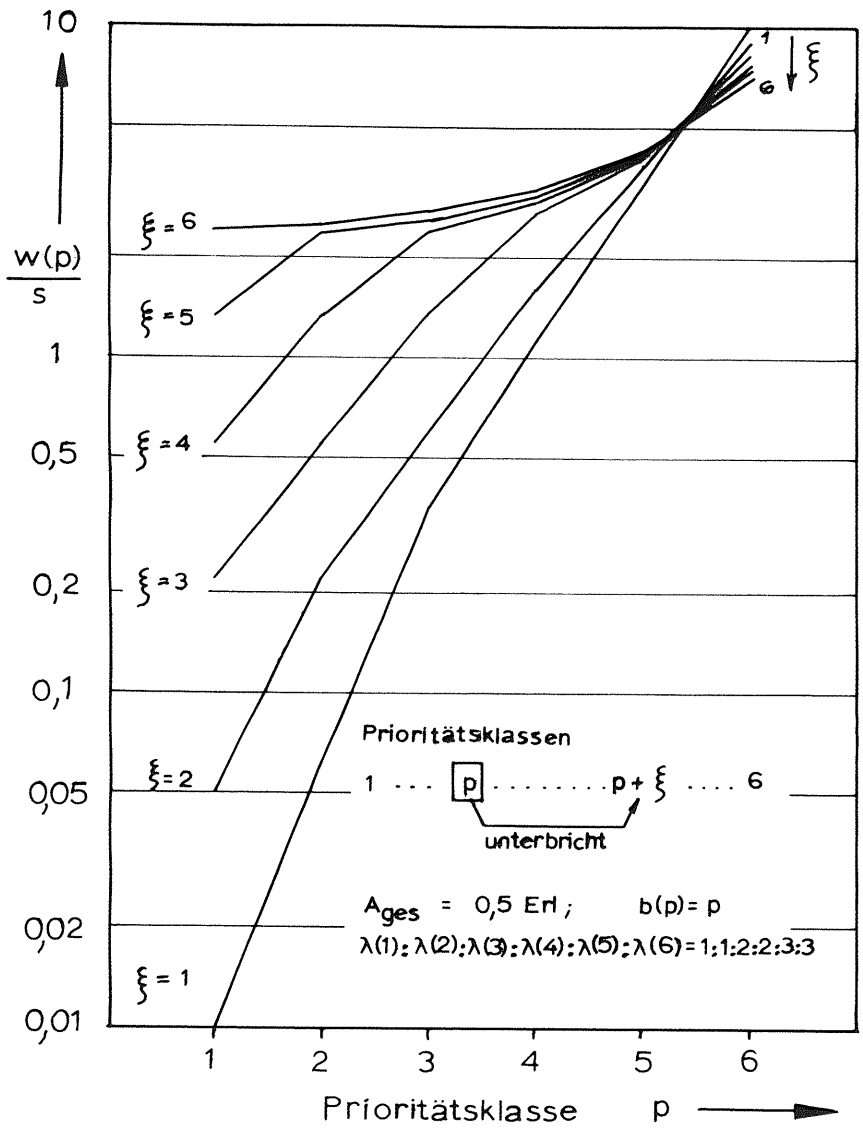


Bild VII.14.: Mittlere Wartezeit w (p) aller Anforderungen für P = 6 Prioritätsklassen und einheitliche Unterbrechungs-Distanz (ξ = 1: rein unterbrechend; ξ = 6: rein nichtunterbrechend). Alle Bedienungszeiten sind in diesem Beispiel negativ exponentiell verteilt, haben jedoch pro Klasse unterschiedliche Mittelwerte.

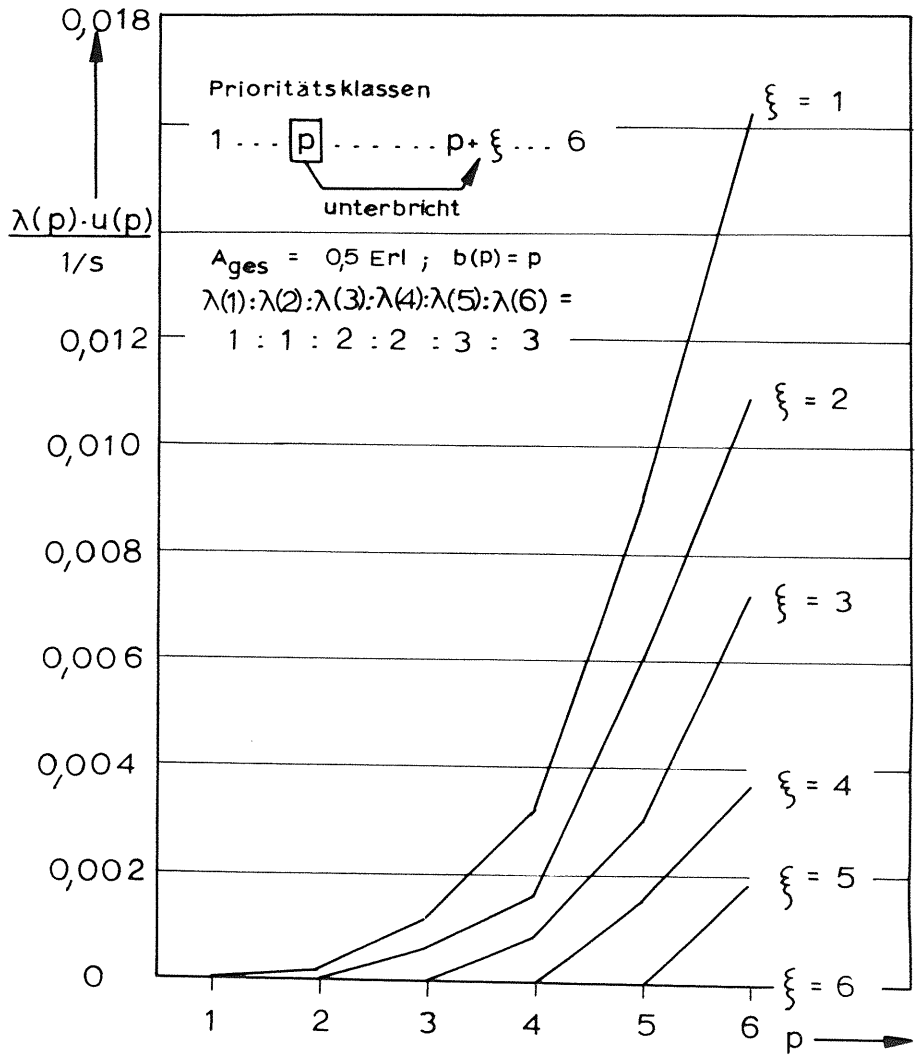


Bild VII.15.: Mittlere Anzahl von Unterbrechungen je Sekunde für P = 6 Prioritätsklassen und einheitliche Unterbrechungs-Distanz (ξ = 1: rein unterbrechend; ξ = 6: rein nichtunterbrechend). Alle Bedienungszeiten sind in diesem Beispiel negativ exponentiell verteilt, haben jedoch pro Klasse unterschiedliche Mittelwerte.

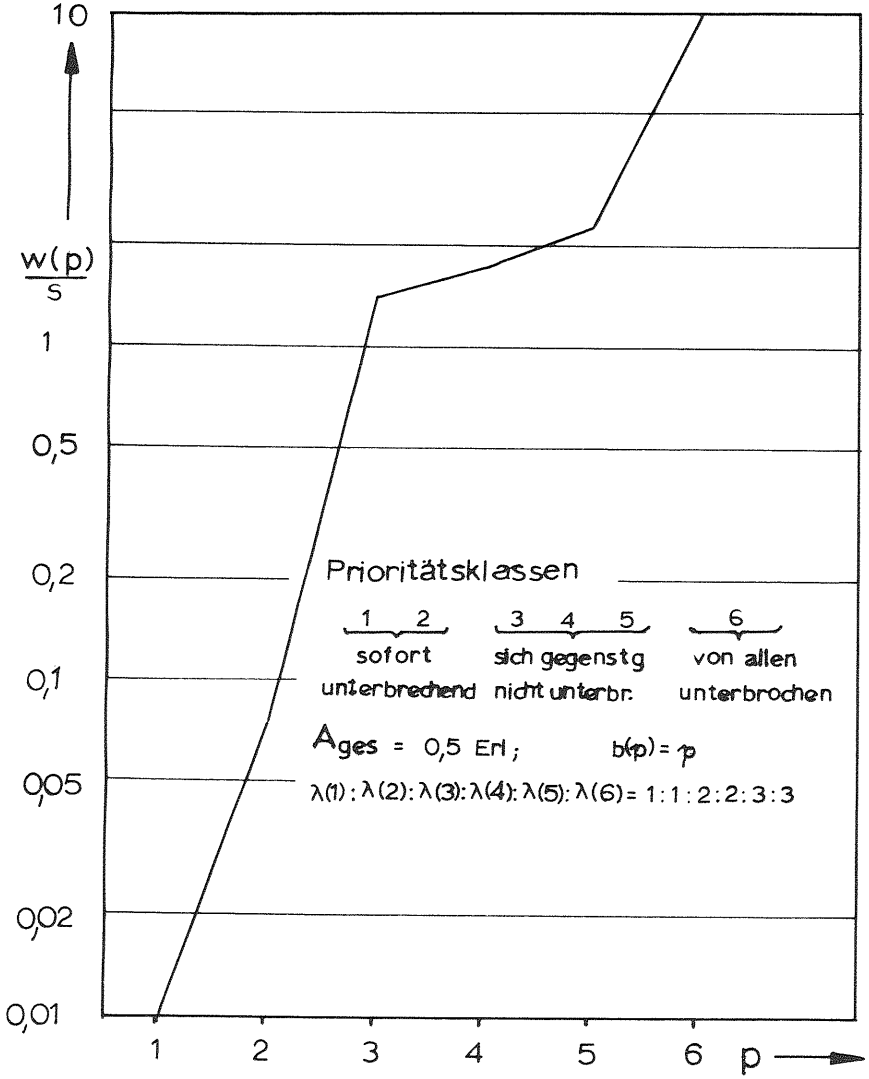


Bild VII.16.: Mittlere Wartezeit w (p) aller Anforderungen für P = 6 Prioritätsklassen bei unterschiedlicher Unterbrechungs-Distanz. Alle Bedienungszeiten sind in diesem Beispiel negativ exponentiell, haben jedoch pro Klasse unterschiedliche Mittelwerte .

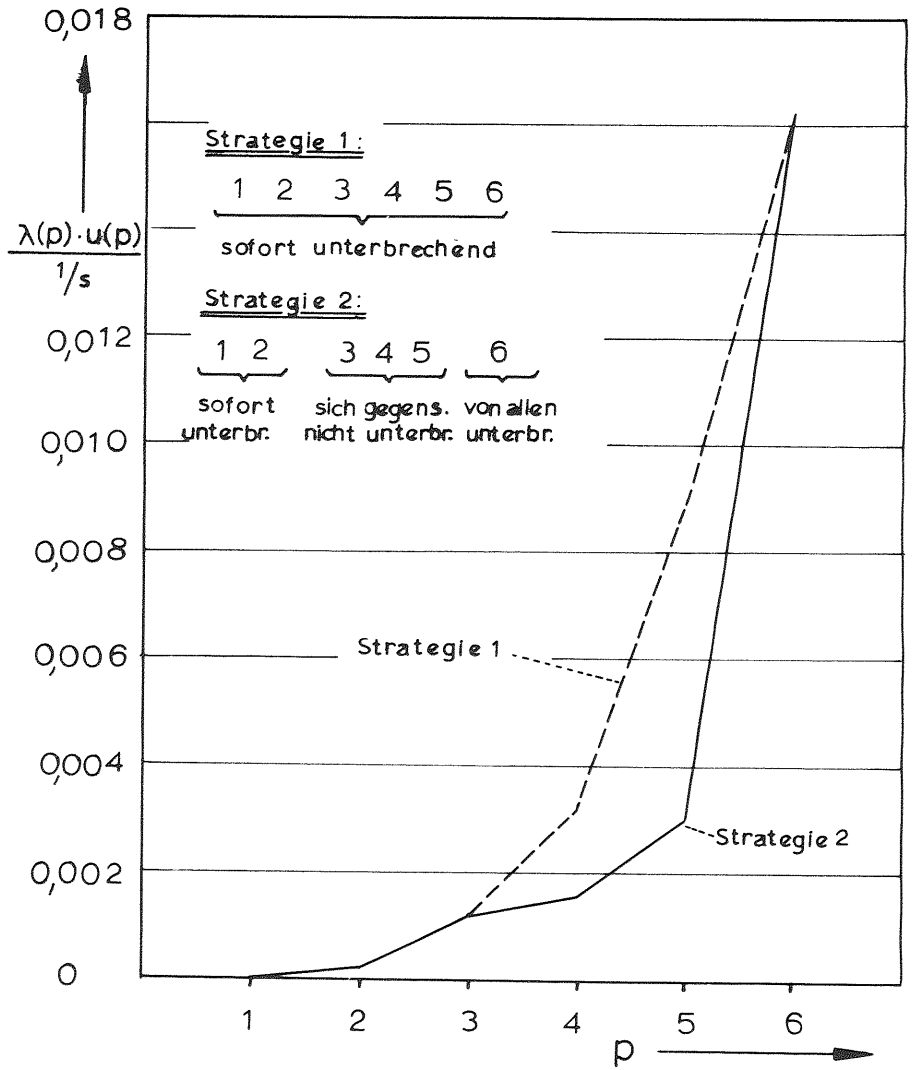


Bild VII.17.: Mittlere Anzahl von Unterbrechungen je Sekunde bei rein unterbrechenden Prioritäten (Strategie 1) und einer Strategie mit variabler Unterbrechungs-Distanz (Strategie 2). Verkehrsparameter siehe Bild VII.16 .

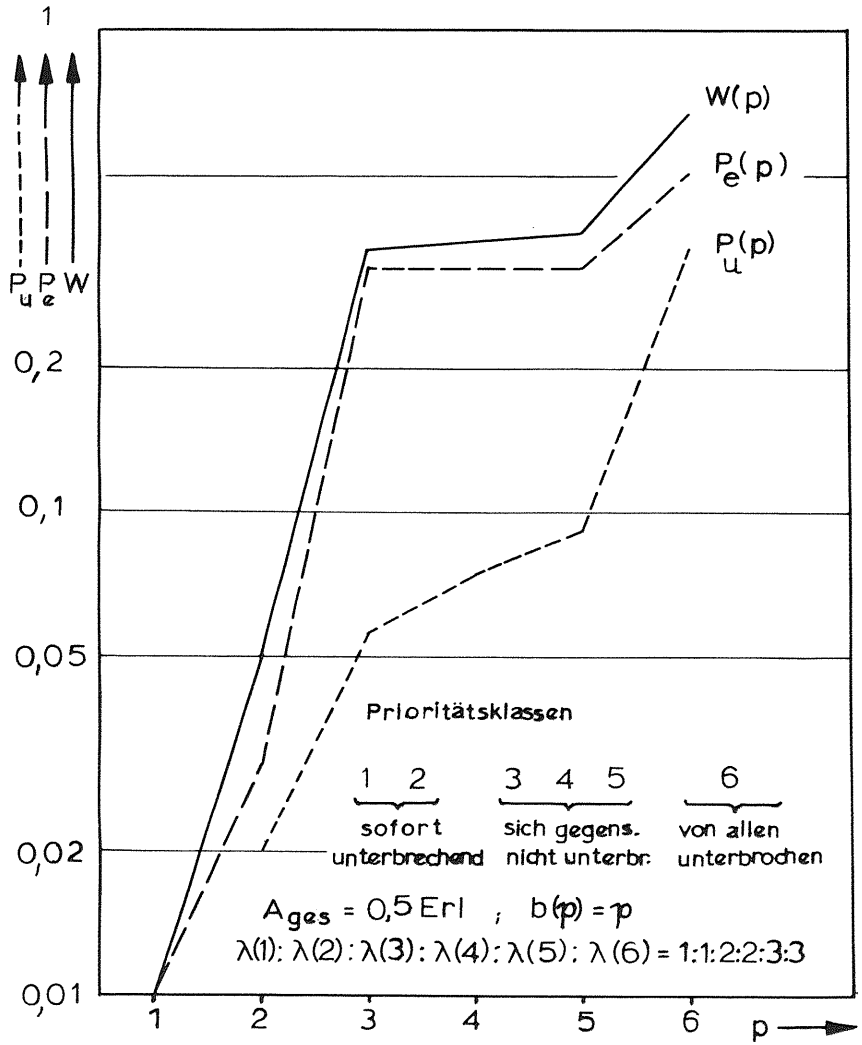


Bild VII.18.: Wartewahrscheinlichkeit W (p), Wartewahrscheinlichkeit beim Eintreffen $P_e(p)$ und Unterbrechungswahrscheinlichkeit $P_u(p)$ für P = 6 Prioritätsklassen. Alle Bedienungszeiten sind in diesem Beispiel negativ exponentiell verteilt, haben jedoch pro Klasse unterschiedliche Mittelwerte .

Prioritätsklasse	Unterbricht nicht Klasse	Unterbricht Klasse	Unterbrechgs-Distanz ξ
1	1	2 , 3	1
2	1, 2 , 3	—	2
3	1, 2 , 3	—	—

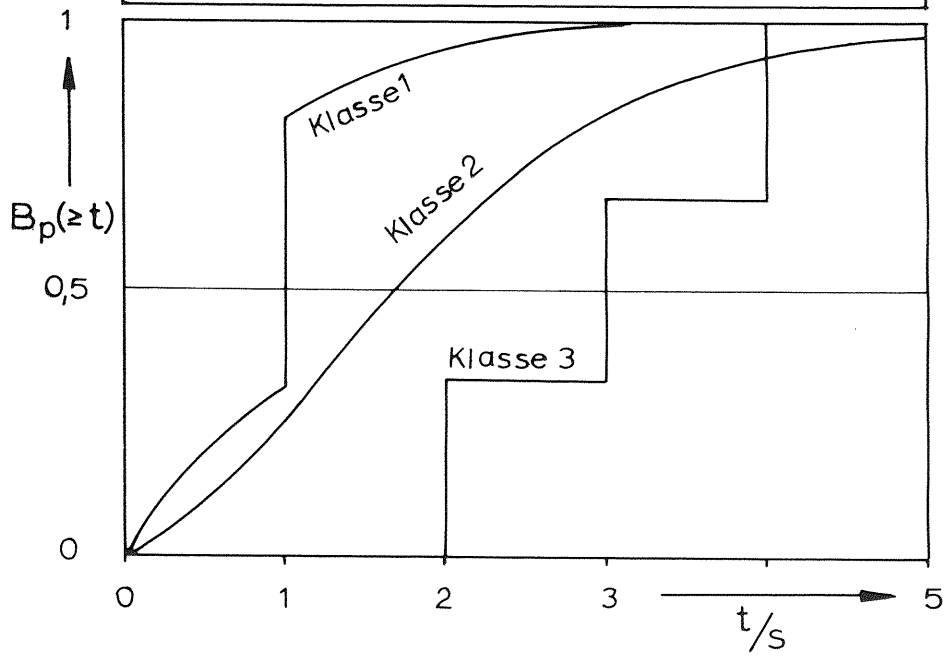


Bild VII.19.: Beispiel für drei verschiedenartige Bedienungszeitverteilungsfunktionen, die durch die "Allgemeine Erlangverteilung" beschreibbar sind. Die mittlere Bedienungszeit ist proportional der Klassennummer (b (1) = 1 s, b (2) = 2 s, b (3) = 3 s), die Unterbrechungs-Distanz für die einzelnen Klassen unterschiedlich .

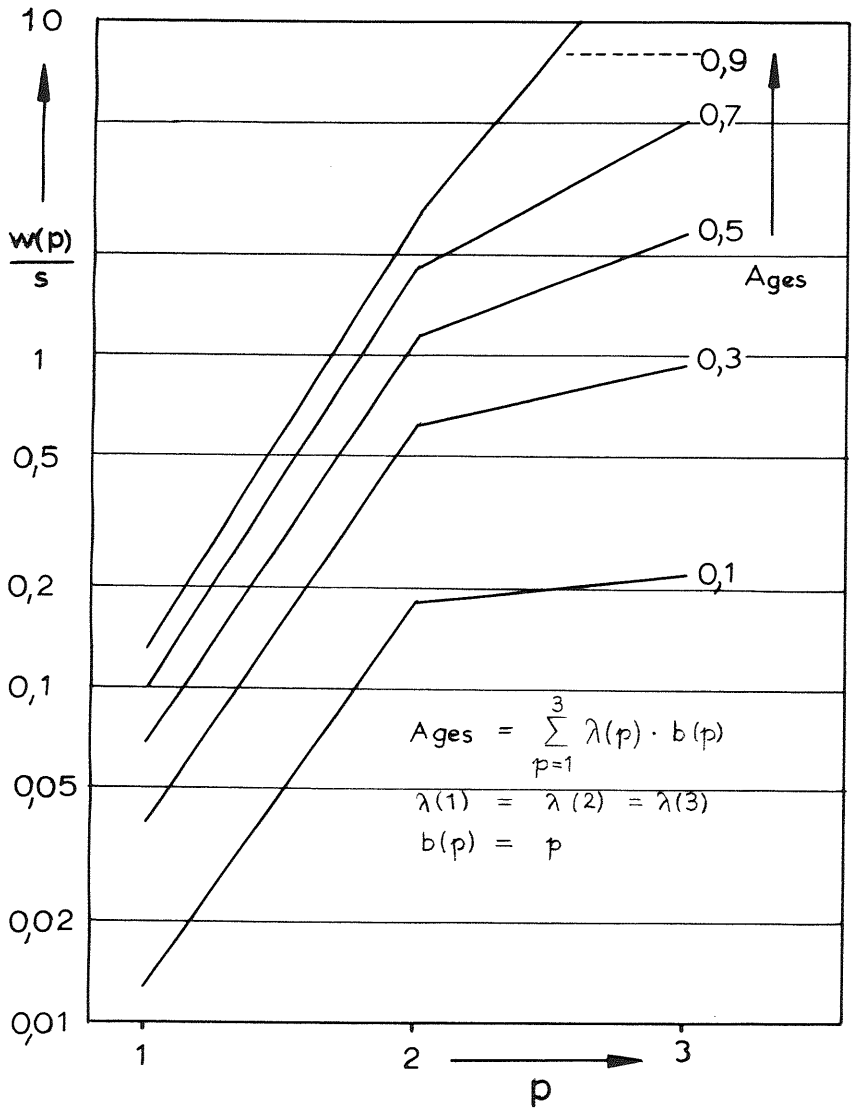


Bild VII.20.: Mittlere Wartezeit w (p) aller Anforderungen für die drei Prioritätsklassen. Unterbrechungsstrategie und Verteilung der Bedienungszeiten siehe Bild VII.19 .

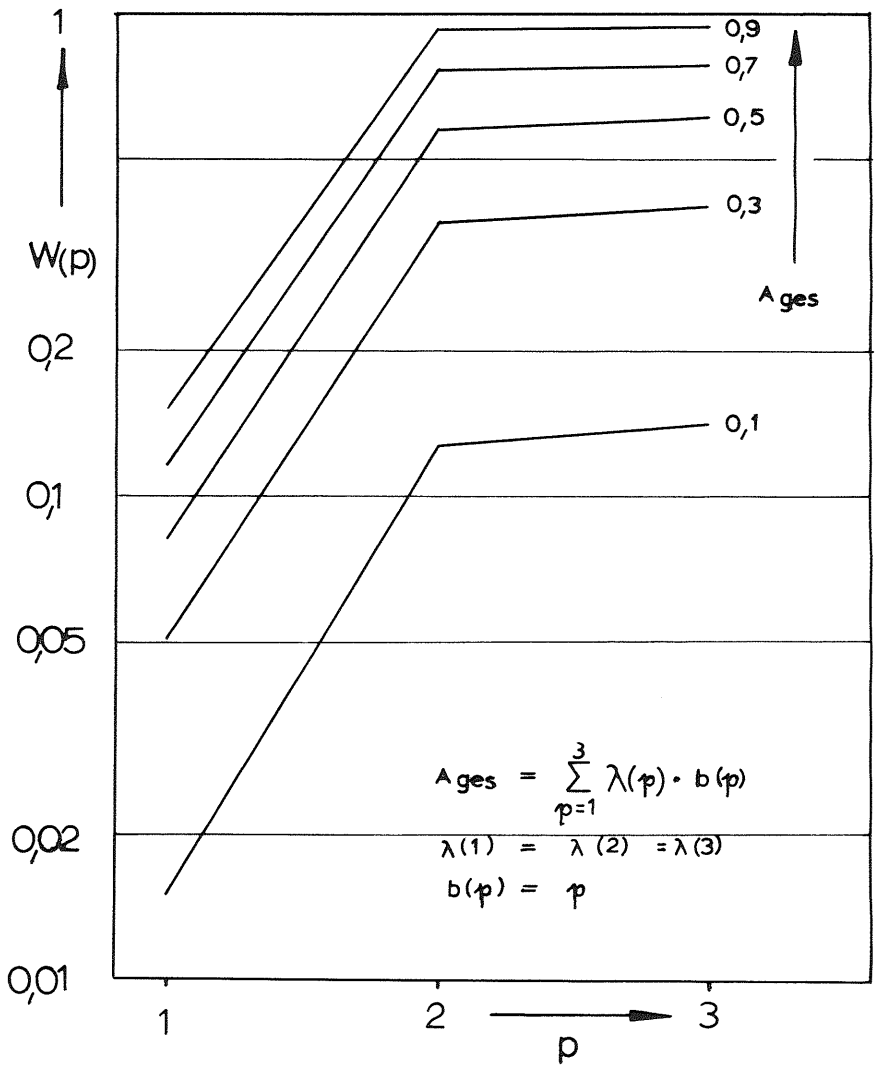


Bild VII.21.: Wartewahrscheinlichkeit W (p) für die drei Prioritätsklassen. Unterbrechungsstrategie und Verteilung der Bedienungszeiten siehe Bild VII. 19 .

Bild VII.20 zeigt den Verlauf der mittleren Wartezeiten, Bild VII.21 jenen der Wartewahrscheinlichkeit für jede Prioritätsklasse.

6. ZUSAMMENFASSUNG UND AUSBLICK

In diesem Kapitel wurde gezeigt, welche Bedeutung UD-Prioritäten bei Echtzeit-Datenverarbeitungssystemen besitzen. Die Vielzahl der möglichen Variationen und Kombinationen von sofort unterbrechenden und nichtunterbrechenden Strategien wurde einheitlich beschrieben und einheitlich mathematisch behandelt.

Individuell pro Prioritätsklasse wurden - bei Bedienungszeiten, die einer "Allgemeinen Erlangverteilung" gehorchen - Formeln abgeleitet zur Bestimmung zahlreicher charakteristischer Größen.

Mittelwerte und charakteristische Einzelwahrscheinlichkeiten geben eine wesentliche Auskunft über das Verhalten eines Systems. In vielen Fällen genügen diese Angaben. Eine allumfassende Auskunft, vor allem über die Schwankungen um diese Mittelwerte, gibt erst die Verteilungsfunktion.

Ein wesentlicher Teil der zukünftigen Untersuchungen wird deshalb beispielsweise der Wartezeitverteilungsfunktion gewidmet sein.

Mit Hilfe der "Allgemeinen Erlangverteilung", die für die Bedienungszeiten angenommen wurde, ist es möglich, jeden Verlauf

einer Verteilungsfunktion beliebig genau nachzubilden. Ein Problem ganz allgemeiner Natur ist deshalb die Frage, welche "Ordnung der Allgemeinen Erlangverteilung" gewählt werden muß, um eine vorgegebene reale Verteilungsfunktion zu beschreiben. Erste Untersuchungen zur Lösung dieses Problems wurden bereits durchgeführt.

Bereits bei der Modellbeschreibung wurde darauf hingewiesen, daß Verwaltungszeiten durch Programmunterbrechungen bei großen Systemen normalerweise vernachlässigbar sind. Bei interessanten UD-Strategien ist die Anzahl der Unterbrechungen - im Vergleich zu rein unterbrechenden Systemen - außerdem gering. Bei kleinen Systemen erfolgt die Unterbrechungsbehandlung jedoch weitgehend durch "software". Die Verwaltungszeiten sind dann nicht mehr oder nur in erster Näherung vernachlässigbar.

Aus diesem Grund beschäftigen wir uns zur Zeit auch mit Modellen, bei denen diese Verwaltungszeiten berücksichtigt werden.

VIII. ZUSAMMENFASSUNG

Durch die stürmische Entwicklung auf dem Gebiet der Datenverarbeitung bekommen Datennetze eine immer größere Bedeutung.

Bei der Planung derartiger Netze ist eine Vielzahl technologischer sowie vermittlungs-, übertragungs- und programmierungstechnischer Gesichtspunkte zu berücksichtigen. Der Entwurf wirtschaftlich optimaler Netze, mit vorgeschriebener Leistungsfähigkeit und vorgeschriebenen Sicherheit, ist jedoch erst dann möglich, wenn neben diesen Kriterien, auch verkehrstheoretische Untersuchungen durchgeführt und bei der Planung beachtet werden.

In dieser Arbeit wurde

- eine Einführung in die verkehrstheoretische Betrachtungsweise und ein kurzer Überblick über deren Methoden gegeben,
- die Vielfalt der verschiedenartigen Strukturen und Betriebsarten von Datennetzen vorgestellt und diskutiert, und
- durch zahlreiche Literaturhinweise, sowohl auf weiterführende Arbeiten hingewiesen, als auch ein gewisser Überblick über den heutigen Stand der Forschung und Entwicklung auf dem Gebiet der Datennetze gegeben.

In einem speziellen Teil wurden drei bisher ungelöste Problemkreise diskutiert, mathematisch beschrieben und mit Hilfe ver-

kehrstheoretischer Methoden untersucht und gelöst:

- die Durchlaufzeit von Nachrichten in Netzen mit Speichervermittlungen und alternativer Leitweglenkung,
- der Verkehrsfluß in einer Satellitenvermittlung für "packet-switching" mit Prioritäten, und
- die einheitliche Beschreibung und Behandlung verschiedenartiger, für die Echtzeit-Datenverarbeitung wichtiger, Prioritätsstrategien, sog. Unterbrechungs-Distanz Prioritäten. Nur für wenige Sonderfälle war bisher überhaupt eine Lösung bekannt.

Ein umfangreiches Kapitel gibt einen Überblick über die verschiedenartigen, meist ungelösten Probleme, die bei der Planung von Datennetzen auftreten. Zahlreiche weitere Hinweise bei der Behandlung der drei speziellen Problemkreise zeigen deutlich, daß eine Fülle interessanter und lohnenswerter Aufgaben auf ihre Lösung wartet.

SCHRIFTTUM

Zu Kapitel I: ÜBERSICHT

/1/ Hough, R.W.: Future Data Traffic Volume. Computer (1970) Sept.-Okt., S. 6 - 12.

/2/ Vollmeyer, W.: Künftige Tendenzen der Datenverarbeitung und des Datenverkehrs. Revue F.I.T.C.E (1970) S. 9 - 14.

/3/ Boehm, B.W.; Mobley, R.L.: Adaptive Routing Techniques for Distributed Communications Systems. IEEE Trans. Commun. Technol. 17 (1969) S. 340 - 349.

/4/ Fultz, G.L.; Kleinrock, L.: Adaptive Routing Techniques for Store - and Forward Computer Communication Networks. Advanced Research Projects Agency Semi-annual Technical Report, June 1971, University of California, S. 83 - 91.

/5/ Martin, J.: Teleprocessing Network Organization. Prentice Hall, Englewood Cliffs, N.J., 1970. Deutsche Übersetzung: Die Organisation von Datennetzen. Hanser Verlag , München, 1972.

/6/ Martin, J.: Design of Real-Time Computer Systems. Prentice Hall, Englewood Cliffs, N.J.,1967.

/7/ Graef,M.; Greiller, R.; Hecht, G.: Datenverarbeitung im Realzeitbetrieb. Oldenbourg Verlag, München, 1970.

Zu Kapitel II: BESCHREIBUNG DES VERKEHRSFLUSSES IN DATENNETZEN

/8/ Brandt, G.J.: Das preemptive Warteverlustsystem. Dissertation Universität Stuttgart, Stuttgart 1971.

/9/ Feller, W.: An Introduction to Probability Theory and its Applications. Wiley and Sons, New York - London, 1961.

/10/ Ferschl,F.: Zufallsabhängige Wirtschaftsprozesse. Physika-Verlag, Wien - Würzburg, 1964.

/11/ Gnedenko, B.W.: Lehrbuch der Wahrscheinlichkeitsrechnung. Akademie - Verlag, Berlin 1968

/12/ Morse, P.M.: Queues, Inventories and Maintenance. Wiley and Sons, London, 1958.

/13/ Riordan, J.: Stochastic Service Systems. Wiley and Sons, New York - London, 1962.

/14 Syski, R.: Introduction to Congestion Theory in Telephone Systems. Oliver and Boyd, Edinburgh and London, 1960.

/15/ Cole, G.D.: Performance Measurements on the ARPA Computer Network. Proc. ACM/IEEE Symp. on Problems in the Optimization of Data Commun. Systems, Palo Alto, Calif., October 1971, S.39-45.

/16/ Jackson, P.E.; Stubbs, C.D.: A Study of Multi - Access Computer Communications. 1969 Spring Joint Computer Conf., AFIPS Conf., Washington, D.C., Vol. 34, S. 491-504.

/17/ Fuchs, E.; Jackson, P.E.: Estimates of Distributions of Random Variables for Certain Computer Communications Traffic Models. Proc. ACM Symp. on Problems in the Optimization of Data Commun. Systems, Pine Mountain, Ga., October 1969, S. 202 - 225.

/18/ Dudick, A.L.; Fuchs, E.; Jackson, P.E.: Data Traffic Measurements for Inquiry-Response Computer Communications Systems. Proc. IFIP, Ljubljana, Jugoslawien, August 1971, TA-4-79 bis 83.

/19/ Hayward, W.S.; Wilkinson, R.I.: Human Factors in Telephone Systems and their Influence on Traffic Theory, especially with regard to Future Facilities. Proc. 6. ITC, München, September 1970, S. 431/1 - 431/10.

/20/ Wilkinson,R.I.: Some Comparisons of Load and Loss Data with Current Traffic Theory. Proc. 6.ITC, München, September 1970, S. 521/1 - 521/7.

/21/ Herzog U.; Kühn, P.; Zeh,A.: Klassifizierung und Analyse von Verkehrsmodellen für das Ablaufgeschehen in Rechnersystemen. NTG/GI Fachtagung "Rechner und Betriebssysteme: Analyse, Simulation und Entwurf", Darmstadt, April 1972. Nachrichtentechn. Fachberichte 44 (1972) S. 181-198.

/22/ Wagner, W.: Wartezeiten und Prioritäten. Aus Teilnehmer-Rechensysteme, Oldenbourg-Verlag, München und Wien, 1968, S. 231 - 253.

/23/ Kühn, P.: Combined Delay and Loss Systems with several Input Queues, full and limited Accessibility. Proc. 6. ITC, München, Sept. 1970 und A.E.Ü 25 (1971) S. 449 - 454.

/24/ Kühn, P.: Über die Berechnung der Wartezeiten in Vermittlungs-und Rechnersystemen. Dissertation, Universität Stuttgart, Stuttgart, 1972.

/25/ Langenbach - Belz, M.: Sampled Queuing Systems. Symp. Comp.-Commun. Networks and Teletraffic. Polytechn. Inst. Brooklyn, New York, N.Y., April 1972. Sess.IV.

/26/ Herzog,U.: Preemption - Distance Priorities in Real-Time Computer - Systems. Nachrichtentechn. Z. 25 (1972) S. 201 - 203.

Zu Kapitel III: STRUKTUR UND BETRIEBSART VON DATENNETZEN

III.1. ÜBERBLICK UND NOMENKLATUR

/27/ Neubauer, R.; Polakowski, G.: Nachrichtenvermittlung mit Digitalrechnern. Elektron. Rechenanlagen 9.(1967) S. 23 - 31.

/28/ White, G.D.: Message Format Principles. Proc. ACM/IEEE Symp. on Problems in the Optimization of Data Commun. Systems, Palo Alto, Calif., October 1971, S. 192 - 198.

/29/ Dell, F.R.E..; Features of a proposed Synchronous Data Network. IEEE Trans. Commun. 20 (1972) S. 499 - 503.

/30/ Chu, W.W.: Optimal Fixed Message Block Size for Computer Communications. Proc. IFIP, Ljubljana, Jugoslawien, August 1971, TA-4-61 bis 68.

/31/ United Kingdom Post Office: New Networks for Data Transmission. CCITT Working Party GM/NDR - No. 11 - E.

/32/ Davies, D.W.: Communication Network for Real - Time Computer Systems. The Radio and Electronic Engineer (1969) S. 47 - 52.

/33/ Gauthier, J.M.: Les Réseaux de Téléinformatique. L'Écho des Recherches, October 1970, S. 4 - 11.

/34/ Holloway, D.G.; Hart, R.I.: Some Design Considerations for High-Speed Circuit Switched Data Networks. Coll. Int. Téléinform., Paris, März 1969, S. 358 - 366.

/35/ Pierce, J.R.: How far can Data Loops go? IEEE Trans.Commun. 20 (1972), S. 527 - 530.

/36/ Closs, F.: Message Delays and Trunk Utilization in Line-Switched and Message Switched Networks. IBM Research Laboratory, Zürich, Juni 1972.

III.2. NETZE

/37/ Silk, D.J.: Routing Doctrines and their Implementation in Message - Switching Networks. Proc. IEE, 116 (1969) S. 1631 - 1638.

/38/ Prosser, R.T.: Routing Procedures in Communications Networks-Part I: Random Procedures - Part II: Directory Procedures. IRE Trans. Commun. Systems (1962) S. 322 -335.

/39/ Butrimenko, A.: Routing Technique for Message Switching Networks with Message Outdating. Symp. Comp.-Commun. Networks and Teletraffic. Polytechn. Inst. Brooklyn, New York, N.Y., April 1972, Session VI.

/40/ Brandt, G.J.; Chrétien, G.J.: Methods to Control and Operate a Message Switching Network. Symp. Comp.-Commun. Networks and Teletraffic. Polytechn. Inst. Brooklyn, New York, N.Y., April 1972, Session VI und S.I.T.A. Monograph No. 1 (1972).

/41/ Beeforth, T.H.; Grimsdale, R.L.; Halsall, F.; Woollons, D.J.; Aspects of a Proposed Data Communication System. Proc. IFIP, Ljubljana, Jugoslawien, August 1971, TA-4-74 bis 78.

/42 Davies, D.W.: The Control of Congestion in Packet-Switching Networks. IEEE Trans. Commun. 20 (1972) S. 546 - 550.

/43/ Frank, H.; Chou, W.: Routing in Computer Networks. Networks 1 (1971) S. 99 - 112.

/44/ Kahn, R.E.; Crowther, W.R.: Flow Control in a Resource-Sharing Computer Network. IEEE Trans. Commun. 20/ 1972/ S. 539 - 546.

/45/ Lotze, A.: Problems of Traffic Theory in the Design of International Direct Distance Dialling Networks. NTZ.-Commun. J. (1968) S. 41 - 46.

/46/ Pedersen, R.D.; Fu, Y.: Optimum Routing Strategies for Noisy Communication Networks. IEEE Trans. Commun. Techn. 17 (1969) S. 104 - 118.

III.3. VERMITTLUNGSEINRICHTUNGEN

/47/ Schomburg, K.: Fernschreib - und Datenvermittlungstechnik. Decker's Verlag, Hamburg - Berlin, 1971.

/48/ Martin, J.: Echtzeit - Programmiertechnik. Verlag Berliner Union GmbH und Verlag Kohlhammer, Stuttgart 1971.

/49/ Stimler, S.: Real - Time Data - Processing Systems. Mc Grow - Hill, New York, 1969.

/50/ Atkin, J.: Entwicklung des öffentlichen Datennetzes. Elektr. Nachrichtenwesen 47 (1972) S. 20 - 23.

/51/ Gabler, H; Staudinger, W.: Das deutsche Daten netz mit dem elektronischen Datenvermittlungssystem (EDS).Fernmelde-Ingenieur 26 (1972) H. 5 und 6.

/52/ Blackburn, R.J.; Muench, P.E.: The Bell System's digital Data System. Proc. ICC, Philadelphia, Pa., Juni 1972, S. 5.1-5.5.

/53/ Melis, A.; Verschueren, G.: Fernschreib - und Datenvermittlungssystem METACONTA 10C. Elektr. Nachrichtenw. 45 (1970) S. 173 - 182.

/54/ Adam, T.W.; Orbell, A.G.: A Digital Data Exchange. Proc. ISS, Cambridge, Mass., Juni 1972, S. 234 - 242.

/55/ Schaffer, B.: Time Division Multiplex in Modern Data Switching Systems. Proc. ICC, Philadelphia, Pa., Juni 1972, S. 5.13 - 5.18.

/56/ Fick, H.; Kammerl, A.: Multiplexing and Switching of Data in Synchrounous Networks and the Realization in the EDS - System. Proc. Zürich Seminar, Zürich, März 1972, S. C5 (1) - C6 (6).

/57/ Austin, D.N.; Huck, J.F.J.: A New Message - Relay System for the Overseas Telegraph Service. POEEJ 62 (1969) S. 44-53.

/58/ Störmer H.: Wartezeiten in Nachrichtenvermittlungen mit Speichern. Oldenbourg - Verlag, München, 1961.

/59/ Swoboda, J.: Verkehrsfragen innerhalb einer Rechenanlage. Elektron. Rechenanl. 12 (1970), S. 249 - 252.

/60/ Becker, W.D.; Aupperle, E.M.: The Communications Computer Hardware of the MERIT Computer Network. IEEE Trans. Commun. 20 (1972) S. 516 - 526.

/61/ Gunderson, D.C.: Some Effects of Advances in Memory System Technology on Computer Organization. Computer (1971) Nov. - Dez., S. 7 - 11.

/62/ Jilek, P.; Jokay Z.: Design and Implementation of Telecommunication Software Systems. Elektron. Rechenanl. 13 (1971) S. 162 - 167.

/63/ Davies, D.W.: Communication Networks to serve Rapid-Response Computers. Inform. Processing 68, North Holland Publ. Comp., Amsterdam (1969) S. 650-657.

/64/ Duke, C.S.; Mc Intyre, M.J.: A New Approach to Data Communications in a Real - Time Environment. Proc. IFIP, Ljubljana, Jugoslawien, August 1971, TA. - 4 - 84 bis 89.

III.4. ÜBERTRAGUNGSWEGE

/65/ Tietz, W.: Betriebsweisen und Verfahren der Datenübertragung. Handb. maschinelle Datenverarbeitung, Forkel - Verlag, Stuttgart, 1969.

/66/ Kaiser, W.: Technische Grundlagen der Datenübertragung. NTG - Fachtagung Datenübertragung, Mannheim, März 1969, Nachrichtentechn. Fachberichte 37 (1969) S. 3 - 22.

/67/ Kaiser, W.; Lang, J.; Ohnsorge, H.: Schnelle Datenübertragung. ETZ - A 89 (1968) S. 550 - 559.

/68/ Voss, H.H.: Technische Möglichkeiten schneller Datenübertragung in eigenständigen Datennetzen. NTG - Fachtagung Datenübertragung, Mannheim, März 1969, Nachrichtentechn. Fachberichte 37 (1969) S. 50 - 58.

/69/ Haas, W.: Möglichkeiten der Datenübertragung mit höheren Geschwindigkeiten in Nachrichtennetzen. VDE - Fachberichte 26 (1970) S. 150 - 155.

/70/ Irmer: Digitale Übertragungssysteme. Professorenkonferenz. 1969 im FTZ, Darmstadt. Nachrichtentechn. Fachberichte 39 (1971) S. 60 - 68.

/71/ Kaiser, W.: Datenübertragungsnetze. VDE - Fachberichte 27 (1972) S. 159.

/72/ Kwan, R.K.: Advances in Digital Radio Systems. Proc. ICC, Philadelphia, Pa., Juni 1972, S. 44.1 - 44.5.

/73/ Poschenrieder, W.: Digitale Nachrichtensysteme: Technischer Stand und Einsatzmöglichkeiten. NTZ 21 (1968) S. 665 - 671.

/74/ Setzer, D.E.: Low Capacity Multipair Cable for 6.3 Megabit per Second Transmission System. Proc. ICC, Philadelphia, Pa. Juni 1972, S. 19.19 - 19.23

/75/ Tietz, W.: Dateldienste. Decker's Verlag. G. Schenck, Hamburg, 1971.

/76/ Vetter, H.: Systemanalyse - Leitungsauswahl und Netzauslegung bei Datenfernverarbeitungssystemen. IBM - Nachrichten (1967) H. 186, S. 777 - 785.

/77/ Gabler, H.: Das Elektronische Datenvermittlungssystem EDS und dessen Einführung in das Netz der Deutschen Bundespost. VDE - Fachberichte 27 (1972) S. 166-168.

/78/ Bocker, P.: Datenübertragung über Fernsprechverbindungen. NTZ 21 (1968) S. 681 - 687.

/79/ Bocker, P.: Nachrichtennetze für Datenfernverarbeitung. VDE - Fachberichte 27 (1972) S. 160 - 165.

/80/ Goßlau, K.: Der Datenverkehr als Nachrichtenverkehr mit digitalen Informationen. NTZ 21 (1968) S. 305 - 312.

/81/ Goßlau, K.; Bacher, A.; Kern, P.; Steigenberger H.; Riedel, E; Hausenblas, A.; Lampe, B.; Gaiser, R.; Voss, H.; Müller, H.: Das Elektronische Datenvermittlungssystem EDS, Ein System für den Datenverkehr. NTZ 22 (1969) S. 444 - 463.

/82/ Oden, H.: Das Telefon im Datendienst.

/83/ Schreiber, F.: Klassifizierung von Aussenstationen und Übermittlungsnetzen für Datenverarbeitung. Elektron. Rechenanl. 12 (1970) S. 20 - 25.

III.5. DATENSTATIONEN

/84/ Letsche, D.: Datenübertragung. IBM-Nachrichten 1969 H. 197, S. 876 - 882; H. 198, S. 959 - 967; H.199, S. 84 - 90; H. 200, S. 176 - 183; H. 201, S.266-272.

/85/ Tietz, W.: Datenendeinrichtungen in Fernmeldewählnetzen. NTZ 25 (1972) S. K 188 - K 192.

/86/ NTG Fachausschuß 12: Endgeräte an Datennetzen. Diskussionstagung der Nachrichtentechn. Ges., Fachausschuß 12 "Datennetze und Datenendgeräte", Stuttgart, Okt. 1972.

III.6. EINIGE BEISPIELE REALER DATENNETZE

/87/ The Infotech State of the Art Report: Computer Networks. Infotech Information Lim. Maidenhead, Berkshire, England, 1971.

/88/ Gibson, M.E.; Millard, D.R.: An Experimental Manually-Switched 48 k bit/s Data Transmission Network. POEEJ 64 (1971) S. 52 - 58.

/89/ Roberts, L.; Wessler, B.: A Computer Network Development to achieve Resource Sharing. Proc. AFIPS SJCC 1970, S. 543 - 549.

/90/ SITA: Annual Report 1968 - 1969. Société International Télécommunications Aéronautiques, Neuilly, France.

/91/ Fisher, C.R.; Slight, R.L.: The Datran Network. Proc. ACM/IEEE Symp. on Problems in the Organization of Data Commun. Systems, Palo Alto, Calif., Oktober 1971, S. 65 - 72.

/92/ Bender, H.M.; Riml, O.: A Synchronous Digital Data Network for Canada. Proc. ICC, Philadelphia, Pa., Juni 1972, S. 13.1 - 13.3.

/93/ Jousset,A.; Profit,A.: The CADUCÉE Project. L'ÉCHO des Recherches 60 (1970) S 10-17. Englische Übersetzung Post Office Research Department, Dollis Hill, London, Hill, London, Transl. No. 2712

/94/ Depres, R.: La Commutation de paquets dans un nouveau réseau de transmissions de données. Proc. Zürich-Seminar, Zürich, März 1972, S. E 2 (1)- E 2 (5).

/95/ Lindberg,G.: Swedish Plans for a Public Data Network. Zürich Seminar, Zürich, März 1972, S. E 5 (1) - E 5 (8).

Weiteres Schrifttum zu Datenvermittlungen und Anschlußnetzen:

/96/ Bell: No. 1 Electronic Switching System Arranged with Data Features. BSTJ 49 (1970) H. 10 mit Beiträgen verschiedener Autoren.

/97/ Caquelin, M.W.: Improved Record Communications in the Terminal Area. Proc. ICC, Philadelphia, Pa., Juni 1972, S. 7.1 - 7.4.

/98/ Kammerl, A.: Eine vollelektronische Fernschreib- und Datenwählvermittlung. NTZ 19 (1966) S. 322-330.

/99/ Huber, J.F.: High Performance Program Request Control. Proc. IFIP, Ljubljana, Jugoslawien, August 1971, S. TA-4-54 bis TA-4-60.

/100/ Gabler, H.: Technik des Elektronischen Datenvermittlungs-Systems EDS. Jahrb. des elektr. Fernmeldewesens (1971) S. 296-337.

/101/ Hummel, E.: Betriebliche Leistungsmerkmale des neuen Elektronischen Datenvermittlungs-Systems EDS. Jahrb. des elektr. Fernmeldewesens (1971) S. 338-358.

/102/ Petersen, J.: Empfang und Übertragung von Daten durch programmgesteuerte Datenvermittlungsanlagen. NTZ 24 (1971), S. 380-384.

/103/ Bartel,W.; Hessenmüller,H.; Schaffers, W.: Ein Verfahren zur synchronen Datenübertragung zwischen EDS-Vermittlungsstellen. Zürich-Seminar, Zürich, März 1972, S. E 6 (1) - E 6 (5).

/104/ Fick, H.: A Switching System for Synchronous Data Networks. Proc. ISS, Cambridge, Mass., Juni 1972, S. 243-247.

/105/ Staudinger,W.: Circuit and Message Switching by the Data Switching System EDS. NTZ 25 (1972) S. 40-42.

/106/ Schwarzer, M.: Zeitmultiplex - gesteuerte Datenlenkung in einer Vermittlung mit Kettenstruktur. NTZ 21 (1968) S. 693-695.

/107/ Atkin, J.; Edwards, B.A.: Phased Implementation of Telex and Data Systems. Proc. ICC, Philadelphia, Pa., Juni 1972, S. 5.12-5.12 D.

/108/ Bell, H.V.; Cook, J.G.: Message Switching: the System at ICI. Systems Technology 11 (1970) S.11-16.

/109/ Scantlebury, R.A.; Wilkinson, P.T.; Bartlett, K.A.: The Design of a Message Switching Centre for a Digital Communication Network. Proc. IFIP, Edinburgh, August 1968.

/110/ Wilkinson, P.T.; Scantlebury,R.A.: The Control Functions in a Local Data Network. Proc. IFIP, Edinburgh, August 1968.

/111/ Bartlet, K.A.: Transmission Control in a Local Data Network. Proc. IFIP, Edinburgh, 1968.

/112/ Benmussa, H.; Girinsky,A.: Das Fernschreibvermittlungssystem DS 4. Elektr. Nachrichtenwesen. 44 (1969) S.41-51.

/113/ Schramel F.J.: DS 714, a Message Switching System for Telegraph and Data Communication Proc. Coll Int. Téléformatique, März 1969, S. 383 - 392.

/114/ Philips: New 'all-electronic' Telegram Retransmission Centre to Handle International Telegram Traffic. Revue F.I.T.C.E., Sept. - Okt. 1971, S.19-21.

/115/ Ogata, K.: Recent Accomplishments and Future Plans for Switching in Japan. Proc. ISS, Cambridge, Mass., Juni 1972, S. 7-10.

Weiteres Schrifttum zu Netzen:

/116/ Elton, M.J.: Evelopment of a Multi-Computer Data Processing Network. SITA-Symposium, Paris, Februar 1968.

/117/ Herzog, B.: Computer Networks. Proc. ICS, Venedig, April 1972, S. 182-199.

/118/ Nowakoski,D.: State Integrated Information Net (IINET). Proc. ACM / IEEEE Symp. on Problems in the Optimization of Data Commun. Systems Palo Alto, Calif. Oktober 1971, S. 137-147.

/119/ Pehrson, D.L. An Engineering view of the LRL Octopus Computer Network. University of California, Livermore 1970.

/120/ Petrasek, A.C.: Data Exchange between Computers-today and tomorrow. SITA-Symposium, Paris, Februar 1968.

/121/ Watts, R.N.: Some Design Considerations for narrow-Band Medical Telemetry over the Switched-Message Network IEEE Trans.Commun. Techn. 19 (1971) S. 246-255.

/122/ Barber,D.L.A.; Bartlett,K.A.; Dewis, I.G.: A Review of the Performance of the N.P. L. Data Communications Network. Tagungsbericht zum Symp. über Computer Netze, Birlinghoven, BRD, Mai 1972.

/123/ Davies, D.W. The Principles of a Data Communication Network for Computers and Remote Peripherals. Information Processing. North-Holland Publ. Comp., Amsterdam, (1969) S. 709-713.

/124/ Davies, D.W.; Bartlett, K.A.; Scantlebury, R.A.; Wilkinson,W.: A Digital Communication Network for Computers giving rapid Response at Remote Terminals. Proc. IFIP, Edinburgh, August 1968.

/125/ Laver,F.J.M.: National Data-Processing Service - the way ahead. POEEJ (19) S.71-73.

/126/ Marsh, H.; Balls, G.C.J.: A Data Transmission Network for the Ministry of Social Security. POEEJ 61 (1969) S. 224-227.

/127/ Williams,M.B.: Studies for a Public Data Communications Network. SITA-Symposium, Paris, Feburar 1970.

/128/ Girardi, S.: SOC-Project: an Experimental Computer Network. Proc. ICS, Venedig, April 1972, S. 210-220.

/129/ Tietz, W.: Datenübertragung und Datenfernverarbeitung in anderen Ländern. ZPF (1971) S. 289-295

sowie Literatur /27,29,31-34,41-44,50-57,60,62-64,77,81,84,86/.

Zu Kapitel IV: VERKEHRSTHEORTEISCHE PROBLEME BEI DATENNETZEN UND IHRE BEHANDLUNG

IV.2. DER WEG ZUR OPTIMIERUNG

/130/ Herzog, U.: Optimization of Data Network with Computerized switching and Processing Centers. Proc. Int. Sem. on Tele Commun. Dipoli, Finnland, 1970, S. 34-70 / 1-13.

IV. 3.1. ANSCHLUSSNETZ

/131/ Avi-Itzhak,B.: Heavy Traffic Characteristics of a Circular Data Network. BSTJ 50 (1971) S. 2521-2549.

/132/ Chu, N.W.; Konheim,A.G.: On the Analysis and Modeling of a Class of Cumputer Communication Systems.IEEE Trans.Commun. 20 (1972), S. 645-660.

/133/ Gall,D.; Mueller, H.R.: Waiting Time Distributions and Buffer Overflow in Priority Systems. Proc. ICC, Philadelphia, Pa., Juni 1972, S. 32.21 - 32.26.

/134/ Hayes, J.F.; Sherman,D.M.: Traffic Analysis of a Ring Switched Data Transmission System.BSTJ 50 (1971) S. 2947-2978.

/135/ Konheim,A.G.; Meister,B.: Service in a Loop System. IBM Research Lab., Zürich, Oktober 1970.

/136/ Spragins, J.D.: Analysis of Loop Transmission Systems. Proc. ACM / IEEE Symp. on Problems in the Optimization of Data Commun. Systems, Palo Alto, Calif., Oktober 1971, S. 175-182.

/137/ Botsch, D.: Die Verlustwahrscheinlichkeit einstufiger Koppelanordnungen der Vermittlungstechnik mit Extern- und Internverkehr. Dissertation, TH Stuttgart, 1966. und A.E.Ü 22 (1968). S. 127 - 132.

/138/ Chu, W.W.: Design. Considerations of Statistical Multiplexors. Proc. ACM Symp. on Problems in the Optimization of Data Commun. Systems, Pine Mountain, Ga., August 1969 S. 39-60.

/139/ Chu, W.W.: A Study of Asynchronous Time Division Multiplexing for Time Sharing Computers. 1969 Spring Joint Computer Conf. AF / PS Conf. Proc. Montvale, N.J., vol. 35, S.669-678.

/140/ Chu, W.W.: Buffer Behavior for Batch Poisson Arrivals and Single Constant Output. IEEE Trans. Commun. Techn. 18 (1970) S. 613-618.

/141/ Rudin , H.R.: Data Transmission: a Direction for Future Development. IEEE Spectrum 7 (1970) S. 79-85.

/142/ Rudin, H.R.: Performance of a Simple Multiplexer Concentrator for Data Communication. IEEE Trans. Commun. 19 (1971) S. 178 - 187.

/143/ Rudin, H.R.: On Economies of Scale and Integration of Services in certain queued Information Transmission Systems Proc. ICC, Philadelphia, Pa., Juni 1972, S. 32.27 - 32.31.

/144/ Weingarten, A.: Response Time in a total Airlines Computer System. American Airlines, Inc., Juli 1969.

/145/ Doll, D.R.: Topology and Transmission Rate Considerations in the Design of Centralized Computer-Communication Networks . IEEE Trans. Commun. Techn. 19 (1971) S. 339-344.

/146/ Gerich, K.: Eine Betrachtung der System-Design Verfahren für Realzeitsysteme. IBM-Nachrichten 205 (1971) S. 661-666.

/147/ Chandy, K.M.; Russell, R.A.: The Design of Multipoint Linkages in a Teleprocessing Tree Network. IEEE Trans. Comp. 21 (1972) S. 1062-1066.

/148/ Esau, L.R.; Williams, K.C.: A Method for Approximating the Optimal Network. IBM Syst. J. 5 (1966) S. 142-147.

/149/ Frank, H.; Frisch, I.T.; Van Slyke, R.; Chou, W.S.: Optimal Design of Centralized Computer Networks. Networks 1 (1971) S. 43-57.

/150/ Bahl, L.R.; Tang, D.T.: Optimization of Concentrator Locations in Teleprocessing Networks. Symp. Com.-Commun. Networks and Teletraffic. Plytechn. Inst. Brooklyn, New York, N.Y., April 1972, Session IX.

/151/ Hyvärinen, L.P.: Optimum Communication Network Design. Proc. Nachrichtentechn. Koll., Universität Bern, 1969/70, S. 93-102.

/152/ Randazzo, F.P.: Optimale Auslegung von Nachrichtennetzen mit Hilfe der Wartezeittheorie. Elektr. Nachrichten wesen. 38 (1963) S. 524-537.

IV. 3.2. DATENVERMITTLUNGEN

/153/ Margopoulos, W.P.; Williams, R.J. Characteristic Problems (on Teleprocessing System Design). IBM Syst. J. 5 (1966) S.134-141.

/154/ Chang, W.: Single Server Queuing Processes in Computing Systems. IBM Syst. J. 9 (1970) S. 36-71.

/155/ IBM: Analysis of some Queuing Models in Real - Time Systems. IBM Data Processing Techniques, White Plains, N.Y., F 20-0007-1.

/156/ Spragins, J.D.: Simulation of a Communication Processor under Extreme Loading Conditions. IEEE Trans. Commun. 20 (1972) S. 609-619.

/157/ Hauth, C.: Turnaround Time for Messages of differing Priorities. IBM. Syst. J., 7 (1968) S.103-122.

/158/ Gay, T.W.: A Technique for Estimating Channel Interference. IBM Syst. J. 5 (1966) S. 171-189.

/159/ Chang, J.H.: An Analysis of Buffering Techniques in Teleprocessing Systems. IEEE Trans Commun. 20 (1972) S. 619-630.

/160/ Bricault, J.P.; Delgavis.: An Analysis of a Request-Queued Buffer Pool. IBM Syst. J. 5 (1966) S. 148-157.

/161/ Seaman, P.H.; Lind, R.A.; Wilson, T.L.: An Analysis of Auxiliary Storage Achivity. IBM Syst. J. 5 (1966) S. 158-170.

/162/ Iazeolla, G.: Computer Controlled Message Switching Systems: Evaluating the Efficiency through the Hierarchical Message Store. Proc. Coll. Ing. Téléinformatique, Paris, März 1969, S.338-347.

/163/ Bovet, D.P.; De Julio, S.: On Queuing Models of Disk Storage Systems. Proc. ICS, Venedig, April 1972, S. 547-560.

/164/ Kümmerle, K.: An Analysis of Loss Approximations for Link Systems. Proc. 5. IIC, New York, N.Y., Juli 1967 S. 327-336 und A.E.Ü 25 (1971) S. 466-471.

/165/ Kümmerle, K.: Berechnungsverfahren für mehrstufige Koppelanordnungen mit konjugierter Durchschlatung-Systematik und Analyse. Dissertation, Universität Stuttgart, Stuttgart, Oktober 1969.

/166/ Bazlen, D.: Call Congestion in Link Systems with Internal and External Traffic. Proc. 6. IIC, München, September 1970, S.244/1-244/7 und A.E.Ü 25 (1971) S.479-483.

/167/ Herzog, U.: A General Variance Theory applied to Link Systems with Alternate Routing. Proc. 5. ITC, New York, N.Y.; Juli 1967, S. 398-406.

/168/ Lörcher,W.: Exakt Calculation of the Probability of Loss for two - Stage Link System with Group Selection. Proc. ITC, München, September 1970, S. 236/1-236/9 und A.E.Ü. 25 (1971) S. 471-476.

/169/ Lotze, A.: Berechnung der Blockerungs- und Verlustwahrscheinlichkeit zwei- und mehrstufiger Linksysteme für Mischwahl oder Gruppenwahl nach dem Näherungsverfahren der "kombinierten Eingangs- und Abnehmerblockierung". In Lotze,A. und Wagner,W.: Tafel der modifizierten Palm-Jacobaeus-Verlustformel, Techn. Hochschule, Stuttgart, Stuttgart 1961.

/170/ Hieber,L. About Multi-Stage Link Systems with Queuing. Proc. ITC, München, Sept. 1970, S. 233/1-233/7 und A.E.Ü. 25 (1971) S. 483-487.

/171/ Huber,M.: On the Congestion in TDM Systems. Proc. ITC, London, Juli 1964, Dokument 104.

/172/ Tanaka, M.: Nonblocking Switching in Integrated PCM Networks. Proc. ITC, München, Sept. 1970, S.443/1-443/3 und A.E.Ü. 25 (1971) S. 477-478.

/173/ Cantor,D.G.: On Non-Blocking Switching Networks. Networks 1 (1971) S. 367-377.

/174/ Lotze,A.: Optimum Link Systems. Proc. 5. ITC, New York, Juli 1967, S. 242-251.

IV. 3.3. VERMITTLUNGSNETZ

/175/ Kroneberg,D.: Cumputer unter sich. Computer-Praxis 1 (1968) S. 237-239.

/176/ Schwartz,M.: Digital Data Networks. IEEE Intern. Convention, New York, N.Y., März 1972.

/177/ Burke,P.J.: The Output of a Queuing System.Operations Res. 4 (1956) S.699-704.

/178/ Jackson,J.R.: Networks of Waiting Lines. Operations Res. 5 (1957) S.518-521.

/179/ Kleinrock,L.: Communication Nets, Stochastic Message Flow and Delay. Mc Graw-Hill Book Company New York, 1964.

/180/ Meister, B.; Müller,H.R.; Rudin,H.R.: New Optimization Criteria for Message Switching Networks. IEEE Trans. Commun. Technol. 19 (1971) S. 256-260.

/181/ Meister, B.; Müller, H.R.; Rudin H.R.: On the Optimization of Message Switching Networks. IEEE Trans. Commun. 20 (1972) S.8-14.

/182/ Fratta,L.; Gerla,M.: The Synthesis of Computer Networks: Properties of the Optimum Solution. Proc. I CS, Venedig, April 1972, S. 200-209.

/183/ Frank,H.; Kahn, R.E.; Kleinrock,L.: Computer Commun. Network Design: Experience with Theory and Practice. Network 2 (1972) S. 135-166.

/184/ Ottermann,J.: Probability Distribution of Delays for store- and forwand Traffic mixed with Direct Traffic. IEEE Trans. Commun. Syst. 10 (1963) S.41-48.

/185/ Thierer,M.: Wartesysteme mit unvollkommener Erreichbarkeit.Dissertation, Universität Stuttgart, Stuttgart 1968 und A.E.Ü. 23 (1969) S. 261-267.

/186/ Silk,D.J.: Queueing Model for Message-switching Networks with Constant -length Messages. IEE 116 (1969) S. 1821 - 1826.

/187/ Lotze, A.: DDD Network Optimization in Field Engineering-from Theory to Application. 7. ITC Stockholm, Juni 1973.

/188/ Katzschner,L.: Loss Systems with Displacing Priorities. Proc. 6 ITC, München, September 1970, S. 224/1-224/8 und A.E.Ü. 25 (1971) S. 416-422.

/189/ Herzog, U.; Kühn, P.: Comparison of some Multi-queue Models with Overflow and Load-Sharing Strategies for Data Transmission and Computer Systems. Symp. Com.-Commun. Networks and Teletraffic. Polytechn. Inst. Brooklyn, New York, N.Y., April 1972, Session XII.

IV.4. METHODEN ZUR BEHANDLUNG VERKEHRSTHEORETISCHER Probleme

/190/ Zurmühl,R.: Praktische Mathematik für Ingenieure und Physiker. Springer-Verlag, Berlin, 1961.

/191/ Busacker,G.R.; Saaty,L.: Endliche Graphen und Netzwerke. Oldenburg-Verlag, München-Wien, 1968.

/192/ Churchman,C.W.; Ackhoff,R.L.; Arnoff, E.L.: Operations Research. Oldenburg-Verlag, Wien-München 1961.

/193/ Elmaghraby,S.E.: Some Network Models in Management Science. Springer-Verlag, Berlin, New York, 1970.

/194/ Hu,T.C.: Ganzzahlige Programmierung und Netzwerkflüsse. Oldenbourg-Verlag, München, Wien 1972.

/195/ Nemhauser,G.L.: Einführung in die Praxis der dynamischen Programmierung. Oldenbourg-Verlag, München-Wien, 1969.

/196/ Zimmermann,H.J. Zielinski,J.: Lineare Programmierung. De Gruyter-Verlag, Berlin New-York, 1971.

/197/ Gordon,G.: System Simulation. Prentice-Hall, Inc., Englewood Cliffs, N.J. 1969.

/198/ Huber,M.; Wagner,W.: Simulation von Nachrichtenvermittlungssystemen. In Nicht-numerische Informationsverarbeitung, Springer-Verlag, Wien-New York, 1968.

/199/ Kampe,G.: SIMSCRIPT. Vieweg-Verlag, Braunschweig, 1971.

/200/ Weisschuh,H.: Theoretical Aspects of Finite Source Input Processes Simulation. Proc. z. ITC, Stockholm, Juni 1973.

/201/ Boehm,B.W.; Mobley,R.L.: A Computer Simulation of Adaptive Routing Techniques for distributed Communications Systems. Rand Corporation,Santa Monika, Calif., Memorandum RM-4782-PR, Feburar 1966.

/202/ Caplan,D.I.; Yonda,A.W.: Discrete Simulation Techniques for Data Communcications Networks. Summer Comp. Simulation Conf. Boston, Mass.; 1971, AFIPS Conf. Proc.

/203/ Price,W.L.: Simulation of Data Transit Networks.National Physical Laboratory, Teddington, Engl., Com Sci 56, April 1972.

/204/ Seaman,P.H.: The Role of Digital Simulation. IBM Syst.J. 5 (1966) S. 175-189.

/205/ Slight, R.L.: A Description of the DATRAN Network Simulator and its Use. Proc. ISS, Cambridge, Mass. Juni 1972, S. 224-232.

/206/ Cox,D.R.; Smith,W.L.: Queues. Methuen & Co Ltd, London; Wiley & Sons Inc., New York, 1961.

/207/ Lahres,H.: Einführung in die diskreten Markoff-Prozesse und ihre Anwendungen. Vieweg-Verlag, 1964.

/208/ Saaty, T.L.: Elements of Queuing Theory. Mc Graw-Hill, New-York, 1961.

/209/ Takács, L.: Stochastische Prozesse. Oldenbourg-Verlag, München und Wien, 1966.

/210/ Boes,D.C.: Note on the Output of a Queuing System. J: Appl. Prob 6 (1969) S. 459-461.

/211/ Burke, P.J.: The Output Process of a Stationary M/M/S Queueing System. Ann. Math. Statistics. 39 (1968) S. 1144-1152.

/212/ Shanbhag,D.N.: Letter to the editor. J.Appl Prob 9 (1972) S. 470.

/213/ Krämer,W.: Total Waiting Time Distribution Function and the Fate of a Customer in a System with two Queues in Series. Proc. 7. ITC, Stockholm, Juni 1973.

Zu Kapitel V: DURCHLAUFZEITEN VON NACHRICHTEN IN NETZEN MIT SPEICHERVERMITTLUNGEN UND ALTERNATIVER LEITWEGLENKUNG

/214/ Bretschneider,G.: Die Berechnung von Leitungsgruppen für überfließenden Verkehr in Fernsprechwählanlagen. Nachrichtentechn. Z. 9 (1956), S.533-540.

/215/ Lotze,A.: A Traffic Variance Method for Gradings of Arbitrary Type. Proc. 4. ITC, London, Juli 1964, Dokument 80 und POEEJ, Special Issue 1966.

/216/ Wilkinson,R.I.: Theories for Toll Traffic Engineering in the U.S.A. BSTJ 35 (1956) S. 421-514.

/217/ Schehrer,R.: Über die exakte Berechnung von Überlaufsystemen der Wählvermittlungstechnik. Dissertation, Universität Stuttgart, Stuttgart 1969.

/218/ Brockmeyer,E.; Halstrom,H.; Jensen,A.: The Life and Works of A.K. Erlang, Acta Polytechnica Scandinav. (1960) Nr. 287.

/219/ Doetsch,G.: Anleitung zum praktischen Gebrauch der Laplace-Transformation und der Z-Transformation. Oldenbourg-Verlag, München-Wien, 1967.

Zusätzliches Schrifttum

/220/ Basharin,G.P.: On Analytical and Numerical Methods of Switching System Investigation. Proc. 6 ITC, München, September 1970, S. 231/1 - 231/11.

/221/ Bhat, U.N.: Customer Overflow in Queues with finite Waiting Space. Austral. J. Statist., 7 (1965) S.15-19.

/222/ Cinlar,E.; Disney,R.L.: Stream of Overflow form a finite Queue. Operations Res. 15 (1967) S. 131-134.

/223/ Herzog,U.: Die exakte Berechnung des Streuwertes von Überlaufverkehr hinter Koppelanordnungen beliebiger Stufenzahl mit vollkommener bzw. unvollkommener Erreichbarkeit A.E.Ü. 20 (1966) S. 180-184.

/224/ Herzog, U.: Die Bemessung ein- und mehrstufiger Koppelanordnungen der Vermittlungstechnik für angebotenen Überlaufverkehr. Dissertation, Universität Stuttgart, Stuttgart, 1968.

/225/ Kümmerle,K.: A Model for the I/o-Channel Traffic in Computer Systems. Nachrichtentechn. Z. 25 (1972) S.137-141.

/226/ Schehrer,R.: On the Calculation of Overflow Systems with a finite Number of Sources and full available Groups. Symp. Comp.-Commun. Networks and Teletraffic. Polytechn. Inst. Brooklyn, New York, N.Y., April 1972, Session V.

/227/ Rosenbohm, W.: Synthese von Fensprechvermittlungsnetzen. Diplomarbeit, Universität Stuttgart, Stuttgart 1969.

/228/ Globas, G.: Programm zur Simulation von Daten - und Fernsprechnetzen nach dem sog. "line-switching" Prinzip. Diplomarbeit, Universität Stuttgart, Stuttgart 1969.

/229/ Rheinwald, S.: Simulationsprogramm für ein Wartesystem mit Überlaufverkehr und alternativer Leitweglenkung. Diplomarbeit, Universität Stuttgart, Stuttgart 1970.

/230/ Schmitz, G.: Programm zur Simulation des Verkehrsflusses in Datennetzen nach dem sog. "message-switching" Prinzip. Diplomarbeit, Universität Stuttgart, Stuttgart 1970.

/231/ Werres, B.: Wartesysteme mit Überlauf. Diplomarbeit, Universität Stuttgart, Stuttgart 1971.

/232/ Ott, R.: Wartesysteme mit angebotenem Überlaufverkehr. Diplomarbeit, Universität Stuttgart, Stuttgart 1972.

/233/ Koch, U.: Untersuchungen des Verkehrsflusses in Datenvermittlungen und Datennetzen mit "message-switching". Diplomarbeit, Universität Stuttgart, Stuttgart 1973.

Zu Kapitel VI: EINE STELLITENVERMITTLUNG FÜR "PACKET-SWITCHING" MIT PRIORITÄTEN

/234/ Konheim,A.G.; Meister,B.: Two-Way Traffic in Loop Service Systems. Networks 1 (1971) S. 291-301.

/235/ Rudin,H.: Buffered Packet-Switching, a Queue with Clustered Arrivals. Proc. ISS, Cambridge, Mass., Juli 1972, S. 259-265.

/236/ Ley, J.: Eine Satellitenvermittlung für "packet-switching" mit Prioritäten. Diplomarbeit, Universität Stuttgart, Stuttgart 1973.

Zu Kapitel VII: UNTERBRECHUNGS-DISTANZ-PRIORITÄTEN

/237/ Jaiswal,N.K.: Priority Queues. Academic Press, New York und London, 1968.

/238/ Chang,W.: Queuing with Nonpreemptive and Preemptive-Resume Priorities. Operations Res. 13 (1965) S. 1020 - 1022.

/239/ Reich,E.: Waiting Times when Queues are in Tandem. Ann. Math. Statist. 28 (1957) S. 768 - 773.

/240/ Meschkowski,H.: Differenzengleichungen. Vandenhoeck und Rupprecht-Verlag, Göttingen, 1959.

Zusätzliches Schrifttum

/241/ Balachandron,K.R.: Parametric Priority Rules: An Approach to Optimization in Priority Queues. Operations Res. 18 (1970). S. 526-540.

/242/ Cobham,A.: Priority Assignment in Waiting Line Problems. Operations Res. 2 (1954) S. 70-76 und 3 (1955) S. 547.

/243/ Gaver,D.P.: On Priority Type Disciplines in Queuing. Proc. Symp. on Congestion Theory (1964), The University of North Carolina Press, Chapel Hill, 1965.

/244/ Takács, L.: Priority Queues. Operations Res. 12 (1964) S. 64-74.

/245/ Vetter, A.: Simulationsprogramm für Wartesysteme mit Unterbrechungs-Distanz Prioritäten. Semesterarbeit, Universität Stuttgart, Stuttgart 1972.

/246/ Röck, H.: Berechnung von Wartesystemen mit Unterbrechungs-Distanz Prioritäten. Diplomarbeit, Universität Stuttgart, Stuttgart 1972.

APPENDIX 1

MESSAGE-SWITCHING NETWORKS WITH ALTERNATE ROUTING

Ulrich Herzog
University of Stuttgart
Stuttgart, Federal Republic of Germany

ABSTRACT

In modern data networks variable routing strategies allow to take into account the instantaneous traffic situation of the network.

Traffics offered to a direct "primary" route can be described realistically by means of Poisson traffic and therefore all characteristic traffic values can be determined by means of well known formulas. However, traffics overflowing from a primary to a secondary route possess quite other stochastic properties than Poisson traffic.

This paper shows, how to dimension store- and forward networks with alternate routing, taking into account the special properties of overflow traffic. Artificial traffic trials show the good accordance between simulated and calculated traffic values.

1. PROBLEM

By reasons of economy, overload and cable-breakdown, networks with variable routing strategies are superior to networks without these facilities.
This paper deals with such networks operated by a store- and forward mode and alternate routing, i.e., traffic is offered at first to a primary channel: if this direct route is overloaded, traffic is diverted to a secondary route, perhaps to further routes.
Such partial traffics rejected by a route possess quite other stochastic properties than "direct traffic" offered to a primary channel (random traffic, pure chance traffic, Poisson traffic).

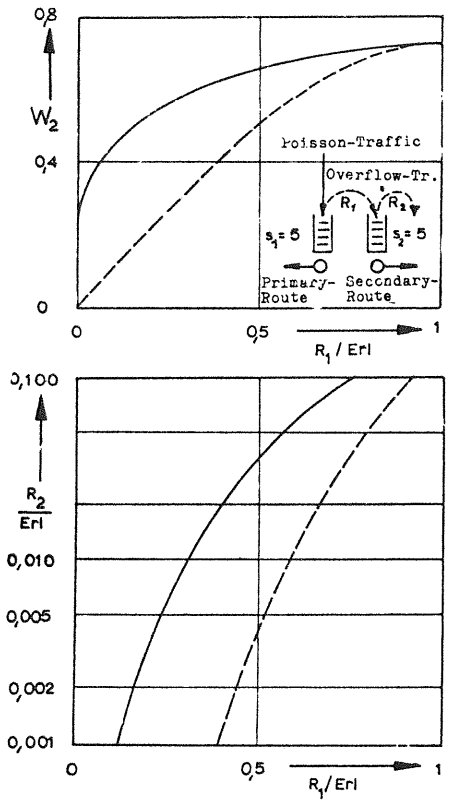


Fig 1: Probability of waiting W_2 and overflow traffic R_2 for a secondary system to which overflow traffic (mean value R_1) is offered. The secondary route is used only (in this example) if there are waiting already $s_1 = 5$ messages in the primary system. The dashed curves are obtained when assuming Poisson traffic with the same mean value.

As an example fig. 1 demonstrates the influence of these special properties on the traffic flow of the overflow system: the solid curves show values calculated exactly whereas the dashed curves are obtained when assuming random traffic instead of overflow traffic.

Therefore, neglecting the special stochastic behavior of overflow traffic, both the probability of waiting W_2 and the overflow traffic R_2 are underestimated considerably. This fact holds for all characteristic traffic values!

This typical property is well known for telephone networks (cf. Bretschneider /1/, Lotze/2/, Wilkinson /3/, etc.) and modern dimensioning methods - e.g. the design tables of the german PTT for local and toll networks - are taking into consideration the special stochastic properties of overflow traffic.
Now, these special stochastic properties have to be taken into account for data networks, too.

Various routing strategies for data networks have been proposed in the past (cf. Boehm and Mobley /4/, Fultz and Kleinrock /5/, Silk /6/, Prosser /7/, see also Butrimenko /8/, Brandt and Chrétien /9/, Beeforth, Grimsdale, Halsall and Woollons /10/, Davies /11/, Frank and Chou /12/, Kahn and Chrowther /13/, Lotze /14/, Petersen and Fu /15/, etc.).
In many papers comparisons have been done by means of simulation. For the first time, comparisons between different routing strategies by means of analytical methods have been made by Herzog and Kühn /16/. Among others there has been presented the basic idea for the dimensioning of data networks with alternate routing.

This method is now extended, i.e.

- the traffic flow within an individual data switching center is investigated by means of a more detailed model of the structur and
- total flowtimes of messages in networks with several nodes are determined.

Besides characteristic mean values and typical probability values, now also results for the waiting time distribution functions are presented.

2. MODELLING

2.1. Structure and Operating mode of Data Switching Centers.

Fig. 2. shows a simplified model for a modern data switching center. Messages arrive from other nodes of the network or from the terminal area of the considered node. These messages are stored either in different input-queues or in a common input-queue, resp.

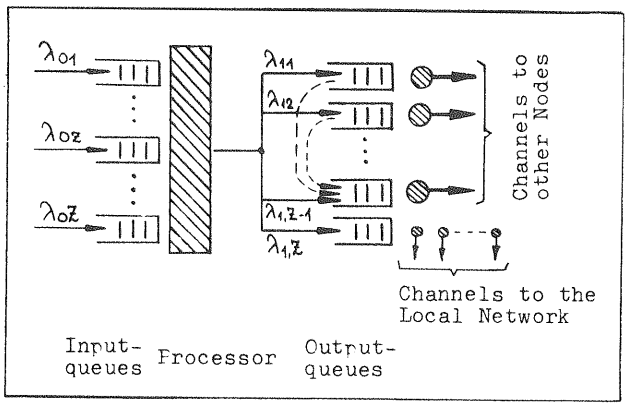


Fig.2.: Model of a data switching center. The dashed arrows indicate overflow traffic (cf. also section 2).

Operating of the input-queues will be done (e.g. in a cyclic mode) by the central processor. The processor interprets the header-information and determines the route to be chosen: if the direct route is overloaded, i.e. if there is waiting a distinct (but arbitrarily selectable) number of messages, traffic is diverted to a secondary route (overflow traffics are indicated by dashed arrows in fig. 2).

According to this routing strategy, messages are put into an output-queue waiting for transmission to the next node.
All input-queues are assumed to be unlimited, i.e. no messages are rejected by the data switching center. Within one distinct queue messages are handled in the order of their arrivals (FIFO).

2.2. Traffic.

Messages arriving from the different directions are assumed to be distributed according to a Poisson-process, i.e. the distribution function of the interarrival times is given by

$$A_{0z}(\leq t) = 1 - e^{-\frac{t}{a_{0z}}} = 1 - e^{-\lambda_{0z} t}$$

with the mean interarrival time a_{0z} or the mean arrival rate λ_{0z}, respectively for each direction z (z = 1,2,....Z).
This assumption approximates very close the actual interarrival times of messages, generated by the terminals. In addition, simulation results show clearly, that - assuming also negativ exponentially distributed message lengths - this holds true also for messages received from the neighbour nodes.

Processing times in the central processor are assumed to be negative exponentially distributed and uniform for each direction:

$$B_{0z}(\leq t) = 1 - e^{-\frac{t}{b_0}}$$

with the mean processing time b_0.
If η_{zj} indicates the probability, that an incoming message of direction z should be transmitted in the outgoing direction j (primary route!), the total arrival rate for "direct" traffic is

$$\lambda_{1j} = \sum_{z=1}^{Z} \lambda_{0z} \cdot \eta_{zj}$$

The transmission speed may be arbitrarily chosen for each outgoing channel. Hence for messages with negative exponentially distributed holding times, the distribution function of the transmission times (for channel j) is:

$$B_{1j}(\leq t) = 1 - e^{-\frac{t}{b_{1j}}}$$

with the mean transmission time b_{1j} per message in direction no. j.
The combination of both negative exponentially distributed interarrival and holding times (processing or transmission) is designated commonly as Poisson traffic, random traffic or pure chance traffic.

3. PRINCIPLE OF SOLUTION

3.1. General remarks.

In principle, an exact investigation of the traffic flow in data networks is possible: multi-dimensional state-probabilities are introduced, the Chapman-Kolmogoroff equations have to be determined and the evaluation has to be done with the aid of a relaxation method (cf. Schehrer /13/). However, for network structures of practical interest, these systems of equations are so large that their evaluation is not possible, even on the largest digital computers.
In this paper an approximate method is suggested, which allows to determine the traffic flow even in large networks. Firstly, the traffic flow through an individual data switching center is investigated and secondly, networks with several nodes are considered.

3.2. Traffic flow through a single data switching center.

When investigating the traffic flow the follow-

ing four subsystems of a data switching center may be distinguished (cf. fig. 2):

- the input-system with the input-queues and the central processor,
- the primary-systems; each primary-system consists of one output-queue and one transmission channel to which direct Poisson traffic is offered only,
- the secondary-systems; each secondary-system consists of one output-queue and one transmission channel to which both offered overflow and direct traffic or overflow traffic only is directed and
- the terminal-system, connecting the data switching center with the local network and the terminals.

3.2.1. The input system.

The input-system corresponds to a single-server queuing system with a common or several parallel inputqueues, respectively, Poisson arrivals and negative exponentially distributed holding times (M/M/1). Such systems have been investigated and mathematically treated in numerous publications. Hence, formulas for all characteristic traffic values are available.
Burke /18/ proved with his famous theorem, that the output process of such an input-system is Poissonian, i.e. the interarrival times of messages offered to the transmission area of the data switching center (primary-, secondary-, terminal-systems) are also negative exponentially distributed.

3.2.2. The primary systems.

Only direct Poissonian traffic is offered to the primary systems - (M/M/1) with limited waiting room -. Therefore, formulas for many characteristic traffic values are available, such as mean waiting times, probability of waiting, waiting time distribution function etc. Traffics rejected by a primary system and offered to a secondary system (overflow traffics) are characterized in this paper not only by the first moment (mean value R_{1i}) but also by the second moment (variance σ_{1i}^2 or variance coefficient $D_{1i} = \sigma_{1i}^2 - R_{1i}$). By doing so the special stochastic character of overflow traffic can be taken into account in close approximation.

3.2.3. The secondary systems.

In the most general case, both

- direct traffics (mean $A_{1j} = \lambda_{1j} \cdot b_{1j}$, variance $\sigma_{1j}^2 = A_{1j}$ and therefore the variance coefficient $D_{1j} = 0$) and
- overflow traffics (mean R_{1i}, variance $\sigma_{1i}^2 > R_{1i}$, variance coefficient $D_{1i} > 0$)

are offered to a secondary system. Therefore, the total traffic offered to the secondary system no. j is described now not only by the sum of all mean values but also by the sum of all variance coefficients:

$$\boxed{\begin{aligned} \overline{R}_{1j} &= A_{1j} + \sum_i R_{1i} \\ \overline{D}_{1j} &= \sum_i D_{1i} \end{aligned}}$$

In order to calculate the characteristic traffic values of the secondary system, a "substitute primary arrangement" and a "generating traffic" A^* are determined such that an overflow traffic is generated with the mean value $\overline{R}_{1j}$ and the variance coefficient $\overline{D}_{1j}$ (cf. fig. 3). In other words: all actual traffics offered to the considered secondary system are prescribed approximately by one substitute overflow traffic with the exact mean value $\overline{R}_{1j}$ and the exact variance coefficient $\overline{D}_{1j}$ (and of course the exact variance $\overline{\sigma}_{1j}^2$). Moments of higher order are neglected.

Then this total arrangement - one substitute primary system offering overflow traffic ($\overline{R}_{1j}$, $\overline{D}_{1j}$) to the actual secondary system - is treated exactly by means of mathematical methods. Having determined the probabilities of state for this total arrangement all characteristic traffic values of the actual secondary system are calculated, too.

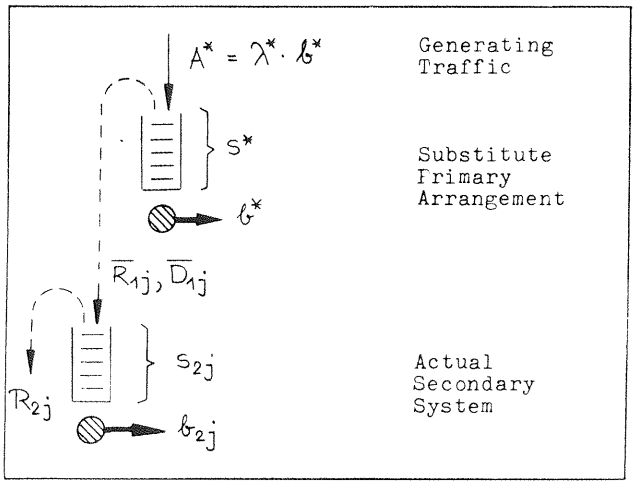


Fig. 3: Total arrangement for the calculation of the characteristic traffic values of secondary systems (cf. section 3.2.3.).

3.2.4. The terminal-system.

The terminal-system connects the data switching center with the local network and the terminals. Different types of local networks are available, such as individual subscriber lines, common bus-lines, concentrators, loop-networks etc. The main characteristic of all structures and operating modes is, that in general, there exists only one fixed route between an individual subscriber and the data switching center, i.e. there is no variable routing. Therefore, all traffics can be assumed to be Poissonian and there is no overflow traffic with special stochastic properties. Detailed studies on the traffic flow in local networks have been published (a summary of many publications is given by Herzog in /19/).

3.3. Traffic flow in a network with several nodes.

By means of the method, outlined in the preceding sections it is possible to investigate the traffic flow in each data switching center for given traffics. In data networks with several nodes the traffics are interdependent because of overflow. Therefore, the traffics offered actually to a specific route have to be determined often by iteration.
Having determined all characteristic traffic values for each connection between two nodes the total flow times for messages (sum of all waiting-, processing-, and transmission times) can be determined. The total flow times from the "originating" node to the destination are composed of the sum of flow times through the individual nodes, weighted with the probability, that a specific route has been chosen.

Fig. 4 shows as an example a small network with four nodes only. Messages from node I to the destination node II are send either on the primary route (I → II) or on the secondary route (I → III → II).

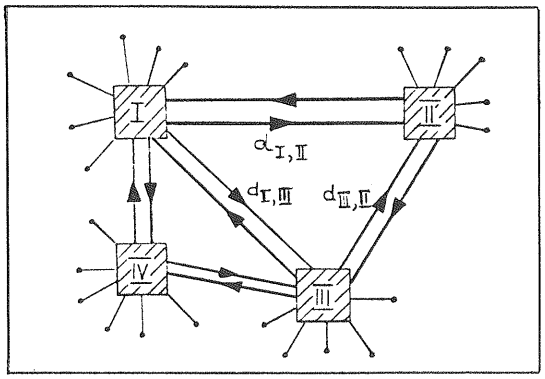


Fig.4.: Example of a small network demonstrating the calculation of total flow times (cf. section 3.3.).

At first we define by

$d_{k,l}$ the mean flow time for all messages transmitted on the channel from node k to node l and

$p_{k,l}(u,z)$ the probability, that when a message - from originating node u to the destination z - has reached the transit node k, it is transmitted via node l.

Then, the total mean flow time d(I,II) for all messages from node I to node II is given in our example by

$$d(I,II) = p_{I,II}(I,II).d_{I,II} + p_{I,III}(I,II)\{d_{I,III} + p_{III,II}(I,II).d_{III,II}\}$$

(where here $p_{III,II}(I,II) = 1$). Analogously, the corresponding equations can be found for all other traffics.

4. CALCULATION OF THE CHARACTERISTIC TRAFFIC VALUES OF A SINGLE DATA SWITCHING CENTER

This section shows in brief some well known formulas for systems with offered Poissonian traffic as well as new results for systems to which overflow traffic is offered.

4.1. The input-system.

The input-system corresponds to a single-server queuing system to which traffic is offered with markovian character (M/M/1). Therefore, formulas for all characteristic traffic values are available. E.g. for the most simple case - one common queue and arbitrary arrival rates λ_{oz} (z = 1....Z) or Z cyclically processed queues with unique arrival rates - the mean waiting time of all messages is given by

$$w_o = \frac{1}{\lambda_o} \cdot \frac{A_o^2}{1-A_o}$$

where

$$\lambda_o = \sum_{z=1}^{Z} \lambda_{oz}$$

$$A_o = \lambda_o \cdot b_o$$

Various queuing desciplines also for different arrival rates and parallel queues have been studied in detail by Kühn /20/.

4.2. The primary systems.

Primary systems are also pure markovian systems - (M/M/1) with $s_{1i} \leq \infty$ waiting places. Therefore, formulas which characterize the waiting process of such a system are well known from literature. In particular, the mean waiting time for all messages carried by a primary system is given by:

$$w_{1i} = \begin{cases} b_{1i} \cdot A_{1i} \cdot \dfrac{1 - A_{1i}^{s_{1i}}[s_{1i}+1 - s_{1i} \cdot A_{1i}]}{1 - A_{1i}} \cdot \dfrac{1 - A_{1i}^{s_{1i}+2}}{(1 - A_{1i}^{s_{1i}+1})^2}; & A_{1i} \gtrless 1 \\ b_{1i} \cdot \dfrac{s_{1i}+2}{2}; & A_{1i} = 1 \end{cases}$$

If there is a limited waiting room in front of a channel ($s_{1i} < \infty$ waiting places) some messages are rejected. This overflow traffic (which is offered to a secondary system, cf. section 3) is characterized by the first two moments, the mean value R_{1i} and the variance coefficient $D_{1i} = \sigma_{1i}^2 - R_{1i}$. Arbitrary moments of overflow traffic have been studied formerly /16/ by means of a two dimensional markovian process. The most interesting results for us, mean value and variance coefficient are given by (cf. fig. 5.):

$$R_1 = \frac{A_1^{s_1+2}}{1 + A_1 \cdot \dfrac{1 - A_1^{s_1+1}}{1 - A_1}}$$

$$D_1 = R_1^2 \cdot \left\{ \frac{c_1 \cdot A_1}{c_2 \cdot R_1} - 1 \right\}$$

where

$$c_1 = \frac{1}{2k_2} \cdot \left\{ [1 + A_1 - (k_1 - k_2)] \cdot (k_1 + k_2)^{s_1+1} - [1 + A_1 - (k_1 + k_2)] \cdot (k_1 - k_2)^{s_1+1} \right\}$$

$$c_2 = 1 + \frac{1}{2k_2} \left\{ [1 + A_1 - (k_1 - k_2)] \sum_{\gamma=0}^{s_1} (k_1 + k_2)^{\gamma+1} - [1 + A_1 - (k_1 + k_2)] \sum_{\gamma=0}^{s_1} (k_1 - k_2)^{\gamma+1} \right\}$$

$$k_1 = 1 + \frac{A_1}{2}$$

$$k_2 = \sqrt{1 - \left(\frac{A_1}{2}\right)^2}$$

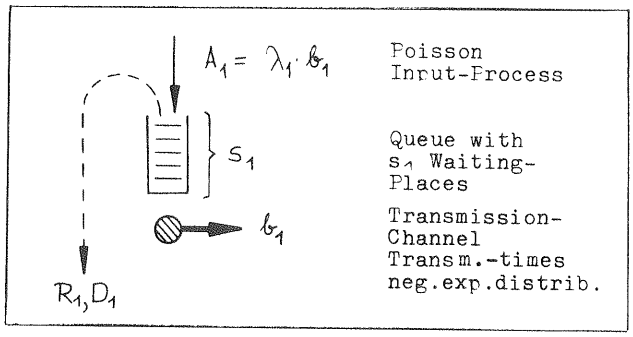


Fig.5.: Primary system with offered Poisson traffic.

4.3. The secondary systems.

As shown before, overflow traffic has quite other statistical properties than Poissonian traffic. However, when assuming a two dimensional Markovian process, it is possible to determine all characteristic traffic values, including the waiting time distribution.
The exact mathematical treatment is outlined for "substitute arrangements" to which all actual systems can be traced back: one primary and one secondary system (cf. section 3.2.3. and fig. 3.).

4.3.1. Probabilities of state.

At first we define by $\{u_1, u_2\}$ the state that there are u_1 messages in the primary system and u_2 in the secondary system simultaneously (waiting or being served). Then, the two-dimensional Chapman-Kolmogoroff equations for the probabilities $p(u_1, u_2)$ of $\{u_1, u_2\}$ can be found. When introducing the Z-transform and systematic

replacement of unknown probability-values by the "basic" states $p(o,u_2)$ it is possible to express all probabilities of state $p(u_1,u_2)$ by the marginal values $p(o,u_2)$, $p(o,u_2+1)$, ... , $p(o,s_2+1)$ (s_2: number of waiting places in the secondary system). In other words: the two dimensional state relations can be reduced to a one dimensional system ($u_1=o$). The solution of such a one dimensional system can be easily found. It is given by (cf. also /16,19/):

$$\boxed{p(u_1,u_2) = \frac{A_1^{u_1}}{1 + A_1 \frac{1 - A_1^{s_1+1}}{1 - A_1}} \cdot \frac{\sum_{\varrho=u_2}^{1+s_2} c_{u_1,u_2,\varrho} \cdot b_\varrho}{\sum_{\nu=0}^{1+s_2} \sum_{\varrho=\nu}^{1+s_2} c_{u_1,\nu,\varrho} \cdot b_\varrho}}$$

where

$$c_{u_1,u_2,\varrho} = -\sum_{\xi=0}^{u_1-1} c_{\xi,u_2+1,\varrho} \cdot \beta_{u_1-\xi} \quad ; u_2 \neq \varrho ; u_2 \neq 0, \varrho > u_2$$

$$c_{u_1,0,\varrho} = -\sum_{\xi=0}^{u_1-1} c_{\xi,1,\varrho} \cdot \beta'_{u_1-\xi} \quad ; u_2 \neq \varrho ; u_2 = 0$$

$$c_{u_1,u_2,\varrho} = (f + A_1)\beta_{u_1} - A_1 \cdot \beta_{u_1-1} \quad ; u_2 = \varrho ; u_2 \neq 0$$

$$c_{u_1,u_2,\varrho} = A_1 \cdot \beta'_{u_1} - A_1 \cdot \beta'_{u_1-1} \quad ; u_2 = \varrho ; u_2 = 0$$

$$c_{u_1,u_2,\varrho} = 0 \quad ; u_2 > \varrho$$

$$b_{u_2} = \sum_{z_1=u_2+1}^{1+s_2} a_{u_2,z_1} \sum_{z_2=z_1+1}^{1+s_2} a_{z_1,z_2} \sum_{z_3=z_2+1}^{1+s_2} a_{z_2,z_3} \cdots \sum_{z_W=z_{W-1}+1}^{1+s_2} a_{z_{W-1},z_W} \quad ; W = 1 + s_2 - u_2$$

$$a_{u_2,\varrho} = \frac{-A_1 \cdot c_{1+s_1,u_2,\varrho} + (1 + f + v \cdot A_1) c_{1+s_1,u_2+1,\varrho} - v \cdot c_{1+s_1,u_2+1,\varrho} - A_1 \cdot c_{s_1,u_2+1,\varrho}}{A_1 \cdot c_{1+s_1,u_2,u_2}}$$

$$v = \begin{cases} 1 & \text{when } 0 \leq u_2 < s_2 \\ 0 & \text{otherwise} \end{cases}$$

$$\beta_{u_1} = \frac{\alpha_1^{u_1} - \alpha_2^{u_1}}{\alpha_1 - \alpha_2} \; ; \; \beta'_{u_1} = \frac{(\alpha'_1)^{u_1} - (\alpha'_2)^{u_1}}{\alpha'_1 - \alpha'_2} \; ; \; f = \frac{\varepsilon_2}{\varepsilon_1} \; ; \; A_1 = \frac{\lambda_1}{\varepsilon_1}$$

$$\alpha_1, \alpha_2 = \frac{(1 + f + A_1) \pm \sqrt{(1 + f + A_1)^2 - 4A_1}}{2} \; ; \; \alpha'_1, \alpha'_2 = \frac{(f + A_1) \pm \sqrt{(f + A_1)^2 - 4A_1}}{2}$$

4.3.2. Characteristic mean values for secondary-systems.

All state probabilities $p(u_1,u_2)$ have been determined in the previous section. Therefore, it is easy to find the characteristic mean values for the secondary system, as there are:

Carried traffic

$$Y_2 = \sum_{u_1=0}^{1+s_1} \sum_{u_2=1}^{1+s_2} p(u_1,u_2) = \frac{\lambda_1}{\varepsilon_2} \cdot \left[p(u_1 = s_1+1) - p(s_1+1, s_2+1)\right]$$

Lost traffic

$$R_2 = \frac{\lambda_1}{\varepsilon_2} \cdot p(1 + s_1, 1 + s_2)$$

Mean number of messages waiting for service

$$\Omega_2 = \sum_{u_1=0}^{1+s_1} \sum_{u_2=1}^{1+s_2} (u_2 - 1) \cdot p(u_1,u_2)$$

Mean waiting time for all messages, not rejected by the system

$$w_{Y2} = \frac{\Omega_2}{\lambda_1 \left[p(u_1 = s_1+1) - p(s_1+1, s_2+1)\right]} = \frac{\Omega_2}{Y_2 \cdot \varepsilon_2}$$

Mean waiting time for all messages, which have to wait in the system

$$t_{W_2} = \frac{\Omega_2}{\lambda_1 \cdot \sum_{u_2=1}^{s_2} p(1 + s_1, u_2)}$$

4.3.3. Probability of waiting for secondary-systems.

The probability that an incoming message has to wait under the condition, that it is offered to the secondary system is given by

$$W_2 = \frac{\sum_{u_2=1}^{s_2} p(1 + s_1, u_2)}{\sum_{u_2=0}^{s_2+1} p(1 + s_1, u_2)}$$

4.3.4. The waiting time distribution function (w.d.f.) for secondary systems.

The w.d.f. $F_2(\leq t)$ is the probability that an arbitrary message has to wait at most the time t under the condition that the considered message was offered to the secondary system. When determining the w.d.f., a <u>test-call</u> is considered. This test-call is offered to the secondary system, if there are u_2 messages in the secondary system ($u_2=o,1, .. s_2+1$).

The probability that the considered test-call has to wait at most the time interval t is defined as $F_{u_2}(\leq t)$.
Therefore, the total w.d.f. is given by

$$F_2(\leq t) = \frac{\sum_{u_2=0}^{s_2+1} \lambda_1 \cdot dt \cdot p(s_1+1, u_2) \, F_{u_2}(\leq t)}{\sum_{u_2=0}^{s_2+1} \lambda_1 \cdot dt \cdot p(s_1+1, u_2)}$$

$$F_2(\leq t) = \frac{1}{p(u_1 = s_1+1)} \cdot \sum_{u_2=0}^{s_2+1} p(s_1+1, u_2) \cdot F_{u_2}(\leq t)$$

The complementary w.d.f. (complement of the waiting time distribution function) is $F_{u_2}(>t)$, i.e. the probability, that the considered test-call has to wait longer than the time t. When calculating this probability, the following three cases have to be distinguished:

<u>$u_2 = o$:</u>

the test-call will be transmitted immediately, i.e.

$$F_o(>t) = 0$$

<u>$o > u_2 \geq s_2$:</u>

the test-call has to wait. This waiting time is longer than the time t, if in this time interval no, one, etc. or at most (u_2-1) messages, waiting before the test-call can be transmitted. In the latter case, the test-call is "sitting" at the first waiting place, i.e. transmission can start not before the instant (t+dt):

$$F_{u_2}(>t) = \sum_{\nu=0}^{u_2-1} \frac{(\varepsilon_2 \cdot t)^\nu}{\nu!} \cdot e^{-\varepsilon_2 t}$$

(The probability, that exactly ν messages can be transmitted in a fixed time interval is given by the well known Poisson-distribution).

<u>$u_2 = s_2+1$:</u>

the test-call is offered to the secondary system. However, because there is no free waiting

place, it is rejected. In other words: the test-call is not allowed to wait

$$F_{s_2+1}(>t) = 0$$

The waiting-time distribution function $F_{u_2}(\leq t)$ is given in each case by the complement of the probability $F_{u_2}(>t)$.
Substituting this quantities into the fundamental equation the w.d.f. for an arbitrarily chosen message is given by

$$F_2(\leq t) = 1 - \frac{e^{-\varepsilon_2 t}}{p(u_1 = s_1+1)} \sum_{u_2=1}^{s_2} p(s_1+1, u_2) \sum_{\nu=0}^{u_2-1} \frac{(\varepsilon_2 t)^\nu}{\nu!}$$

5. NUMERICAL RESULTS, COMPARISON WITH SIMULATION RESULTS

As shown before, all characteristic traffic values for the input-system and also for primary systems can be determined exactly.
All characteristic traffic values referring to secondary systems are approximations because the actually offered overflow traffics are substituded by a ficticious traffic generating exactly only the most important moments of the actual traffics.
Therefore, a large number of simulation runs have been performed by means of two simulation programs. Comparison of both simulated and calculated values shows the accuracy of the new method (cf. fig. 6 to 11).

6. CONCLUDING REMARKS AND FUTURE PLANS

By reasons of economy and reliability networks with flexible routing strategies are superior to networks without this capability. Then however, overflow traffics occur with special stochastic properties.
In order to avoid wrong estimates of traffic capacity and flow times this special stochastic behaviour has to be taken into consideration when planning such networks.
Today, the exact mathematical treatment of complex networks is not possible even on the largest digital computers. Therefore a handy approximate method has been proposed which allows to determine important traffic characteristics for individual data switching centers as well as for complex store- and forward networks with alternate routing. Comparisons show the good accordance between calculated values and simulation results.
The above presented investigations dealt with

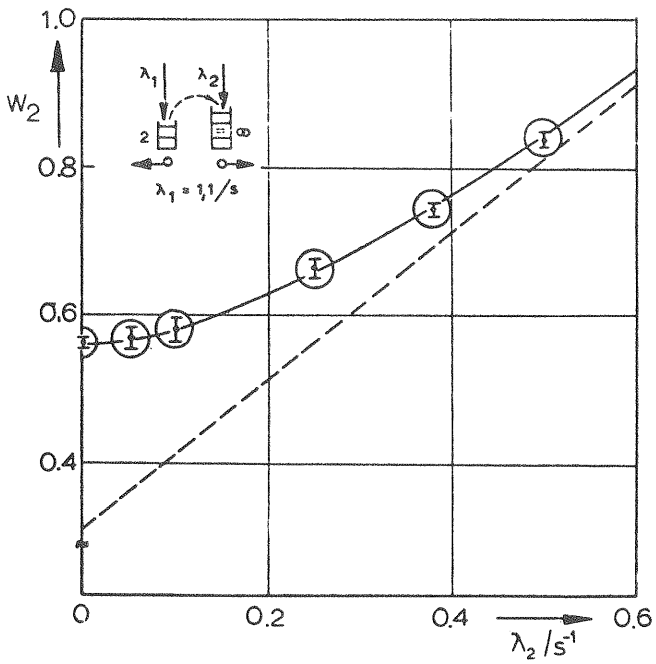


Fig.6.: Probability of waiting W_2 for a secondary system, to which overflow traffic and direct traffic (with call rate λ_2) is offered. The mean transmission time per message is 1 second for both primary and secondary channel. Comparison between calculation (——) and simulation (I; 95 % - confidence intervals). Neclecting the special character of the overflow traffic, the dashed line is obtained.

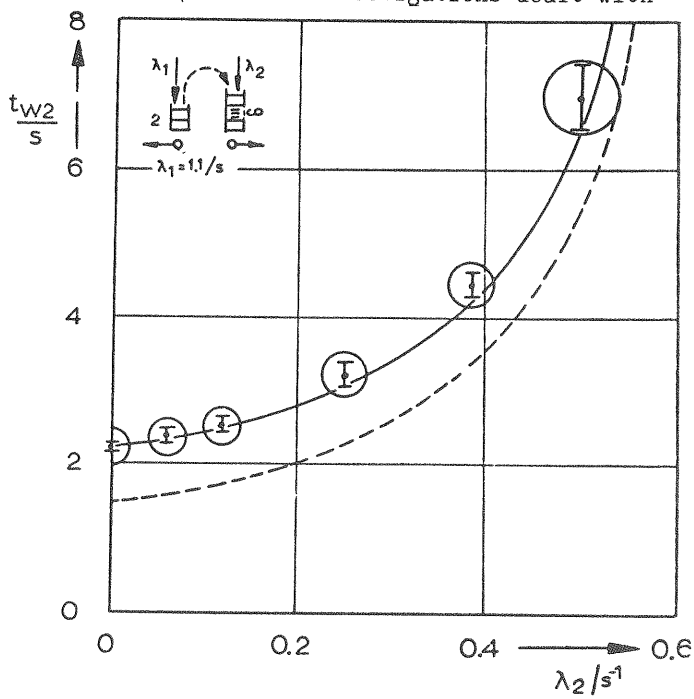


Fig.7.: Mean waiting time t_{W2} for messages which have to wait in the secondary system. Comparison between calculation and simulation (cf. fig.6.).

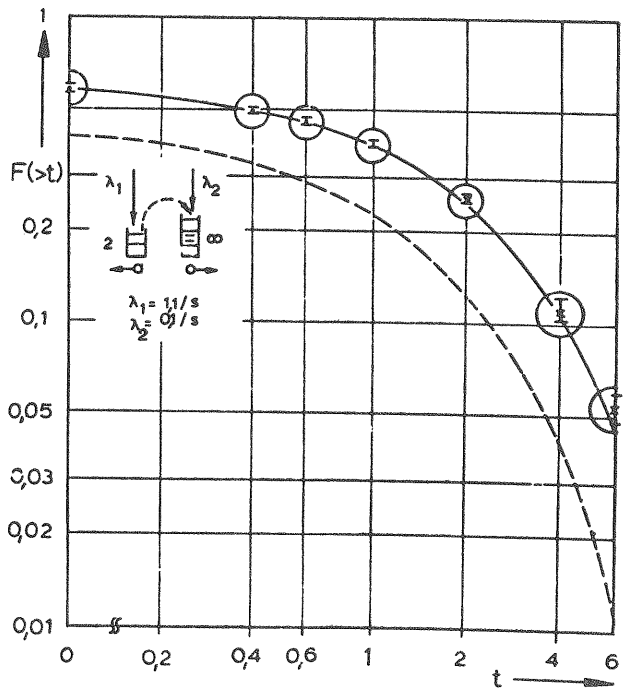


Fig.8.: Complementary w.d.f. for all messages offered to the secondary system. Comparison between calculation and simulation (cf.fig.6.).

networks with at most a single overflow for each direction. It could be advantageous for some networks to have more than one overflow possibility or to allow mutual overflow. Investigations are planned to include such possibilities in the new method.

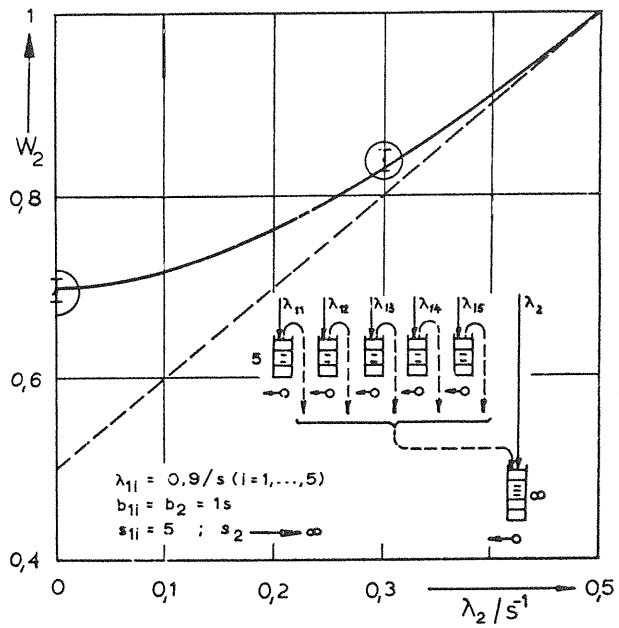


Fig.9.: Probability of waiting W_2 for a secondary system, to which five overflow traffics and direct traffic are offered. Comparison between calculation (——) and simulation (I; 95 %-confidence intervals). Neclecting the special character of overflow traffic, the dashed line is obtained.

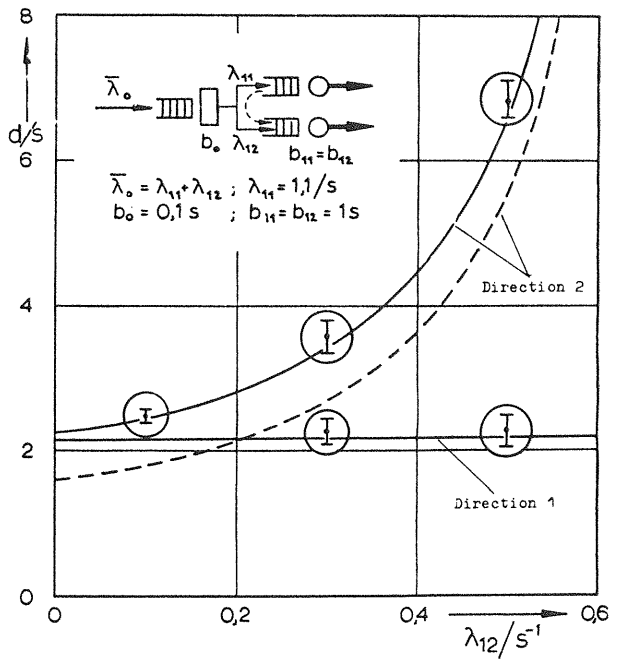
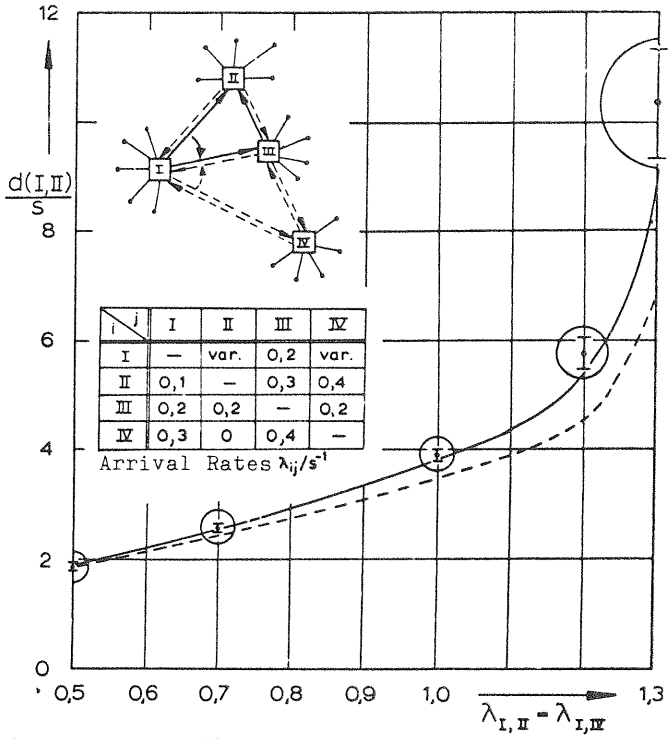


Fig.10.: Flow time through a data switching center for messages transmitted via direction 1 or 2 respectively. Comparison between calculation (——) and simulation (I; 95%-confidence intervals). Neglecting the special character of overflow traffic the dashed curve is obtained.



i \ j	I	II	III	IV
I	—	var.	0,2	var.
II	0,1	—	0,3	0,4
III	0,2	0,2	—	0,2
IV	0,3	0	0,4	—

Fig.11.: Total flow time for all messages from "originating" node I to the destination II. Example of a network with four nodes. Comparison between calculation (——) and simulation (I; 95%-confidence intervals). Neglecting the special character of overflow traffic the dashed curve is obtained.

In the last few years store- and forward networks with packet-switching became more important. For such networks, with constant transmission time for each packet an analogous solution seems to be possible.

The above investigated routing strategy is only one flexible routing strategy which can be realized most easily and without great expenditure. Numerous other strategies have been proposed and can be investigated only by means of simulation programs. If one could find for such strategies also good approximate calculation methods, it would be possible to compare different routing strategies analytically to a great extend.

ACKNOWLEDGEMENT

The author wishes to express his thanks to Professor Dr.- Ing. A. Lotze - Director of the Institute for Switching and Data Technics, University of Stuttgart - which has initiated the above presented investigations.
He is also grateful to W. Rosenbohm, G. Globas, S. Rheinwald, G. Schmitz, B. Werres, R. Ott and U. Koch contributing to this investigations while preparing their diploma thesises (cf. also lit./19/). Furthermore, the author is thankful to Dr.-Ing. M. Langenbach-Belz and Dipl.-Ing. H. Weisschuh for valuable discussions.

BIBLIOGRAPHY

/1/ Bretschneider,G.: Die Berechnung von Leitungsgruppen für überfließenden Verkehr in Fernsprechwählanlagen. Nachrichtentechn. Z. 9 (1956), pp. 533-540.

/2/ Lotze,A.: A Traffic Variance Method for Gradings of Arbitrary Type. Proc.4. ITC, London, Juli 1964, Dokument 80 und POEEJ, Special Issue 1966.

/3/ Wilkinson,R.I.: Theories for Toll Traffic Engineering in the U.S.A. BSTJ 35 (1956) S. 421-514.

/4/ Boehm,B.W.; Mobley,R.L.: Adartive Routing Techniques for Distributed Communications Systems. IEEE Trans.Commun.Technol. 17 (1969) S. 340-349.

/5/ Fultz,G.L.; Kleinrock,L.: Adartive Routing Techniques for Store- and Forward Computer Communication Networks. Advanced Research Projects Agency Semiannual Technical Report, June 1971, University of California, S. 83-91.

/6/ Silk,D.J.: Routing Doctrines and their Implementation in Message-Switching Networks. Proc. IEE, 116(1969) S.1631-1638.

/7/ Prosser,R.T.: Routing Procedures in Communications Networks-Part I: Random Procedures-Part II: Directory Procedures. IRE Trans.Commun.Systems (1962) S.322-335.

/8/ Butrimenko,A.: Routing Technique for Message Switching Networks with Message Outdating. Symp.Comp.-Commun.Networks and Teletraffic. Polytechn.Inst.Brooklyn, New York, N.Y.,April 1972,Session VI.

/9/ Brandt,G.J.; Chrétien,G.J.: Methods to Control and Operate a Message Switching Network.Symp.Comp.-Commun.Networks and Teletraffic.Polytechn.Inst.Brooklyn,New York,N.Y., April 1972, Session VI and S.I.T.A.Monograph No.1(1972).

/10/ Beeforth,T.H.; Grimsdale,R.L.; Halsall,F.; Woollons,D.J.: Aspects of a Proposed Data Communication System.Proc. IFIP,Ljubljana, Jugoslawien,August 1971, TA-4-74 bis 78.

/11/ Davies,D.W.: The Control of Congestion in Packet-Switching Networks. IEEE Trans. Commun.20(1971) S.546-550.

/12/ Frank,H.; Chou,W.: Routing in Computer Networks. Networks 1(1971) S.99-112.

/13/ Kahn,R.E.; Crowther,W.R.: Flow Control in a Resource-Sharing Computer Network.IEEE Trans.Commun.20(1972) S.539-546.

/14/ Lotze,A.: Problems of Traffic Theory in the Design of International Direct Distance Dialling Networks. NTZ.-Commun.J. (1968) S. 41-46.

/15/ Petersen,R.D.: Fu,Y.: Optimum Routing Strategies for Noisy Communication Networks. IEEE Trans.Commun.Techn.17 (1969) S. 104-118.

/16/ Herzog,U.; Kühn,P.: Comparison of some Multi-queue Models with Overflow and Load-Sharing Strategies for Data Transmission and Computer Systems. Symp.Com.-Commun.Networks and Teletraffic.Polytechn. Inst.Brooklyn, New York, N.Y. April 1972, Session XII.

/17/ Schehrer,R.: Über die exakte Berechnung von Überlaufsystemen der Wählvermittlungstechnik.Dissertation, Universität Stuttgart, Stuttgart 1969.

/18/ Burke,P.J.: The Output of a Queuing System.Operations Res. 4(1956) S.699-704.

/19/ Herzog,U.: Verkehrsfluß in Datennetzen. Habilitation-thesis, University of Stuttgart, submitted january 1973.

/20/ Kühn,P.: Über die Berechnung der Wartezeiten in Vermittlungs- und Rechnersystemen.Dissertation, Universität Stuttgart, Stuttgart, 1972.

APPENDIX 2

EFFICIENT PRIORITY STRATEGIES FOR SWITCHING CENTERS IN COMMUNICATION NETWORKS

Ulrich Herzog

Institute for Switching and Data Technics, University of Stuttgart

Federal Republic of Germany, currently at

IBM Thomas J. Watson Research Center, Yorktown Heights, N. Y.

Abstract : Various priority strategies for real-time computer systems are discussed, uniformly described and mathematically treated, as well. Numerical examples illustrate the importance of so called preemption-distance priorities.

1. PROBLEM

The most important demand to switching centers in communication networks is their fast reaction to urgent signals. Typical examples are the processing of

- alarms indicating the breakdown of transmission channels, of other switching centers or of system components in the considered switching center itself,
- routing informations concerning the instantaneous traffic load in the network,
- urgent messages (or telephone calls) between different terminals (or telephone subscribers) in the network.

Such demands might be fulfilled by means of pure preemptive priorities for all priority classes. However, each interruption needs some additional amount of overhead:

- the interrupt must be analyzed,
- the contents of registers have to be saved,
- the queues have to be reorganized,

etc. In order to keep this additional system load as small as possible modern real-time systems use reasonable combinations of preemptive and non-preemptive (head-of-the-line) priorities.

Two examples may illustrate the manifold of strategies to be considered and to be analyzed:

EWS 1 (cf. [1])

In the new electronic telephone switching system of the Federal German Post Office the following strategy is used:

All signals are classified into some few priority groups which interrupt each other (preemptive priorities). Within one group several non-preemptive (head-of the-line) classes are to be distinguished (e.g. alarms of different importance, various switching demands, etc.). Figure 1 shows a small

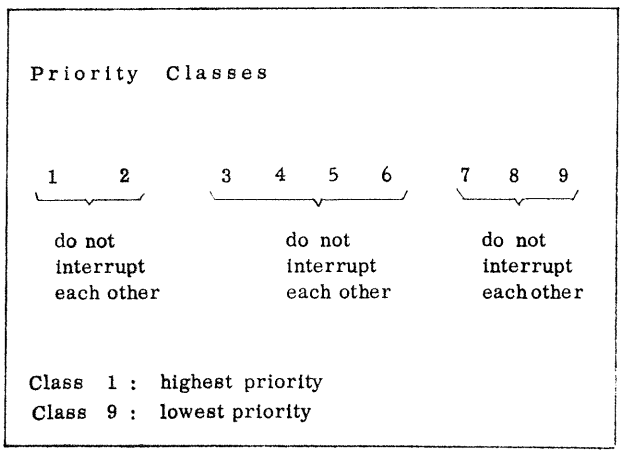


Fig. 1: First example for a combination of both, preemptive and non-preemptive (head-of-the-line) priorities (EWS 1, cf. text).

New arriving demand of class	does not interrupt service of class	interrupts service of class	Preemption Distance
1	1,2	3,4,5,6,7,8,9	2
2	1,2	3,4,5,6,7,8,9	1
3	1,2,3,4,5,6	7,8,9	4
4	1,2,3,4,5,6	7,8,9	3
5	1,2,3,4,5,6	7,8,9	2
6	1,2,3,4,5,6	7,8,9	1
7	1,2,3,4,5,6,7,8,9	-	-
8	1,2,3,4,5,6,7,8,9	-	-
9	1,2,3,4,5,6,7,8,9	-	-

Fig. 2: Service mechanism and Preemption-Distance corresponding to the priority strategy of figure 1.

but typical example for these types of priority strategies. Dependent on the importance of a demand, the distance to the next class of priorities to be interrupted - the so called Preemption-Distance - varies (cf. also fig. 2).

EDS (cf. [2])

EDS is the new electronic data switching system of the German Post Office and Western Union, as well. For the I/O-Control of the storage unit the following remarkable strategy is used: The priorities can be controlled such

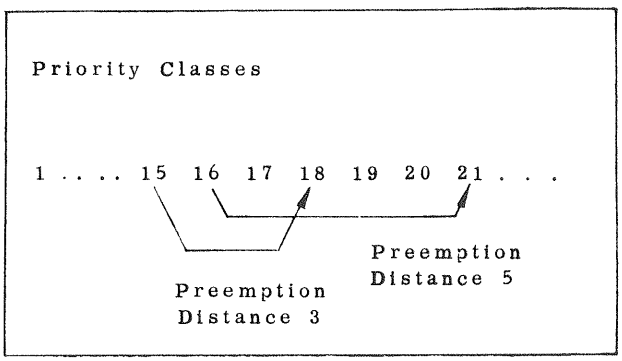


Fig. 3: Second example (EDS) for a combination of preemptive and non-preemptive priorities (cf. text).

that, e.g., priority class 15 interrupts demands of class 18, 19, etc., however not the intermediate classes 16 and 17 whereas class 16 interrupts only class 21, 22, etc. The strategy is illustrated in fig. 3 and allows generally interpreted a sliding passage from preemptive to non-preemptive priorities.

These are just two important examples, numerous other reasonable combinations of preemptive and non-preemptive priorities are possible and implemented.

In the following section it is shown how this variety of combinations can be uniformly described by means of the Preemption-Distance ξ. Therefore, it seems to be reasonable to introduce the unifying term "Preemption -Distance Priorities".

Results presented in [3] for Preemption-Distance priorities are generalized by introducing a General Erlangian distribution. Moreover, additional characteristic performance values are presented. Numerical examples demonstrate the advantages of Preemption-Distance priorities.

2. DESCRIPTION OF PREEMPTION-DISTANCE PRIORITIES.

2.1. Uniform Preemption-Distance for all Classes, Preemptive priorities, Non-Preemptive Priorities.

Already in the first section the preemption-distance was introduced: The distance between an arbitrary but distinct priority class and the next priority class being interrupted.

Fig. 4 illustrates this definition: Demands of class p ($p \in 1, 2, \ldots P$; class 1 most urgent) interrupt only demands of class ($p+\xi$) to P, however not the intermediate classes (p+1) to ($p+\xi-1$). On the other hand demands of

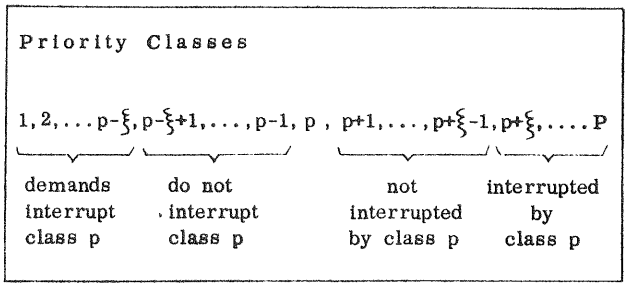


Fig. 4: Introduction of the Preemption-Distance . The special cases, preemptive priorities (ξ=1) and non-preemptive priorities (ξ= P), are included.

the considered class p can be interrupted by class 1 to ($p-\xi$), however not by classes ($p-\xi+1$) to (p-1).

It is seen easily that two wellknown special cases of Preemption-Distance priorities are included:

$\xi = 1$: Preemptive priorities

$\xi = P$: Non-preemptive (head of the line) priorities.

2.2. Arbitrary, Non-Uniform Preemption-Distance for each Priority Class.

Introducing for each class p ($p \in 1, 2, \ldots, P$) the preemption-distance $\xi(p)$ with a definition analogeous to section 2.1. arbitrary combinations of pure preemptive and non-preemptive priorities are possible. Although the uniform representation and analysis is possible, this way of solution is rather complex.

A much more elegant solution is possible using a uniform preemption-distance and introducing "empty" priority classes: ficticious priority classes (with the arrival rate null) are interleaved between the actual ones.

This trick allows us to generate all strategies of practical interest easily (the only two special cases known from literature [4, 5] are included). Furthermore, it facili-

tates the investigation of their influence on the waiting process.

As an example fig. 5 shows how the priority strategies of the EWS-type (cf. fig. 1 and section 1) can be obtained.

Uniformly prescribed Preemption-Distance	actual and "empty" priority classes	mean arrival rate	actual priority classes	actual Preemption-Distance
4	1	λ_1	1	2
4	2	λ_2	2	1
4	3	-	-	-
4	4	-	-	-
4	5	λ_5	5	4
4	6	λ_6	6	3
4	7	λ_7	7	2
4	8	λ_8	8	1
4	9	-	-	-
4	10	-	-	-
4	11	-	-	-
4	12	-	-	-
4	13	λ_{13}	13	-
4	14	λ_{14}	14	-
4	15	λ_{15}	15	-

Fig. 5: Example for the generation of a non-uniform Preemption-Distance by means of "empty" classes (cf. fig. 1, 2 and text).

3. ANALYSIS

3.1 Structur and Operating Mode of the Investigated System.

Fig. 6 shows schematically the system to be investigated: Arriving demands are classified into P parallel queues according to their priority. All queues are assumed to be

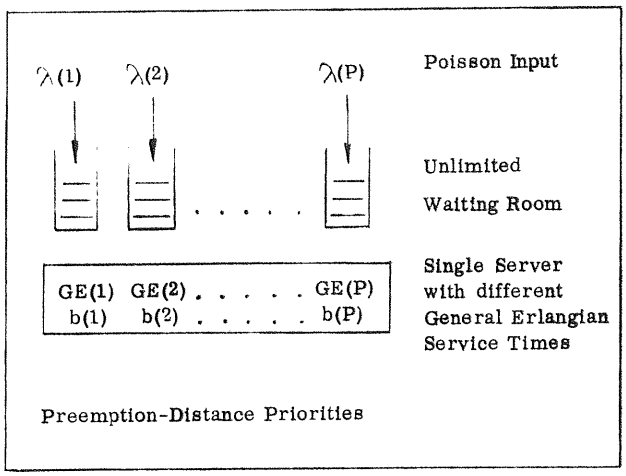


Fig. 6: The investigated system (details cf. text).

unlimited, i.e. every arriving demand will be stored and processed. This assumption is almost always fulfilled, especially in systems with dynamic core allocation.

All demands are served according to an arbitrary Preemption-Distance strategy, treated in the previous sections; FIFO is assumed within each priority class.

3.2. Traffic Parameters.

Demands of each priority class $p(p \in 1, 2, \ldots, P)$ are distributed according to a Poisson process with the mean arrival rate $\lambda(p)$:

$$A_p(\leq t) = 1 - e^{-\frac{t}{a(p)}} = 1 - e^{-\lambda(p)\cdot t}$$

Service times follow -individually for each priority class- a General Erlangian (GE) distribution:

$$B_p(\leq t) = \sum_{\nu=1}^{l(p)} q_\nu(p)\left(1 - e^{-\frac{t}{b_\nu(p)/k_\nu(p)}} \cdot \sum_{\eta=0}^{k_\nu(p)-1} \frac{\left(\frac{t}{b_\nu(p)/k_\nu(p)}\right)^\eta}{\eta!}\right)$$

with the mean value

$$b(p) = \sum_{\nu=1}^{l(p)} q_\nu(p) \cdot b_\nu(p)$$

and the variance

$$\sigma^2(p) = \sum_{\nu=1}^{l(p)} \frac{k_\nu(p)+1}{k_\nu(p)} \cdot b_\nu(p)^2 \cdot q_\nu(p) - b(p)^2$$

where

l : number of (ficticious) parallel "chains" of exponential "stages"

q_ν: probability that chain no ν $(\nu \in 1, 2, \ldots, l)$ is passed

k_ν: number of stages for chain no ν

b_ν: mean service time for one stage of chain no ν

It is worthwhile to notice that this distribution allows to approximate any type of distribution function of service times with any required accuracy. Obviously, it includes the hyperexponential $(k_\nu(p)=1)$ as well as the Erlangian distribution $(l(p)=1)$ both most important for many applications.

Neglected is the time to handle interrupts. This assumption is also allowed because of large real-time computers have duplexed register sets and hardware for interrupt handling (for small systems without these facilities, cf. section 5).

3.3. Analysis.

3.3.1. General remarks.

The most famous methods to investigate the stochastic behavior of such non-Markovian queuing systems are

- the method of imbedded markov chains [6]
- the phase method [7]
- the integral method [8]
- the method of substitut variables [9]

When investigating arbitrary kinds of Preemption-Distance priorities, all methods failed because of the complex interdependencies between different priority classes. However, a general solution was possible by means of the method of moments: The fate of an individual demand of priority class p is persued from its arrival up to the point where it leaves the system. All possibilities of interruptions, processing, pushing back in the queue, etc. are considered. Finally, when introducing expectation values the presented solution can be obtained.

3.3.2. Characteristic Performance Values.

The expected response time (time spend in the system, waiting and being processed) d(p) for a demand of priority class p ($p \in 1, 2, \ldots, P$) is composed of the following five terms:

1) The expectation $b_R(p+\xi-1)$ of the remaining rest-service time for demands of the priority classes 1 to $(p+\xi-1)$ present at its time of arrival in the server and not being interrupted by the considered p-demand.

2) The expected time $w_I(p)$ necessary to serve demands of the priority classes 1 to p waiting in the system at its time of arrival.

3) Its expected time in service b(p).

4) The expected time $w_{II}(p)$ necessary to serve demands of preemptive priority classes 1 to $(p-\xi)$ which enter the system, while the considered p-demand is still in the system.

5) The expected time $w_{III}(p)$ necessary to serve demands of the non-preemptive priority classes $(p-\xi+1)$ to $(p-1)$ which enter the system, while the considered p-demand is still in the system, however before its last interruption.

A detailed study of these five terms presented in [10] leads to the following recursive solution for the expected response time for priority class p ($p \in 1, 2, \ldots, P$):

(5)
$$d(p) = \left\{ \sum_{i=1}^{p-1} A(i)\cdot d(i) + b(p) + b_R(\leq p+\xi-1) - \sum_{i=1}^{p} A(i)\cdot b(i) - \sum_{i=1}^{p} \Omega_U(i)\left\{b(i)-b_g(i)\right\} - b_n(p)\sum_{i=p-\xi+1}^{p-1} A(i) \right\} \cdot \left\{ 1 - \sum_{i=1}^{p} A(i) \right\}^{-1}$$

Where
$$b_R(\leq p+\xi-1) = \sum_{v=1}^{p} \frac{A(v)}{2\,b(v)} \sum_{r=1}^{l(v)} \frac{k_r(v)+1}{k_r(v)} \cdot b_r(v)^2 \cdot q_r(v) + \sum_{v=p+1}^{p+\xi-1} \frac{A(v)}{\lambda_U(v)^2 \cdot b(v)} \cdot \left\{ \lambda_U(v)\cdot b(v) - 1 + \sum_{r=1}^{l(v)} \frac{q_r(v)}{\left(\lambda_U(v)\cdot b_r(v)/k_r(v)+1\right)^{k_r(v)}} \right\}$$

is the expected time a demand of class p has to wait until demands of lower priority which can not be interrupted leave the server.

$$\Omega_U(i) = \lambda(i)\cdot \frac{b(i)\sum_{j=1}^{i-1} A(j) - b_n(i)\sum_{j=i-\xi+1}^{i-1} A(j)}{1 - \sum_{j=1}^{i-1} A(j)}$$

is the mean number of demands of class i waiting however being interrupted at least once.

$$b_n(p) = \frac{1}{\lambda_U(p)} \left\{ 1 - \sum_{r=1}^{l(p)} \frac{q_r(p)}{\left(\lambda_U(p)\cdot b_r(p)/k_r(p)+1\right)^{k_r(p)}} \right\}$$

is the remaining rest-service time of a demand of class p after its last interruption.

$$b_g(v) = \frac{1}{2\cdot b(v)} \sum_{r=1}^{l(v)} \frac{k_r(v)+1}{k_r(v)} \cdot b_r(v)^2 \cdot q_r(v)$$

is the remaining service time for a demand of class i.

Additionally,

$$A(p) = \lambda(p)\cdot b(p)$$

$$\lambda_U(p) = \sum_{i=1}^{p-\xi} \lambda(i)$$

Remark: It should be mentioned that an explicit solution has also been found. However, from the computational viewpoint the presented recursive solution is more practical.

Besides the expected response time and waiting time w(p) = d(p) - b(p) the following characteristic performance measures have been derived:

Probability that a demand of class p is interrupted at least once:

(6)
$$P_U(p) = 1 - \sum_{r=1}^{l(p)} \frac{q_r(p)}{\left(\lambda_U(p)\cdot b_r(p)/k_r(p)+1\right)^{k_r(p)}}$$

Mean number of interrupts per demand of priority class p:

(7)
$$u(p) = \lambda_U(p)\cdot b(p)$$

Probability of waiting for demands of priority class p:

(8)
$$W(p) = \sum_{i=1}^{p+\xi-1} A(i) + \left\{ 1 - \sum_{i=1}^{p+\xi-1} A(i) \right\} \cdot P_U(p)$$

4. NUMERICAL RESULTS

Three examples may show the various kinds of operating strategies and service times to be prescribed and analyzed now uniformly.

The examples also show how advantageous Preemption-

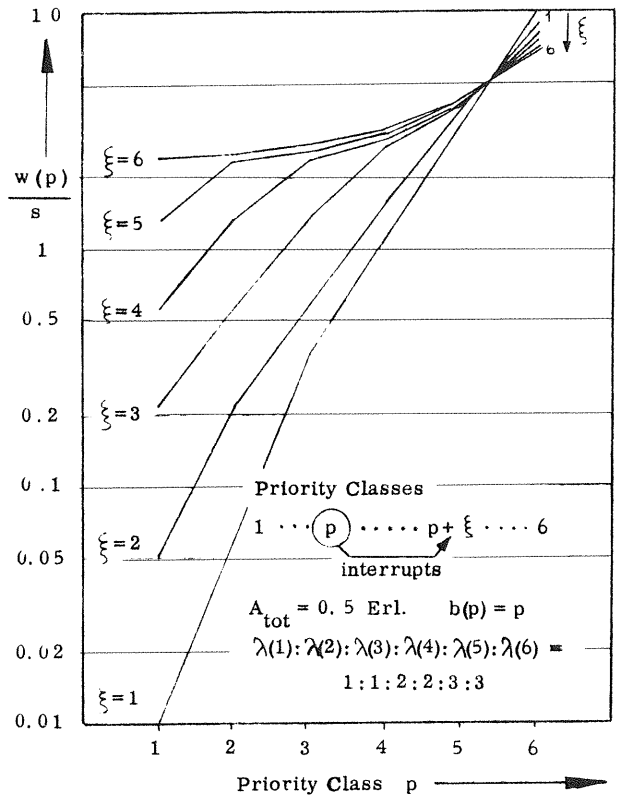


Fig. 7: Mean waiting time w(p). For this example all service-times are assumed to be exponentially distributed, however with different mean values per class.

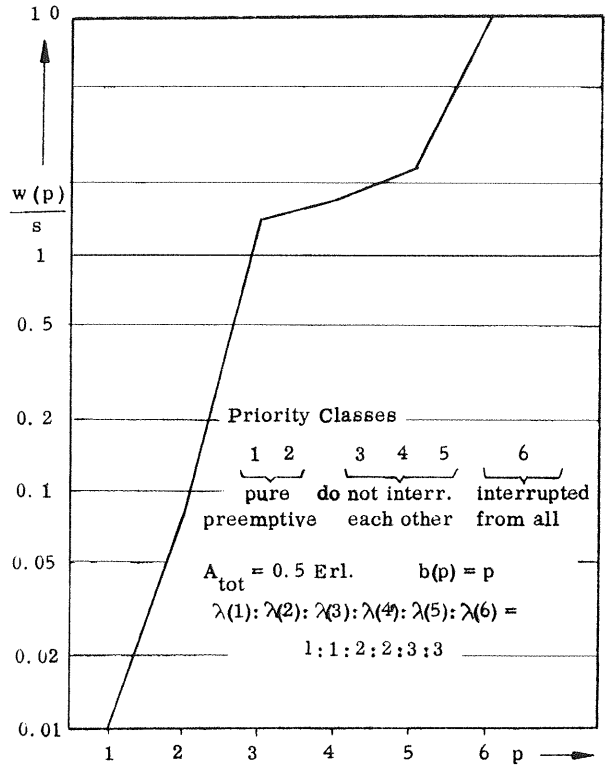


Fig. 9: Typical example for an efficient combination of pre-emptive and non-preemptive priorities (compare it with the priority strategies of figure 7 !).

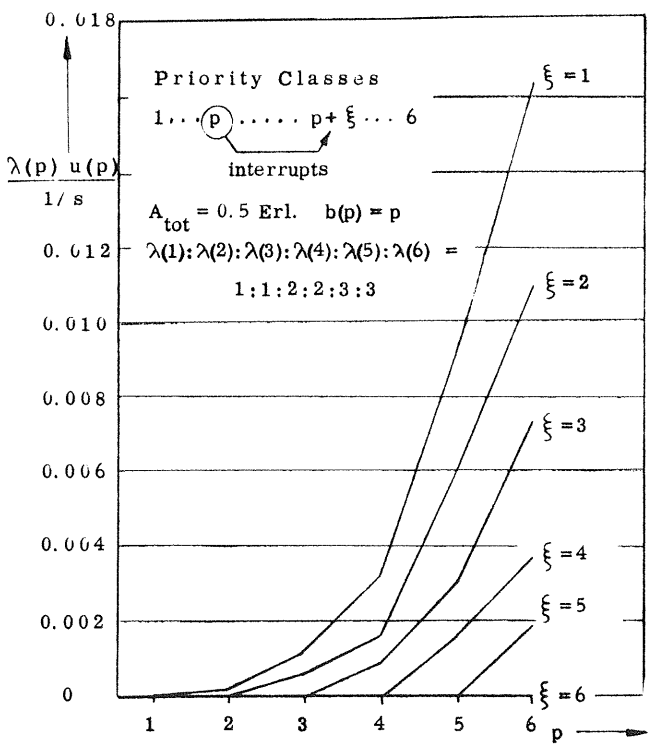


Fig. 8: Mean number of interrupts per second for uniform Pre-emption-Distance (cf. fig. 7).

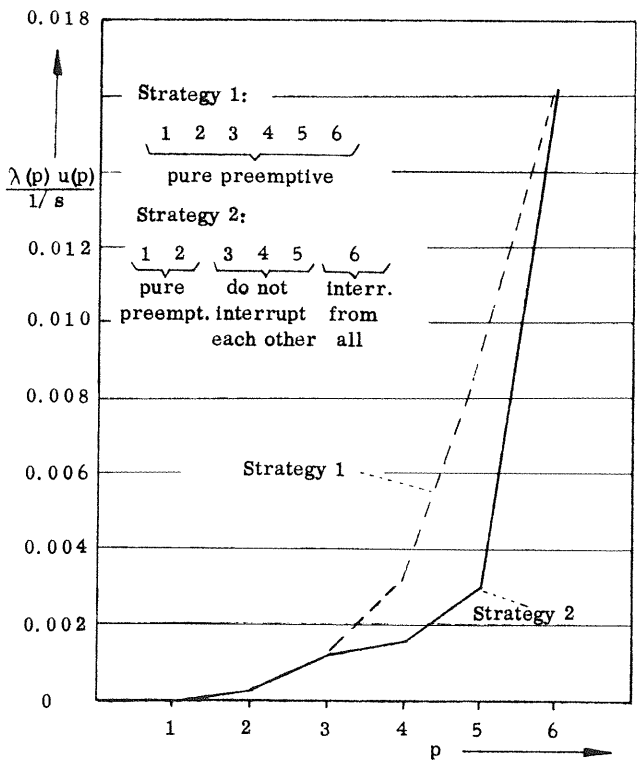


Fig. 10: Mean number of interrupts per second for non-uniform Preemption-Distance (cf. fig. 9).

Distance priorities are compared to pure preemptive or pure non-preemptive (head-of-the-line) priorities.

4.1. Uniform Preemption-Distance.

Fig. 7 shows the influence of the Preemption-Distance on the mean waiting time. Traffic intensity and traffic character are constant.

The response time is 250 times smaller for $\xi = 1$ than for $\xi = 6$ (head-of-the-line). However, the price to be paid is shown in fig. 8: an immense amount of interruptions occur.

4.2. Non-Uniform Preemption-Distance.

Fig. 9 shows a reasonable and often used "mixed" strategy between the two extremes leading for the urgent demands exactly to the same fast response time as preemptive priorities, however saving a remarkable amount of interrupts (cf. fig. 10).

4.3. Extreme Distribution Functions for the Processing Times.

Fig. 11 demonstrates which extreme types of distribution functions are included in the solution presented above. Fig. 12 shows for this example some values for the probability of waiting.

5. CONCLUSION

Reasonable combinations of preemptive and head of the line priorities are of major interest when operating switching centers in communication networks. They guarantee fast reaction to urgent signals avoiding large overhead. All these strategies are uniformly described by introducing the Preemption-Distance. The only two special cases known from literature are included in the description and analysis, as well.

The modelling with service times according to a General Erlangian distribution allows the accurate description of any type of processing times.

Mean values (response time, waiting time,...) and characteristic probability values (probability of interruption,...) show the main feature of a destinct strategy. For many practical applications these results are sufficient. A more detailed analysis is possible by means of the distribution function and a first step in this direction is the variance being taken just into consideration.

Modern switching centers have duplexed register sets and hardware for interrupt handling. Therefore, the time for interrupt handling is very small compared to the processing times. However, it is proposed to analyze control-

Priority Class	Does Not Interrupt	Interrupts Class	Preemption Distance
1	1	2, 3	1
2	1, 2, 3	-	2
3	1, 2, 3	-	-

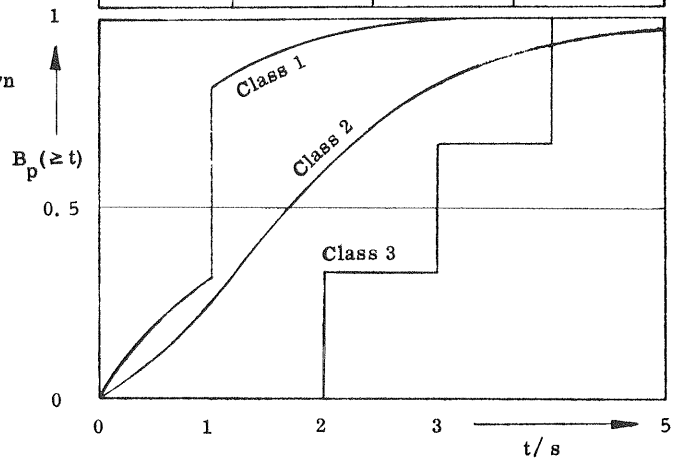


Fig. 11: Example for three different types of service time distribution functions included in the General Erlangian distribution. The mean values are assumed to be proportional to the class-number (b(1)=1sec, b(2) = 2 sec, b(3) = 3 sec), the Preemption-Distance is non-uniform.

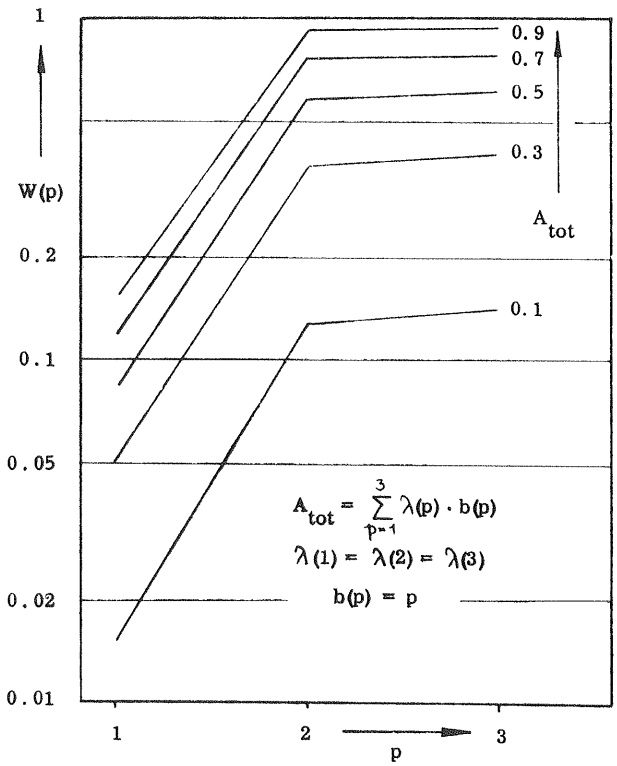


Fig. 12: Probability of waiting W(p) for all three priority classes. Priority strategy and service time distributions cf. figure 11.

computers also with the method presented above. And for small control-computers interrupt handling may be done by software adding a remarkable overhead. First results for these systems are already available and will be published at some future time.

ACKNOWLEDGEMENT

The author wishes to express his thanks to Professor Dr.-Ing. A. Lotze, Director of the Institute for Switching and Data Technics, University of Stuttgart having supported continuously these investigations.

He is also grateful to H. Roeck and A. Vetter contributing to this investigations while preparing their thesis. Furthermore, the author is thankful to his collegues Dr.-Ing. M. Langenbach-Belz, Dipl.-Ing. W. Kraemer and Dipl.-Ing. H. Weisschuh for their valuable help and many discussions.

REFERENCES

/1/ Siemens: System IV, ein Fernsprech-Vermittlungssystem mit gespeichertem Steuerprogramm. Reprints from "Informationen Fernsprechvermittlungstechnik", Siemens A.G., Muenchen, 1970.

/2/ Gabler, H.G.: The German EDS Network. Proc. ACM/IEEE Second Symp. on Problems in the Optimization of Data Commun. Systems. Palo Alto. Calif. Oct. 1971, pp. 80-85.

/3/ Herzog, U.: Preemption-Distance Priorities in Real-Time Computer Systems. Nachrichtentechn. Z. 25 (1972) pp. 201-203.

/4/ Chang, W.: Queuing with Non-Preemptive and Preemptive Resume Priorities. Operations Res. 13 (1965) pp. 1021-1022.

/5/ Dukhovnyi, I.M.; Kolin, K.K.: Optimization of the Structure of a Digital Computer Program as a Priority Queuing System. Proc. Symp. on Computer-Commun. Networks and Teletraffic, Polytechn. Inst. of Brooklyn, New York (1972), pp. 145-156.

/6/ Kendall, D.G.: Stochastic Processes Occuring in the Theory of Queues and their Analysis by the Method of the Imbedded Markov Chain. Ann. math. Statist 24 (1953), pp. 338-354.

/7/ Brockmeyer, E.; Halström, H.L.; Jensen, A.: The life and works of A.K. Erlang. Acta Polytechnica Scandinavia (1960), No. 287.

/8/ Lindley, D.V.: The Theory of Queues with a Single Server. Proc. Cambridge Phil. Soc. Vol. 48, part 2, 1952.

/9/ Cox, D.R., Miller, H.D.: The Theory of Stochastic Processes. J. Wiley & Sons Inc., New York, 1965.

/10/ Herzog, U.: Verkehrsfluss in Datennetzen. Habilitation Thesis, University of Stuttgart, submitted in January 1973.