

Institut für Nachrichtenvermittlung und Datenverarbeitung

Universität Stuttgart

Prof. Dr.-Ing. A. Lotze

15. Bericht über verkehrstheoretische Arbeiten

**Über die Berechnung der Wartezeiten in
Vermittlungs- und Rechnersystemen**

von

PAUL KÜHN

Institute for Switching and Data Technics

University Stuttgart

Prof. Dr.-Ing. A. Lotze

15th Report on Studies in Congestion Theory

**On the Calculation of Waiting Times
in Switching and Computer Systems**

by

PAUL KÜHN

ABSTRACT

The paper deals with the analytical investigation of single-stage service systems having n fully or limited accessible servers and $g \geq 1$ limited input queues with s_j waiting places, $J = 1, 2, \dots, g$. The input processes are Poissonian with arrival rates λ_j , $J = 1, 2, \dots, g$. Service times are negative exponentially distributed with termination rate ϵ_I for the I -th server, $I = 1, 2, \dots, n$ (Markovian assumptions). As queue disciplines FIFO (D1), RANDOM (D2), and LIFO (D3) are assumed, the latter with two variations: a) with and b) without displacing priorities, respectively. For the service of a certain queue, a general probabilistic interqueue discipline is introduced defined by a conditional probability p_{IJ} for service of the J -th queue on condition that server I terminates its service, $I = 1, 2, \dots, n$; $J = 1, 2, \dots, g$. In the general case, p_{IJ} depends on the whole state of the system, including various special cases as fixed or variable nonpre-emptive priorities, service according to probabilities proportional to the momentary queue lengths, and random selection. For the hunting of free servers, two hunting modes are considered: sequential hunting and random hunting.

Based on the Kolmogorov equations, the stationary state probabilities and the conditional probability distribution functions (cpdf) of waiting time are investigated by considering the corresponding random processes of system states and waiting. Main interest is drawn to the methodically uniform construction of the waiting processes from the process of system states and their investigation on the basis of eigenvalue theory. Characteristic values as probabilities for waiting and loss, mean queue lengths, mean waiting times, and the carried traffic are derived from the probabilities of state. From the cpdfs of waiting time the total pdf of waiting time and its higher moments are also derived. Numerical results are presented to demonstrate the influences of system structure, queue, and interqueue disciplines on the characteristic values and the pdf of waiting time. In the final chapter, various methods are suggested for approximate calculations.

In the following, a brief review of the various chapters will be given.

Chapter 1. General Remarks on Service Systems (pp.13-14)

Introducing remarks are given on classifications of service systems and their appearance in practice. It is referred to general problems of traffic theory, its basic assumptions and the possible outcome.

Chapter 2. Assumptions and Stating of the Problem (pp.15-26)

The basic assumptions about the input process (2.1), the termination process (2.2), the structure of the service system (2.3), and the organization (disciplines) of the service system (2.4) are explained in detail and briefly summed up (2.5). In Section 2.6, the traffic characteristics are defined which characterize the grade of service of the service systems in question. A short reference to the calculation of the various traffic characteristics and the contents of the following chapters ends the chapter.

Chapter 3. Fields of Application (pp.27-31)

Two fields of application of the investigated service systems are inspected: switching systems (3.1), as well as computer structures and operating systems (3.2). For traffic models in switching systems, a single-stage connecting array for the speech path network, and a connecting network between junctors and centralized registers are referred to. Traffic models in computer systems can be found in batch processing computers, real-time computers for process-control and measurement processing, multiprogrammed computers with scheduling routines for the central processing unit, as well as in computerized data networks with message switching, overflow, and load-sharing strategies.

Chapter 4. Introduction to the Theory of Markovian Service Systems (pp.32-59)

In Section 4.1, the principal Kolmogorov equations are derived for the random process of system states. For the probabilities of state, the basic forward-type differential equation Eq.(4.12) for the non-stationary case, and the corresponding equilibrium state equation Eq.(4.14) for the stationary case are derived, respectively. The fundamental characteristic values are defined for the example of a single-queue system (4.2). Section 4.3 handles the problem of waiting and response times. The process of waiting of an arbitrary "test call" is constructed uniformly from the process of system states by

consideration of all those transitions which influence the fate of the waiting test call either directly or indirectly. By application of the Kolmogorov-backward equation to this waiting process, the basic backward-type differential equation Eq.(4.32) for the cpdfs of waiting time is derived. The theoretical treatment of the differential equations is carried out by investigation of the underlying eigenvalue problem. The residual parts of Section 4.3 deal with the total pdf of waiting time, the moments of the cpdfs and the total pdf of waiting time, and the calculation of response times, as well. In Section 4.4, the methods are extended to multiqueue service systems with more than one input process. Finally, some basic properties of multidimensional Markovian processes are derived.

Chapter 5. Single-Queue Service Systems (pp.60-98)

This chapter deals with the fully accessible n-server system with a limited input queue. In Section 5.1, the probabilities of state are investigated for sequential hunting (5.1.1) as well as for random hunting (5.1.2); for the latter, an explicit solution is derived. From the stationary probabilities of state, the most important characteristic values are derived (5.2). In Section 5.3, the waiting processes are investigated systematically for the above mentioned queue disciplines FIFO (D1,5.3.1), RANDOM (D2,5.3.2), LIFO with (D3a,5.3.3) and LIFO without displacing priorities (D3b,5.3.4). The solution steps are as follows: Construction of the waiting process and its graphical representation by random walk diagrams, matrix of the transition coefficients of the waiting process, investigation of the eigenvalues (cf. also the Appendix), conditional and total pdf of waiting time, conditional mean waiting times and conditional higher moments, total mean waiting times and total higher moments. For D1, the well-known results are reproduced by this method; for D3a, the explicit solution of the total pdf of waiting time could also be derived. For all disciplines, explicit solutions or recursion algorithms are given for calculation of the higher moments. In Section 5.4, some limiting cases of the arrival rate are considered.

Section 5.5 presents some numerical results for a single-server system operating under the four different queue disciplines to demonstrate the effects of queue disciplines and load on the character-

istic values and the total pdf of waiting time $W(>t)$, cf. Table 1 and Figs. 15a,b,c. It should be mentioned that the effect of queue disciplines and load on the waiting times is also reflected by the distribution of eigenvalues, cf. Fig. 16.

Chapter 6. Multiqueue Service Systems with Full Accessibility (pp.99-134)

This chapter deals with the fully accessible n-server system with $g > 1$ limited input queues (for simplicity, in most cases $g = 2$). The state space and the calculation of the stationary probabilities of state are discussed in general as well as for some special cases of the interqueue discipline (6.1). The characteristic values are defined in Section 6.2. The waiting times are treated in Section 6.3. This section is subdivided into two main parts according to two main groups of interqueue disciplines which are independent (type A, 6.3.1) or dependent (type B, 6.3.2) of the momentary lengths of the (nonempty) queues. For each type A and B, the four queue disciplines D1, D2, D3a, and D3b are considered. The waiting processes and their corresponding multidimensional random walk diagram representations are investigated with respect to eigenvalues and division into subsystems, respectively.

Section 6.4 presents a numerical example of a double-queue service system operating under three different interqueue disciplines. The results are shown in Fig. 24a for the probabilities of waiting and loss, W_J and B_J , in Fig. 24b for the mean waiting times t_{WJ} , and in Fig. 25 for the total pdfs of waiting time $W_J(>t)$, $J = 1,2$. The interqueue disciplines are: priority service for queue one (solid lines), service according to probabilities proportional to the momentary queue lengths (dashed lines), and random selection of queues (dotted lines). For Fig.25, FIFO was assumed as queue discipline. By these results, the influence of interqueue disciplines on the traffic characteristics can be studied.

Chapter 7. Multiqueue Service Systems with Limited Accessibility (pp.135-152)

In this chapter, the methods of Chapter 6 are extended to limited accessibility. Section 7.1 deals with the probabilities of state. In Section 7.2, the characteristic values are defined by the example of a simple grading to show the differences compared with full

accessibility. Section 7.3 deals with the waiting times. This section follows the same line as Section 6.3 for fully accessible servers. Therefore, the main interest is drawn to the more complicated construction of waiting processes which is additionally demonstrated by a more extensive practical example, cf. Fig.28 (7.3.1.1c).

The final Section 7.4 presents two numerical examples. In the first example, a double-queue, 3-server system with different mean service times is considered working under two types of interqueue disciplines, cf. Fig. 30: priority service for queue one (solid lines), and service according to probabilities proportional to the momentary queue lengths (dashed lines). This example stands for a three-computer system, where the third computer takes the overflow traffic of the first and second one, and demonstrates the effects of unbalanced load as well as interqueue disciplines on the characteristic values. The second example is drawn from switching systems and shows the influence of the accessibility k on the traffic characteristics. Two 6-server systems with three input queues and limited accessibility ($k = 4$) or full accessibility ($k = 6$), cf. Fig.31, are compared with each other. Table 2 shows the characteristic values for 3 cases of offered traffic $A_j = 1, 2, 3$, $J = 1, 2, 3$. In Fig. 31, the total pdfs of waiting time are shown. The queue discipline was FIFO, nonempty queues were selected randomly for service.

Chapter 8. Approximate Calculation Methods (pp.153-178)

According to the two basic equations for the probabilities of state and the cpdfs of waiting time, various approximation methods are reviewed in Sections 8.1 and 8.2, respectively. Each section starts with a short recapitulation of the exact solution and refers to some special cases, cf. Sections 8.1.1 and 8.2.1, respectively. Numerical examples of the various approximation methods are given in Section 8.3.

For the approximate solution of the equations for the multidimensional probabilities of state, two conceptions of "half-symmetry" (8.1.2) and "full-symmetry" (8.1.3) are proposed which allow the recursive calculation of all probabilities of state. Another conception introduces "macrostates", especially to calculate large connecting arrays (8.1.4). An example of such a description by "macrostates" in connection with the conception of "half-symmetry" is given for

demonstration. The accuracy of this method can be seen from the examples given in Figs. 34 and 35 where the obtained results are compared with simulation results. Finally, an approved approximation method based on given distributions with iterated parameters is also referred to, cf. Section 8.1.5.

For the approximate solution of large systems of differential equations for the cpdfs of waiting time, the conceptions of "half-symmetry" (8.2.2) and "full-symmetry" (8.2.3) are introduced analogously. In Section 8.2.4, a method of successive approximations by power series expansions is proposed which works recursively and can be applied to systems of differential equations of extremely high order. The accuracy of this method can be prescribed arbitrarily. By this method, systems of differential equations can be solved with an order up to one third of the available computer memory. The results of Figs. 25 and 31 have been calculated by this method. Besides recursivity and prescribed accuracy, another aspect of this method is formed by the interchangeability of necessary computer memory and computing time, which can be adjusted by the order of the power series expansion, cf. the example given in Table 4. In Section 8.2.5, the well-known method of approximation by exponential sums matching the first moments of the total pdf of waiting time has been extended to the approximation of the cpdfs of waiting time. By this method, a more accurate approximation was obtained as it is shown in Table 3 by comparison of exact results with those results obtained from various approximations of the total or conditional pdfs with one or two exponential functions, respectively.

Appendix (pp.181-190)

In the Appendix, various theorems stated in Chapter 5 are proved concerning properties of matrices and eigenvalues of the waiting processes for queue disciplines D2 (Appendix A1), D3a (Appendix A2), and D3b (Appendix A3).

<u>INHALTSVERZEICHNIS</u>	Seite
Schrifttumsverzeichnis	6
Bezeichnungen	9
1. Allgemeines über Bedienungssysteme	13
2. Voraussetzungen und Aufgabenstellung	15
2.1 Der Anrufprozeß	15
2.2 Der Endeprozeß	16
2.3 Die Struktur des Bedienungssystems	17
2.4 Die Organisation des Bedienungssystems	19
2.4.1 Allgemeines	19
2.4.2 Abfertigungsdisziplinen innerhalb der Warteschlangen	20
2.4.3 Abfertigungsdisziplinen zwischen den Warteschlangen	21
2.4.4 Absuchdisziplin der erreichbaren Leitungen	23
2.5 Zusammenfassung der Fälle von Systemstrukturen und Abfertigungsdisziplinen	23
2.6 Gütemerkmale und Kenngrößen	23
3. Anwendungsbereiche	27
3.1 Wählvermittlungssysteme	27
3.1.1 Koppelanordnungen im Verbindungsweg	27
3.1.2 Anschaltung zentralisierter Baugruppen	28
3.2 Rechnerstrukturen und Betriebssysteme	29
3.2.1 "Closed Shop"-Betrieb im Rechenzentrum	29
3.2.2 Prozeßsteuerung und Meßwertverarbeitung	29
3.2.3 Rechnerkernvergabe in Realzeit- und Großrechenanlagen	29
3.2.4 Vermaschte Rechnersysteme	30
4. Einführung in die Theorie Markoff'scher Bedienungssysteme	32
4.1 Zustände und Übergänge	32
4.1.1 Zustände als Zufallsprozeß	32
4.1.2 Die Markoff'sche Eigenschaft	32
4.1.3 Übergangs- und Zustandswahrscheinlichkeiten	33
4.1.4 Die Kolmogoroff'schen Gleichungen	35
4.1.5 Der stationäre Fall	35
4.2 Charakteristische Größen	36
4.3 Warte- und Antwortzeiten	38
4.3.1 Definitionen	38
4.3.2 Der Warteprozeß	39
4.3.3 Bedingte Wartezeitverteilungsfunktionen	42
4.3.4 Absolute Wartezeitverteilungsfunktionen	45
4.3.5 Laplace-Transformierte der Wartezeitverteilungsfunktionen und Eigenwertproblem	47

	Seite
4.3.6 Momente der Wartezeitverteilungsfunktionen	48
4.3.7 Mittlere Wartezeiten	50
4.3.8 Antwortzeiten	51
4.4 Markoff'sche Bedienungssysteme mit mehreren Ein- gangs-Warteschlangen	52
4.4.1 Zustände als Zufallsprozeß	53
4.4.2 Der Warteprozess	53
4.4.3 Einige Eigenschaften mehrdimensionaler Prozesse	57
5. Bedienungssysteme mit einer Warteschlange	60
5.1 Zustandswahrscheinlichkeiten	60
5.1.1 Geordnetes Absuchen der Leitungen	60
5.1.2 Zufallsmäßiges Absuchen der Leitungen	63
5.1.3 Sonderfall gleicher mittlerer Enderaten	65
5.2 Charakteristische Größen	66
5.3 Wartezeiten	68
5.3.1 Abfertigung in der Reihenfolge des Eintreffens	68
5.3.2 Abfertigung in zufälliger Reihenfolge	73
5.3.3 Abfertigung in der inversen Reihenfolge des Eintreffens mit verdrängender Priorität	78
5.3.4 Abfertigung in der inversen Reihenfolge des Eintreffens mit nichtverdrängender Priorität	84
5.4 Grenzfälle der mittleren Anrufrate	89
5.4.1 Verschwindende Anrufrate	89
5.4.2 Unbegrenzt wachsende Anrufrate	89
5.5 Numerische Beispiele	91
6. Bedienungssysteme mit vollkommener Erreichbarkeit und mehreren Warteschlangen	99
6.1 Zustandswahrscheinlichkeiten	99
6.1.1 Allgemeine Zustandsgleichungen	99
6.1.2 Sonderfälle	102
6.2 Charakteristische Größen	105
6.3 Wartezeiten	106
6.3.1 Abfertigung zwischen den Warteschlangen nach Wahrscheinlichkeiten unabhängig von den momentanen Warteschlangenlängen	107
6.3.1.1 Abfertigung in der Reihenfolge des Eintreffens	107
6.3.1.2 Abfertigung in zufälliger Reihenfolge	113
6.3.1.3 Abfertigung in der inversen Reihen- folge des Eintreffens mit ver- drängender Priorität	116
6.3.1.4 Abfertigung in der inversen Reihen- folge des Eintreffens mit nicht- verdrängender Priorität	119
6.3.2 Abfertigung zwischen den Warteschlangen nach Wahrscheinlichkeiten abhängig von den momentanen Warteschlangenlängen	121

	Seite
6.3.2.1 Abfertigung in der Reihenfolge des Eintreffens	121
6.3.2.2 Abfertigung in zufälliger Reihenfolge	126
6.3.2.3 Abfertigung in der inversen Reihenfolge des Eintreffens mit verdrängender Priorität	126
6.3.2.4 Abfertigung in der inversen Reihenfolge des Eintreffens mit nichtverdrängender Priorität	130
6.4 Numerische Beispiele	130
7. Bedienungssysteme mit unvollkommener Erreichbarkeit und mehreren Warteschlangen	135
7.1 Zustandswahrscheinlichkeiten	135
7.2 Charakteristische Größen	138
7.3 Wartezeiten	139
7.3.1 Abfertigung zwischen den Warteschlangen nach Wahrscheinlichkeiten unabhängig von den momentanen Warteschlangenlängen	140
7.3.1.1 Abfertigung in der Reihenfolge des Eintreffens	140
7.3.1.2 Abfertigung in der inversen Reihenfolge des Eintreffens mit verdrängender Priorität	144
7.3.2 Abfertigung zwischen den Warteschlangen nach Wahrscheinlichkeiten abhängig von den momentanen Warteschlangenlängen	147
7.3.2.1 Abfertigung in der Reihenfolge des Eintreffens	147
7.3.2.2 Abfertigung in der inversen Reihenfolge des Eintreffens mit verdrängender Priorität	147
7.4 Numerische Beispiele	148
8. Approximative Berechnungsmethoden	153
8.1 Zustandswahrscheinlichkeiten	153
8.1.1 Die exakte Lösung	153
8.1.2 Das Konzept der Halbsymmetrie	157
8.1.3 Das Konzept der Vollsymmetrie	157
8.1.4 Das Konzept der Makrozustände	158
8.1.5 Das Konzept vorgegebener Verteilungen mit iteriertem Parameter	159
8.2 Wartezeitverteilungsfunktionen	159
8.2.1 Die exakte Lösung	159
8.2.2 Das Konzept der Halbsymmetrie	162
8.2.3 Das Konzept der Vollsymmetrie	164
8.2.4 Das Konzept der sukzessiven Approximation durch Potenzreihenentwicklungen	165
8.2.5 Das Konzept der Approximation durch Exponentialsummen (Momentenmethode)	167
8.3 Numerische Beispiele	171

	Seite
8.3.1 Zustandswahrscheinlichkeiten	171
8.3.2 Wartezeitverteilungsfunktionen	175
Zusammenfassung	179
Anhang	
A1 Eigenschaften der Matrix der Sprungkoeffizienten bei einer Warteschlange und Abfertigung in zufälliger Reihenfolge	181
A2 Wartezeitverteilungsfunktion bei einer Warteschlange und Abfertigung in der inversen Reihenfolge des Eintreffens mit verdrängender Priorität	186
A3 Eigenschaften der Matrix der Sprungkoeffizienten bei einer Warteschlange und Abfertigung in der inversen Reihenfolge des Eintreffens mit nicht-verdrängender Priorität	189

SCHRIFTTUMSVERZEICHNIS:

- [1] BROCKMEYER, E.
HALSTRØM, H.L.
JENSEN, A. The life and works of A.K.ERLANG
Acta Polytech.Scandinavica, Copenhagen,
1960.
- [2] SYSKI, R. Introduction to congestion theory in
telephone systems
Oliver and Boyd, Edinburgh,London,1960.
- [3] COHEN, J.W. The single server queue
North Holland, 1969.
- [4] POLLACZEK, F. Problèmes stochastiques posés par le phé-
nomène de formation d'une queue d'attente
à un guichet et par des phénomènes
apparentés
Mém. Sc. Math., fasc. 136, Paris, 1957.
- [5] LOTZE, A. History and development of grading theory
a) 5th International Teletraffic Congress
(ITC), New York, 1967.
b) Arch. für Elektronik und Übertragungs-
tech. (A.E.Ü.) 25, 1971.
- [6] WAGNER, W. Über ein kombiniertes Warte-Verlust-
system mit Prioritäten
Dissertation Universität Stuttgart,1968.
- [7] KATZSCHNER, L. Verlustsysteme der Wählermittlungstech-
nik mit Prioritäten
Dissertation Universität Stuttgart,1970.
- [8] BRANDT, G.J. Das preemptive Warteverlustsystem
Dissertation Universität Stuttgart,1971.
- [9] JAISWAL, N.K. Priority Queues
Academic Press, New York und London,
1968.
- [10] THIERER, M. Wartesysteme mit unvollkommener Erreich-
barkeit
Dissertation Universität Stuttgart,1968.
- [11] KÜHN, P. Combined delay and loss systems with
several input queues, full and limited
accessibility
a) Proceedings of the 6th International
Teletraffic Congress (ITC), München,
1970.
b) Arch. für Elektronik und Übertragungs-
tech.(A.E.Ü.) 25, 1971.

- [12] KINGMAN, J.F.C. The effect of queue discipline on waiting time variance
Proc. Phil. Soc. 58, 1962.
- [13] TAMBOURATZIS, D.G. On a property of the variance of the waiting time of a queue
Journ. Appl. Prob. 5, 1968.
- [14] TAKÁCS, L. Introduction to the theory of queues
Oxford University Press, New York, 1962.
- [15] SYSKI, R. Markovian queues
Proceedings of the Symp. on Congestion Theory, University of North Carolina, 1964.
- [16] PALM, C. Research on telephone traffic carried by full availability groups
Tele, v.1, 1957.
- [17] DOETSCH, G. Anleitung zum praktischen Gebrauch der Laplace-Transformation
R. Oldenbourg Verlag, München, 1961.
- [18] ZURMÜHL, R. Matrizen und ihre technischen Anwendungen
Springer-Verlag, Berlin/Göttingen/Heidelberg, 1964.
- [19] SCHWARZ, H.R.
RUTISHAUSER, H.
STIEFEL, E. Matrizen-Numerik
B.G.Teubner Verlag, Stuttgart, 1968.
- [20] EBERLEIN, P.J.
BOOTHROYD, J. Solution to the eigenproblem by a norm reducing Jacobi type method
Numerische Mathematik 11, 1968.
- [21] STÖRMER, H. Wartezeitlenkung in handbedienten Vermittlungsanlagen
Arch.d.Elekt.r.Übertr.(A.E.Ü.)10, 1956.
- [22] SEGAL, M. A multiserver system with a finite number of sources and delayed requests served at random
Proceedings of the 6th International Teletraffic Congress (ITC), München, 1970.
- [23] FELLER, W. Introduction to probability theory and its applications
J. Wiley, New York/London, 1961.
- [24] RIORDAN, J. Stochastic service systems
J. Wiley, New York/London, 1962.

- [25] LOTZE, A. Verluste und Güteigenschaften einstufiger Mischungen
Nachrichtentechn.Zeitschrift (NTZ) 14, 1961.
- [26] LOTZE, A. Berechnung von Zeit- und Rufblockierung in Linksystemen mit zwei und mehr Stufen, Mischwahl und Richtungswahl nach einem Näherungsverfahren "Kombinierte Eingangs- und Abnehmerblockierung" (CIRB)
3. Arbeitsbericht über verkehrstheoretische Arbeiten, Institut für Nachrichtenvermittlung und Datenverarbeitung Universität Stuttgart, 1963.
- [27] LOTZE, A. Berechnung der Verkehrsgrößen im Wartezeitsystem aus den Verkehrsgrößen eines Verlustsystems
Fernmeldetechn.Zeitschr.(FTZ) 7, 1954.
- [28] MELLOR, J.W. Delayed call formulae when calls are served in a random order
P.O.E.E.J. 25, 1942.
- [29] RIORDAN, J. Delay curves for calls served at random
B.S.T.J. 32, 1953.

BEZEICHNUNGEN

1. Strukturparameter

g	Zahl der Teilgruppen bzw. Wartespeicher
n	Zahl der Leitungen
k_J	Zahl der von der Teilgruppe J aus erreichbaren Leitungen
k	Erreichbarkeit
s_J	Zahl der Wartepplätze des J-ten Wartespeichers
I	Zählvariable für Leitungen
J	Zählvariable für Teilgruppen

2. Organisationsparameter

$D...$	Abfertigungsdisziplinen innerhalb einer Warteschlange
A, B	Abfertigungsdisziplinen zwischen den Warteschlangen
P_J	Bedingte Wahrscheinlichkeit für das Abfertigen der J-ten Warteschlange bei einem Belegungsende (vollk. Erreichbarkeit)
P_{IJ}	Bedingte Wahrscheinlichkeit für das Abfertigen der J-ten Warteschlange bei Belegungsende der Leitung I (unvollk. Erreichbarkeit)

3. Verkehrsparameter

λ_J	Mittlere Anrufrate des J-ten Anrufprozesses
λ	Gesamte mittlere Anrufrate
ϵ_I	Mittlere Enderate der I-ten Leitung
ϵ	Mittlere Enderate einer beliebigen Leitung
μ	Gesamte mittlere Enderate aller Leitungen
ϵ_{IJ}	Bedingte Abfertigungsrate der J-ten Warteschlange bezüglich der Leitung I
μ_J	Bedingte Abfertigungsrate der J-ten Warteschlange bezüglich aller Leitungen
A_J	Verkehrsangebot des J-ten Anrufprozesses
A	Gesamtes Verkehrsangebot
ρ_J	Verkehrsangebot des J-ten Anrufprozesses bezogen auf eine Leitung
ρ	Gesamtes Verkehrsangebot bezogen auf eine Leitung

4. Zufallsvariable

T_A	Anrufabstand (interarrival time)
T_S	Belegungsdauer (service time)

T_W	Wartezeit (waiting time)
T_W^*	Erfolgreiche Wartezeit
T_W^{**}	Erfolglose Wartezeit
T_R	Antwortzeit (response time)
i, j, k	Allgemeine Systemzustände
x_I	Belegungszustand der I-ten Leitung
x	Zahl insgesamt belegter Leitungen
i_J, z_J	Zahl der Wartenden in der J-ten Warteschlange
u_r	Zahl der auf dem r-ten Wartepplatz Wartenden

5. Zufallsprozesse

$\xi(t)$	Zufallsprozeß der Systemzustände
$\xi(t)$	Warteprozeß
$\xi_J(t)$	Warteprozeß im J-ten Wartespeicher
Z	Menge der Systemzustände des $\xi(t)$ -Prozesses
H	Menge der Absorptionszustände des $\xi(t)$ -Prozesses
H_J	Menge der Absorptionszustände des $\xi_J(t)$ -Prozesses

6. Verteilungen

$A(\leq t)$	Verteilungsfunktion (VF) der Anrufabstände
$S(\leq t)$	VF der Belegungsauern (Bedienungsdauern)
$p_r(t)$	VF für r Ereignisse
$P(t; j i)$	Bedingte Übergangswahrscheinlichkeit des $\xi(t)$ -Prozesses
$H^P(t; j i)$	Bedingte Übergangswahrscheinlichkeit des $\xi(t)$ -Prozesses
$P(t; i)$	Zeitabhängige Zustandswahrscheinlichkeit
$p(i)$	Stationäre Zustandswahrscheinlichkeit
q_{1j}, q_i	Übergangswahrscheinlichkeitsdichten des $\xi(t)$ -Prozesses
$q(i, j), q(1)$	Übergangswahrscheinlichkeitsdichten des $\xi(t)$ -Prozesses
$\epsilon(i)$	Endewahrscheinlichkeitsdichte des $\xi(t)$ -Prozesses
$P(t)$	Matrix der bedingten Übergangswahrscheinlichkeiten
Q	Matrix der Übergangswahrscheinlichkeitsdichten des $\xi(t)$ -Prozesses
A	Matrix der Übergangswahrscheinlichkeitsdichten des $\xi(t)$ -Prozesses
$P(i)$	Antreffwahrscheinlichkeit

$w(t i)$	Bedingte Wartezeitverteilungsfunktion
$h(t i)$	Bedingte Wartezeitverteilungsdichtefunktion
$W(s i)$	Laplace-Transformierte von $w(t i)$
$H(s i)$	Laplace-Transformierte von $h(t i)$
$W(>t)$	(komplementäre) Wartezeitverteilungsfunktion
$R(>t)$	(komplementäre) Antwortzeitverteilungsfunktion
J	Index bezüglich der J-ten Teilgruppe
*	Index bezüglich erfolgreich wartender Rufe
**	Index bezüglich erfolglos wartender Rufe

7. Mittelwerte

B	Verlustwahrscheinlichkeit
W	Wartewahrscheinlichkeit
t_A	Mittlerer Anrufabstand
t_S	Mittlere Belegungsdauer
$t_W(i)$	Mittlere bedingte Wartezeit
t_W	Mittlere Wartezeit
$M_K(i)$	K-tes Moment der bedingten Wartezeitverteilungsfunktion
M_K	K-tes Moment der gesamten Wartezeitverteilungsfunktion
Y	Mittlere Zahl belegter Leitungen
Ω	Mittlere Warteschlangenlänge
J	Index bezüglich der J-ten Teilgruppe
*	Index bezüglich erfolgreich wartender Rufe
**	Index bezüglich erfolglos wartender Rufe

8. Allgemeine Größen

h, t, u	Zeitvariable
dh, dt, du	Zeitdifferentiale
τ, η, ξ	Hilfsvariable
s	Komplexe Bildvariable
σ, ω	Reelle bzw. imaginäre Komponente von s
e_ν	Eigenwert
N	Rang, Anzahl aller Eigenwerte
N'	Anzahl unterschiedlicher Eigenwerte
K_ν	Vielfachheit des ν -ten Eigenwerts e_ν
ν, α	Zählgrößen

$a_{i\nu}^{(\alpha)}$	Residuum
A, B, C, T	Matrizen
$D_i, D_i(s)$	Hauptabschnittsdeterminanten
$\sigma_{i,j}$	Nullstellen von $D_i(s)$
G, G'	Einschränkungsgebiete für Eigenwerte

1. ALLGEMEINES ÜBER BETRIEBSSYSTEME

Betriebssysteme haben die Aufgabe, Anfragende möglichst schnell mit einem freien Abfertigungsorgan zu verbinden. Die Zahl der Abfertigungsorgane ist in der Regel kleiner als die maximale Zahl der Anfragenden, wodurch in Zeiten von Anfragenhäufungen ein Hemmnis bei der Abfertigung auftreten kann. Trifft eine Anfrage zu einem Zeitpunkt ein, in dem alle erreichbaren Abfertigungsorgane belegt sind, so darf sie entweder auf ihre Abfertigung warten (Wartesystem) oder sie wird sofort abgewiesen (Verlustsystem). Bei einer beschränkten Zahl von Warteplätzen darf in diesem Falle die Anfrage nur dann warten, wenn noch mindestens ein freier Warteplatz angetroffen wurde. Sind auch sämtliche verfügbaren Warteplätze belegt, so kann die Anfrage entweder sofort abgewiesen werden oder sie wird akzeptiert und verdrängt damit eine andere bereits wartende Anfrage (Kombiniertes Warte-Verlustsystem). Sind die Anfragen in Gruppen von warte- und nichtwarteberechtigten aufgeteilt, so wird eine nichtwarteberechtigte Anfrage abgewiesen, wenn sie alle erreichbaren Bedienungsorgane belegt vorfindet, während eine warteberechtigte Anfrage in diesem Falle einen Warteplatz belegen darf (Gemischtes Warte-Verlustsystem).

Wartezeitprobleme beobachtet man u.a. an Schaltern, Kassen von Supermärkten sowie bei der Abfertigung von Transportmitteln wie Lastwagen, Schiffen und Flugzeugen. Eine sehr wichtige Rolle spielen Wartezeitprobleme in Vermittlungssystemen für Ferngespräche und Fernschreiben, in Prozeßrechnersystemen, innerhalb von modernen Großrechenanlagen und in vermaschten Datenübertragungsnetzen für Auskunftszwecke. Bei der Massenbedienung in derartigen Systemen stellt man fest, daß i.a. sowohl die Anfrageabstände als auch die Bedienungszeiten statistischen Gesetzmäßigkeiten unterliegen.

Die Aufgabe der Bedienungstheorie (Verkehrstheorie) besteht darin, optimale Strukturen und optimale Organisationsformen von Betriebssystemen zu finden, um einerseits eine schnelle Abwicklung aller Aufgaben zu erreichen, und um andererseits die erforderlichen Bedienungseinheiten gut auszunutzen. Zur Lösung dieser Aufgabe gibt es prinzipiell zwei Möglichkeiten: analytische Methoden und Simulationsmethoden. In der vorliegenden Arbeit werden nur analytische Methoden behandelt.

Die Optimierungsaufgabe kann in vielen Fällen nicht direkt gelöst werden. Man geht vielmehr so vor, daß vorgegebene Strukturen und Organisationsformen mit Hilfe geeigneter Kenngrößen, welche entweder durch analytische Rechnung oder Simulation gewonnen werden, beurteilt werden. Hieraus gewinnt man Anhaltspunkte für eine iterative Lösung der gestellten Aufgabe.

Zur Charakterisierung des Problems werden folgende vier Angaben gemacht:

- Verteilung der Anfrageabstände (Anrufprozeß)
- Verteilung der Bedienungszeiten (Endeprozeß)
- Struktur des Betriebssystems
- Organisation des Betriebssystems.

Als Kenngrößen werden Wahrscheinlichkeiten für das Auftreten der verschiedenen Systemzustände, für das Warten, das Abgewiesenwerden sowie für beliebige Wartezeiten berechnet. Ferner werden Mittelwerte für Warteschlangenlängen, Belegung der Abfertigungsorgane sowie auftretende Wartezeiten bestimmt.

Die ersten wichtigen Arbeiten auf dem Gebiet der Bedienungstheorie gehen auf A. K. ERLANG zu Beginn dieses Jahrhunderts zurück [1]. In der Zwischenzeit ist eine sehr große Zahl von Arbeiten auf diesem Gebiet erschienen, siehe [2,3,4]. Die meisten dieser Arbeiten behandeln Systeme mit einem oder mehreren Abfertigungsorganen und einem unbeschränkt großen Warteraum unter sehr allgemeinen Voraussetzungen bezüglich Anruf- und Endeprozeß. In jüngerer Zeit haben jedoch auch Untersuchungen über komplizierter strukturierte und komplizierter organisierte Betriebssysteme an Bedeutung gewonnen. Als Beispiele seien Systeme mit seriellem Warten, unvollkommen erreichbaren Abfertigungsorganen bzw. Systeme mit Prioritätsdisziplinen angeführt.

In der vorliegenden Arbeit werden kombinierte Warte-Verlustsysteme mit einer Warteschlange wie auch mit mehreren Warteschlangen behandelt. Die betrachteten Systeme finden vornehmlich ihre Anwendung in Wahlvermittlungen, Prozeßsteuerungen und modernen Großrechenanlagen.

2. VORAUSSETZUNGEN UND AUFGABENSTELLUNG

Im folgenden sollen die Voraussetzungen über Anrufprozeß, Endeprozeß, Struktur und Organisation eines Bedienungssystems im einzelnen erläutert werden. Die wichtigsten Kenngrößen zur Beurteilung der Güte eines Bedienungssystems werden diskutiert.

Im Verlauf der weiteren Arbeit werden Anfragen kurz als "Rufe" und Bedienungsgorgane als "Leitungen" bezeichnet.

2.1 Der Anrufprozeß

Der Anrufprozeß beschreibt die zeitliche Aufeinanderfolge von Rufen. Ist ..., t_{i-1}, t_i, ... die Folge von Anrufzeitpunkten, so kann der Anrufprozeß durch die Wahrscheinlichkeits-Verteilungsfunktion (VF) der Anrufabstände T_A = t_i - t_{i-1} beschrieben werden. Im allgemeinen Falle des rekurrenten Anrufprozesses sind die Anrufabstände T_A (interarrival times) unabhängige Zufallsvariable, welche entsprechend einer allgemeinen VF

A(≤t) = P{T_A ≤ t} (2.1)

verteilt sind.

Der in der Praxis wichtigste Fall eines Anrufprozesses ist der POISSON-Prozeß. Er hat in der Theorie und in den Anwendungen deshalb die größte Bedeutung erreicht und soll in dieser Arbeit durchweg vorausgesetzt werden. Der POISSON-Prozeß ist dadurch gekennzeichnet, daß ein Ruf im Zeitintervall (t, t+dt) mit der Wahrscheinlichkeit

p_1(dt) = λdt + o(dt) (2.2)

eintrifft, wobei λ eine Konstante ist, welche die mittlere Anzahl von Rufen pro Zeiteinheit angibt (mittlere Anrufrate oder Rufintensität). o(dt) ist eine Größe von höherer Ordnung in dt mit der Eigenschaft lim_{dt→0} o(dt)/dt = 0. Ebenso ist die Wahrscheinlichkeit für das Eintreffen von ν > 1 Rufen im Intervall (t, t+dt) von höherer Ordnung in dt.

Durch Betrachtung aller möglichen Ereignisse im Intervall der Länge dt läßt sich eine Differentialgleichung für die Wahrscheinlichkeit p_ν(t), daß im Intervall (0, t) genau ν Rufe eintreffen, aufstellen:

p_ν(t+dt) = p_ν(t) · [1-λdt] + p_{ν-1}(t) · λdt + o(dt), ν ≥ 0. (2.3a)

p_ν(t+dt) ist die Wahrscheinlichkeit, daß im Intervall (0, t+dt) ν Rufe eintreffen. Sie bestimmt sich entsprechend Gl.(2.3a) aus der Wahrscheinlichkeit, daß im Intervall (0, t) schon alle ν Rufe eingetroffen sind und sich im Intervall (t, t+dt) nichts ereignet, p_ν(t) · [1-λdt], und der Wahrscheinlichkeit, daß im Intervall (0, t) nur ν-1 Rufe eingetroffen sind und im Intervall (t, t+dt) der ν-te Ruf eintrifft, p_{ν-1}(t) · λdt.

Durch Grenzübergang dt→0 folgt aus Gl.(2.3a)

dp_ν(t)/dt = -λ · p_ν(t) + λ · p_{ν-1}(t), ν ≥ 0. (2.3b)

Mit der Anfangsbedingung

p_ν(0) = 0, ν ≥ 0, (2.3c)

erhält man als Lösung von Gl.(2.3b)

p_ν(t) = (λt)^ν / ν! · exp(-λt), ν ≥ 0. (2.4)

Die Zahl der Rufe in einem festen Intervall (0, t) folgt also einer POISSON-Verteilung.

Wegen der konstanten mittleren Anrufrate λ kann der POISSON-Anrufprozeß durch eine unendlich große Quellenzahl erzeugt gedacht werden, wobei jede einzelne Quelle einen verschwindenden Anteil liefert, derart, daß die Summe aller Anteile die endliche Rufintensität λ ergibt.

Die VF der Anrufabstände folgt aus Gl.(2.4) und ist negativ-exponentiell:

A(≤t) = 1 - p_0(t) = 1 - exp(-λt). (2.5)

Der mittlere Anrufabstand t_A ist der Erwartungswert von T_A und bestimmt sich zu

t_A = ∫_{t=0}^∞ t · A'(≤t) dt = 1/λ. (2.6)

2.2 Der Endeprozeß

Der Endeprozeß beschreibt die Bedienungszeiten (Belegungsauern, service times oder holding times) der einzelnen Rufe. Die Belegungsauern T_S sind voneinander unabhängige Zufallsvariable mit der VF

S(≤t) = P{T_S ≤ t}. (2.7)

Die Zahl der Koppelpunkte beträgt

$$\sum_{j=1}^g k_j \quad \left\{ \begin{array}{l} = g \cdot n \quad \text{für vollkommene Erreichbarkeit} \\ < g \cdot n \quad \text{für unvollkommene Erreichbarkeit.} \end{array} \right.$$

Die Zuordnung der einzelnen Leitungen auf die Zubringerteilgruppen wird als "Mischung" bezeichnet. Koppelanordnungen mit unvollkommener Erreichbarkeit haben vor allem in Wählvermittlungen aus Gründen der Koppelpunkteersparnis große Bedeutung erreicht. Mischungen sind insbesondere bei Verlustsystemen intensiv untersucht worden ([5] und alle dort angegebenen Literaturstellen).

In dieser Arbeit sollen die im Vorangehenden beschriebenen Systemstrukturen zugrundegelegt werden. Sie können als "einstufige Bedienungssysteme mit Wartemöglichkeit" bzw. bei $g > 1$ als "Bedienungssysteme mit parallelem Warten" zu einem Oberbegriff zusammengefaßt werden.

Es sei an dieser Stelle noch erwähnt, daß es auch Bedienungssysteme gibt, bei denen ein Ruf mehrere Abfertigungsorgane nacheinander durchlaufen muß (serielles Warten), bei denen ein Ruf mehrere Male dasselbe Bedienungssystem durchlaufen muß (Bedienungssysteme mit Rückkopplung) bzw. bei denen ein Ruf eine mehrstufige Koppelanordnung mit konjugierter Durchschaltung (Linksystem) durchlaufen muß. Schließlich sei noch auf vermaschte Netze mit mehreren Knotenpunkten hingewiesen, bei denen ein Ruf über mehrere Koppelanordnungen entweder auf einmal (line-switching) oder stafettenweise (Speicher- vermittlungen, message-switching) durchgeschaltet wird.

2.4 Die Organisation des Bedienungssystems

2.4.1 Allgemeines

Unter der Systemorganisation soll allgemein die Art und Weise der Zuordnung von Warteplätzen bzw. Leitungen zu neu eintreffenden Rufen sowie deren Abfertigungsreihenfolge verstanden werden. Die Systemorganisation wird häufig auch als "Abfertigungsdisziplin" oder "Abfertigungspolitik" bezeichnet. Sie regelt allgemein die Prioritäten bei der Abfertigung von Rufen.

Zunächst lassen sich zwei Hauptgruppen von Prioritäten unterteilen:

- Unterbrechende (pre-emptive) Prioritäten
- Nichtunterbrechende (nonpre-emptive) Prioritäten.

Ein Ruf mit unterbrechender Priorität darf eine bestehende Belegung geringerer Priorität unterbrechen und die freiwerdende Leitung zur eigenen Abfertigung belegen.

Bei nichtunterbrechenden Prioritäten werden die wartenden Rufe im Wartespeicher entsprechend einer vorgegebenen Prioritätsregel für die Abfertigungsreihenfolge eingeordnet. Man unterscheidet bei Bedienungssystemen mit begrenztem Wartespeicher noch zwei Untergruppen:

- Verdrängende Priorität
- Nichtverdrängende Priorität.

Bei verdrängender Priorität wird ein neuer Ruf im Wartespeicher auf einen bevorrechtigten Platz gesetzt; er verschiebt somit einen Teil der wartenden Rufe auf ungünstigere Warteplätze. Ist jedoch bei Ankunft des neuen Rufes der Wartespeicher schon voll, so wird dadurch ein anderer wartender Ruf aus dem Wartespeicher verdrängt und geht verloren.

Bei nichtverdrängender Priorität wird einem neuen Ruf entsprechend seiner Wichtigkeit ein bevorrechtigter Warteplatz zugeteilt, solange noch freie Warteplätze verfügbar sind. Bei vollem Wartespeicher wird der neue Ruf jedoch abgewiesen und geht sofort verloren.

Den allgemeinsten Fall gemischter Prioritäten erhält man als Kombination von unterbrechenden mit nichtunterbrechenden Prioritäten, wobei die Unterbrechungsstrategie entweder von der Zeit, dem Zustand oder der Distanz der Prioritätsklasse eines momentan bedienten Rufes zur Klasse des neu eintreffenden Rufes abhängig ist.

In dieser Arbeit werden nur nichtunterbrechende Prioritäten vorausgesetzt; sobald ein Ruf von einer Leitung abgefertigt wird, so wird er auch vollständig abgefertigt.

2.4.2 Abfertigungsdisziplinen innerhalb der Warteschlangen

Nach den allgemeinen Bemerkungen werden nun spezielle Voraussetzungen über die Abfertigungsreihenfolge getroffen. Als Disziplinen innerhalb der Warteschlangen (queue disciplines) werden zugrundegelegt:

- D1 Abfertigung in der Reihenfolge des Eintreffens
- D2 Abfertigung in zufälliger Reihenfolge

- D3 Abfertigung in der inversen Reihenfolge des Eintreffens
- a) mit verdrängender Priorität
 - b) mit nichtverdrängender Priorität.

Bei D1, D2 und D3b geht ein neuer Ruf sofort verloren, wenn er seinen Wartespeicher voll belegt vorfindet. Bei D3a hingegen wird in diesem Falle der neue Ruf trotzdem vorrangig auf den 1. Warteplatz gesetzt; er kann jedoch durch nachfolgende Rufe nachträglich aus dem Wartespeicher verdrängt werden.

Es gibt noch sehr viele weitere Abfertigungsdisziplinen, welche jedoch nicht im Rahmen dieser Arbeit behandelt werden können. Eine sehr bekannte dieser Art ist jene, welche die Rufe in der inversen Reihenfolge zur erwarteten Bedienungszeit abfertigt, wodurch die Gesamtwartezeit bezüglich aller Rufe ein Minimum wird.

Häufig sind die Rufe eines Anrufprozesses in Klassen verschiedener Priorität unterteilt. Hierbei kann es sich um unterbrechende oder nichtunterbrechende Prioritäten handeln. Innerhalb der einzelnen Prioritätsklassen kann dabei jeweils noch nach verschiedenartigen Abfertigungsdisziplinen (z.B. D1 bis D3) abgefertigt werden. Kombinierte Warte-Verlustsysteme mit nichtunterbrechenden Prioritäten wurden von W. WAGNER [6] untersucht. Für unterbrechende Prioritäten wurde das Verlustsystem von L. KATZSCHNER [7], das kombinierte Warte-Verlustsystem von G. J. BRANDT [8] behandelt. Für Bedienungssysteme mit nur einer Leitung ist eine Vielzahl von Arbeiten über Warteschlangen mit Prioritäten bekannt geworden (siehe z.B. [9]).

2.4.3 Abfertigungsdisziplinen zwischen den Warteschlangen

Bei Systemen mit $g > 1$ Wartespeichern müssen noch weitere Voraussetzungen bezüglich der Auswahl der abzufertigenden nichtleeren Warteschlange im Falle des Freiwerdens einer bestimmten Leitung getroffen werden. Als mögliche Disziplinen zwischen den Warteschlangen (interqueue disciplines) seien folgende genannt:

- Zyklische Abfertigung von Warteschlangen
 - a) Es wird jeweils nur ein Wartender abgefertigt
 - b) Es wird jeweils die ganze Warteschlange abgefertigt
- Wahrscheinlichkeitsmäßige Abfertigung von Warteschlangen
 - a) Entsprechend einer Aufteilungsregel, die unabhängig von den Warteschlangenlängen ist
 - b) Entsprechend einer Aufteilungsregel, die von den momentanen Warteschlangenlängen abhängt

- Abfertigung einzelner Rufe innerhalb des von der freiwerdenden Leitung aus erreichbaren Gesamtwarteraumes nach Kriterien der Ankunftsreihenfolge (D1 bis D3).

Im weiteren Verlauf dieser Arbeit soll nur die wahrscheinlichkeitsmäßige Abfertigung von Warteschlangen vorausgesetzt werden. Endet die Belegung einer bestimmten Leitung, so wird unter allen diese Leitung erreichenden nichtleeren Wartespeichern (Warteschlangen) einer entsprechend einer Wahrscheinlichkeitsregel herausgesucht. Aus später noch genauer zu erläuternden Gründen soll in zwei Untergruppen unterteilt werden:

- A Wahrscheinlichkeitsmäßige Abfertigung nach einer Aufteilungsregel, welche unabhängig von den momentanen Warteschlangenlängen ist
- B Wahrscheinlichkeitsmäßige Abfertigung nach einer Aufteilungsregel, welche von den momentanen Warteschlangenlängen abhängt.

Als Aufteilungsregeln der Disziplin A können z.B. gelten:

- Aufteilungswahrscheinlichkeiten alle gleich groß
- Aufteilungswahrscheinlichkeiten entsprechend den mittleren Anrufraten
- Aufteilungswahrscheinlichkeiten nach Prioritätsregeln zwischen den Warteschlangen.

Als Aufteilungsregeln der Disziplin B seien beispielsweise genannt:

- Aufteilungswahrscheinlichkeiten proportional zu momentanen Warteschlangenlängen
- Aufteilungswahrscheinlichkeiten nach Prioritätsregeln, welche der momentan längsten Warteschlange Priorität geben.

Die Abfertigungsdisziplin zwischen den Schlangen wird allgemein bei unvollkommener Erreichbarkeit durch eine Matrix (p_{IJ}) angegeben, wobei p_{IJ} die Wahrscheinlichkeit angibt, mit der Warteschlange J abgefertigt wird, vorausgesetzt, daß die Belegung der Leitung I endet, $I=1,2,\dots,n$; $J=1,2,\dots,g$.

Im allgemeinen muß für jeden Systemzustand (d.h. für jedwede Kombination aus belegten Leitungen und Warteplätzen) eine Matrix (p_{IJ}) angegeben werden.

Im Falle vollkommener Erreichbarkeit vereinfachen sich die Matrizen (p_{IJ}) zu Spaltenvektoren (p_J) , wobei p_J die Wahrscheinlichkeit für das Abfertigen der Warteschlange J bedeutet, vorausgesetzt, daß eine beliebige Belegung des voll belegten Leitungsbündels endet, $J=1,2,\dots,g$.

Für die Wahrscheinlichkeiten p_{IJ} bzw. p_J gilt in jedem Zustand:

$$\sum_{J=1}^g p_{IJ} = 1, \quad I = 1, 2, \dots, n, \quad (2.12a)$$

bzw.

$$\sum_{J=1}^g p_J = 1. \quad (2.12b)$$

2.4.4 Absuchdisziplin der erreichbaren Leitungen

Die von einer Teilgruppe J aus erreichbaren k_J Leitungen sollen der Reihenfolge nach von oben nach unten belegt werden ("geordnetes Absuchen") bzw. unter den erreichbaren freien Leitungen soll eine beliebige Leitung ausgesucht werden ("zufällsmäßiges Absuchen"). Diese Voraussetzungen bedeuten keine Einschränkung der Allgemeinheit und können beliebig modifiziert werden.

2.5 Zusammenfassung der Fälle von Systemstrukturen und Abfertigungsdisziplinen

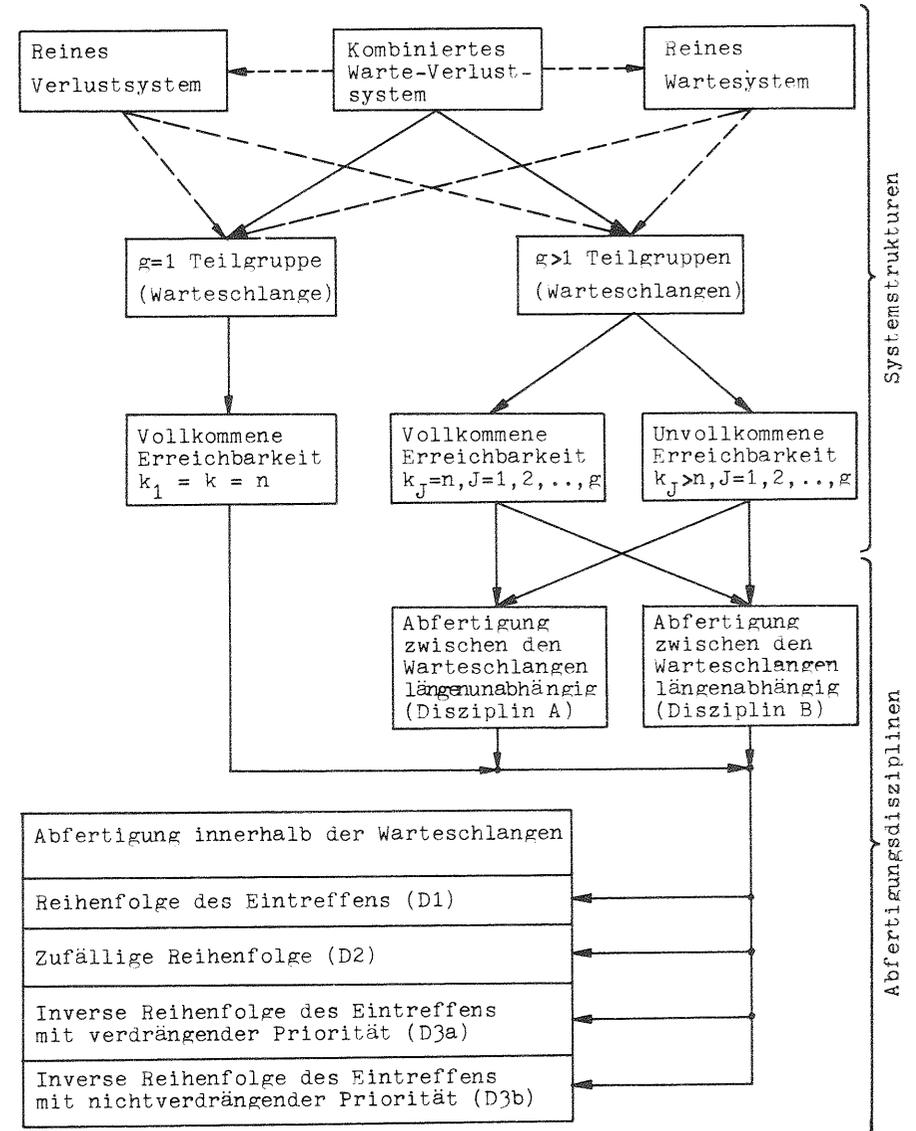
Nachfolgend seien die in der Arbeit zu untersuchenden Fälle von Systemstrukturen und Abfertigungsdisziplinen in übersichtlicher Weise zusammengestellt (siehe Seite 24).

Das reine Verlustsystem bzw. das reine Wartesystem folgen als Sonderfälle aus dem allgemeineren kombinierten Warte-Verlustsystem durch verschwindende bzw. unbeschränkte Zahl von Warteplätzen. Dabei ist zu beachten, daß beim reinen Verlustsystem die Unterscheidung der Abfertigungsdisziplinen allgemein und beim reinen Wartesystem die Unterscheidung bei inverser Abfertigungsreihenfolge in verdrängende und nichtverdrängende Priorität bedeutungslos werden.

Es werde ferner vorausgesetzt, daß wartende Rufe nicht vorzeitig das Warten aufgeben ("no defections") und daß abgewiesene bzw. verdrängte Rufe das System verlassen und keinen weiteren Einfluß auf das System haben ("lost calls cleared").

2.6 Güteerkmale und Kenngrößen

Nach Herleitung der allgemeinen Gleichungen zur exakten Beschreibung der stochastischen Prozesse wird eine Reihe von Kenngrößen ermittelt, welche für die Güte des Bedienungssystems nach außen kennzeichnend sind:



Fälle von Systemstrukturen und Abfertigungsdisziplinen (zu Abschnitt 2.5)

1. Stationäre Zustandswahrscheinlichkeiten
2. Wartewahrscheinlichkeiten
3. Verlustwahrscheinlichkeiten
4. Mittlere Zahl belegter Leitungen
5. Mittlere Warteschlangenlängen
6. Mittlere Wartezeiten
7. Wartezeitverteilungsfunktion
8. Momente der Wartezeitverteilungsfunktion
9. Antwortzeitverteilungsfunktion
10. Mittlere Antwortzeiten

Diese Kenngrößen sind Funktionen der

- Systemstruktur ($g, n, k_J, s_J; J=1,2,\dots,g$)
- Systemorganisation ($A, B; D1, D2, D3a, D3b$)
- Parameter der Anrufprozesse ($\lambda_J; J=1,2,\dots,g$)
- Parameter der Endprozesse ($\epsilon_I; I=1,2,\dots,n$)

und gestatten daher die Einflüsse dieser Parameter auf die Güte des Bedienungssystems zu beurteilen. Wie einleitend schon erwähnt wurde, ist i.a. eine explizite Optimierung aller Parameter nicht möglich (hierzu müßten außerdem Kostenfunktionen beispielsweise für Wartezeiten und Bedienungszeiten als auch Kosten des erforderlichen Aufwands berücksichtigt werden) und zumeist auch nicht erforderlich. Teilprobleme lassen sich jedoch mit Hilfe der Kenngrößen entweder explizit oder zumindest iterativ optimieren.

Die Kenngrößen 1. bis 6. lassen sich, unter obigen Voraussetzungen, über die Lösung eines linearen Gleichungssystems ermitteln. Die weitreichendste Aussage über das Systemverhalten erhält man jedoch aus der VF der Wartezeiten. Hierzu ist i.a. die Lösung größerer linearer Differentialgleichungssysteme erforderlich. Wenn dies nicht möglich oder zu umfangreich ist, so können die Momente der Wartezeitverteilungsfunktion zur Beurteilung herangezogen werden, welche wieder aus linearen Gleichungssystemen gewonnen werden. Mit Hilfe der Momente ist außerdem approximativ eine beliebig genaue Lösung für die Wartezeitverteilungsfunktion möglich (siehe Abschnitt 8.2.5).

In Rechneranwendungen ist häufig die VF der Antwortzeiten von Interesse. Die zufällige Antwortzeit setzt sich aus einer eventuellen zufälligen Wartezeit und der sich daran anschließenden Bedienungszeit zusammen. Sie kann unter der Voraussetzung der Unabhängigkeit zwischen zufälliger Wartezeit und Bedienungszeit, welche hier gegeben ist, durch eine Faltungsoperation berechnet werden.

Anknüpfend an die Voraussetzungen werden im 3. Kapitel der Arbeit zunächst einige Bereiche für die Anwendung der untersuchten Modelle aufgezeigt.

Im 4. Kapitel werden in knapper Form die theoretischen Grundlagen Markoff'scher Bedienungssysteme dargestellt. Es wird gezeigt, daß sich die wesentlichsten Kenngrößen aus der Untersuchung der Zufallsprozesse für Systemzustände und Wartezeiten ableiten lassen. Insbesondere wird hierbei auf eine methodisch einheitliche Konstruktion und Formulierung der Warteprozesse eingegangen, welche dazu geeignet ist, Bedienungssysteme mit mehreren Warteschlangen und Abfertigungsarten systematisch zu beschreiben.

In den Kapiteln 5 bis 7 werden für Bedienungssysteme mit einer Warteschlange, mehreren Warteschlangen und vollkommener bzw. unvollkommener Erreichbarkeit die Zufallsprozesse für Systemzustände und Wartezeiten für verschiedene Abfertigungsarten im einzelnen untersucht. Für Systeme mit $g = 1$ Warteschlange sind die Verhältnisse am übersichtlichsten und lassen sich daher weitestgehendst entwickeln. Die Beschreibung von Systemen mit allgemeiner Zahl g von Teilgruppen führt häufig auf unübersichtliche Darstellungen; in solchen Fällen werden wir uns auf $g = 2$ Teilgruppen beschränken, woran jedoch die Unterschiede von Systemen mit $g > 1$ Teilgruppen zu Systemen mit $g = 1$ Teilgruppe gezeigt werden können.

Das abschließende 8. Kapitel befaßt sich mit verschiedenen approximativen Berechnungsmethoden, welche für die praktische Auswertung der Theorie von Bedeutung sind.

Die Kapitel 5 bis 8 werden jeweils durch eine Reihe numerischer Beispiele abgeschlossen, welche den Einfluß von Systemstruktur und Abfertigungsart auf die Kenngrößen des Systems aufzeigen.

In der vorliegenden Arbeit wird überwiegend auf das Problem der Berechnung der Wartezeitverteilungsfunktionen eingegangen, worüber es in der Literatur nur in vereinzelt Fällen Lösungen gibt. Wartezeitverteilungsfunktionen bei Systemen mit unvollkommener Erreichbarkeit sind bislang noch nicht untersucht worden.

3. ANWENDUNGSBEREICHE

In diesem Abschnitt wird das Auftreten der zu untersuchenden Bedienungssysteme anhand typischer Beispiele aus der Wählvermittlungstechnik sowie von Rechnerstrukturen und Betriebssystemen aufgezeigt.

3.1 Wählvermittlungssysteme

3.1.1 Koppelanordnungen im Verbindungsweg

In Wählvermittlungen für Ferngespräche und Fernschreiben ist prinzipiell die Zahl der internen Verbindungswege kleiner als die Zahl der Teilnehmer. Bei der Verkehrskonzentration als auch bei der Verteilung des Verkehrs auf verschiedene Richtungen treten Engpässe auf. In den meisten Realisierungen arbeiten die Koppelanordnungen als Verlustsysteme: ist im Falle des Verbindungswunsches keine erreichbare Leitung frei, so geht der Ruf verloren. Lässt man nun die Möglichkeit des Wartens zu, so können die Leitungen noch besser ausgenutzt werden.

In Bild 2 ist eine typische Koppelanordnung mit unvollkommener Erreichbarkeit gezeigt. Vor den erreichbaren Leitungen einer Teilgruppe ist jeweils ein begrenzter Wartespeicher vorgesehen. Koppelanordnungen der Wählvermittlungstechnik zeichnen sich i.a. durch Symmetrie in Struktur und Verkehrsparametern aus, siehe Bild 2. Sie wurden bereits näherungsweise für unendlich große [10] und endlich große Wartespeicher [11] untersucht.

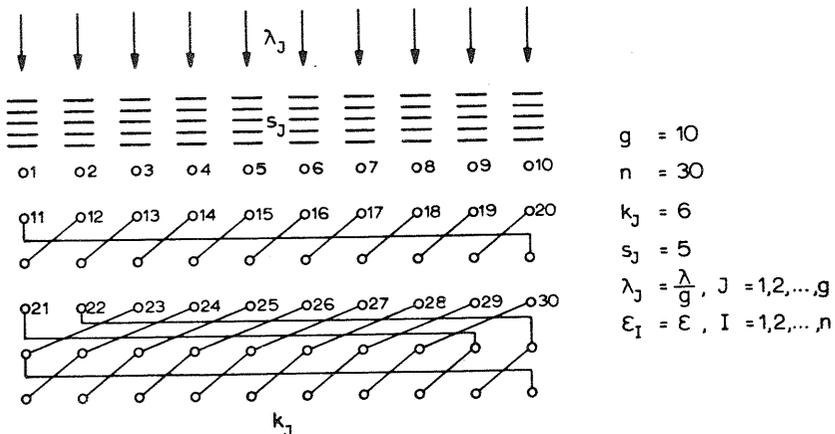


Bild 2. Koppelanordnung der Wählvermittlungstechnik mit Wartespeichern

3.1.2 Anschaltung zentralisierter Baugruppen

Moderne Wählvermittlungssysteme zeichnen sich gegenüber klassischen Lösungen durch weiter zentralisierte elektronisch arbeitende Baugruppen aus, welche nur sehr kurze Zeit zur Aufnahme bzw. Verarbeitung von Steuerinformationen benötigt werden. Als Beispiele seien Register zur Aufnahme von Wählinformationen, Zuordner zur Umwertung von Wählinformationen, zentrale Prozessoren zur Ermittlung der Einstellinformationen für die Koppelnetze sowie Markierer für die Einstellung der Koppelnetze selbst genannt. Um derartige teure Baugruppen gut auszunutzen, stellt man - je nach Belegungszeit - nur wenige im Verhältnis zur Zahl innerer Verbindungswege zur Verfügung. Die Anschaltung erfolgt über Koppelnetze entweder im Raum- oder im Zeitmultiplex.

Als Beispiel ist in Bild 3 die Anschaltung zentralisierter Register R ("Leitungen") an die Verbindungssätze VS der inneren Verbindungswege ("Rufe" und zugleich "Warteplätze") angedeutet.

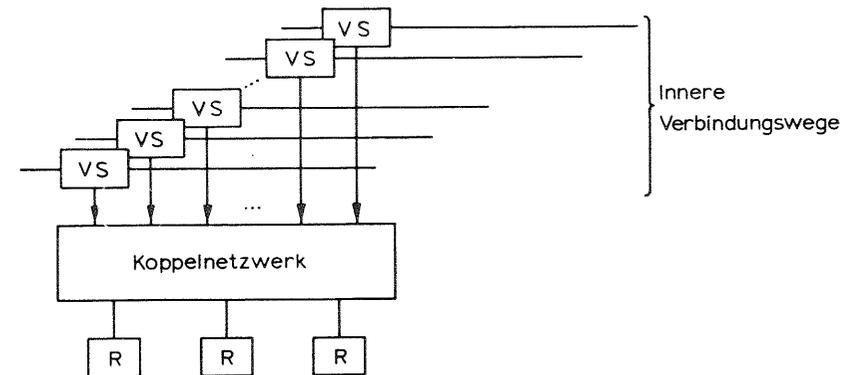


Bild 3. Anschaltung zentralisierter Register

Je nach Ausführung des Koppelnetzwerkes (eine oder mehrere Zubringerteilgruppen, vollkommene oder unvollkommene Erreichbarkeit, Linksystem) kommen verschiedene Bedienungssystemtypen in Betracht.

3.2 Rechnerstrukturen und Betriebssysteme

3.2.1 "Closed Shop"-Betrieb im Rechenzentrum

In Rechenzentren, die vielen Benutzern zur Verfügung stehen, werden die einzelnen Programme ("Rufe") z.B. in der Reihenfolge des Eintreffens in einer Bänderinheit gepuffert ("Warteschlange"), bis sie von dem Rechner ("Leitung") verarbeitet werden. Da ein Programm nach dem anderen von einem "Stapel" abgearbeitet wird, spricht man auch von "Stapelverarbeitung" (batch processing). Zur Berechnung der Wartezeiten kann das Modell (a) aus Bild 1 benutzt werden.

3.2.2 Prozeßsteuerung und Meßwertverarbeitung

Moderne Prozeßrechner gestatten die simultane Steuerung mehrerer Prozesse bzw. die simultane Verarbeitung von Meßwerten mehrerer Vorgänge. Meßwertänderungen können beispielsweise in regelloser Folge eintreffen und werden bis zur Verarbeitung gespeichert. Diese gespeicherten "Rufe" bilden je nach Herkunft Warteschlangen, welche nach einer Prioritätsregel abgefertigt werden. Zur theoretischen Behandlung solcher Probleme eignet sich das Modell (b) aus Bild 1.

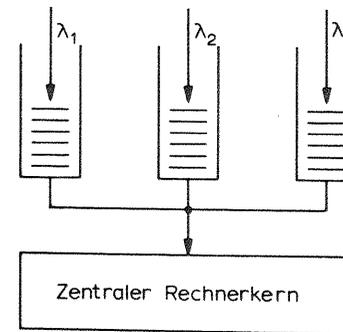
3.2.3 Rechnerkernvergabe in Realzeit- und Großrechenanlagen

Rechenanlagen für Realzeitbetrieb (z.B. Dialogverkehr) müssen extremen Anforderungen bezüglich der Antwortzeiten gerecht werden. Dies wird erreicht, indem wichtigere Anfragen höhere Prioritäten zur vorrangigen Abarbeitung erhalten. Auf Änderungen der Anfrangeprioritäten müssen solche Systeme sehr schnell reagieren, u.U. durch Unterbrechung gerade bearbeiteter Aufträge geringerer Priorität.

Moderne Großrechenanlagen arbeiten heute ausnahmslos nach dem Prinzip der Vielfachprogrammierung (multiprogramming). Hiermit wird erreicht, daß mehrere Anfragen gleichzeitig von der Rechenanlage bearbeitet werden. Während der Rechnerkern beispielsweise gerade eine Anfrage i bearbeitet, kann die Eingabe/Ausgabe (E/A)-Einheit parallel hierzu entweder die Ergebnisse der Anfrage i-1 ausspeichern oder eine neue Anfrage i+1 einspeichern.

Die Organisation derartiger Aufgaben obliegt sog. Betriebsprogrammen höchster Priorität (Betriebssystem oder operating system). In Bild 4

ist an einem Beispiel angedeutet, wie vom Betriebssystem einzelne Aktivitäten prioritätsgerecht in mehreren Listen (Warteschlangen) zusammengefaßt werden. Der zentrale Rechnerkern (central processor unit CPU) bedient die Warteschlangen entsprechend einer vorgegebenen Abfertigungsdisziplin (scheduling routine). Die Prioritäten nehmen in diesem Beispiel von 1 nach 3 ab.



Bedeutungen:

- 1 Realzeit-Anfragen
- 2 Durch E/A-Aktionen zu peripheren Speichern unterbrochene Programme
- 3 Stapelverarbeitung von "background"-Programmen

Bild 4. Rechnerkernvergabe durch das Betriebssystem

3.2.4 Vermaschte Rechnersysteme

In Auskunftssystemen, wie z.B. das weltweite Netz zur Platzreservierung bei Flügen, treten Rechner über ein vermaschtes Datenübertragungsnetz miteinander in Verbindung. In den Knotenpunkten des Netzes treten infolge beschränkter Übertragungs-Bitraten Warteschlangen vor den einzelnen Datenfernübertragungsleitungen auf. Für die Berechnung der zu erwartenden Warte- und Antwortzeiten können Modelle nach Bild 1 herangezogen werden.

Reicht die vorhandene Rechnerkapazität zur Überbrückung eines Engpasses nicht aus oder sollen Rechner gegen Totalausfall geschützt werden, so kommt man zum Prinzip der Parallelverarbeitung (parallel processing, load sharing). In Bild 5 ist ein einfaches Beispiel angegeben, wie man zwei einzelne Rechner C_1 und C_2 gegen Totalausfall der Bedienung schützen kann, indem beispielsweise der Anschluß an einen fernen schnellen Rechner C_3 (remote computer) zur Aushilfe vorgesehen wird. Die Berechnung solcher Strukturen kann mit Hilfe von Bedienungssystemen vollkommener und unvollkommener Erreichbarkeit vorgenommen werden.

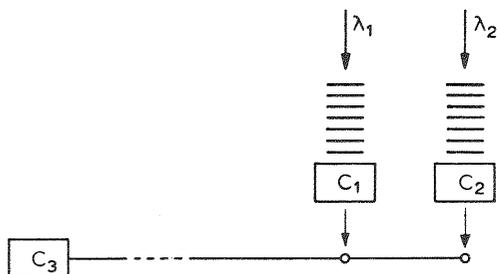


Bild 5. Parallelverarbeitung zur Entlastung und Sicherung gegen Ausfall

Abschließende Bemerkungen

1. Bei der Untersuchung realer Systeme werden nicht immer die Voraussetzungen über Anruf- und Endeprozeß exakt erfüllt sein. Sieht man von VF extrem großer Varianz (z.B. hyperexponentielle Verteilung) ab, so erhält man unter MARKOFF'schen Voraussetzungen meistens zu pessimistische Ergebnisse; die Systeme arbeiten dann in Wirklichkeit besser als vorausberechnet.
2. Die Voraussetzung der wahrscheinlichkeitmäßigen Abfertigung der verschiedenen Warteschlangen ist gegenüber den anderen Abfertigungsarten sehr flexibel in der Anpassung an die unterschiedlichen Arbeitsweisen und erfordert im Einzelfalle keine neue Berechnungsmethode.
3. Innerhalb der Warteschlangen wird in vielen Fällen die "gerechte" Abfertigungsdisziplin D1 verwirklicht sein. Oft bedingt die technische Realisierung jedoch mehr eine zufallsmäßige Abfertigung (D2). In Prozeßsteuerungen kann es vorkommen, daß die Abfertigung in der inversen Reihenfolge des Eintreffens (D3) erfolgt, da die zuletzt eingetroffene Nachricht die aktuellste Information enthält.

Für jedwede Abfertigungsdisziplin, welche nicht die VF der Anzahl von Wartenden in der Warteschlange beeinflusst, haben J.F.C. KINGMAN [12] bzw. D.G. TAMBOURATZIS [13] gezeigt, daß die Varianz der VF der Wartezeiten für D1 ein Minimum bzw. für D3 ein Maximum annimmt. Mit der Behandlung dieser zwei Fälle haben wir also gleichzeitig die Grenzfälle erfaßt.

4. EINFÜHRUNG IN DIE THEORIE MARKOFF'SCHER BEDIENTUNGSSYSTEME

In diesem Abschnitt soll in kurzer Form die Theorie Markoff'scher Bedienungssysteme dargestellt werden. Die Ausführungen gelten zunächst für ein Bedienungssystem mit n vollkommen erreichbaren Leitungen und einem Wartespeicher mit endlich vielen Warteplätzen entsprechend Bild 1(a). Über die Systemorganisation wird lediglich vorausgesetzt, daß nichtunterbrechende Prioritäten (verdrängende und nichtverdrängende) zugelassen sind. Erweiterungen dieser Theorie auf Systeme mit mehreren Warteschlangen und vollkommen bzw. unvollkommen erreichbaren Leitungen werden abschließend diskutiert.

4.1 Zustände und Übergänge

4.1.1 Zustände als Zufallsprozeß

Der Zustand des Systems zur Zeit t ist durch das zufällige Muster $\xi(t)$ aus Belegungen von Leitungen und Warteplätzen gekennzeichnet. $\xi(t)$ ist i.a. ein Vektor, gebildet aus mehreren Zustandskomponenten. Die Menge aller Zustände sei Z.

Im einfachsten Falle kann das System durch einen eindimensionalen Vektor beschrieben werden, welcher die Zahl der im System weilenden Rufe zur Zeit t angibt; es ist dann $\xi(t)=i, i=0,1,\dots,n+s_1$, wobei Z aus $n+s_1+1$ Zuständen besteht.

Der Zustand ξ des Systems besteht so lange, bis er durch einen einfallenden Ruf oder eine endende Belegung in einen anderen Zustand übergeht. Die zeitliche Folge der (diskontinuierlichen) Zufallsvariablen ξ bildet den Zufallsprozeß

$$\{ \xi(t), t \geq 0 \}, \xi \in Z,$$

(kurz " $\xi(t)$ -Prozeß").

4.1.2 Die Markoff'sche Eigenschaft

Der $\xi(t)$ -Prozeß werde an k aufeinanderfolgenden Zeitpunkten t_1, t_2, \dots, t_k betrachtet, wobei $0 \leq t_1 < t_2 < \dots < t_k, k \geq 2$. Der Prozeß besitzt dann die Markoff'sche Eigenschaft, wenn

$$P\{ \xi(t_k) = \xi_k | \xi(t_1) = \xi_1, \xi(t_2) = \xi_2, \dots, \xi(t_{k-1}) = \xi_{k-1} \} = P\{ \xi(t_k) = \xi_k | \xi(t_{k-1}) = \xi_{k-1} \} \quad (4.1)$$

gilt [14, 15]. Gl.(4.1) besagt, daß der Verlauf des $\xi(t)$ -Prozesses nur vom gegenwärtigen Zustand abhängig ist, nicht jedoch davon, wie dieser Zustand zustandekam. Der Markoff'sche Zufallsprozeß besitzt kein "Gedächtnis"; er ist der einfachste aller denkbaren Zufallsprozesse und wird oft auch als "Markoffkette" bezeichnet.

4.1.3 Übergangs- und Zustandswahrscheinlichkeiten

Der Prozeß befinde sich zu einem Zeitpunkt $u \geq 0$ im Zustand i . Die Wahrscheinlichkeit dafür, daß der Prozeß zu einem späteren Zeitpunkt $u+t$, $t > 0$, den Zustand j annimmt, vorausgesetzt, daß er sich zum Zeitpunkt u im Zustand i befunden hat, heißt bedingte Übergangswahrscheinlichkeit

$$P(t; j|i) = P\{\xi(u+t)=j | \xi(u)=i\}, \quad u \geq 0, t > 0, i, j \in Z. \quad (4.2)$$

Der Prozeß wird außerdem als "homogen" bezeichnet, da die bedingte Übergangswahrscheinlichkeit $P(t; j|i)$ nicht vom Ausgangszeitpunkt u abhängt.

Die absolute Wahrscheinlichkeit für das Auftreten eines Zustands wird als Zustandswahrscheinlichkeit bezeichnet und ist wie folgt definiert:

$$P(t; i) = P\{\xi(t)=i\}, \quad t \geq 0, i \in Z. \quad (4.3)$$

Ist die Anfangsverteilung der Zustände $P(0; i), i \in Z$, bekannt, so lassen sich die Zustandswahrscheinlichkeiten $P(t; i), i \in Z$, mit Hilfe der bedingten Übergangswahrscheinlichkeiten ausdrücken:

$$P(t; j) = \sum_{i \in Z} P(0; i) \cdot P(t; j|i), \quad t \geq 0, j \in Z. \quad (4.4)$$

Da sich die Zustände gegenseitig ausschließen und ein vollständiges Ereignissystem bilden, gelten folgende Vollständigkeitsrelationen:

$$\sum_{j \in Z} P(t; j|i) = 1, \quad t \geq 0, \quad (4.5a)$$

$$\sum_{j \in Z} P(t; j) = 1, \quad t \geq 0. \quad (4.5b)$$

Den Zusammenhang zwischen den bedingten Übergangswahrscheinlichkeiten erhält man, indem zwei aufeinanderfolgende Zeitabschnitte der Längen h und t betrachtet werden. Zu Beginn befinde sich der Prozeß

im Zustand i ; er entwickelt sich während der Spanne h in einen Zustand k und während der anschließenden Spanne t in den Zustand j :

$$P(h+t; j|i) = \sum_{k \in Z} P(h; k|i) \cdot P(t; j|k), \quad h, t > 0. \quad (4.6a)$$

Gl.(4.6a) wird auch CHAPMAN-KOLMOGOROFF-Gleichung genannt.

Ist $\mathbf{P}(t) = (P(t; j|i))$ die Matrix der bedingten Übergangswahrscheinlichkeiten ("Übergangsmatrix" oder "Stochastische Matrix"), so läßt sich Gl.(4.6a) einfach in Form des Matrizenprodukts darstellen:

$$\mathbf{P}(h+t) = \mathbf{P}(h) \cdot \mathbf{P}(t), \quad h, t > 0. \quad (4.6b)$$

Als bedingte Übergangswahrscheinlichkeitsdichten werden definiert:

$$\lim_{h \rightarrow 0} \frac{P(h; j|i)}{h} = q_{ij}, \quad i, j \in Z, i \neq j, \quad (4.7a)$$

$$\lim_{h \rightarrow 0} \frac{1 - P(h; i|i)}{h} = q_i, \quad i \in Z. \quad (4.7b)$$

Die Koeffizienten q_{ij} geben die Wahrscheinlichkeitsdichte an, mit welcher der $\xi(t)$ -Prozeß vom Zustand i in den Zustand j wechselt; sie heißen auch "Intensitäten", "Übergangskoeffizienten" oder "Sprungkoeffizienten". Der Koeffizient q_i ist die Wahrscheinlichkeitsdichte für die augenblickliche Änderung des Zustandes i , und es gilt wegen Gl.(4.5a)

$$q_i = \sum_{j \neq i} q_{ij} \quad (4.7c)$$

Die Übergangswahrscheinlichkeitsdichten werden in der Matrix

$$\mathbf{Q} = (-q_{ij}) \quad (4.8)$$

zusammengefaßt, wobei $q_{ii} = -q_i$. Die Koeffizienten der Matrix \mathbf{Q} lassen sich für jede Systemstruktur aus den mittleren Anrufraten, den mittleren Enderaten und der Systemorganisation (Absuch- und Abfertigungsdisziplinen) bestimmen. Insbesondere folgt aus (4.7a) und (4.7b) für die Übergangswahrscheinlichkeiten im Intervall $(t, t+dt)$:

$$P(dt; j|i) = q_{ij} dt + o(dt), \quad i, j \in Z, i \neq j, \quad (4.9a)$$

$$1 - P(dt; i|i) = q_i dt + o(dt), \quad i \in Z. \quad (4.9b)$$

4.1.4 Die Kolmogoroff'schen Gleichungen

Aus der CHAPMAN-KOLMOGOROFF-Gleichung (4.6a) lassen sich zwei verschiedene Differentialgleichungen ableiten, indem einmal der Differentialquotient für $t \rightarrow 0$, zum anderen für $h \rightarrow 0$ gebildet wird:

d/dh P(h, j|i) = -q_j P(h, j|i) + sum_{k != j} q_kj P(h, k|i), ij in Z. (4.10a) (KOLMOGOROFF-VORWÄRTS-Gleichung)

d/dt P(t, j|i) = -q_i P(t, j|i) + sum_{k != i} q_ik P(t, j|k), ij in Z. (4.11a) (KOLMOGOROFF-RÜCKWÄRTS-Gleichung)

In Matrixschreibweise erhält man hierfür

d/dh P(h) = -P(h)Q, (4.10b)

d/dt P(t) = -Q P(t). (4.11b)

Diese zwei Differentialgleichungssysteme sind grundlegend für die Behandlung Markoff'scher Bedienungssysteme.

Mit Hilfe der Definition für die Zustandswahrscheinlichkeiten Gl. (4.4) läßt sich aus der KOLMOGOROFF-VORWÄRTS-Gleichung (4.10a) eine entsprechende Differentialgleichung für die Zustandswahrscheinlichkeiten ableiten:

d/dh P(h, j) = -q_j P(h, j) + sum_{k != j} q_kj P(h, k), j in Z. (4.12)

Die KOLMOGOROFF-RÜCKWÄRTS-Gleichung eignet sich übrigens nicht zur Bestimmung der Zustandswahrscheinlichkeiten. Sie wird jedoch später bei der Bestimmung von Wartezeitverteilungsfunktionen zentrale Bedeutung haben (siehe Abschnitt 4.3).

4.1.5 Der stationäre Fall

Für die meisten praktischen Anwendungen ist das Verhalten des Bedienungssystems im Falle einer längere Zeit andauernden Spitzenbelastung interessant. Es ist aus vielen Untersuchungen bekannt, daß der Verkehr in diesen Zeiten (sog. "Hauptverkehrsstunde") als stationär angenommen werden kann. Die allgemein hergeleiteten

Merkmale zur Beschreibung des Markoff'schen Prozesses werden nun für den "eingeschwungenen Fall" (d.h. stationären Fall) angegeben. In diesem Falle gilt:

lim_{t -> infinity} P(t, j|i) = p(j), ij in Z, (4.13)

d.h. die stationäre Zustandswahrscheinlichkeit p(j) ist unabhängig von der Zeit wie auch vom Ausgangszustand i.

Wird Gl.(4.13) auf Gl.(4.12) angewendet, so erhält man

q_j p(j) = sum_{k != j} q_kj p(k), j in Z. (4.14)

Gl.(4.14) wird auch als Gleichung für das "Statistische Gleichgewicht" bezeichnet und ist sehr anschaulich zu deuten: q_j p(j) ist die Wahrscheinlichkeitsdichte für das Verschwinden des bestehenden Zustands j; sie ist gleich der Wahrscheinlichkeitsdichte für das Entstehen des Zustands j aus allen Zuständen k != j, jeweils mit der Wahrscheinlichkeitsdichte q_kj p(k). (Ist kein direkter Übergang von einem Zustand k != j in den Zustand j möglich, so ist der betreffende Übergangskoeffizient q_kj=0).

Gl.(4.14) ist ein lineares homogenes Gleichungssystem, bestehend aus sovielen Gleichungen wie es Zustände gibt und mit verschwindender Systemdeterminante. Es darf eine beliebige Gleichung weggelassen und durch die aus Gl.(4.5b) resultierende "Normalisierungsbedingung"

sum_{j in Z} p(j) = 1 (4.15)

ersetzt werden. Das Gleichungssystem (4.14) und (4.15) ist dann inhomogen und eindeutig auflösbar.

Im weiteren Verlauf der Arbeit soll durchweg nur der stationäre Fall betrachtet werden.

4.2 Charakteristische Größen

In den meisten Fällen sind die stationären Zustandswahrscheinlichkeiten allein zur Beurteilung der Abfertigungsgüte nicht geeignet (z.B. bei sehr großen Zustandszahlen). Man leitet daher "charakteristische Größen" aus den Zustandswahrscheinlichkeiten ab. Im folgenden werden für das System mit einer begrenzten Warteschlange

einige Definitionen angegeben.

Die Zustände seien durch die Zahl der Rufe im System definiert, $\xi(t)=i, i=0,1,\dots,n+s_1$.

Die Wartewahrscheinlichkeit W ist die Wahrscheinlichkeit, mit der ein eintreffender Ruf warten muß (Wahrscheinlichkeit für Wartebeginn). Dies ist immer dann der Fall, wenn ein Zustand angetroffen wird, welcher weder sofortige Abfertigung noch sofortiges Verlorengehen bedeutet.

$$W = \sum_{i=n}^{\nu} p(i) \quad , \quad \text{wobei } \nu = \begin{cases} n+s_1-1 & \text{für D1,D2,D3b} \\ n+s_1 & \text{für D3a.} \end{cases} \quad (4.16)$$

Die Verlustwahrscheinlichkeit B ist die Wahrscheinlichkeit, mit der ein eintreffender Ruf bei Ankunft verlorengeht. Es ist

$$B = \begin{cases} p(n+s_1) & \text{für D1,D2,D3b} \\ 0 & \text{für D3a.} \end{cases} \quad (4.17)$$

Die Erfolgs-Wartewahrscheinlichkeit W* ist die Wahrscheinlichkeit, mit der ein eintreffender Ruf warten muß und seine Wartezeit erfolgreich, d.h. mit der Belegung einer Leitung, beendet.

$$W^* = \begin{cases} W & \text{für D1,D2,D3b} \\ \sum_{i=n}^{n+s_1-1} p(i) & \text{für D3a.} \end{cases} \quad (4.18)$$

Die Mißerfolgs-Wartewahrscheinlichkeit W** ist die Wahrscheinlichkeit, mit der ein eintreffender Ruf warten muß und seine Wartezeit erfolglos, d.h. durch Verdrängtwerden, beendet; sie wird auch als Verdrängungswahrscheinlichkeit bezeichnet.

$$W^{**} = \begin{cases} 0 & \text{für D1,D2,D3b} \\ p(n+s_1) & \text{für D3a.} \end{cases} \quad (4.19)$$

Zwischen den Größen W, W* und W** gilt stets die Beziehung

$$W = W^* + W^{**} \quad (4.20)$$

Der Verkehrswert Y ist die mittlere Zahl belegter Leitungen (Erwartungswert)

$$Y = \sum_{i=0}^{n-1} i \cdot p(i) + n \cdot \sum_{i=n}^{n+s_1} p(i) \quad (4.21)$$

Der Warteverkehrswert Ω (auch Wartebelastung genannt) ist die mittlere Warteschlangenlänge und berechnet sich als Erwartungswert belegter Warteplätze:

$$\Omega = \sum_{i=n}^{n+s_1} (i-n) \cdot p(i) \quad (4.22)$$

Die mittlere Wartezeit t_W bezogen auf alle Rufe, die bei Ankunft warten müssen (ohne Ansehen, ob sie erfolgreich oder erfolglos warten), läßt sich aus der Bilanz

$$\Omega = (\lambda \cdot W) \cdot t_W \quad (4.23)$$

berechnen. Die mittlere Warteschlangenlänge ist dem Produkt gleich, gebildet aus der mittleren Zahl von Rufen, welche pro Zeiteinheit einfallen und warten müssen, $\lambda \cdot W$, und der mittleren Aufenthaltsdauer im Wartespeicher t_W . Aus Gl.(4.23) folgt

$$t_W = \frac{\Omega}{\lambda \cdot W} \quad (4.24)$$

Die VF der Wartezeiten und ihre Momente, mittlere Wartezeiten erfolgreich bzw. erfolglos wartender Rufe und die VF der Antwortzeiten sind Gegenstand des nächsten Abschnittes.

4.3 warte- und Antwortzeiten

4.3.1 Definitionen

Ein beliebiger Ruf kommt in das System und beginnt zu warten. Es können drei Typen von Zufallsvariablen der Wartezeiten (waiting times) unterschieden werden:

- Wartezeit T_W = Aufenthaltsdauer im Wartespeicher
- Erfolgreiche Wartezeit T_W^* = Zeitintervall zwischen Ankunft und Bedienungsbeginn
- Erfolglose Wartezeit T_W^{**} = Zeitintervall zwischen Ankunft und Verdrängung.

Als zufällige Aufenthaltsdauer erfolgreicher Rufe im System wird definiert:

- Antwortzeit T_R = Zeitintervall zwischen Ankunft und Bedienungsende.

Die zufällige Antwortzeit T_R (response time) eines eintreffenden Rufes setzt sich aus einer zufälligen erfolgreichen Wartezeit T_W^* und

der sich daran anschließenden Bedienungsdauer T_S (service time) zusammen:

$$T_R = \begin{cases} T_S & \text{falls der Ruf sofort bedient wird} \\ T_W^* + T_S & \text{falls der Ruf warten muß.} \end{cases} \quad (4.25)$$

4.3.2 Der Warteprozeß

Zur Bestimmung der Wartezeit wird ein sog. "Testruf" betrachtet. Der Testruf komme zu einem beliebigen Zeitpunkt in das System und beginnt zu warten. Es wird durch Betrachtung aller möglichen Ereignisse das Schicksal des Testrufs im Wartespeicher verfolgt.

R. SYSKI [15] hat für Markoff'sche Bedienungssysteme ohne unterbrechende bzw. verdrängende Disziplinen eine allgemeine Theorie der Wartezeiten entwickelt, indem die Phase des Wartens des Testrufs im Wartespeicher mit Hilfe eines aus dem $\xi(t)$ -Prozeß konstruierten sog. "modifizierten Prozesses" beschrieben wird. Diese Methode des modifizierten Prozesses soll im folgenden angewendet werden. Der modifizierte Prozeß heiße kurz "Warteprozeß".

Der Testruf kommt zu einem beliebigen Zeitpunkt ($t=0$) in das System und startet einen speziellen Warteprozeß. Dieser Warteprozeß endet, wenn der Testruf entweder abgefertigt oder verdrängt wird. Die Zufallsvariable $\xi(t)$ des Warteprozesses $\{\xi(t), t \geq 0\}$ gibt das Muster aller jener Rufe an, welche zur Zeit t im System sind und die Wartezeit des Testrufs beeinflussen können (der Testruf selbst ist in diesem Muster *n i c h t* enthalten).

$\xi(t)$ ist i.a. ein Vektor, welcher aus mehreren Zustandskomponenten zusammengesetzt ist. Die spezielle Definition von $\xi(t)$ hängt von der Systemstruktur und der Systemorganisation ab.

Im einfachsten Falle ist $\xi(t)$ wieder ein eindimensionaler Vektor, welcher eine Zahl von beeinflussenden Rufen angibt. Als Beispiel werde das System mit einer Warteschlange bei der Disziplin D1 betrachtet. Befindet sich der Testruf nach der Wartezeit t auf dem i -ten Warteplatz, so ist $\xi(t) = n+1-i$ die Zahl derjenigen Rufe, welche die Wartezeit des Testrufs noch beeinflusst, gebildet aus den n Rufen, welche gerade abgefertigt werden und den $(i-1)$ Rufen, welche vor dem Testruf in der Warteschlange warten und mit

Sicherheit *v o r* dem Testruf abgefertigt werden. Im Gegensatz dazu haben alle Rufe, welche *n a c h* dem Testruf angekommen sind und sich hinter dem Testruf in der Warteschlange eingeordnet haben, keinen Einfluß auf dessen Wartezeit; sie sind deshalb in $\xi(t)$ nicht enthalten.

Mehrdimensionale Vektoren $\xi(t)$ erhält man bei komplizierter strukturierten bzw. organisierten Systemen. Betrachten wir als Beispiel wieder das einfache System mit einer begrenzten Warteschlange und der Disziplin D3b. Ein Testruf treffe bei Ankunft z Wartende an und werde während einer Wartezeit t durch nachfolgende Rufe auf den i -ten Warteplatz zurückgeschoben. Nun ist $\xi(t) = (n+1-i, z)$ ein zweidimensionaler Vektor: die Komponente $n+1-i$ gibt die Zahl der Rufe an, welche gerade abgefertigt werden bzw. in der Warteschlange *v o r* dem Testruf warten (direkt beeinflussende Rufe). Die bei Ankunft angetroffenen z Wartenden können erst *n a c h* dem Testruf abgefertigt werden; sie haben jedoch auch einen Einfluß auf die Wartezeit des Testrufs, da sie das Abweisen nachfolgender Rufe mitbestimmen, wenn die Warteschlange voll ist (indirekt beeinflussende Rufe). Mehrdimensionale Vektoren $\xi(t)$ gibt es vor allem bei Systemen mit mehreren Warteschlangen.

Allgemein können die Rufe, welche die Wartezeit des Testrufs beeinflussen, in zwei Kategorien eingeteilt werden:

- Direkt beeinflussende Rufe
Dies sind Rufe, welche entweder gerade abgefertigt werden oder warten und *v o r* dem Testruf abgefertigt werden bzw. *v o r* dem Testruf abgefertigt werden können ("Konkurrenten", z.B. bei D2).
- Indirekt beeinflussende Rufe
Dies sind Rufe, welche entweder mit Sicherheit erst *n a c h* dem Testruf abgefertigt werden oder auf den Testruf keine Konkurrenz bezüglich einer Leitung ausüben, deren Anwesenheit im System jedoch das Schicksal des Testrufs mitbestimmt (z.B. bei D3b).

Der $\xi(t)$ -Prozeß kann aus dem $\xi(t)$ -Prozeß konstruiert werden, indem nur diejenigen Übergänge des $\xi(t)$ -Prozesses berücksichtigt werden, welche einen Einfluß auf die Wartezeit des betrachteten Testrufs haben. Da der $\xi(t)$ -Prozeß die Markoff'sche Eigenschaft

besitzt und die Übergänge des $\xi(t)$ -Prozesses eine Untermenge der Übergänge des $\xi(t)$ -Prozesses sind, besitzt auch der $\xi(t)$ -Prozeß die Markoff'sche Eigenschaft.

Der $\xi(t)$ -Prozeß startet nur dann, wenn der Testruf bei seiner Ankunft weder sofort bedient noch abgewiesen wird. Er endet entweder durch Abfertigen (Bedienungsbeginn) oder Verdrängen des Testrufs. Alle Zustände, welche das Enden des $\xi(t)$ -Prozesses bedeuten (sie "absorbieren" den Testruf), werden in einer "Tabu-Menge" H zusammengefaßt. Die "Lebenszeit" des $\xi(t)$ -Prozesses dauert also so lange an, wie seine Zustände außerhalb der Tabu-Menge H bleiben. Die Wartezeit des Testrufs kann auf diese Weise als "Zeit für den ersten Eintritt in die Tabu-Menge H " aufgefaßt werden.

Für die Übergänge des $\xi(t)$ -Prozesses werden nun - in entsprechender Weise wie für den $\xi(t)$ -Prozeß - bedingte Übergangswahrscheinlichkeiten und Sprungkoeffizienten definiert (die Zustände i, j beziehen sich nun auf den $\xi(t)$ -Prozeß).

${}_H P(t; j|i)$ sei die bedingte Übergangswahrscheinlichkeit dafür, daß der $\xi(t)$ -Prozeß in der Zeit t vom Ausgangszustand i unter Meidung der Zustände der Tabu-Menge H in den Zustand j übergeht:

$${}_H P(t; j|i) = P\{\xi(u) \notin H, 0 < u < t, \xi(t) = j | \xi(0) = i\}, \quad i, j \notin H. \quad (4.26)$$

Gehören i bzw. j oder beide der Menge H an, so ist ${}_H P(t; j|i) = 0$. Eine Vollständigkeitsrelation für die bedingten Übergangswahrscheinlichkeiten - entsprechend Gl.(4.5a) für den $\xi(t)$ -Prozeß - gilt für den $\xi(t)$ -Prozeß $n \text{ i c h t}$. Da der $\xi(t)$ -Prozeß während der Zeit t enden kann, gilt

$$\sum_{j \notin H} {}_H P(t; j|i) < 1, \quad t \geq 0, \quad i \notin H. \quad (4.27)$$

Die Vollständigkeitsrelation kann jedoch wieder hergestellt werden, wenn die Bedingung $j \notin H$ fallengelassen wird.

Entsprechend Gl.(4.7a) und (4.7b) erhalten wir die Sprungkoeffizienten für den $\xi(t)$ -Prozeß:

$$\lim_{t \rightarrow 0} \frac{{}_H P(t; j|i)}{t} = q(i, j), \quad i \neq j, \quad i, j \notin H, \quad (4.28a)$$

$$\lim_{t \rightarrow 0} \frac{1 - {}_H P(t; i|i)}{t} = q(i), \quad i \notin H. \quad (4.28b)$$

Bedeutet $\varepsilon(i)$ die bedingte Übergangswahrscheinlichkeitsdichte für das Enden des $\xi(t)$ -Prozesses vom Zustand i aus, so gilt die Vollständigkeitsrelation

$$q(i) = \sum_{\substack{j \neq i \\ j \notin H}} q(i, j) + \varepsilon(i), \quad i \notin H. \quad (4.28c)$$

Da der $\xi(t)$ -Prozeß die Markoff'sche Eigenschaft besitzt, gehorchen die bedingten Übergangswahrscheinlichkeiten wieder den beiden Kolmogoroff'schen Gleichungen. Für die Bestimmung der Wartezeiten ist insbesondere die KOLMOGOROFF-Rückwärts-Gleichung interessant:

$$\frac{d}{dt} {}_H P(t; j|i) = -q(i) {}_H P(t; j|i) + \sum_{\substack{k \neq i \\ k \notin H}} q(i, k) {}_H P(t; j|k), \quad i, j \notin H. \quad (4.29)$$

4.3.3 Bedingte Wartezeitverteilungsfunktionen

a) Bedingte Wartezeitverteilungsfunktionen aller wartenden Rufe
Ein Ruf kommt in das System und treffe den Zustand $\xi(0)=i, i \notin H$, an. Mit welcher Wahrscheinlichkeit $w(t|i)$ wartet er länger als die Zeit t ($t \geq 0$)? Die Wahrscheinlichkeit $w(t|i)$ heißt bedingte (komplementäre) Wartezeitverteilungsfunktion und ist wie folgt definiert:

$$w(t|i) = P\{T_W > t | \xi(0) = i\}, \quad i \notin H. \quad (4.30)$$

Es wird nun der Zusammenhang zwischen den bedingten Wartezeitverteilungsfunktionen und den bedingten Übergangswahrscheinlichkeiten des $\xi(t)$ -Prozesses hergestellt. Wie weiter oben schon ausgeführt wurde, kann die Wartezeit eines Testrufs als "Zeit für den ersten Eintritt in die Tabu-Menge H " aufgefaßt werden. Der Testruf wartet länger als die Zeit t , solange der $\xi(t)$ -Prozeß beliebige Zustände außerhalb H annimmt, d.h. mit Gl.(4.26)

$$\begin{aligned} w(t|i) &= P\{\xi(u) \notin H, 0 < u < t | \xi(0) = i\} \\ &= \sum_{j \notin H} P\{\xi(u) \notin H, 0 < u < t, \xi(t) = j | \xi(0) = i\} \\ &= \sum_{j \notin H} {}_H P(t; j|i). \end{aligned} \quad (4.31)$$

Durch Summation von Gl.(4.29) entsprechend Gl.(4.31) erhält man die KOLMOGOROFF-Rückwärts-Gleichung für die bedingten Wartezeitverteilungsfunktionen:

$$\frac{d}{dt}w(t|i) = -q(i)w(t|i) + \sum_{\substack{k \neq i \\ k \in H}} q(i,k)w(t|k), \quad i \in H. \quad (4.32)$$

Gemäß Gl.(4.32) sind die bedingten Wartezeitverteilungsfunktionen aus einem System gekoppelter linearer Differentialgleichungen 1. Ordnung zu bestimmen. Zur eindeutigen Auflösung müssen jedoch noch die Anfangsbedingungen angegeben werden.

Betrachten wir hierzu zunächst nur die bedingten Wartezeitverteilungsfunktionen der Rufe, welche nach ihrer Ankunft warten, ohne Rücksicht auf das spätere Schicksal (Abfertigung bzw. Verdrängung). Aus der Definitionsgleichung (4.30) entnehmen wir für $t = 0$, daß $w(0|i)$, $i \in H$, identisch ist mit der Wahrscheinlichkeit des Auftretens einer positiven Wartezeit. Wenn jedoch der Ruf warten muß, was vorausgesetzt wird ($i \in H$), so tritt eine positive Wartezeit mit der Wahrscheinlichkeit 1 auf:

$$w(0|i) = 1, \quad i \in H. \quad (4.33)$$

Als nächstes werde der Grenzwert

$$\lim_{t \rightarrow 0+} \frac{dw(t|i)}{dt}$$

untersucht. Dieser Grenzwert wird aus

$$\frac{w(dt|i) - w(0|i)}{dt}$$

erzeugt für $dt \rightarrow 0$.

Mit

$$w(dt|i) - w(0|i) = P\{T_W > dt | \xi(0) = i\} - P\{T_W > 0 | \xi(0) = i\} = -P\{0 < T_W \leq dt | \xi(0) = i\}$$

folgt

$$\lim_{t \rightarrow 0+} \frac{dw(t|i)}{dt} = -E(i), \quad (4.34)$$

wobei $E(i)$ die (bedingte) Übergangswahrscheinlichkeitsdichte für das Ende des $\xi(t)$ -Prozesses vom Zustand i aus ist. Wird nun Gl.(4.32) an der Stelle $t = 0+$ betrachtet, wobei $w(0|i)$ nach Gl.(4.33) eingesetzt wird, so ergibt sich gerade die Vollständigkeitsrelation Gl.(4.28c).

b) Bedingte Wartezeitverteilungsfunktionen erfolgreich bzw. erfolglos wartender Rufe

Bei Abfertigungsdisziplinen, welche Verdrängung aus dem Wartespeicher zulassen, endet die zufällige Wartezeit eines Testrufs entweder erfolgreich (Bedienungsbeginn) oder erfolglos (Verdrängung). Es werde nun nach den bedingten Wartezeitverteilungsfunktionen der erfolgreich (*) bzw. erfolglos (**) wartenden Rufe gefragt:

$$w^*(t|i) = P\{T_W^* > t | \xi(0) = i\}, \quad i \in H, \quad (4.35a)$$

bzw.

$$w^{**}(t|i) = P\{T_W^{**} > t | \xi(0) = i\}, \quad i \in H, \quad (4.35b)$$

Durch Betrachtung der Warteprozesse für die erfolgreich bzw. erfolglos wartenden Rufe ergibt sich, daß für $w^*(t|i)$ bzw. $w^{**}(t|i)$, $i \in H$, jeweils dasselbe Differentialgleichungssystem gilt wie für $w(t|i)$, Gl.(4.32). Zwischen $w(t|i)$, $w^*(t|i)$ und $w^{**}(t|i)$ gilt die Beziehung

$$w(t|i) = w^*(t|i) + w^{**}(t|i), \quad i \in H. \quad (4.36)$$

Die drei in ihrer Koeffizientenstruktur gleichen Differentialgleichungssysteme unterscheiden sich in den Anfangsbedingungen $w(0|i)$, $w^*(0|i)$ und $w^{**}(0|i)$, $i \in H$. Für $w(0|i)$ gilt stets Gl.(4.33). Für $w^*(0|i)$ bzw. $w^{**}(0|i)$ kann das Differentialgleichungssystem an der Stelle $t = 0+$ benutzt werden. Hierzu sind die Grenzwerte

$$\lim_{t \rightarrow 0+} \frac{dw^*(t|i)}{dt} = -E^*(i), \quad (4.37a)$$

bzw.

$$\lim_{t \rightarrow 0+} \frac{dw^{**}(t|i)}{dt} = -E^{**}(i), \quad i \in H, \quad (4.37b)$$

erforderlich. $E^*(i)$ bzw. $E^{**}(i)$ bedeuten die bedingten Übergangswahrscheinlichkeitsdichten für das Ende des jeweiligen Warteprozesses vom Zustand i aus. Man gewinnt diese Größen unmittelbar aus dem Ablauf des Warteprozesses für die erfolgreichen bzw. erfolglosen Rufe unter Beachtung, daß erfolgreiche Rufe nur durch Bedienungsbeginn, erfolglose Rufe nur durch Verdrängung den Warteprozeß beenden.

Die Anfangsbedingungen $w^*(0|i)$ bzw. $w^{**}(0|i)$, $i \in H$, bestimmen sich also jeweils aus einem linearen Gleichungssystem. Wegen Gl. (4.33) und (4.36) gilt $w^*(0|i) + w^{**}(0|i) = 1$, $i \in H$; es braucht aus diesem Grunde jedoch nur eines der beiden linearen Gleichungssysteme aufgelöst zu werden. Die Anfangsbedingungen $w^*(0|i)$ bzw. $w^{**}(0|i)$ sind gleichzeitig auch die Wahrscheinlichkeiten für Erfolg bzw. Mißerfolg vom Startzustand $\xi(0) = i$ aus.

Abschließend sei noch auf ein praktisch sehr einfaches Beweisverfahren für die KOLMOGOROFF-Rückwärts-Gleichung der bedingten Wartezeitverteilungsfunktionen Gl.(4.32) hingewiesen, welches auf C. PALM [16] zurückgeht.

Ein Testruf kommt in das System und trifft den Zustand $\xi(0) = i$, $i \notin H$, an. Es wird das differentielle Intervall der Länge dt nach der Ankunft betrachtet. Es gibt zwei Möglichkeiten im Intervall dt :

1. Es fällt ein neuer (beeinflussender) Ruf ein oder es endet eine der Belegungen.

Der Zustand $\xi(0) = i$ des Warteprozesses geht hierdurch in einen Zustand $\xi(dt) = k$ mit der Wahrscheinlichkeit $q(i,k)dt$ über, $k \neq i$, $k \notin H$. Von dort aus wartet der Testruf länger als die Zeit $(t-dt)$ mit der Wahrscheinlichkeit $w(t-dt|k)$.

2. Es ereignet sich nichts.

Der angetroffene Zustand $\xi(0) = i$ bleibt erhalten, sodaß $\xi(dt) = i$. Es ereignet sich nichts im Intervall der Länge dt mit der Wahrscheinlichkeit $1-q(i)dt$. Der Testruf wartet sodann vom Zustand $\xi(dt) = i$ aus länger als die Zeit $(t-dt)$ mit der Wahrscheinlichkeit $w(t-dt|i)$.

Der gesamte Zeitintervall ist in zwei aufeinanderfolgende Abschnitte zerlegt, die Spanne dt und die Restzeit $(t-dt)$. Wie bei der Ableitung der CHAPMAN-KOLMOGOROFF-Gleichung (4.6a) können wir deshalb ansetzen:

$$w(t|i) = \sum_{\substack{k \neq i \\ k \notin H}} q(i,k)dt \cdot w(t-dt|k) + [1-q(i)dt] \cdot w(t-dt|i), \quad i \notin H,$$

oder

$$\frac{w(t|i) - w(t-dt|i)}{dt} = -q(i)w(t|i) + \sum_{\substack{k \neq i \\ k \notin H}} q(i,k)w(t-dt|k), \quad i \notin H.$$

Durch Grenzübergang $dt \rightarrow 0$ folgt hieraus sofort Gl.(4.32).

4.3.4 Absolute Wartezeitverteilungsfunktionen

Die absolute VF der Wartezeiten bezüglich aller Rufe ohne Rücksicht auf den angetroffenen Zustand,

$$W(>t) = P\{T_w > t\}, \quad (4.38)$$

erhält man durch Mittelbildung über die bedingten Wartezeitverteilungsfunktionen aller antreffbaren Zustände $i \notin H$, verbunden mit den "Antreffwahrscheinlichkeiten" $P(i)$

$$W(>t) = \sum_{i \notin H} P(i)w(t|i), \quad (4.39)$$

wobei

$$P(i) = P\{\xi(0) = i\}, \quad i \notin H. \quad (4.40)$$

Die Antreffwahrscheinlichkeiten $P(i)$ können aus den stationären Zustandswahrscheinlichkeiten gewonnen werden.

Gl.(4.39) gilt in entsprechender Weise auch für erfolgreich bzw. erfolglos wartende Rufe:

$$W^*(>t) = \sum_{i \notin H} P(i)w^*(t|i), \quad (4.41a)$$

bzw.

$$W^{**}(>t) = \sum_{i \notin H} P(i)w^{**}(t|i) \quad (4.41b)$$

Insbesondere erhält man für $t = 0$ die Wartewahrscheinlichkeiten W , W^* bzw. W^{**} :

$$W(>0) = W, \quad (4.42a)$$

$$W^*(>0) = W^*, \quad (4.42b)$$

$$W^{**}(>0) = W^{**}. \quad (4.42c)$$

Zwischen den Größen W , W^* und W^{**} gilt außerdem die Beziehung Gl.(4.20).

Die VF der Wartezeiten bezogen auf die wartenden Rufe, erfolgreich wartenden Rufe bzw. erfolglos wartenden Rufe sind dann durch $W(>t)/W$, $W^*(>t)/W^*$ bzw. $W^{**}(>t)/W^{**}$ gegeben. Diese (normierten) VF nehmen für $t = 0$ den Wert 1 an.

In den folgenden Abschnitten wird auf die prinzipiellen mathematischen Methoden zur Behandlung der Differentialgleichungssysteme für die bedingten Wartezeitverteilungsfunktionen bzw. der linearen Gleichungssysteme für die entsprechenden bedingten Momente eingegangen.

4.3.5 Laplace-Transformierte der Wartezeitverteilungsfunktionen und Eigenwertproblem

Zur weiteren Behandlung des Differentialgleichungssystems (4.32) wird die Laplace-Transformierte der bedingten Wartezeitverteilungsfunktion eingeführt [17]:

$$W(s|i) = \int_{t=0}^{\infty} w(t|i) \exp(-st) dt \quad (4.43)$$

wobei $s = \sigma + j\omega$ die komplexe Bildvariable bedeutet. Wird die Laplace-Transformation auf Gl.(4.32) angewendet, so erhält man

$$[s+q(i)]W(s|i) - \sum_{\substack{k \neq i \\ k \in H}} q(i,k)W(s|k) = w(0|i), \quad i \notin H. \quad (4.44)$$

Es sei vermerkt, daß Gl.(4.44) auch für erfolgreich bzw. erfolglos wartende Rufe gilt, wenn alle Wahrscheinlichkeiten bzw. deren Transformierte mit * bzw. ** versehen werden.

Die Sprungkoeffizienten des $\xi(t)$ -Prozesses werden in der Matrix

$$A = (-q(i,k)) \quad (4.45)$$

zusammengefaßt, wobei $-q(i,i) = q(i)$. Die Koeffizienten des Gleichungssystems (4.44) lauten dann in Matrixschreibweise

$$K = A + sI, \quad (4.46)$$

wobei I die Einheitsmatrix bedeutet.

Als Lösung des Systems (4.44) erhält man gebrochene rationale Funktionen in s , wobei das Nennerpolynom mit $\det(A + sI)$ identisch ist. Durch Partialbruchentwicklung folgt

$$W(s|i) = \sum_{\nu=1}^{N'} \sum_{\kappa=1}^{K_{\nu}} \frac{a_{i\nu}^{(\kappa)}}{(s-e_{\nu})^{\kappa}}, \quad i \notin H, \quad (4.47)$$

wobei e_{ν} die Nullstelle der Vielfachheit K_{ν} bedeutet, $\nu=1,2,\dots,N'$. Die Größe N' kennzeichnet die Anzahl verschiedener Nullstellen, während N den Rang des Systems (4.44) angibt. $a_{i\nu}^{(\kappa)}$ sind die Residuen der Partialbruchentwicklung, $\nu=1,2,\dots,N'$; $\kappa=1,2,\dots,K_{\nu}$, $i \notin H$.

Zwischen N , N' und K_{ν} gilt außerdem die Beziehung

$$\sum_{\nu=1}^{N'} K_{\nu} = N.$$

Die Laplace-Rücktransformation [17] von $W(s|i)$ nach Gl.(4.47) liefert den gesuchten Ausdruck für die bedingte Wartezeitverteilungsfunktion

$$w(t|i) = \sum_{\nu=1}^{N'} \exp(e_{\nu}t) \sum_{\kappa=1}^{K_{\nu}} a_{i\nu}^{(\kappa)} \cdot \frac{t^{\kappa-1}}{(\kappa-1)!}, \quad i \notin H. \quad (4.48)$$

Mit Gl.(4.39), (4.40) und (4.48) ist die Wartezeitverteilungsfunktion damit prinzipiell bekannt.

Die Hauptschwierigkeit bei der Bestimmung der Wartezeitverteilungsfunktion entsteht bei der Bestimmung der Nullstellen (Eigenwerte) von

$$\det(A+sI) = 0, \quad (4.49)$$

welche auch bei der sog. Eigenwertaufgabe [18] auftritt: Man finde einen Vektor X derart, daß

$$AX + sX = 0. \quad (4.50)$$

Gl.(4.50) hat dann und nur dann nichttriviale Lösungen, wenn Gl.(4.49) erfüllt ist.

Liegt die Matrix der Sprungkoeffizienten Gl.(4.45) des $\xi(t)$ -Prozesses vor, so kann das Wartezeitproblem bei Markoff'schen Bedienungssystemen auf ein Eigenwertproblem reduziert werden. Die Untersuchung der Wartezeiten beginnt deshalb zunächst mit der Untersuchung der Eigenwerte. Es gibt hierfür Kriterien, welche von der "charakteristischen Gleichung" (4.49) ausgehen (HURWITZ, STURM). In dieser Arbeit wird die Kenntnis der Koeffizienten des "charakteristischen Polynoms" $\det(A + sI)$ jedoch nicht vorausgesetzt; es werden nur Kriterien benutzt, welche von den Elementen der Matrix A ausgehen.

4.3.6 Momente der Wartezeitverteilungsfunktionen

In vielen Fällen ist die exakte Berechnung der Wartezeitverteilungsfunktion mit sehr großem Rechenaufwand verbunden. Zur Beurteilung

der Eigenschaften des Warteprozesses werden dann die Momente der Wartezeitverteilungsfunktion herangezogen.

Es sei $h(t|i)$ die Dichte der bedingten (nichtkomplementären) Wartezeitverteilungsfunktion $1-w(t|i)$, d.h.

$$h(t|i) = -\frac{d}{dt} w(t|i). \quad (4.51)$$

Das K-te Moment der bedingten Wartezeitverteilungsfunktion ist definiert als

$$M_K(i) = \int_{t=0}^{\infty} t^K h(t|i) dt. \quad (4.52)$$

Ist $H(s|i)$ die Laplace-Transformierte von $h(t|i)$, so gilt die Korrespondenz [17]

$$(-1)^K t^K h(t|i) \circ \longrightarrow \frac{d^K}{ds^K} H(s|i). \quad (4.53)$$

Aus der Definition der Laplace-Transformation sowie Gl.(4.52) und (4.53) folgt

$$M_K(i) = \lim_{s \rightarrow 0} (-1)^K \frac{d^K}{ds^K} H(s|i). \quad (4.54)$$

Wird Gl.(4.32) nach t differenziert und die Laplace-Transformation angewendet, so folgt

$$[s+q(i)]H(s|i) - \sum_{\substack{k \neq i \\ k \in H}} q(i,k)H(s|k) = \varepsilon(i), \quad i \in H. \quad (4.55)$$

Sukzessive Differentiation von Gl.(4.55) nach s und Grenzübergang $s \rightarrow 0$ führen mit Gl.(4.54) auf folgende Rekursionsgleichungen für die bedingten Momente

$$q(i)M_K(i) - \sum_{\substack{k \neq i \\ k \in H}} q(i,k)M_K(k) = KM_{K-1}(i), \quad i \in H, \quad K=1,2,\dots \quad (4.56)$$

wobei $M_0(i) = w(0|i)$.

Für erfolgreich bzw. erfolglos wartende Rufe gilt das Gl.(4.56) entsprechende System, wenn $M_K(i)$ durch $M_K^*(i)$ bzw. $M_K^{**}(i)$ ersetzt wird.

Aus Gl.(4.56) ist zu ersehen, daß für jedes $K=1,2,\dots$ ein lineares Gleichungssystem aufzulösen ist. Die Lösungen für das $(K-1)$ -te System erscheinen als "rechte Seiten" im K -ten Gleichungssystem, womit eine rekursive Auflösung bezüglich K möglich ist.

Die absoluten Momente der Wartezeitverteilungsfunktion ergeben sich mit Gl.(4.39) zu

$$M_K = \sum_{i \in H} P(i)M_K(i), \quad K=1,2,\dots \quad (4.57)$$

Es sei an dieser Stelle noch vermerkt, daß die Momente für die approximative Lösung der Wartezeitverteilungsfunktion mit Vorteil angewendet werden können, worauf wir in 8.2.5 noch zurückkommen.

4.3.7 Mittlere Wartezeiten

Die mittlere Wartezeit bezüglich aller bei Ankunft wartenden Rufe, t_W , wurde bereits mit Hilfe der mittleren Warteschlangenlänge in Gl.(4.23) berechnet. Mit dieser Methode ist es jedoch nicht möglich, bedingte mittlere Wartezeiten oder mittlere Wartezeiten erfolgreich bzw. erfolglos wartender Rufe anzugeben, wenn verdrängende Prioritäten zugelassen werden.

Für die (bedingten) mittleren Wartezeiten für Warten vom Startzustand $\xi(0) = 1$ aus gilt

$$\begin{aligned} t_W(i) &= \int_{t=0}^{\infty} t^1 h(t|i) dt = M_1(i) \\ &= \int_{t=0}^{\infty} w(t|i) dt = W(0|i), \quad i \in H. \end{aligned} \quad (4.58)$$

Die bedingten mittleren Wartezeiten $t_W(i)$, $i \in H$, errechnen sich also aus einem linearen Gleichungssystem (entweder Gl.(4.56) für $K=1$ oder Gl.(4.44) für $s=0$).

Gl.(4.58) gilt entsprechend auch für erfolgreich bzw. erfolglos wartende Rufe. In jedem Falle hat wegen Gl.(4.36) die Beziehung

$$t_W(i) = t_W^*(i) + t_W^{**}(i), \quad i \in H, \quad (4.59)$$

Gültigkeit.

Die absolute mittlere Wartezeit bezüglich a l l e r wartenden Rufe, t_W (bzw. t_W^* , t_W^{**}), wird durch Integration der Wartezeitverteilungsfunktion der wartenden Rufe, $W(>t)/W$ (bzw. $W^*(>t)/W^*$, $W^{**}(>t)/W^{**}$), gewonnen. Mit Gl.(4.39) und (4.58) folgt

$$t_W = \frac{1}{W} \cdot \sum_{i \in H} P(i) t_{W(i)}. \quad (4.60)$$

Ein Vergleich von Gl.(4.57) mit Gl.(4.60) lehrt, daß

$$M_y = W \cdot t_W. \quad (4.61)$$

4.3.8 Antwortzeiten

Die Antwortzeit T_R beschreibt die erfolgreiche Aufenthaltszeit eines eintreffenden Rufes im System. Der eintreffende Ruf kann folgende Schicksale erleiden:

Schicksal	Antwortzeit	Wahrscheinlichkeit
Bedienung ohne Warten	T_S	$1-B-W$
Verlust	0	B
Erfolgreiches Warten	$T_W^* + T_S$	W^*
Erfolgloses Warten	0	W^{**}

Die Verteilungsfunktion der Antwortzeiten eines beliebigen Rufes

$$R(>t) = P\{T_R > t\} \quad (4.62)$$

stellt sich demnach dar als

$$R(>t) = (1-B-W) \cdot S(>t) + W^* \cdot P\{T_W^* + T_S > t\} + (B+W^{**}) \cdot 0 \quad (4.63)$$

In Gl.(4.63) ist als einzige unbekannté Größe noch die VF einer Summe von Zufallsvariablen zu bestimmen. Nach einem Satz der Wahrscheinlichkeitstheorie ist die Wahrscheinlichkeitsdichtefunktion einer Summe von Zufallsvariablen gleich der Faltung der Wahrscheinlichkeitsdichtefunktionen der einzelnen Zufallsvariablen. Hierbei muß vorausgesetzt werden, daß die Zufallsvariablen voneinander unabhängig sind.

Mit
$$d_{W^*}(t) = \frac{d}{dt} \frac{W^*(\leq t)}{W^*} = - \frac{d}{dt} \frac{W^*(>t)}{W^*} \quad (4.64a)$$

und
$$d_S(t) = \frac{dS(\leq t)}{dt} = \varepsilon \cdot \exp(-\varepsilon t) \quad (4.64b)$$

folgt zunächst für die Dichte von $P\{T_W^* + T_S \leq t\}$

$$\frac{d}{dt} P\{T_W^* + T_S \leq t\} = d_{W^*}(t) * d_S(t) = \int_{\tau=0}^{\infty} d_{W^*}(\tau) \cdot d_S(t-\tau) d\tau \quad (4.65)$$

und aus Gl.(4.65)

$$P\{T_W^* + T_S > t\} = 1 - \int_{\eta=0}^t \int_{\tau=0}^{\infty} d_{W^*}(\tau) \cdot d_S(\eta-\tau) d\eta d\tau. \quad (4.66)$$

Mit der Kenntnis der Verteilungsfunktionen der erfolgreichen Wartezeiten und der Belegungsauern kann nach Gl.(4.63) mit Gl.(4.66) die Verteilungsfunktion der Antwortzeiten berechnet werden (Dieses Verfahren läßt sich allerdings nur dann anwenden, wenn alle Leitungen dieselbe Belegungsauerverteilung besitzen).

Der Faltungsoperation zweier Zeitfunktionen im Zeitbereich Gl.(4.65) entspricht im Bildbereich die Multiplikation der entsprechenden Laplace-Transformierten. Auf diese Weise gelangt man zu dem Ergebnis

$$P\{T_W^* + T_S \leq t\} = \mathcal{L}^{-1} \left\{ W^*(s) \cdot \frac{\varepsilon}{s+\varepsilon} \right\}, \quad (4.67)$$

wobei $W^*(s) = \mathcal{L}\{W^*(\leq t)/W^*\}$.

Die mittlere Antwortzeit bezüglich aller erfolgreichen Rufe, t_R , erhält man durch Integration der Antwortzeitverteilungsfunktion der erfolgreichen Rufe $R(>t)/R(>0)$ zu

$$t_R = t_S + \frac{W^*}{1-B-W^{**}} \cdot t_W^*. \quad (4.68)$$

(t_S =mittlere Belegungsauer, t_W^* =mittlere Wartezeit der erfolgreich wartenden, $R(>0) = 1-B-W^{**} =$ Erfolgswahrscheinlichkeit).

4.4 Markoff'sche Bedienungssysteme mit mehreren Eingangswarteschlangen

Im Vorangehenden wurde die Theorie der Markoff'schen Bedienungssysteme für den Fall einer einzigen Warteschlange entsprechend Bild 1(a) dargelegt. Es sollen nun die Erweiterungen auf die Fälle $g > 1$ Warteschlangen entsprechend Bild 1(b) und 1(c) diskutiert werden. Der Allgemeinheit wegen werde das Bedienungssystem mit unvollkommen erreichbaren Leitungen entsprechend Bild 1(c) zugrundegelegt.

4.4.1 Zustände als Zufallsprozeß

Die Zustände des Systems müssen durch ein hinreichend differenziertes Muster aus Belegungen von Leitungen und Wartepätzen charakterisiert werden, damit eine exakte Beschreibung des $\xi(t)$ -Prozesses möglich ist. "Hinreichend differenziert" bedeutet hierbei, daß - mit Rücksicht auf Struktur und Organisation des Systems - die Matrix Q der Übergangswahrscheinlichkeitsdichten des $\xi(t)$ -Prozesses in eindeutiger Weise angegeben werden kann.

Die Zustände des Systems werden durch einen $(n+g)$ -dimensionalen Vektor gekennzeichnet:

$$\xi(t) = (x_1, \dots, x_I, \dots, x_n; z_1, \dots, z_J, \dots, z_g), \quad (4.69)$$

wobei $x_I \in [0,1]$, je nachdem Leitung I frei oder belegt ist, und $z_J \in [0, s_J]$ die Zahl belegter Wartepätze in der Warteschlange J angibt, $I = 1, 2, \dots, n$, $J = 1, 2, \dots, g$.

Die Darstellungen von 4.1 haben für beliebig vieldimensionale Zustände Gültigkeit. Die Definition der charakteristischen Größen folgt prinzipiell den Ausführungen in 4.2; es muß jedoch zwischen Kenngrößen unterschieden werden, welche sich auf das gesamte System bzw. nur auf eine spezielle Warteschlange beziehen.

4.4.2 Der Warteprozeß

Entsprechend den g verschiedenen Anrufprozessen bzw. Warteschlangen gibt es g verschiedene Warteprozesse. Zur allgemeinen Betrachtung wird ein Testruf des Anrufprozesses J ("J-Testruf") angenommen; der Warteprozeß bezüglich des J-Testrufs heiÙe $\xi_J(t)$ -Prozeß, $J = 1, 2, \dots, g$ (Ohne Einschränkung der Allgemeinheit kann später $J = 1$ gesetzt werden).

Bild 6a zeigt die Verhältnisse der J-ten Teilgruppe. Der J-Testruf wird vom J-ten Anrufprozeß mit der mittleren Anrufrate λ_J erzeugt. J-Rufen ist ein Warteraum von insgesamt s_J Wartepätzen zugeordnet. Es werden $k_J < n$ Leitungen zur Abfertigung erreicht, die zugehörigen mittleren Enderaten sind $\epsilon_1, \epsilon_2, \dots, \epsilon_{k_J}$ (Bei vollkommener Erreichbarkeit ist $k_J = n$).

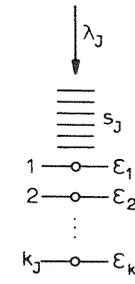


Bild 6a. Zur Definition des Warteprozesses bezüglich eines J-Testrufs

Trifft ein J-Testruf ein, so startet der $\xi_J(t)$ -Prozeß nur dann, wenn mindestens alle k_J Leitungen belegt sind und wenn der J-Testruf nicht sofort abgewiesen wird. Die k_J Rufe, welche gerade abgefertigt werden, haben in jedem Falle einen direkten Einfluß auf die Wartezeit des J-Testrufs. Aus den wartenden Rufen innerhalb der J-ten Warteschlange wird - nach Maßgabe der Abfertigungsdisziplinen innerhalb u n d zwischen den Warteschlangen - ein Muster gebildet, welches nur diejenigen Rufe enthält, die einen direkten oder indirekten Einfluß auf die Wartezeit des J-Testrufs ausüben. Der J-Testruf ist dabei in diesem Muster nicht enthalten.

Unter den k_J erreichbaren Leitungen gibt es k'_J ($0 < k'_J \leq k_J$) Leitungen, welche außerdem von Rufen anderer Anrufprozesse erreicht werden (Für $k'_J = 0$ zerfällt das Bedienungssystem in zwei unabhängige Teilsysteme; dieser triviale Fall sei durch $k'_J > 0$ ausgeschlossen). Hierdurch wird die Wartezeit des J-Testrufs auch von Rufen aus jenen g_J Teilgruppen, welche mit Ausnahme der Teilgruppe J noch eine oder mehrere der k'_J Leitungen erreichen, direkt beeinflusst. Über die restlichen $(n - k_J)$ Leitungen sind i.a. auch noch indirekte Beeinflussungen der Wartezeit des J-Testrufs durch Rufe der anderen Teilgruppen möglich.

Beispiel:

Die Konstruktion der Zustände des $\xi_J(t)$ -Prozesses soll im folgenden Beispiel veranschaulicht werden. Gegeben sei ein Bedienungssystem mit unvollkommen erreichbaren Leitungen entsprechend Bild 6b.

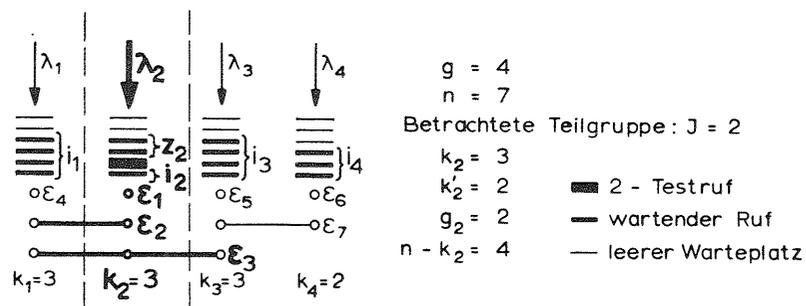


Bild 6b. Zur Konstruktion des $\xi_J(t)$ -Prozesses bei Bedienungssystemen mit mehreren Warteschlangen

Es werde der Warteprozess bezüglich der Teilgruppe $J = 2$ betrachtet. Die Abfertigungsdisziplin innerhalb der 2. Warteschlange sei D1, zwischen den Warteschlangen werde nach Disziplin B mit Wahrscheinlichkeiten abgefertigt, welche proportional zu den momentanen Warteschlangenlängen sind.

Der $\xi_2(t)$ -Prozess existiert überhaupt nur, wenn die Leitungen 1, 2 und 3 belegt sind, d.h. wenn $x_1=x_2=x_3=1$. Wir brauchen den Belegungszustand dieser drei Leitungen deshalb nicht extra im Zustandsmuster zu kennzeichnen, wir müssen aber den Einfluß dieser Belegungen bei den Sprungkoeffizienten berücksichtigen. Die Leitungen 4, 5, 6 und 7 können ein beliebiges Belegungsmuster (x_4, x_5, x_6, x_7) bilden; in den Wartespeichern 1, 3 bzw. 4 können sich jedoch nur dann Warteschlangen aufbauen, wenn jeweils die erreichbaren Leitungen der zugehörigen Teilgruppe belegt sind.

Wartende Rufe der 1. und 3. Warteschlange (i_1 bzw. i_3) sind Konkurrenten zum 2-Testruf bezüglich der Leitungen 2 bzw. 3, sie sind daher direkt beeinflussende Rufe. Wartende Rufe der 4. Warteschlange (i_4) sind keine Konkurrenten zum 2-Testruf; sie beeinflussen über Leitung 7 jedoch Rufe der 3. Teilgruppe und damit indirekt auch den 2-Testruf.

Die außer dem 2-Testruf in der 2. Warteschlange wartenden Rufe können in zwei Klassen eingeteilt werden. Die i_2 vor dem 2-Testruf wartenden Rufe werden wegen D1 vor dem 2-Testruf abgefertigt, sie sind also direkt beeinflussende Rufe. Die z_2 Rufe, welche nach dem Testruf eingetroffen sind und sich hinter dem 2-Testruf

in die Warteschlange eingeordnet haben, werden mit Sicherheit erst nach dem 2-Testruf abgefertigt; ihre Anwesenheit wirkt wegen der Disziplin B jedoch beschleunigend auf die Abfertigung des 2-Testrufs (Akzelerationseffekt), sie sind daher indirekt beeinflussende Rufe.

Das "hinreichend differenzierte" Zustandsmuster für den $\xi_2(t)$ -Prozess läßt sich in diesem Beispiel dann folgendermaßen definieren:

$$\xi_2(t) = (x_4, x_5, x_6, x_7; i_2, z_2; i_1, i_3, i_4). \quad (4.70)$$

Nachdem die Zustände des $\xi_J(t)$ -Prozesses definiert sind, kann die Matrix der Sprungkoeffizienten A_J entsprechend Gl.(4.45) angegeben werden.

Stellvertretend hierfür sei die Differentialgleichung für den Wartezustand $(1, 1, 1, 1; i_2, z_2; i_1, i_3, 0)$ angegeben:

$$\begin{aligned} \frac{d}{dt} w_2(t|1,1,1,1; i_2, z_2; i_1, i_3, 0) = & - \left(\sum_{j=1}^4 \lambda_j + \sum_{I=1}^7 \epsilon_I \right) w_2(t|1,1,1,1; i_2, z_2; i_1, i_3, 0) \\ & + \lambda_1 \cdot w_2(t|1,1,1,1; i_2, z_2; i_1+1, i_3, 0) \\ & + \lambda_2 \cdot w_2(t|1,1,1,1; i_2, z_2+1; i_1, i_3, 0) \\ & + \lambda_3 \cdot w_2(t|1,1,1,1; i_2, z_2; i_1, i_3+1, 0) \\ & + \lambda_4 \cdot w_2(t|1,1,1,1; i_2, z_2; i_1, i_3, 1) \\ & + (\epsilon_1 + \epsilon_{22} + \epsilon_{32}) w_2(t|1,1,1,1; i_2-1, z_2; i_1, i_3, 0) \\ & + (\epsilon_{21} + \epsilon_{31} + \epsilon_4) w_2(t|1,1,1,1; i_2, z_2; i_1-1, i_3, 0) \\ & + (\epsilon_{33} + \epsilon_5 + \epsilon_7) w_2(t|1,1,1,1; i_2, z_2; i_1, i_3-1, 0) \\ & + \epsilon_6 \cdot w_2(t|1,1,0,1; i_2, z_2; i_1, i_3, 0), \end{aligned} \quad (4.71)$$

wobei

$$\begin{aligned} \epsilon_{21} &= \epsilon_2 \cdot \frac{i_1}{i_1 + i_2 + z_2 + 1}, & \epsilon_{22} &= \epsilon_2 \cdot \frac{i_2 + z_2 + 1}{i_1 + i_2 + z_2 + 1}, \\ \epsilon_{31} &= \epsilon_3 \cdot \frac{i_1}{i_1 + i_2 + z_2 + 1 + i_3}, & \epsilon_{32} &= \epsilon_3 \cdot \frac{i_2 + z_2 + 1}{i_1 + i_2 + z_2 + 1 + i_3}, & \epsilon_{33} &= \epsilon_3 \cdot \frac{i_3}{i_1 + i_2 + z_2 + 1 + i_3}. \end{aligned}$$

Die weitere Behandlung von Systemen aus Differentialgleichungen dieser Art erfolgt entsprechend dem in 4.3.5 beschriebenen Weg. In den Kapiteln 6 bzw. 7 werden für Systeme vollkommener bzw. unvollkommener Erreichbarkeit die Fälle im einzelnen näher behandelt.

4.4.3 Einige Eigenschaften mehrdimensionaler Prozesse

Die linearen Differentialgleichungssysteme zur Beschreibung des Zufallsprozesses der Systemzustände, entspr. der Kolmogoroff-Vorwärts-Gleichung (4.12), bzw. des Warteprozesses, entspr. der Kolmogoroff-Rückwärts-Gleichung (4.32), werden für allgemein mehrdimensionale Zufallsvariable mit Hilfe sog. "Zustandsräume" bzw. "Wartezustandsräume", welche die Zustände des Prozesses sowie deren Übergänge mit den zugehörigen Übergangskoeffizienten enthalten, übersichtlich dargestellt, vergl. Kap. 5 bis 7. Dabei wird aufgezeigt, wie sich die Struktur der geometrischen Zustandsräume in der Matrix-Koeffizientenstruktur der Gleichungssysteme niederschlägt. Beide Darstellungsarten - lineare Differentialgleichungssysteme und Zustands- bzw. Wartezustandsräume - werden im weiteren Verlauf der Arbeit als völlig gleichwertige Beschreibungsweisen für die Zufallsprozesse nebeneinander benutzt.

Die geometrische Darstellung läßt u.a. eine schnellere Einsicht in den Aufbau der Gleichungssysteme und deren Lösungsstrategie zu. Im folgenden sollen noch einige Eigenschaften mehrdimensionaler Prozesse aufgezeigt werden, welche zur Reduzierung des Grades der Gleichungssysteme von Bedeutung sind.

a) Instationärer Prozeß der Systemzustände

Kann man allgemein den Zustandsraum Z der Systemzustände durch einen Schnitt in zwei Teilräume Z' bzw. Z'' zerlegen, derart, daß beide Teilräume nur über einseitig gleichsinnig gerichtete Übergänge Verbindung miteinander haben, so kann derjenige Teilraum für sich betrachtet werden, von welchem die einseitig gerichteten Übergänge ausgehen, vergl. Bild 7a und Gl.(4.72a,b).

Gl.(4.72a,b) folgt aus der Kolmogoroff-Vorwärts-Gleichung (4.12), angewendet auf den Teilraum Z'. Dieser Fall kommt nur im Zusammenhang mit instationären Prozessen vor, bei denen transiente Systemzustände auftreten.

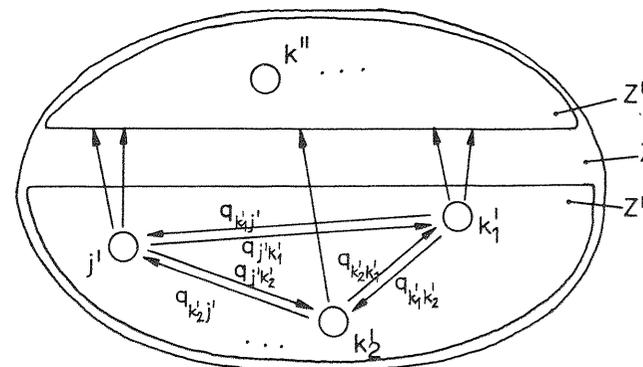


Bild 7a. Instationärer Teil-Prozeß der Systemzustände

$$\frac{d}{dt} P(t, j') = -q_{j'} P(t, j') + \sum_{k' \neq j'} q_{k'j'} P(t, k') \quad (4.72a)$$

wobei

$$q_{j'} = \sum_{k' \neq j'} q_{j'k'} + \sum_{k''} q_{j'k''} \quad , \quad j', k' \in Z' \quad (4.72b)$$

b) Stationärer Prozeß der Systemzustände

Kann man allgemein den Zustandsraum Z der Systemzustände durch einen Schnitt in zwei Teilräume Z' bzw. Z'' derart zerlegen, daß beide Teilräume nur über einen Schlüsselzustand j, welcher beiden Teilräumen als zugehörig angesehen werden kann, Verbindung miteinander haben, so ist dieser Schlüsselzustand bereits mit den Nachbarzuständen der Teilräume Z' bzw. Z'' für sich im statistischen Gleichgewicht, vergl. Bild 7b und Gl.(4.73a,b).

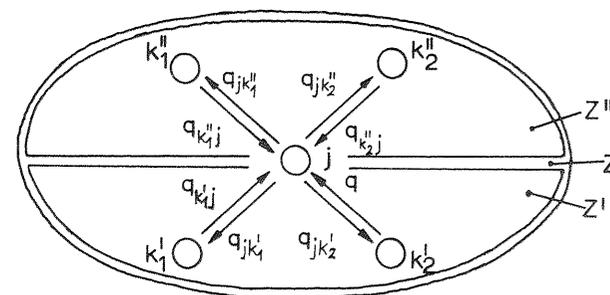


Bild 7b. Stationäre Teil-Prozesse der Systemzustände

$$P(j) \sum_{k' \neq j} q_{jk'} = \sum_{k' \neq j} P(k') q_{k'j} \quad , \quad k' \in Z' \quad (4.73a)$$

$$P(j) \sum_{k'' \neq j} q_{jk''} = \sum_{k'' \neq j} P(k'') q_{k''j} \quad , \quad k'' \in Z'' \quad (4.73b)$$

Zum Beweis von Gl.(4.73a,b) beachte man, daß von dem System der Kolmogoroff-Vorwärts-Gleichungen stets eine weggelassen und durch die Normalisierungsbedingung ersetzt werden darf. Die eindeutige Lösung des gesamten Gleichungssystems kann beispielsweise dadurch erfolgen, daß die Gleichung für den Zustand j weggelassen wird. Nun ersetze man z.B. alle Zustände $k'' \in Z''$ durch einen Makrozustand K'' , welcher sich mit dem Zustand j im statistischen Gleichgewicht befinden muß (eindimensionale Zustandskette!):

$$\sum_{k'' \neq j} p(k'') q_{k''j} = \left[\sum_{k'' \neq j} p(k'') \right] \cdot q_{K''j} = p(j) \sum_{k'' \neq j} q_{jk''} \quad (4.73c)$$

Setzt man die Gleichgewichtsgleichung (4.73c) in die Gleichgewichtsgleichung für den Zustand j ein, so folgt daraus Gl.(4.73a).

c) Stationärer Warteprozeß

Kann man allgemein den Wartezustandsraum eines Warteprozesses durch einen Schnitt in zwei Teilräume \bar{H}' und \bar{H}'' zerlegen, derart, daß beide Teilräume nur über einseitig gleichsinnig gerichtete Übergänge Verbindung miteinander haben, so kann stets derjenige Teilraum für sich, d.h. unabhängig, betrachtet werden, auf welchen die Übergänge gerichtet sind, vergl. Bild 7c und Gl.(4.74a,b).

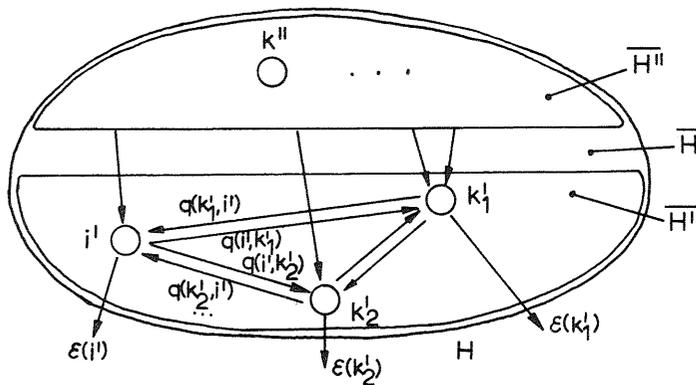


Bild 7c. Stationärer Teil-Warteprozeß

$$\frac{d}{dt} w(t|i') = -q(i') w(t|i') + \sum_{k' \neq i', k' \in \bar{H}'} q(i', k') w(t|k'), \quad (4.74a)$$

wobei
$$q(i') = \sum_{k' \neq i', k' \in \bar{H}'} q(i', k') + \epsilon(i'), \quad i' \in \bar{H}'. \quad (4.74b)$$

Gl.(4.74a,b) folgt aus der Kolmogoroff-Rückwärts-Gleichung (4.32), angewendet auf den Teilraum \bar{H}' .

5. BEDIENUNGSSYSTEME MIT EINER WARTESCHLANGE

In diesem Abschnitt wird das Bedienungssystem mit n vollkommen erreichbaren Leitungen und einer Warteschlange der maximalen Länge s_1 entsprechend Bild 1(a) untersucht. Die Rufe werden von einem Anrufprozeß mit der mittleren Anrufrate $\lambda_1 = \lambda$ erzeugt. Die mittlere Enderate der Leitung I sei ϵ_I , $I = 1, 2, \dots, n$. Zwei verschiedene Absuchdisziplinen für das Belegen von Leitungen werden unterschieden: geordnetes und zufallsmäßiges Absuchen. Als Abfertigungsdisziplinen innerhalb der Warteschlange werden D1, D2, D3a und D3b zugrunde gelegt.

5.1 Zustandswahrscheinlichkeiten

Zur exakten Beschreibung des $\xi(t)$ -Prozesses wird der Zustand

$$\xi(t) = (x_1, x_2, \dots, x_n; z) \quad (5.1)$$

definiert, worin $x_I \in [0, 1]$ den Belegungszustand der Leitung I charakterisiert, $I = 1, 2, \dots, n$, und $z_1 = z \in [0, s_1]$ die Anzahl belegter Wartepplätze angibt. Damit $z > 0$ werden kann, müssen in jedem Falle alle Leitungen belegt sein, d.h. $x_I = 1$, $I = 1, 2, \dots, n$.

Die Beschreibung nach Gl.(5.1) genügt den Disziplinen D1, D2, D3a und D3b, da die Verteilung der Anzahl Wartender durch diese Disziplinen nicht beeinträchtigt wird.

Der (n+1)-dimensionale Zustandsraum läßt sich grundsätzlich in zwei Unterräume aufteilen, nämlich einen n-dimensionalen Unterraum, welcher die Belegungsmuster der n Leitungen wiedergibt, und einen eindimensionalen Unterraum, welcher die Anzahl der wartenden Rufe charakterisiert. Die Menge Z umfaßt insgesamt $2^n + s_1$ Zustände.

5.1.1 Geordnetes Absuchen der Leitungen

Bei geordnetem Absuchen wird einem eintreffenden Ruf - sofern er nicht warten muß bzw. nicht sofort abgewiesen wird - eine durch den angetroffenen Zustand fest bestimmte Leitung zugewiesen und zwar die erste freie Leitung in der Reihenfolge von oben nach unten. In Bild 8 ist der Zustandsraum für den Fall $n = 3$ Leitungen angegeben.

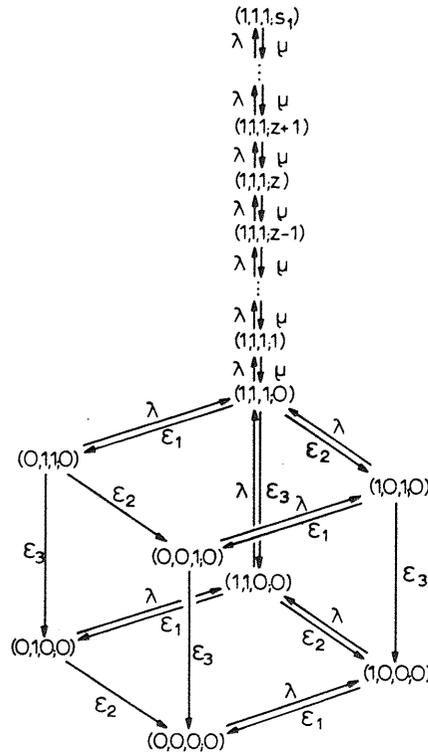


Bild 8. Zustandsraum und Übergangskoeffizienten für das Bedienungssystem vollkommener Erreichbarkeit mit $n = 3$ Leitungen und einer Warteschlange (geordnetes Absuchen der Leitungen)

Die Pfeile zwischen den Zuständen zeigen die Übergänge an mit den zugehörigen Übergangskoeffizienten. Dabei wurde für die mittlere Enderate des voll belegten Leitungsbündels die Abkürzung μ gewählt, d.h.

$$\mu = \sum_{I=1}^n \epsilon_I \quad (5.2)$$

Die stationären Zustandswahrscheinlichkeiten berechnen sich nach Abschnitt 4.1.5 aus einem linearen Gleichungssystem. Man erhält dieses System, indem für die einzelnen Zustände jeweils das "statische Gleichgewicht" angeschrieben wird. Außer diesem System von

Zustandsgleichungen gilt noch die "Normalisierungsbedingung" zwischen den Zustandswahrscheinlichkeiten (Wegen der Homogenität des Systems der Zustandsgleichungen kann immer eine Zustandsgleichung weggelassen werden; sie wird durch die Normalisierungsbedingung ersetzt).

Das lineare Gleichungssystem zur Bestimmung der stationären Zustandswahrscheinlichkeiten lautet im allgemeinen Falle von n Leitungen und geordnetem Absuchen der Leitungen:

$$\left. \begin{aligned} \left(\lambda + \sum_{I=1}^n x_I \epsilon_I \right) p(x_1, \dots, x_I, \dots, x_n, 0) &= \lambda \cdot \sum_{I=1}^n \left(\prod_{y=1}^I x_y \right) \cdot p(\dots, x_I - 1, \dots, 0) \\ &+ \sum_{I=1}^n (1 - x_I) \epsilon_I \cdot p(\dots, x_I + 1, \dots, 0), \quad \prod_{I=1}^n x_I = 0, \end{aligned} \right\} (5.3a)$$

$$\mu \cdot p(x_1, \dots, x_I, \dots, x_n, 0) = \lambda \cdot \sum_{I=1}^n x_I \cdot p(\dots, x_I - 1, \dots, 0), \quad \prod_{I=1}^n x_I = 1.$$

$$\mu \cdot p(1, \dots, 1, z) = \lambda \cdot p(1, \dots, 1, z - 1), \quad z = 1, 2, \dots, s_1. \quad (5.3b)$$

$$\sum_{x_1=0}^1 \dots \sum_{x_I=0}^1 \dots \sum_{x_n=0}^1 p(x_1, \dots, x_I, \dots, x_n, 0) + \sum_{z=1}^{s_1} p(1, \dots, 1, z) = 1. \quad (5.3c)$$

Innerhalb des eindimensionalen Unterraumes sind schon jeweils zwei Nachbarzustände im statistischen Gleichgewicht. Dies läßt sich sofort zeigen, wenn man nacheinander die Zustandsgleichungen für die Zustände $(1, \dots, 1; s_1)$, $(1, \dots, 1; s_1 - 1)$ usw. aufstellt und die sich ergebenden Zustandsgleichungen in die jeweils nächste Gleichung einsetzt. Dadurch ist es allgemein möglich, die Zustände des eindimensionalen Unterraumes rekursiv zu bestimmen:

$$p(1, \dots, 1, z) = p(1, \dots, 1, 0) \cdot \left(\frac{\lambda}{\mu} \right)^z, \quad z = 0, 1, \dots, s_1. \quad (5.4)$$

Die Zustände des n -dimensionalen Unterraumes müssen i.a. durch Lösung des linearen Gleichungssystems bestimmt werden. Eine explizite Lösung ist wegen der Unsymmetrie der Übergänge des n -dimensionalen Zustandsraumes (vergl. Bild 8) nicht zu erwarten.

5.1.2 Zufallsmäßiges Absuchen der Leitungen

Unter den freien Leitungen wird dem eintreffenden Ruf eine zufallsmäßig herausgesuchte Leitung zugewiesen. Sind mehrere Leitungen bei Ankunft des Rufs frei, so erreicht der Ruf eine bestimmte Leitung nur mit einer gewissen Wahrscheinlichkeit. Sind insgesamt $0 \leq x < n$ Leitungen belegt, so ist die Wahrscheinlichkeit für das Erreichen einer bestimmten Leitung $1/(n-x)$.

Gegenüber Bild 8 ändern sich nur einige Übergänge innerhalb des $3(n)$ -dimensionalen Unterraumes, vergl. Bild 9.

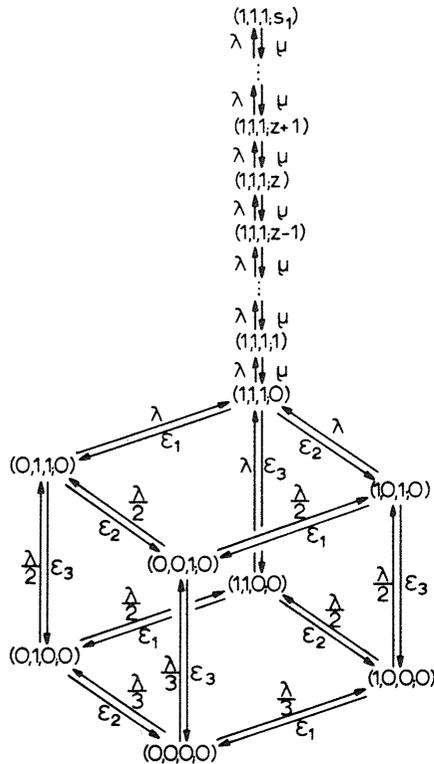


Bild 9. Zustandsraum und Übergangskoeffizienten für das Bedienungssystem vollkommener Erreichbarkeit mit $n = 3$ Leitungen und einer Warteschlange (zufallsmäßiges Absuchen der Leitungen)

Die Zustandswahrscheinlichkeiten werden aus nachfolgendem Gleichungssystem bestimmt:

$$\left. \begin{aligned} (\lambda + \sum_{I=1}^n x_I \epsilon_I) \cdot p(x_1, \dots, x_I, \dots, x_n, 0) &= \frac{\lambda}{n-x+1} \sum_{I=1}^n x_I p(\dots, x_I-1, \dots, 0) \\ &+ \sum_{I=1}^n (1-x_I) \epsilon_I p(\dots, x_I+1, \dots, 0), \quad \prod_{I=1}^n x_I = 0, \end{aligned} \right\} (5.5a)$$

$$\mu \cdot p(x_1, \dots, x_I, \dots, x_n, 0) = \lambda \cdot \sum_{I=1}^n x_I p(\dots, x_I-1, \dots, 0), \quad \prod_{I=1}^n x_I = 1. \quad (5.5b)$$

$$\mu \cdot p(1, \dots, 1, z) = \lambda \cdot p(1, \dots, 1, z-1), \quad z = 2, \dots, s_1. \quad (5.5c)$$

$$\sum_{x_1=0}^1 \dots \sum_{x_I=0}^1 \dots \sum_{x_n=0}^1 p(x_1, \dots, x_I, \dots, x_n, 0) + \sum_{z=1}^{s_1} p(1, \dots, 1, z) = 1, \quad (5.5c)$$

wobei

$$x = \sum_{I=1}^n x_I \quad (5.6)$$

die Zahl der Leitungsbelegungen in einem beliebigen Zustand kennzeichnet.

In Bild 10 ist ein Ausschnitt aus dem n -dimensionalen Unterraum für einen beliebigen Zustand angegeben. Er besitzt demnach $(n-x)$ "obere" und x "untere" Nachbarzustände.

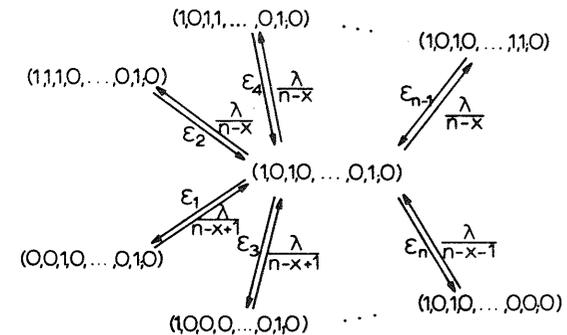


Bild 10. Ausschnitt aus dem n -dimensionalen Unterraum bei zufallsmäßigem Absuchen der Leitungen

Die Zustandsgleichung für den in Bild 10 gezeigten Zustand $(1,0,1,0,\dots,0,1;0)$ lautet:

$$(\lambda + \varepsilon_1 + \varepsilon_3 + \dots + \varepsilon_n) p(1,0,1,0,\dots,0,1;0) = \frac{\lambda}{n-x+1} [p(0,0,1,0,\dots,0,1;0) + p(1,0,0,0,\dots,0,1;0) + \dots + p(1,0,1,0,\dots,0,0;0)] + \varepsilon_2 \cdot p(1,1,1,0,\dots,0,1;0) + \varepsilon_4 \cdot p(1,0,1,1,\dots,0,1;0) + \dots + \varepsilon_{n-1} \cdot p(1,0,1,0,\dots,1,1;0) \quad (5.7)$$

Die allgemeine Zustandsgleichung (5.5a) wird von der folgenden Lösung befriedigt:

$$p(x_1, x_2, \dots, x_n; 0) = p(0, \dots, 0; 0) \cdot \frac{(n-x)!}{n!} \cdot \prod_{I=1}^n \left(\frac{\lambda}{\varepsilon_I} \right)^{x_I}, \quad x = 0, 1, \dots, n, \quad (5.8)$$

worin $p(0,0,\dots,0;0)$ ein multiplikativer Parameter ist, welcher mit Hilfe der Normalisierungsbedingung (5.5c) bestimmt wird. Es sei noch bemerkt, daß mit Gl.(5.8) gezeigt werden kann, daß auch im n-dimensionalen Unterraum sich schon jeweils 2 Nachbarzustände im statistischen Gleichgewicht befinden.

Mit der weiterhin gültigen Lösung Gl.(5.4) für den eindimensionalen Unterraum ergibt sich mit Gl.(5.8)

$$p(1, \dots, 1; z) = p(0, \dots, 0; 0) \cdot \frac{(\lambda/z)^z}{n!} \cdot \prod_{I=1}^n \left(\frac{\lambda}{\varepsilon_I} \right), \quad z = 0, 1, \dots, s_1. \quad (5.9)$$

Die Normalisierungsbedingung entsprechend Gl.(5.5c) ergibt mit Gl.(5.8) und (5.9) schließlich

$$p(0, \dots, 0; 0)^{-1} = \sum_{x_1=0}^1 \sum_{x_2=0}^1 \dots \sum_{x_n=0}^1 \frac{(n-x)!}{n!} \prod_{I=1}^n \left(\frac{\lambda}{\varepsilon_I} \right)^{x_I} + \frac{\lambda}{n!} \cdot \frac{1 - (\lambda/\mu)^{s_1}}{1 - \lambda/\mu} \cdot \prod_{I=1}^n \left(\frac{\lambda}{\varepsilon_I} \right). \quad (5.10)$$

Mit Gl.(5.8), (5.9) und (5.10) sind damit sämtliche stationären Zustandswahrscheinlichkeiten bekannt.

5.1.3 Sonderfall gleicher mittlerer Enderaten

Haben alle Leitungen dieselben mittleren Enderaten, d.h.

$$\varepsilon_I = \varepsilon, \quad I = 1, 2, \dots, n, \quad (5.11a)$$

$$\mu = n \cdot \varepsilon, \quad (5.11b)$$

so brauchen keine Belegungsmuster mehr unterschieden werden. Zur Kennzeichnung des Belegungszustandes des Leitungsbündels genügt

eine Variable x , welche die Anzahl der belegten Leitungen angibt, $x = 0, 1, \dots, n$. Auch die Unterscheidung der Absuchdisziplin verliert ihre Bedeutung.

Die Lösung für die Zustandswahrscheinlichkeiten $p(x; z)$ kann entweder direkt aus zwei eindimensionalen Zustandsräumen (welche zu einem eindimensionalen Zustandsraum zusammengefaßt werden können) oder aus der allgemeinen Lösung (5.8-10) gewonnen werden. Letzteres Verfahren werde beispielsweise angewendet.

Alle Zustände mit der gleichen Anzahl x belegter Leitungen und $z = 0$ besitzen nach Gl.(5.8) gleiche Wahrscheinlichkeiten. Es gibt davon genau $\binom{n}{x}$ Zustände, welche zu einem "Makrozustand" $(x; 0)$ zusammengefaßt werden können mit der Zustandswahrscheinlichkeit

$$p(x; 0) = \binom{n}{x} \cdot p(0; 0) \cdot \frac{(n-x)!}{n!} \left(\frac{\lambda}{\varepsilon} \right)^x = p(0; 0) \cdot \frac{A^x}{x!}, \quad x = 0, 1, \dots, n, \quad (5.12)$$

wobei $A = \frac{\lambda}{\varepsilon}$. Aus Gl.(5.9) folgt mit Gl.(5.12)

$$p(n; z) = p(0; 0) \cdot \frac{A^n}{n!} \rho^z, \quad (5.13)$$

wobei $\rho = \frac{\lambda}{\mu} = \frac{A}{n}$ (5.11c)

Aus Gl.(5.10) folgt schließlich

$$p(0; 0)^{-1} = \sum_{r=0}^n \frac{A^r}{r!} + \frac{A^n}{n!} \cdot \rho \cdot \frac{1 - \rho^{s_1}}{1 - \rho} \quad (5.14)$$

5.2 Charakteristische Größen

Ausgehend von den stationären Zustandswahrscheinlichkeiten können entsprechend 4.2 die wichtigsten charakteristischen Größen berechnet werden. Die Wahrscheinlichkeiten $p(1, \dots, 1; z)$, $z = 0, 1, \dots, s_1$, sollen kurz mit $p(n; z)$ bezeichnet werden.

a) Wartewahrscheinlichkeit W

$$W = \sum_{z=0}^{s_1-1} p(n; z) = p(n; 0) \cdot \frac{1 - \rho^{s_1}}{1 - \rho} \quad \text{für D1, D2, D3b} \quad (5.15a)$$

$$W = \sum_{z=0}^{s_1} p(n; z) = p(n; 0) \cdot \frac{1 - \rho^{s_1+1}}{1 - \rho} \quad \text{für D3a.} \quad (5.15b)$$

b) Verlustwahrscheinlichkeit B

$$B = p(n_i, s_i) = p(n_i, 0) \cdot \rho^{s_i} \quad \text{für D1, D2, D3b} \quad (5.16a)$$

$$B = 0 \quad \text{für D3a.} \quad (5.16b)$$

c) Erfolgs-Wartewahrscheinlichkeit W*

$$W^* = W \quad \text{für D1, D2, D3b} \quad (5.17a)$$

$$W^* = \sum_{z=0}^{s_i-1} p(n_i, z) = p(n_i, 0) \cdot \frac{1-\rho^{s_i}}{1-\rho} \quad \text{für D3a.} \quad (5.17b)$$

d) Mißerfolgs-Wartewahrscheinlichkeit W**

$$W^{**} = 0 \quad \text{für D1, D2, D3b} \quad (5.18a)$$

$$W^{**} = p(n_i, s_i) = p(n_i, 0) \cdot \rho^{s_i} \quad \text{für D3a.} \quad (5.18b)$$

Zwischen W, W* und W** gilt außerdem Gl.(4.20).

e) Verkehrswert Y

$$Y = \sum_{x_1=0}^1 \sum_{x_2=0}^1 \dots \sum_{x_n=0}^1 (x_1 + x_2 + \dots + x_n) p(x_1, x_2, \dots, x_n, 0) + n \cdot \sum_{z=1}^{s_i} p(n_i, z) \quad (5.19)$$

f) Mittlere Warteschlangenlänge Ω

$$\Omega = \sum_{z=0}^{s_i} z p(n_i, z) = p(n_i, 0) \rho \cdot \left[\frac{1-\rho^{s_i}}{(1-\rho)^2} - \frac{s_i \rho^{s_i}}{1-\rho} \right] \quad (5.20)$$

g) Mittlere Wartezeit der Wartenden t_W

$$t_W = \frac{1}{\mu} \left[\frac{1}{1-\rho} - \frac{s_i \rho^{s_i}}{1-\rho^{s_i}} \right] \quad \text{für D1, D2, D3b} \quad (5.21a)$$

$$t_W = \frac{1}{\mu} \left[\frac{1}{1-\rho} \cdot \frac{1-\rho^{s_i}}{1-\rho^{s_i+1}} - \frac{s_i \rho^{s_i}}{1-\rho^{s_i+1}} \right] \quad \text{für D3a.} \quad (5.21b)$$

5.3 Wartezeiten

Für die vier Abfertigungsdisziplinen D1, D2, D3a und D3b wird im folgenden der $\xi(t)$ -Prozeß in systematischer Weise untersucht. Neben den Wartezeitverteilungsfunktionen wird außerdem auf mittlere Wartezeiten und die Momente der Wartezeitverteilungsfunktionen eingegangen.

Nach den Ausführungen in 4.3.2 haben die n Leitungsbelegungen in jedem Falle einen direkten Einfluß auf die Wartezeit eines Testrufs. Da aber der $\xi(t)$ -Prozeß bei Systemen mit vollkommener Erreichbarkeit immer nur dann existieren kann, wenn alle n Leitungen belegt sind, braucht das Muster der Leitungsbelegungen nicht explizit in den Zustandsbeschreibungen des $\xi(t)$ -Prozesses enthalten zu sein. Die Komponenten der Zustandsvariablen des $\xi(t)$ -Prozesses setzen sich daher ausschließlich aus wartenden Rufen zusammen.

5.3.1 Abfertigung in der Reihenfolge des Eintreffens (D1)

a) Der Warteprozeß

Bei der Abfertigung wartender Rufe in der Reihenfolge des Eintreffens haben dem Testruf nachfolgende Rufe keinen Einfluß auf dessen Wartezeit. Als Zustandsvariable des $\xi(t)$ -Prozesses werde die Zahl i der vor dem Testruf in der Warteschlange wartenden Rufe definiert:

$$\xi(t) = i, \quad i = 0, 1, \dots, s_i - 1. \quad (5.22)$$

Für die Antreffwahrscheinlichkeit P(i) gilt:

$$P(i) = P\{\xi(0) = i\} = p(n_i, i), \quad i = 0, 1, \dots, s_i - 1. \quad (5.23)$$

Um das Geschehen des Warteprozesses anschaulich darzustellen, werden in den folgenden Beispielen sog. "Wartezustandsräume" aufgestellt. Es wird dabei angenommen, daß sich der $\xi(t)$ -Prozeß von einem allgemeinen Startzustand aus entwickelt, welcher bei Ankunft des Testrufs angetroffen wurde ("Beginn des Warteprozesses"). Der Wartezustandsraum beschreibt alle möglichen Zustände, die der Warteprozeß während der Anwesenheit des betrachteten Testrufs annehmen kann, wie auch deren Übergänge mit den zugehörigen Sprungkoeffizienten. Die zeitliche Folge der Zustände des $\xi(t)$ -Prozesses kann als

sog. "Irrfahrt" ("random walk") durch diesen Wartezustandsraum aufgefaßt werden. Schließlich werden auch die Endungen des $\xi(t)$ -Prozesses mit den zugehörigen Endekoeffizienten angegeben ("Ende des Warteprozesses" durch "Abfertigung" bzw. "Verdrängung").

Die Beschreibung des $\xi(t)$ -Prozesses durch die Kolmogoroff-Rückwärts-Gleichung bezieht sich jeweils auf den Startzustand und diejenigen Nachbarzustände, welche durch einen Übergang aus dem Startzustand erreicht werden können.

Bild 11 zeigt den Wartezustandsraum mit den Übergängen, Sprung- und Endekoeffizienten des $\xi(t)$ -Prozesses. Im Beispiel Bild 11 ist die "Irrfahrt" eindimensional und sehr einfach, indem der Testruf nacheinander die unteren Zustände bis zur Abfertigung durchläuft.

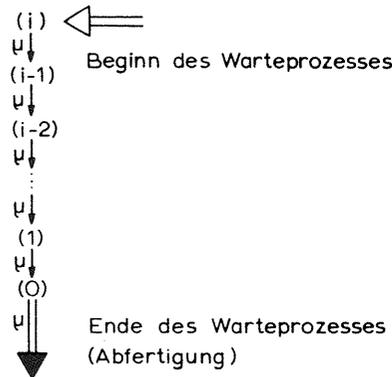


Bild 11. Wartezustandsraum für Warten nach D1

Da die Zahl der in der Warteschlange vor dem Testruf wartenden Rufe nur abnehmen kann, ist

$$q(i,j) = \begin{cases} \mu & , j = i-1 \\ 0 & , j < i-1 \end{cases} \quad , i = 1, 2, \dots, s_1-1. \quad (5.24a)$$

Der Koeffizient für eine Änderung des Zustands (i) ist

$$q(i) = \mu \quad , i = 0, 1, \dots, s_1-1. \quad (5.24b)$$

Die Sprungkoeffizienten hängen nicht von der mittleren Anrufrate λ ab, da nachfolgende Rufe keinen Einfluß auf die Wartezeit des Testrufs ausüben.

b) Wartezeitverteilungsfunktionen

Als bedingte Wartezeitverteilungsfunktion werde $w(t|i)$ eingeführt, wobei nach Gl.(4.30) $\xi(0) = i$. Da kein wartender Ruf aus dem Wartespeicher verdrängt wird, gilt $w(t|i) = w^*(t|i)$, $w^{**}(t|i) = 0$, $i = 0, 1, \dots, s_1-1$.

Für die bedingten Wartezeitverteilungsfunktionen gilt entsprechend Gl.(4.32) und mit Gl.(5.24a,b) das folgende Differentialgleichungssystem:

$$\frac{dw(t|0)}{dt} = -\mu w(t|0) \quad (5.25a)$$

$$\frac{dw(t|i)}{dt} = -\mu w(t|i) + \mu w(t|i-1), \quad i = 1, 2, \dots, s_1-1. \quad (5.25b)$$

Die Anfangsbedingungen sind nach Gl.(4.33) einheitlich $w(0|i) = 1$, $i = 0, 1, \dots, s_1-1$. Dies folgt auch aus Gl.(5.25a,b) mit Gl.(4.34) für die Endekoeffizienten des $\xi(t)$ -Prozesses

$$\varepsilon(i) = \begin{cases} \mu & , i = 0 \\ 0 & , i > 0 \end{cases} \quad (5.26)$$

Die Anwendung der Laplace-Transformation auf das lineare Differentialgleichungssystem (5.25a,b) ergibt Gl.(5.27). Die am oberen Rand angeschriebenen Zustände (i) deuten die Zugehörigkeit der in dieser Spalte stehenden Koeffizienten zu den jeweiligen "Unbekannten" $W(s|i)$ an. Die Matrix der Sprungkoeffizienten **A** ist identisch mit der Koeffizientenmatrix von Gl.(5.27) für $s = 0$. **A** ist in diesem Beispiel bidiagonal und deshalb besonders einfach zu behandeln.

$$\begin{array}{cccccc|c} (0) & (1) & (2) & \dots & (s_1-2) & (s_1-1) & \\ \hline (s+\mu) & & & & & & 1 \\ -\mu & (s+\mu) & & & & & 1 \\ & -\mu & (s+\mu) & & & & 1 \\ & & & \dots & & & \vdots \\ & & & & -\mu & (s+\mu) & 1 \\ & & & & & -\mu & (s+\mu) & 1 \end{array} \quad , N = s_1. \quad (5.27)$$

Durch sukzessive Entwicklung der Determinante der Koeffizientenmatrix $\det(\mathbf{A} + s\mathbf{I})$ folgt:

$$\det(\mathbf{A} + s\mathbf{I}) = (s + \mu)^{s_1} \quad (5.28)$$

Die Matrix \mathbf{A} hat demnach nur $N = 1$ Eigenwert der Vielfachheit $K_1 = s_1$:

$$e_i = -\mu \quad (5.29)$$

Die Laplace-Transformierten der bedingten Wartezeitverteilungsfunktionen $W(s|i)$, $i = 0, 1, \dots, s_1 - 1$, können durch Rekursion aus Gl. (5.27) gewonnen werden:

$$W(s|i) = \sum_{r=1}^{i+1} \frac{\mu^{r-1}}{(s+\mu)^r}, \quad i = 0, 1, \dots, s_1 - 1. \quad (5.30)$$

Die Laplace-Rücktransformation von Gl.(5.30) ergibt

$$w(t|i) = \sum_{r=0}^i \frac{(\mu t)^r}{r!} \exp(-\mu t), \quad i = 0, 1, \dots, s_1 - 1. \quad (5.31)$$

$w(t|i)$ nach Gl.(5.31) ist offensichtlich eine ERLANG-($i+1$)-Verteilung. Dieses Ergebnis läßt sich anschaulich deuten: Trifft der Testruf i Rufe an, so muß er bis zum Beginn seiner Abfertigung genau ($i+1$) negativ-exponentiell verteilte Abfertigungsphasen abwarten; zusammengesetzt aus einer ersten, bis eine der n Leitungsbelegungen endet, und den i sich daran anschließenden Phasen der vor dem Testruf wartenden Rufe. Ein vollkommen erreichbares und belegtes Leitungsbündel mit der Summen-Endedichte

$$\mu = \sum_{I=1}^n \epsilon_I$$

verhält sich wie eine einzige Leitung der Endedichte μ . Wegen der Voraussetzung negativ-exponentiell verteilter Belegungsdauern ist die Verteilung der Restphase einer bestehenden Belegung wieder negativ-exponentiell mit demselben Mittelwert, unabhängig davon, wie lange die Belegung schon angedauert hatte. Die ($i+1$) abzuwartenden Phasen sind also alle negativ-exponentiell verteilt mit dem gleichen Mittelwert $1/\mu$. Die VF einer Summe von ($i+1$) Zufallsvariablen, welche jeweils negativ-exponentiell verteilt sind mit gleichem Mittelwert $1/\mu$, ergibt aber gerade die ERLANG-($i+1$)-Verteilung nach Gl.(5.31).

Die Wartezeitverteilungsfunktion bezüglich aller Rufe, $W(>t)$, folgt entsprechend Gl.(4.39) mit $P(i)$ nach Gl.(5.23) und $w(t|i)$ nach Gl.

(5.31) nach einigen Umformungen

$$W(>t) = \frac{W}{1 - \rho^{s_1}} \left[\sum_{r=0}^{s_1-1} \frac{(\lambda t)^r}{r!} - \rho^{s_1} \sum_{r=0}^{s_1-1} \frac{(\mu t)^r}{r!} \right] \exp(-\mu t). \quad (5.32)$$

Das Ergebnis Gl.(5.32) wurde bereits 1956 auf eine direktere Art von H. STÖRMER [21] abgeleitet.

c) Mittlere Wartezeiten

Die mittlere Wartezeit der Rufe, welche i wartende Rufe angetroffen haben, folgt aus Gl.(4.58) und (5.30)

$$t_W(i) = \frac{i+1}{\mu}, \quad i = 0, 1, \dots, s_1 - 1. \quad (5.33)$$

Die mittlere Wartezeit bezüglich aller wartenden Rufe erhält man nach Gl.(4.60) zu

$$t_W = \frac{1}{W} \sum_{i=0}^{s_1-1} p(n;i) t_W(i) = \frac{1}{\mu} \left[\frac{1}{1-\rho} - \frac{s_1 \rho^{s_1}}{1-\rho^{s_1}} \right]. \quad (5.34)$$

t_W nach Gl.(5.34) stimmt mit dem Ergebnis für t_W nach Gl.(5.21a) überein.

d) Momente der Wartezeitverteilungsfunktion

Die K -ten bedingten Momente lassen sich aus einem linearen Gleichungssystem entsprechend Gl.(4.56) berechnen. Aus Gl.(5.25a,b) läßt sich das folgende System von Gleichungen ableiten:

$$\mu M_K(0) = K M_{K-1}(0) \quad (5.35a)$$

$$\mu M_K(i) - \mu M_K(i-1) = K M_{K-1}(i), \quad i = 1, 2, \dots, s_1 - 1, \quad K = 1, 2, \dots \quad (5.35b)$$

Die Lösung dieses Systems wird durch Rekursion erhalten:

$$M_K(i) = \frac{1}{\mu^K} \prod_{r=1}^K (i+r) = \frac{1}{\mu^K} \cdot \frac{(i+K)!}{i!}. \quad (5.36)$$

Die absoluten Momente der Wartezeitverteilungsfunktion berechnen sich entsprechend Gl.(4.57) zu

$$M_K = p(n;0) \cdot \frac{1}{\mu^K} \sum_{i=0}^{s_1-1} \frac{(i+K)!}{i!} \rho^i = p(n;0) \cdot \frac{1}{\mu^K} \cdot \frac{d^K}{ds^K} \left(\rho^K \frac{1-\rho^{s_1}}{1-\rho} \right), \quad K = 1, 2, \dots \quad (5.37)$$

5.3.2 Abfertigung in zufälliger Reihenfolge (D2)

a) Der Warteprozess

Bei der Abfertigung in zufälliger Reihenfolge konkurrieren s_1 Rufe im Wartespeicher wartenden Rufe um eine freiwerdende Leitung. Aus diesem Grunde haben neben den bei Ankunft des Testrufs angetroffenen Wartenden auch jene Rufe einen Einfluß auf das Schicksal des Testrufs, welche nach dem Testruf eingetroffen sind und akzeptiert wurden. Als Zustandsvariable werde

$$\xi(t) = i, \quad i = 0, 1, \dots, s_1 - 1, \quad (5.38)$$

definiert. In Gl.(5.38) gibt i die Zahl der mit dem Testruf wartenden Rufe im Wartespeicher zur Zeit t an. Für die Antreffwahrscheinlichkeiten $P(i)$ gilt weiterhin Gl.(5.23).

In Bild 12 sind die Zustände des $\xi(t)$ -Prozesses mit den Übergängen und zugehörigen Sprung- bzw. Endekoeffizienten angegeben.

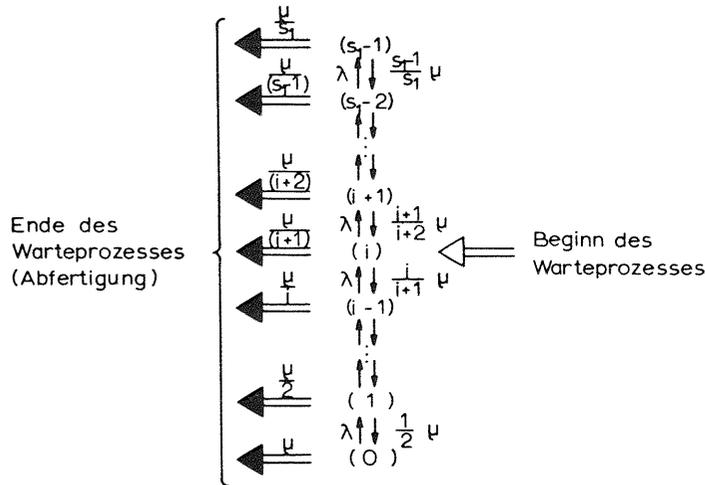


Bild 12. Wartezustandsraum für Warten nach D2

Wie aus Bild 12 ersichtlich ist, kann der $\xi(t)$ -Prozess von jedem Zustand i aus durch ein Belegungsende mit dem Endekoeffizienten

$$\varepsilon(i) = \frac{1}{i+1} \mu, \quad i = 0, 1, \dots, s_1 - 1, \quad (5.39)$$

beendet werden. Der Faktor $1/(i+1)$ steht dafür, daß gerade der Testruf aus den $(i+1)$ Wartenden (zufallsmäßig) ausgesucht wird.

Die Sprungkoeffizienten des $\xi(t)$ -Prozesses sind

$$q(i, j) = \begin{cases} \lambda & \text{für } j = i+1, i = 0, 1, \dots, s_1 - 2 \\ \frac{i}{i+1} \mu & \text{für } j = i-1, i = 1, 2, \dots, s_1 - 1 \\ 0 & \text{für } j > i+1 \text{ bzw. } j < i-1. \end{cases} \quad (5.40a)$$

Der Zustand $\xi(t) = i$ ändert sich mit dem Koeffizienten

$$q(i) = \begin{cases} \lambda + \mu & \text{für } i = 0, 1, \dots, s_1 - 2 \\ \mu & \text{für } i = s_1 - 1. \end{cases} \quad (5.40b)$$

b) Wartezeitverteilungsfunktionen

Es sei wieder $w(t|i)$ die bedingte Wartezeitverteilungsfunktion, wobei $\xi(0) = 1$. Die bedingten Wartezeitverteilungsfunktionen $w(t|i)$ gehorchen nach Gl.(4.32) mit Gl.(5.40a,b) dem folgenden Differentialgleichungssystem:

$$\frac{dw(t|i)}{dt} = -(\lambda + \mu)w(t|i) + \lambda w(t|i+1) + \frac{i}{i+1} \mu w(t|i-1), \quad i = 0, 1, \dots, s_1 - 2, \quad (5.41a)$$

$$\frac{dw(t|s_1-1)}{dt} = -\mu w(t|s_1-1) + \frac{s_1-1}{s_1} \mu w(t|s_1-2), \quad (5.41b)$$

mit den Anfangsbedingungen $w(0|i) = 1, i = 0, 1, \dots, s_1 - 1$. Letzteres kann wieder durch Gl.(5.41a,b) mit Gl.(5.39) nachgeprüft werden.

Gl.(5.42) stellt das transformierte Gleichungssystem (5.41a,b) dar. Die Matrix A der Sprungkoeffizienten ist tridiagonal und unsymmetrisch.

$$\begin{array}{cccccc|c} (0) & (1) & (2) & \dots & (s_1-2) & (s_1-1) & \\ \hline (s_1 + \lambda + \mu) & -\lambda & & & & & 1 \\ -\frac{1}{2} \mu & (s_1 + \lambda + \mu) & -\lambda & & & & 1 \\ & -\frac{2}{3} \mu & (s_1 + \lambda + \mu) & -\lambda & & & 1 \\ & & \ddots & \ddots & \ddots & \ddots & \vdots \\ & & & -\frac{s_1-2}{s_1-1} \mu & (s_1 + \lambda + \mu) & -\lambda & 1 \\ & & & & -\frac{s_1-1}{s_1} \mu & (s_1 + \lambda + \mu) & 1 \end{array}, \quad N = s_1. \quad (5.42)$$

Im Anhang A1 werden drei Eigenschaften dieser Matrix bewiesen:

1. Die Eigenwerte sind negativ-reell.
2. Die Eigenwerte sind verschieden, d.h. $N' = N = s_1$, $K_{\nu} = 1$, $\nu = 1, 2, \dots, N$.
3. Die Eigenwerte liegen alle im Intervall

$$\left[-\max\left(2\lambda + \frac{2s_1-3}{s_1-1}\mu, \frac{2s_1-1}{s_1}\mu\right), -\frac{\mu}{s_1} \right], \quad s_1 > 2.$$

Für die Sonderfälle $s_1 = 1$ und $s_1 = 2$ sind die Eigenwerte explizit angebar:

$$e_1 = -\mu, \quad s_1 = 1$$

$$e_{1,2} = -\frac{1}{2} \left[(\lambda + 2\mu) \mp \sqrt{\lambda^2 + 2\lambda\mu} \right], \quad s_1 = 2.$$

Die Eigenwerte e_{ν} , $\nu = 1, 2, \dots, s_1$, müssen i.a. durch eine numerische Rechnung bestimmt werden, ebenso die Residuen $a_{i\nu}$, $i = 0, 1, \dots, s_1-1$, $\nu = 1, 2, \dots, s_1$.

Die allgemeine Lösung des Systems (5.41a,b) ist somit

$$w(t|i) = \sum_{\nu=1}^{s_1} a_{i\nu} \exp(e_{\nu}t), \quad i = 0, 1, \dots, s_1-1, \quad (5.43)$$

wobei wegen der Anfangsbedingungen außerdem

$$\sum_{\nu=1}^{s_1} a_{i\nu} = 1, \quad i = 0, 1, \dots, s_1-1, \quad (5.44)$$

gilt. Gl.(5.44) kann bei numerischen Rechnungen zu Kontrollzwecken benutzt werden.

Die absolute Wartezeitverteilungsfunktion bezüglich aller Rufe ist entsprechend Gl.(4.39) mit Gl.(5.23)

$$W(>t) = \sum_{i=0}^{s_1-1} p(n, i) w(t|i). \quad (5.45)$$

c) Mittlere Wartezeiten

Die bedingten mittleren Wartezeiten $t_w(i)$ für das Warten vom Anfangszustand i an, $i = 0, 1, \dots, s_1-1$, berechnen sich entsprechend Gl.(4.58), (4.44) und (5.41a,b) aus dem linearen Gleichungssystem

$$(\lambda + \mu)t_w(i) - \lambda t_w(i+1) - \frac{\mu}{i+1} t_w(i-1) = 1, \quad i = 0, 1, \dots, s_1-2, \quad (5.46a)$$

$$\mu t_w(s_1-1) - \frac{s_1-1}{s_1} \mu t_w(s_1-2) = 1. \quad (5.46b)$$

Für die rekursive Berechnung der bedingten mittleren Wartezeiten wird im folgenden Teilabschnitt d) ein Verfahren angegeben.

Die mittlere Wartezeit bezüglich aller wartenden Rufe, t_w , kann mit Hilfe der bedingten mittleren Wartezeiten $t_w(i)$ entsprechend Gl.(4.60) mit P(i) nach Gl.(5.23) bestimmt werden. t_w wurde bereits mit Hilfe der mittleren Warteschlangenlänge in Gl.(5.21a) berechnet und kann daher bei numerischen Rechnungen ebenfalls zu Kontrollzwecken benutzt werden.

d) Momente der Wartezeitverteilungsfunktion

Für die K-ten bedingten Momente $M_K(i)$ gilt entsprechend Gl.(4.56) und Gl.(5.41a,b) das Gleichungssystem

$$(\lambda + \mu)M_K(i) - \lambda M_K(i+1) - \frac{\mu}{i+1} M_K(i-1) = KM_{K-1}(i), \quad i = 0, 1, \dots, s_1-2, \quad (5.47a)$$

$$\mu M_K(s_1-1) - \frac{s_1-1}{s_1} \mu M_K(s_1-2) = KM_{K-1}(s_1-1), \quad (5.47b)$$

$K = 1, 2, \dots$

Zur Ermittlung der einzelnen bedingten Momente muß im allgemeinen für jedes $K = 1, 2, \dots$ jeweils das lineare Gleichungssystem aufgelöst werden. Die Koeffizientenmatrix bleibt für jedes K gleich; es ändern sich jeweils nur die Absolutglieder auf der rechten Seite, welche durch den K-fachen Lösungsvektor der (K-1)-ten Momente gegeben sind.

Zur Berechnung der ersten bedingten Momente (bedingte mittlere Wartezeiten) kann folgender rekursiver Algorithmus aus Gl.(5.47a,b) abgeleitet werden:

$$M_1(i) = M_1(0) \cdot P_i(\xi) - \frac{1}{\lambda} Q_i(\xi), \quad i = 1, 2, \dots, s_1-1, \quad (5.48a)$$

wobei

$$M_1(0) = \frac{1}{\mu} \cdot \frac{1 + \xi \cdot [Q_{s_1-1}(\xi) - \frac{s_1-1}{s_1} Q_{s_1-2}(\xi)]}{P_{s_1-1}(\xi) - \frac{s_1-1}{s_1} P_{s_1-2}(\xi)}, \quad \xi = \frac{1}{s}. \quad (5.48b)$$

Die "erzeugenden" Polynome $P_i(\xi)$ und $Q_i(\xi)$ können ebenfalls rekursiv berechnet werden:

$$P_i(\xi) = (1+\xi)P_{i-1}(\xi) - \frac{i-1}{\xi} \xi \cdot P_{i-2}(\xi), \quad P_0(\xi) = 1, \quad (5.48c)$$

$$Q_i(\xi) = (1+\xi)Q_{i-1}(\xi) - \frac{i-1}{\xi} \xi \cdot Q_{i-2}(\xi) + 1, \quad Q_0(\xi) = 0, \quad (5.48d)$$

$$i = 1, 2, \dots, s_1 - 1.$$

Für die höheren bedingten Momente lassen sich entsprechende rekursive Berechnungsverfahren ableiten, welche jedoch komplizierter aufgebaut sind.

Das K-te absolute Moment der Wartezeitverteilungsfunktion berechnet sich nach Gl.(4.57). Nach einem Verfahren von M. SEGAL [22] kann das K-te absolute Moment M_K durch die (K-1)-ten bedingten Momente $M_{K-1}(i)$, $i = 0, 1, \dots, s_1 - 1$, ausgedrückt werden. Wird Gl.(5.47a,b) mit $(i+1)p(n;i)$ multipliziert und über i aufsummiert, so folgt

$$M_K = p(n;0) \cdot \frac{1}{\mu} \sum_{i=0}^{s_1-1} (i+1) K M_{K-1}(i) \cdot s^i, \quad K = 1, 2, \dots \quad (5.49)$$

Für $K = 1$ ergibt Gl.(5.49) dasselbe Ergebnis wie für M_1 nach D1, Gl.(5.37), da das erste Moment (mittlere Wartezeit!) von der Abfertigungsdisziplin unabhängig ist. Für $K = 2$ erhält man aus Gl.(5.49)

$$M_2 = p(n;0) \cdot \frac{1}{\mu} \sum_{i=0}^{s_1-1} (i+1) 2 M_1(i) s^i \quad (5.50a)$$

Für D1 wurde mit Gl.(5.37) abgeleitet:

$$M_2 = p(n;0) \cdot \frac{1}{\mu} \sum_{i=0}^{s_1-1} (i+1) \frac{i+2}{\mu} s^i \quad (5.50b)$$

Durch Betrachtung eines Hilfsfalles kann nach M. SEGAL [22] gezeigt werden, daß $2M_1(1) > \frac{1+2}{\mu}$, woraus

$$M_2^{(D2)} > M_2^{(D1)} \quad (5.50c)$$

folgt: Das zweite Moment (und damit auch die Streuung der Wartezeiten) ist für zufällmässiges Abfertigen größer als bei Abfertigung in der Reihenfolge des Eintreffens.

5.3.3 Abfertigung in der inversen Reihenfolge des Eintreffens mit verdrängender Priorität (D3a)

a) Der Warteprozess

Sind alle Leitungen blockiert, so belegt ein eintreffender Ruf grundsätzlich den ersten Warteplatz und verschiebt alle angetroffenen Wartenden um einen Platz nach hinten. Findet der eintreffende Ruf auch den Wartespeicher voll besetzt vor, so wird der Wartende mit der größten Wartezeit verdrängt. Die bei Ankunft angetroffenen Wartenden haben also keinen Einfluß auf die Wartezeit eines Testrufs. Dem Testruf nachfolgende Rufe haben dagegen einen direkten Einfluß auf das Schicksal des Testrufs: sie werden - wenn überhaupt - vor dem Testruf abgefertigt und bestimmen außerdem mit, ob der Testruf verdrängt wird oder nicht.

Die Zustandsvariable

$$\xi(t) = i, \quad i = 0, 1, \dots, s_1 - 1, \quad (5.51)$$

kennzeichne die Zahl der in der Warteschlange vor dem Testruf wartenden Rufe. Jeder wartende Testruf beginnt das Warten vom 1. Warteplatz aus. Für die gesamte Wartezeit des Testrufs gilt also $\xi(0) = 0$. Die Antreffwahrscheinlichkeit hierfür ist gleich der Wahrscheinlichkeit, daß der Testruf warten muß:

$$P(0) = W, \quad (5.52)$$

mit W nach Gl.(5.15b).

Bei D1 und D2 konnten vom Testruf alle Zustände i bei Ankunft angetroffen werden, d.h. der $\xi(t)$ -Prozeß konnte sich von jedem Startzustand aus entwickeln. Bei D3a hingegen gibt es nur einen Startzustand $\xi(0) = 0$. Von den Zuständen $i > 0$ aus entwickelt sich nur ein Prozeß für eine partielle Wartezeit. Die Warteprozesse für partielle Wartezeiten können jedoch in gleicher Weise wie die Warteprozesse für die gesamten Wartezeiten behandelt werden.

In Bild 13 ist der Zustandsraum der Wartezustände und Übergänge des $\xi(t)$ -Prozesses angegeben.

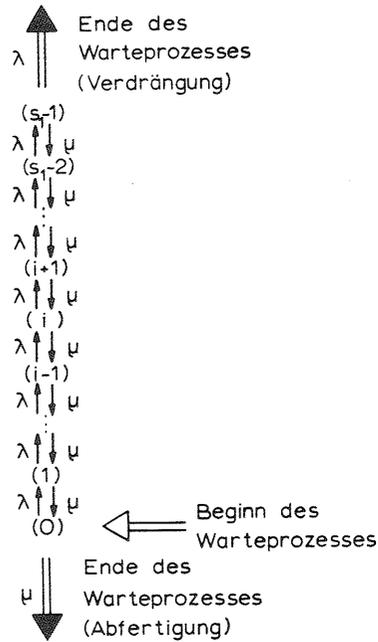


Bild 13. Wartezustandsraum für Warten nach D3a

Die Sprungkoeffizienten sind

$$q(i, j) = \begin{cases} \lambda & \text{für } j = i+1, \quad i = 0, 1, \dots, s_1-2 \\ \mu & \text{für } j = i-1, \quad i = 1, 2, \dots, s_1-1 \\ 0 & \text{für } j > i+1 \text{ bzw. } j < i-1 \end{cases} \quad (5.53a)$$

Der Zustand $\xi(t) = i$ ändert sich mit dem Koeffizienten

$$q(i) = \lambda + \mu \quad \text{für } i = 0, 1, \dots, s_1-1 \quad (5.53b)$$

b) Wartezeitverteilungsfunktionen

$w(t|i)$ sei die bedingte VF der Wartezeiten für das Warten vom Zustand i aus. $w(t|0)$ ist die VF der Gesamtwartezeit des Testrufs; die Größen $w(t|i)$, $i > 0$, sind hingegen nur VF partieller Wartezeiten.

Für die bedingten Wartezeitverteilungsfunktionen $w(t|i)$ gilt entsprechend Gl.(4.32) und mit Gl.(5.53a,b) das folgende Differentialgleichungssystem:

$$\frac{dw(t|0)}{dt} = -(\lambda + \mu)w(t|0) + \lambda w(t|1) \quad (5.54a)$$

$$\frac{dw(t|i)}{dt} = -(\lambda + \mu)w(t|i) + \lambda w(t|i+1) + \mu w(t|i-1), \quad i = 1, 2, \dots, s_1-2, \quad (5.54b)$$

$$\frac{dw(t|s_1-1)}{dt} = -(\lambda + \mu)w(t|s_1-1) + \mu w(t|s_1-2). \quad (5.54c)$$

Gl.(5.54a-c) gilt zunächst für Rufe, welche nach Ankunft warten, ohne Rücksicht darauf, ob sie erfolgreich oder erfolglos warten. Entsprechend den Ausführungen in 4.3.3 gilt für die bedingten Wartezeitverteilungsfunktionen erfolgreich bzw. erfolglos wartender Rufe, $w^*(t|i)$ bzw. $w^{**}(t|i)$, $i = 0, 1, \dots, s_1-1$, dasselbe Differentialgleichungssystem. Die Anfangsbedingungen erhält man aus Gl.(5.54a-c) entsprechend 4.3.3 für die Endkoeffizienten

$$\varepsilon(i) = \begin{cases} \mu & \text{für } i = 0 \\ 0 & \text{für } i = 1, 2, \dots, s_1-2 \\ \lambda & \text{für } i = s_1-1 \end{cases} \quad (5.55a)$$

$$\varepsilon^*(i) = \begin{cases} \mu & \text{für } i = 0 \\ 0 & \text{für } i = 1, 2, \dots, s_1-1 \end{cases} \quad (5.55b)$$

$$\varepsilon^{**}(i) = \begin{cases} 0 & \text{für } i = 0, 1, \dots, s_1-2 \\ \lambda & \text{für } i = s_1-1 \end{cases} \quad (5.55c)$$

Als Lösungen für die Anfangsbedingungen stellen sich ein:

$$w(0|i) = 1 \quad (5.56a)$$

$$w^*(0|i) = \frac{1 - \rho^{s_1-i}}{1 - \rho^{s_1+1}}, \quad (5.56b)$$

$$w^{**}(0|i) = \frac{\rho^{s_1-i} - \rho^{s_1+1}}{1 - \rho^{s_1+1}}, \quad i = 0, 1, \dots, s_1-1. \quad (5.56c)$$

$w^*(0|i)$ bzw. $w^{**}(0|i)$ sind die Wahrscheinlichkeiten für erfolgreiches bzw. erfolgloses Warten vom Zustand i aus, sie wurden bereits an anderer Stelle bei Betrachtung des "Ruin-Problems" abgeleitet [6, 23].

Das Gl.(5.54a-c) entsprechende transformierte System wird durch Gl.(5.57) wiedergegeben. Die Matrix **A** der Sprungkoeffizienten besitzt eine spezielle Form der tridiagonalen Matrix (siehe Anhang A2).

$$\begin{array}{cccc|c}
 (0) & (1) & (2) & \dots & (s_1-2) & (s_1-1) & \\
 \hline
 (s_1+\mu) & -\lambda & & & & & w(0|0) \\
 -\mu & (s_1+\mu) & -\lambda & & & & w(0|1) \\
 & -\mu & (s_1+\mu) & -\lambda & & & w(0|2) \\
 & & \dots & \dots & \dots & & \vdots \\
 & & & -\mu & (s_1+\mu) & -\lambda & w(0|s_1-2) \\
 & & & & -\mu & (s_1+\mu) & w(0|s_1-1)
 \end{array} , N = s_1 \quad (5.57)$$

Gl.(5.57) gilt entsprechend für erfolgreich bzw. erfolglos wartende Rufe, wenn der Vektor der Anfangsbedingungen nach Gl.(5.56a) durch den entsprechenden Vektor nach Gl.(5.56b) bzw. (5.56c) ersetzt wird.

In Anhang A2 werden folgende Sätze bewiesen:

1. Die Matrix **A** nach Gl.(5.57) hat s_1 verschiedene negativ-reelle Eigenwerte

$$e_{\nu} = 2\sqrt{\lambda\mu} \cos \frac{\nu\pi}{s_1+1} - (\lambda+\mu), \quad \nu = 1, 2, \dots, s_1. \quad (5.58)$$

2. Die Wartezeitverteilungsfunktion bei D3a ist durch folgenden Ausdruck gegeben:

$$W(>t) = W \cdot \sum_{\nu=1}^{s_1} \left(\frac{\sum_{i=1}^{s_1} w(0|i-1) \lambda^{i-1} D_{s_1-i}(e_{\nu})}{\prod_{i \neq \nu} (e_{\nu} - e_i)} \right) \exp(e_{\nu} t), \quad (5.59)$$

wobei

$$D_{\ell}(e_{\nu}) = \sqrt{\lambda\mu}^{\ell} \cdot \frac{\sin(\ell+1) \frac{\nu\pi}{s_1+1}}{\sin \frac{\nu\pi}{s_1+1}}, \quad \begin{array}{l} \nu = 1, 2, \dots, s_1 \\ \ell = 0, 1, \dots, s_1-1. \end{array}$$

Die VF der erfolgreichen bzw. erfolglosen Wartezeiten bezüglich aller Rufe, $W^*(>t)$ bzw. $W^{**}(>t)$, wird aus Gl.(5.59) gewonnen, indem die Anfangsbedingungen $w(0|i)$ durch $w^*(0|i)$ bzw. $w^{**}(0|i)$, $i = 0, 1, \dots, s_1-1$, ersetzt werden.

c) Mittlere Wartezeiten

Die bedingten mittleren Wartezeiten $t_W(i)$ können entsprechend Gl.(4.58), (4.44) und (5.54a-c) aus dem folgenden linearen Gleichungssystem berechnet werden:

$$(\lambda+\mu)t_W(0) - \lambda t_W(1) = w(0|0) \quad (5.60a)$$

$$(\lambda+\mu)t_W(i) - \lambda t_W(i+1) - \mu t_W(i-1) = w(0|i), \quad i = 1, 2, \dots, s_1-2, \quad (5.60b)$$

$$(\lambda+\mu)t_W(s_1-1) - \mu t_W(s_1-2) = w(0|s_1-1) \quad (5.60c)$$

Gl.(5.60a-c) gelten entsprechend für erfolgreiche bzw. erfolglose bedingte mittlere Wartezeiten, $t_W^*(i)$ bzw. $t_W^{**}(i)$, wenn $w(0|i)$ durch $w^*(0|i)$ bzw. $w^{**}(0|i)$ ersetzt wird, $i = 0, 1, \dots, s_1-1$.

Die Lösung des Gleichungssystems (5.60a-c) kann nach der Determinantenmethode, durch Lösung einer Differenzgleichung oder rekursiv vorgenommen werden. Hier soll das letztgenannte Verfahren am allgemeineren Fall der K -ten bedingten Momente demonstriert werden, siehe folgenden Teilabschnitt d).

d) Momente der Wartezeitverteilungsfunktion

Die K -ten bedingten Momente $M_K(i)$ berechnen sich aus dem Gleichungssystem

$$(\lambda+\mu)M_K(0) - \lambda M_K(1) = KM_{K-1}(0) \quad (5.61a)$$

$$(\lambda+\mu)M_K(i) - \lambda M_K(i+1) - \mu M_K(i-1) = KM_{K-1}(i), \quad i = 1, 2, \dots, s_1-2, \quad (5.61b)$$

$$(\lambda+\mu)M_K(s_1-1) - \mu M_K(s_1-2) = KM_{K-1}(s_1-1), \quad (5.61c)$$

$K = 1, 2, \dots$

Die allgemeine Lösung des Gleichungssystems (5.61a-c) soll nach einem rekursiven Verfahren vorgenommen werden. Aus Gl.(5.61a,b) lässt sich zunächst folgende Rekursionsformel ableiten:

$$M_K(i) = M_K(0) \sum_{\nu=1}^i \xi^{\nu} - \frac{K}{\lambda} \left[\sum_{\nu=0}^{i-1} M_{K-1}(\nu) \cdot \sum_{\mu=0}^{i-\nu-1} \xi^{\mu} \right], \quad i = 1, 2, \dots, s_1-1, \quad K = 1, 2, \dots \quad (5.62a)$$

wobei $\xi = 1/\rho$. Sind alle Momente $M_K(0)$ bekannt, so können mit Gl. (5.62a) sämtliche bedingten Momente $M_K(i)$ rekursiv berechnet werden, $i = 1, 2, \dots, s_1-1$; $K = 1, 2, \dots$.

Die Momente $M_K(0)$ können entweder durch Integration aus der (bekannten) bedingten Wartezeitverteilungsfunktion $w(t|0)$ oder mit Gl. (5.62a) aus Gl.(5.61c) gewonnen werden.

Mit Hilfe der bedingten Wartezeitverteilungsfunktion (siehe Anhang A2)

$$w(t|0) = \sum_{\nu=1}^{s_1} a_{\nu} \exp(-e_{\nu} t) \quad (A2.12)$$

folgt das Moment $M_K(0)$ entsprechend Gl.(4.51) und (4.52) zu

$$M_K(0) = \sum_{\nu=1}^{s_1} a_{\nu} \cdot \frac{K!}{(-e_{\nu})^K}, \quad K = 1, 2, \dots \quad (5.62b)$$

Diese Berechnungsmethode verwendet die Kenntnis der Eigenwerte nach Gl.(5.58).

Setzt man die Rekursionsformel Gl.(5.62a) in die zu ihrer Ableitung nicht benutzte Gl.(5.61c) ein, so gewinnt man eine weitere Rekursionsformel:

$$M_K(0) = \frac{K}{\mu} \cdot \frac{1}{1-\rho^{s_1+1}} \cdot \sum_{\nu=0}^{s_1-1} M_{K-1}(\nu) \cdot [\rho^{\nu} - \rho^{s_1}], \quad K = 1, 2, \dots \quad (5.62c)$$

Zur Berechnung von $M_K(0)$ werden hierbei (anstelle der Eigenwerte) sämtliche bedingten Momente niedriger Ordnung benötigt.

Bemerkung:

Die abgeleiteten Gleichungen gelten zunächst nur für Rufe, welche nach Ankunft warten ohne Ansehen ihres späterer Schicksals. Die Gleichungen gelten ebenfalls für erfolgreich bzw. erfolglos wartende Rufe, wenn die Größen $M_0(i) = w(0|i) = 1$ durch die Größen $M_0^*(i) = w^*(0|i)$ bzw. $M_0^{**}(i) = w^{**}(0|i)$ nach Gl.(5.56b,c) ersetzt werden.

Das K-te absolute Moment der Wartezeitverteilungsfunktion folgt aus Gl.(A2.4)

$$M_K = W \cdot M_K(0), \quad K = 1, 2, \dots \quad (5.63a)$$

Für erfolgreich bzw. erfolglos wartende Rufe gilt entsprechend

$$M_K^* = W \cdot M_K^*(0), \quad (5.63b)$$

bzw.
$$M_K^{**} = W \cdot M_K^{**}(0), \quad K = 1, 2, \dots \quad (5.63c)$$

Zur Berechnung der absoluten Momente M_K , M_K^* bzw. M_K^{**} sind also nur die bedingten Momente $M_K(0)$, $M_K^*(0)$ bzw. $M_K^{**}(0)$ erforderlich.

Unter Verwendung der Rekursionsformeln Gl.(5.62a) und (5.62c) können leicht geschlossene Ausdrücke für die Momente abgeleitet werden. Als Beispiel seien die ersten beiden Momente der Wartezeitverteilungsfunktion alle Rufe entspr. Gl.(5.63a) angegeben:

$$M_1 = p(n_1, 0) \frac{1}{\mu} \cdot \frac{1}{(1-\rho)^2} [1 - (s_1+1)\rho^{s_1} + s_1 \rho^{s_1+1}], \quad (5.64a)$$

$$M_2 = p(n_1, 0) \frac{2}{\mu^2} \cdot \frac{1}{(1-\rho)^2} \cdot \frac{1}{1-\rho^{s_1+1}} \left[\frac{1+s_1 \rho^{s_1+1} - (s_1+1)\rho^{s_1}}{(1-\rho)^2} - \rho^{s_1+2} \frac{1+(s_1+1)\rho^{s_1} - s_1 \rho^{s_1+1}}{(1-\rho)^2} - \frac{1}{2} s_1 (s_1+1) \rho^{s_1} (3+\rho^{s_1+1}) + (2s_1+1) \rho^{s_1+1} \frac{1-\rho^{s_1}}{1-\rho} \right]. \quad (5.64b)$$

Unter Beachtung obiger Bemerkung können entsprechende Ausdrücke auch für erfolgreich bzw. erfolglos wartende Rufe abgeleitet werden. M_1/W nach Gl.(5.64a) und Gl.(5.15b) stimmt mit der mittleren Wartezeit bezogen auf alle Wartenden überein, welche in Gl.(5.21b) mit Hilfe der mittleren Warteschlangenlänge berechnet wurde.

5.3.4 Abfertigung in der inversen Reihenfolge des Eintreffens mit nichtverdrängender Priorität (D3b)

a) Der Warteprozess

Ein eintreffender Ruf belegt grundsätzlich den ersten Warteplatz, wenn alle Leitungen blockiert sind und mindestens noch ein freier Warteplatz verfügbar ist. Sind bei Ankunft eines Rufes alle Warteplätze belegt, so wird dieser Ruf abgewiesen (Verlustruf). Ein einmal akzeptierter Ruf wird nicht mehr verdrängt und mit Sicherheit abgefertigt.

Dem Testruf nachfolgende Rufe werden sich - sofern sie akzeptiert werden - in der Warteschlange vor dem Testruf einordnen. Da diese Rufe mit Sicherheit vor dem Testruf abgefertigt werden, beeinflussen sie direkt die Wartezeit des Testrufs. Die bei Ankunft des Testrufs angetroffenen Wartenden warten während des ganzen Warteprozesses in der Warteschlange hinter dem Testruf. Diese Rufe werden mit Sicherheit erst nach dem Testruf abgefertigt, beeinflussen

durch ihre Speicherbelegung jedoch, wieviele neue Rufe akzeptiert werden. Sie sind also indirekt beeinflussende Rufe.

Für jede Zahl z der bei Ankunft des Testrufs angetroffenen Wartenden entwickelt sich ein spezifischer Warteprozeß. Der Testruf kann nur noch Plätze innerhalb der (s_1-z) Stellen des verbleibenden Wartespeichers einnehmen.

Als Zustandsmuster werde definiert

$$\xi(t) = (i, z), \quad \begin{matrix} i = 0, 1, \dots, s_1-z-1, \\ z = 0, 1, \dots, s_1-1, \end{matrix} \quad (5.65)$$

wobei i die Zahl der Wartenden angibt, welche zur Zeit t vor dem Testruf in der Warteschlange warten. Für die Antreffwahrscheinlichkeit gilt

$$P(0, z) = p(n; z), \quad z = 0, 1, \dots, s_1-1. \quad (5.66)$$

Bild 14 zeigt den Wartezustandsraum des Warteprozesses für eine allgemeine Zahl z von angetroffenen Wartenden.

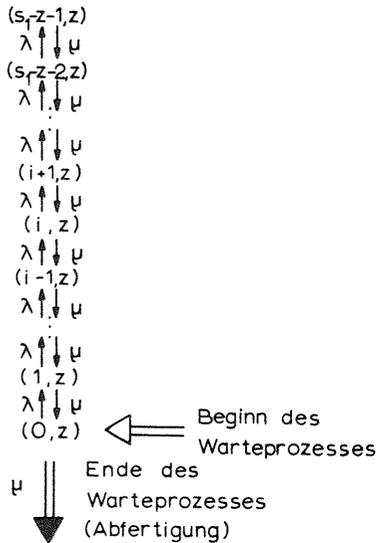


Bild 14. Wartezustandsraum für Warten nach D3b

Die Sprungkoeffizienten lauten

$$q(i, z; j, z) = \begin{cases} \lambda & \text{für } j = i+1, i = 0, 1, \dots, s_1-z-2 \\ \mu & \text{für } j = i-1, i = 1, 2, \dots, s_1-z-1 \\ 0 & \text{für } j > i+1 \text{ bzw. } j < i-1 \end{cases} \quad (5.67a)$$

Der Zustand (i, z) ändert sich mit dem Koeffizienten

$$q(i, z) = \begin{cases} \lambda + \mu & \text{für } i = 0, 1, \dots, s_1-z-2 \\ \mu & \text{für } i = s_1-z-1 \end{cases} \quad (5.67b)$$

b) Wartezeitverteilungsfunktionen

$w(t|i, z)$ sei die bedingte Wartezeitverteilungsfunktion für Warten vom $(i+1)$ -ten Warteplatz aus, unter der Bedingung, daß der Testruf bei Ankunft z Wartende vorgefunden hat.

Die bedingten Wartezeitverteilungsfunktionen $w(t|i, z)$ gehorchen entsprechend Gl.(4.32) mit Gl.(5.67a,b) dem folgenden System von Differentialgleichungen:

$$\frac{dw(t|0, z)}{dt} = -(\lambda + \mu)w(t|0, z) + \lambda w(t|1, z) \quad (5.68a)$$

$$\frac{dw(t|i, z)}{dt} = -(\lambda + \mu)w(t|i, z) + \lambda w(t|i+1, z) + \mu w(t|i-1, z), \quad i = 1, 2, \dots, s_1-z-2, \quad (5.68b)$$

$$\frac{dw(t|s_1-z-1, z)}{dt} = -\mu w(t|s_1-z-1, z) + \mu w(t|s_1-z-2, z). \quad (5.68c)$$

Gl.(5.68a-c) gelten für $z = 0, 1, \dots, s_1-3$; für $z = s_1-2$ und $z = s_1-1$ ergeben sich entsprechend einfachere Systeme.

Da ein akzeptierter Ruf mit Sicherheit abgefertigt wird, gilt als Anfangsbedingung $w(0|i, z) = 1, i = 0, 1, \dots, s_1-z-1, z = 0, 1, \dots, s_1-1$. Dieses folgt außerdem aus Gl.(5.68a-c) für die Endkoeffizienten

$$\varepsilon(i, z) = \begin{cases} \mu & \text{für } i = 0, z = 0, 1, \dots, s_1-1 \\ 0 & \text{für } i > 0, z = 0, 1, \dots, s_1-i-1 \end{cases} \quad (5.69)$$

Das Gl.(5.68a-c) entsprechende transformierte System ist mit Gl.(5.70) in Matrixdarstellung angegeben. Der Typ der Matrix A des Systems Gl.(5.70) stimmt - bis auf das letzte Glied der Hauptdiagonale - mit dem Typ der Matrix nach Gl.(5.57) überein. Die Unsymmetrie im letzten Glied der Hauptdiagonale verhindert die geschlossene Auflösung der Eigenwerte dieser Matrix.

$$\begin{array}{cccc|c}
 (0,z) & (1,z) & (2,z) & \dots & (s_1-z-2,z) & (s_1-z-1,z) & \\
 \hline
 (s+\lambda+\mu) & -\lambda & & & & & 1 \\
 -\mu & (s+\lambda+\mu) & -\lambda & & & & 1 \\
 & -\mu & (s+\lambda+\mu) & -\lambda & & & 1 \\
 & & \ddots & \ddots & \ddots & & \vdots \\
 & & & -\mu & (s+\lambda+\mu) & -\lambda & 1 \\
 & & & & -\mu & (s+\mu) & 1
 \end{array}
 \quad , N = s_1 - z. \quad (5.70)$$

Im Anhang A3 werden drei Eigenschaften dieser Matrix bewiesen:

1. Die Eigenwerte $e_{z,\nu}$ sind negativ-reell.
2. Die Eigenwerte sind verschieden, d.h. $N' = N = s_1 - z$, $K_{\nu} = 1$, $\nu = 1, 2, \dots, N$.
3. Die Eigenwerte liegen alle im Intervall $[-2(\lambda+\mu), 0)$, $s_1 > 2$.

Für die Sonderfälle $s_1 = 1$ und $s_1 = 2$ sind die Eigenwerte wie bei D2 explizit angebar:

$$\begin{aligned}
 e_{01} &= -\mu, & s_1 &= 1, z = 0 \\
 e_{01,2} &= -\frac{1}{2} \left[(\lambda+2\mu) \mp \sqrt{\lambda^2+4\lambda\mu} \right], & s_1 &= 2, z = 0 \\
 e_{11} &= -\mu, & s_1 &= 2, z = 1.
 \end{aligned}$$

Für $s_1 > 2$ müssen die Eigenwerte wieder mittels einer numerischen Rechnung bestimmt werden.

Entsprechend den ersten zwei Eigenschaften der Matrix **A** stellen sich die bedingten Wartezeitverteilungsfunktionen wie folgt dar:

$$w(t|i,z) = \sum_{\nu=1}^{s_1-z} a_{i,z,\nu} \exp(e_{z,\nu} t), \quad \begin{matrix} i = 0, 1, \dots, s_1-z-1, \\ z = 0, 1, \dots, s_1-1. \end{matrix} \quad (5.71)$$

Die absolute Verteilungsfunktion der Gesamtwartezeiten bezüglich aller Rufe ist nach Gl.(4.39) mit Gl.(5.66)

$$W(>t) = \sum_{z=0}^{s_1-1} p(n;z) w(t|0,z) \quad (5.72)$$

c) Mittlere Wartezeiten

Für die bedingten mittleren Wartezeiten $t_w(i,z)$ gilt entsprechend Gl.(4.58), (4.44) und (5.68a-c) das lineare Gleichungssystem Gl.(5.73a-c).

$$(\lambda+\mu)t_w(0,z) - \lambda t_w(1,z) = 1 \quad (5.73a)$$

$$(\lambda+\mu)t_w(i,z) - \lambda t_w(i+1,z) - \mu t_w(i-1,z) = 1, \quad i=1,2,\dots,s_1-z-2, \quad (5.73b)$$

$$\mu t_w(s_1-z-1,z) - \mu t_w(s_1-z-2,z) = 1. \quad (5.73c)$$

Nach einer längeren Rechnung erhält man als Lösung der Gl.(5.73a-c)

$$t_w(i,z) = \frac{1}{\mu} \cdot \frac{1}{(1-\rho)^2} \cdot \left[(i+1)(1-\rho) + \rho^{s_1-z-i} \cdot (\rho^{i+1}-1) \right]. \quad (5.74)$$

Die mittlere Wartezeit der wartenden Rufe berechnet sich entsprechend Gl.(4.60) und Gl.(5.72) zu

$$t_w = \frac{1}{W} \sum_{z=0}^{s_1-1} p(n;z) t_w(0,z) = \frac{1}{\mu} \left[\frac{1}{1-\rho} - \frac{s_1 \rho^{s_1}}{1-\rho^{s_1}} \right]. \quad (5.75)$$

t_w wurde bereits über die mittlere Warteschlangenlänge in Gl.(5.21a) berechnet und stimmt mit Gl.(5.75) überein.

d) Momente der Wartezeitverteilungsfunktion

Die K-ten bedingten Momente $M_K(i,z)$ gehorchen dem folgenden linearen Gleichungssystem:

$$(\lambda+\mu)M_K(0,z) - \lambda M_K(1,z) = K M_{K-1}(0,z) \quad (5.76a)$$

$$(\lambda+\mu)M_K(i,z) - \lambda M_K(i+1,z) - \mu M_K(i-1,z) = K M_{K-1}(i,z), \quad i=1,2,\dots,s_1-z-2, \quad (5.76b)$$

$$\mu M_K(s_1-z-1,z) - \mu M_K(s_1-z-2,z) = K M_{K-1}(s_1-z-1,z), \quad K=1,2,\dots \quad (5.76c)$$

Für $K = 1$ stimmt das Gleichungssystem (5.76a-c) mit dem Gleichungssystem für die bedingten mittleren Wartezeiten (5.73a-c) überein, denn es gilt wegen Gl.(4.58) $t_w(i,z) = M_1(i,z)$.

Ausgehend von den bekannten ersten bedingten Momenten $M_1(i,z)$ können aus Gl.(5.76a-c) die zweiten bedingten Momente $M_2(i,z)$ berechnet werden:

$$M_2(i,z) = M_2(0,z) \sum_{\nu=0}^i \xi^{\nu} - \frac{2}{\lambda} \left[\sum_{\nu=0}^{i-1} M_1(\nu,z) \cdot \sum_{x=0}^{i-\nu-1} \xi^x \right], \quad \xi = 1/\rho, \quad (5.77a)$$

wobei

$$M_2(0,z) = \frac{2}{\mu^2} \cdot \frac{1}{(1-\rho)^2} \left[\frac{1-\rho^{2(s_1-z)+1}}{1-\rho} - [2(s_1-z)+1] \cdot \rho^{s_1-z} \right] \quad (5.77b)$$

Ausgehend von den bedingten Momenten $M_2(i,z)$ können die bedingten Momente $M_3(i,z)$ berechnet werden usw.

Das K-te absolute Moment der Wartezeitverteilungsfunktion ist entsprechend Gl.(5.72)

$$M_K = \sum_{z=0}^{s_1-1} p(n,z) M_K(0,z), \quad K=1,2,\dots \quad (5.78)$$

Das Moment M_1 stimmt mit den entsprechenden ersten Momenten für D1 und D2 überein. Für $K=2$ erhält man mit Gl.(5.77a,b) und (5.78) den Ausdruck

$$M_2 = p(n,0) \cdot \frac{2}{\mu^2} \cdot \frac{1}{(1-\rho)^2} \left[\frac{1-\rho^{s_1} + \rho^{2s_1+2} - \rho^{s_1+2}}{(1-\rho)^2} - s_1(s_1+2)\rho^{s_1} \right]. \quad (5.79)$$

5.4 Grenzfälle der mittleren Anrufrate

Für die zwei Sonderfälle $\lambda \rightarrow 0$ und $\lambda \rightarrow \infty$ können weitere Aussagen über die Eigenwerte, und damit über die Wartezeitverteilungsfunktion, gemacht werden.

5.4.1 Verschwindende Anrufrate

Für $\lambda \rightarrow 0$ fallen die Eigenwerte der Matrizen A für D2 nach Gl.(5.42), D3a nach Gl.(5.57) und D3b nach Gl.(5.70) mit dem s_1 -fachen Eigenwert $e_1 = -\mu$ der Matrix A nach Gl.(5.27) für D1 zusammen. Die Verteilungsfunktion der Wartezeiten der Wartenden wird einheitlich

$$\frac{W(>t)}{W} \rightarrow \exp(-\mu t). \quad (5.80)$$

Das Ergebnis Gl.(5.80) läßt sich anschaulich interpretieren: Für $\lambda \rightarrow 0$ wird - wenn überhaupt - höchstens ein Ruf warten müssen. Für einen einzelnen Wartenden spielt aber die Abfertigungsreihenfolge keine Rolle, weil er nur mit Wahrscheinlichkeit Null zurückgedrängt wird bzw. einen Konkurrenten bekommt. Die Wartezeitverteilungsfunktion der Wartenden ist somit identisch mit der Verteilungsfunktion der Phase, bis eine der n belegten Leitungen frei wird (Exponentialverteilung).

5.4.2 Unbegrenzt wachsende Anrufrate

Bei der Disziplin D1 bleibt der s_1 -fache Eigenwert $e_1 = -\mu$ für $\lambda \rightarrow \infty$ erhalten, da die Matrix A nach Gl.(5.27) nicht von λ abhängt. Aus Gl.(5.32) folgt durch Grenzübergang $\rho \rightarrow \infty$

$$\frac{W(>t)}{W} \rightarrow \exp(-\mu t) \cdot \sum_{r=0}^{s_1-1} \frac{(\mu t)^r}{r!}. \quad (5.81)$$

Gl.(5.81) stellt offenbar eine ERLANG- s_1 -Verteilung dar. Ein einfallender Ruf wird - wenn überhaupt - höchstens den letzten (s_1 -ten) Warteplatz belegen können. Er wird dann auch genau s_1 Endungen, d.h. s_1 negativ-exponentiell verteilte Phasen mit gleichem Mittelwert $1/\mu$ abwarten müssen. Die Verteilungsfunktion hierfür ist aber gerade die ERLANG- s_1 -Verteilung.

Für die Disziplin D2 folgt aus der Matrix A nach Gl.(5.42) für den Grenzübergang $\lambda \rightarrow \infty$, daß für den ersten Eigenwert $e_1 = -\mu/s_1$ gilt, während alle weiteren Eigenwerte betragsmäßig über alle Grenzen wachsen. Es folgt hieraus

$$\frac{W(>t)}{W} \rightarrow \exp\left(-\frac{\mu}{s_1} t\right). \quad (5.82)$$

Die gleichbleibende Konkurrenz von (s_1-1) Mitwartenden wirkt sich so auf die Wartezeit des Testrufs aus, wie wenn der Testruf allein warten würde, die Endedichte der Leitungen aber nur den s_1 -ten Teil betragen würde.

Bei der Disziplin D3a wachsen sämtliche Eigenwerte für $\lambda \rightarrow \infty$ betragsmäßig über alle Grenzen. Die Verteilungsfunktion der erfolgreich wartenden Rufe verschwindet identisch, und es gilt für die erfolglos wartenden Rufe

$$\frac{W(>t)}{W} = \frac{W^{**}(>t)}{W^{**}} \rightarrow \exp(-\alpha t) \Big|_{\alpha \rightarrow \infty} \quad (5.83)$$

wobei $W = W^{**} \rightarrow 1$. Jeder neue Rufe belegt in diesem Falle den ersten Warteplatz, wird aber mit Wahrscheinlichkeit 1 durch nachfolgende Rufe aus dem Wartespeicher verdrängt.

Bei der Disziplin D3b entwickelte sich für jede Zahl z angetroffener Wartender ein spezieller Warteprozess. Mit Hilfe von Matrix Gl.(5.70) kann gezeigt werden, daß für $\lambda \rightarrow \infty$ in allen Fällen $z = 0, 1, \dots, s_1-2$ jeweils ein Eigenwert gegen Nullstrebt, während alle anderen Eigenwerte über alle Grenzen wachsen. Im Falle $z = s_1-1$ hingegen bleibt der einzige Eigenwert $e_{s_1-1,1} = -\mu$ erhalten. Für $\lambda \rightarrow \infty$ wird der Fall $z = s_1-1$ dominierend sein, so daß für endliche Wartezeiten

$$\frac{W(>t)}{W} \rightarrow \exp(-\mu t) \quad (5.84)$$

zu erwarten ist. Dies ist dasselbe Ergebnis wie für $\lambda \rightarrow 0$, da es sich auch in beiden Fällen um Warten vom 1. Warteplatz aus handelt.

Nun folgt aber aus Gl.(5.21a), daß die mittlere Wartezeit der Wartenden, welche nach Gl.(4.60) der Fläche unter der Wartezeitverteilungsfunktion der Wartenden gleich ist, für $S \rightarrow \infty$ gegen den oberen Grenzwert s_1/μ strebt. Für D1 bzw. D2 folgt dieser Grenzwert durch Integration der Gl.(5.81) bzw. (5.82). Für D3b erhält man aus Gl.(5.84) nur die mittlere Wartezeit der Rufe, welche vom 1. Platz aus warten. Die (s_1-1) Wartenden des 2. bis s_1 -ten Warteplatzes verursachen jedoch praktisch unendlich große Wartezeiten, aber mit verschwindenden Antreffwahrscheinlichkeiten, derart, daß die mittlere Wartezeit aller Wartenden gegen s_1/μ strebt. Durch eine genauere Grenzwertbetrachtung, welche den Einfluß der verschwindenden Eigenwerte berücksichtigt, läßt sich dies auch analytisch nachweisen.

5.5 Numerische Beispiele

Der Einfluß des Angebots und der Abfertigungsdisziplinen auf die charakteristischen Größen sowie die Wartezeitverteilungsfunktionen soll am Beispiel eines Bedienungssystems mit $n = 1$ Leitung und $s_1=12$ Warteplätzen demonstriert werden.

In Tabelle 1, S.94, sind für die vier Abfertigungsdisziplinen D1,D2, D3a und D3b die charakteristischen Größen angegeben. Es wurden dabei die Angebotswerte $A = 0,8$ (Normalbelastung), $A = 1,0$ (Grenzbelastung) und $A = 1,5$ (Überbelastung) zugrunde gelegt.

Die drei Abfertigungsdisziplinen D1, D2 und D3b besitzen denselben Mechanismus bezüglich der Akzeptierung von Rufen und können deshalb miteinander verglichen werden. Wie aus Gl.(5.21a) einerseits und Gl. (5.37), (5.49) sowie (5.78) andererseits folgt, stimmen die mittleren Wartezeiten, und damit die ersten Momente, überein:

$$M_1^{(D1)} = M_1^{(D2)} = M_1^{(D3b)}. \tag{5.85}$$

Die Abfertigungsdisziplin drückt sich in den Unterschieden der zweiten und höheren Momente aus:

$$M_K^{(D1)} < M_K^{(D2)} < M_K^{(D3b)}, \quad K \geq 2. \tag{5.86}$$

Die linke Beziehung dieser Ungleichung erfuhr für $K = 2$ in Ungl. (5.50c) ihren Beweis, die rechte Beziehung kann mit Hilfe der Lösung für $M_2^{(D3b)}$ nach Gl.(5.79) ebenfalls nachgewiesen werden. Die Streuungen der Wartezeiten nehmen von D1 über D2 nach D3b zu, was auch schon allgemein gezeigt wurde [12, 13].

In den Bildern 15a,b, S. 95 und 96, sind die auf die jeweiligen Anfangswerte W, W^* , bzw. W^{**} normierten Wartezeitverteilungsfunktionen dargestellt mit den Parametern $A = 0,8$, $A = 1,0$ und $A = 1,5$ bzw. D1, D2 (Bild 15a) und D3a, D3b (Bild 15b). Bei D3a sind die Wartezeitverteilungsfunktionen für die erfolgreich und erfolglos Wartenden angegeben.

Für D1 und D2 erkennt man aus Bild 15a die Zunahme der Wartezeitverteilungsfunktion mit dem Angebot A . Für D3a läßt sich aus Bild 15b der Effekt feststellen, daß die Wartezeitverteilungsfunktion der erfolgreich Wartenden für mittlere Angebotswerte am höchsten liegt. Für D3b stellt man für kleine Wartezeiten denselben Effekt fest; wegen der Zunahme der mittleren Wartezeit der Wartenden mit dem Angebot müssen die Kurven mit höherem Angebotswert für größere Wartezeiten jedoch am höchsten liegen (Im Beispiel Bild 15b schneiden sich die Kurven für $A = 1,0$ und $A = 1,5$ bei $t/t_S \approx 30$).

In Bild 15c, S. 97, sind für den Angebotswert $A = 1,0$ die Wartezeitverteilungsfunktionen für alle Abfertigungsdisziplinen D1 bis D3b aufgetragen. Man erkennt deutlich, daß die Wartezeiten nach der durch Ungleichung (5.86) beschriebenen Tendenz streuen. Für kleine Wartezeiten garantieren die Abfertigungsdisziplinen nach der inversen Ankunftsreihenfolge raschen Erfolg, während in Bezug auf große Wartezeiten die Abfertigung in Ankunftsreihenfolge am günstigsten ist. Die zufallsmäßige Abfertigungsreihenfolge nimmt in jeder Hinsicht eine Mittelstellung zwischen diesen beiden Extremen ein.

Einen interessanten Aspekt weisen die Eigenwertverteilungen für die verschiedenen Angebotswerte und Abfertigungsdisziplinen auf, Bild 16, S. 98. Wie schon durch die Eigenwertschranken nach dem Theorem von S. GERSCHGORIN gezeigt wurde, nimmt für D2, D3a und D3b die Streuung der Eigenwerte mit wachsendem Angebot zu. Außerdem nimmt die Streuung der Eigenwerte von D1 über D2 nach D3a bzw. D3b zu. Aus Bild 16 ist weiter zu ersehen, daß das Theorem von S. GERSCHGORIN eine relativ enge Einschränkung für die Lage der Eigenwerte auf der negativ-reellen Achse liefert.

Die Berechnung der Momente der Wartezeitverteilungsfunktion erfolgte bei D1 direkt nach der Lösung Gl.(5.37), bei D2, D3a bzw. D3b durch iterative Auflösung der linearen inhomogenen Gleichungssysteme (5.47a,b), (5.61a,b,c) bzw. (5.76a,b,c) nach dem Verfahren der

"Sukzessiven Überrelaxation" [19]. Die Lösungen dieser Gleichungssysteme, d.h. die bedingten Momente, weisen mit zunehmendem Wert K sehr große Unterschiede auf, so daß i.a. sehr viele Iterationsschritte für eine vorgeschriebene Genauigkeit erforderlich sind. Außerdem beobachtet man eine starke Empfindlichkeit der Konvergenz bezüglich des Relaxationsfaktors.

Zur "exakten" Berechnung der Eigenwerte der (unsymmetrischen) Matrizen wurde ein iteratives Verfahren angewendet, welches die Matrix sukzessive auf Diagonalform transformiert. In Erweiterung zum klassischen JACOBI'schen Verfahren, welches symmetrische Matrizen durch elementare orthogonale zweidimensionale Drehungen auf Diagonalform transformiert [19], wurde von P.J. EBERLEIN und J. BOOTHROYD [20] eine Methode angegeben, welche durch elementare zweidimensionale Drehungen und Scherungen eine unsymmetrische Matrix auf Diagonalform transformiert. Mit Hilfe dieses Verfahrens wurden die Eigenwerte und damit die Wartezeitverteilungsfunktionen berechnet. Das Verfahren ist verhältnismäßig aufwendig und konvergiert für extrem unterschiedliche Matrixelemente entweder nur sehr langsam oder überhaupt nicht.

Auf approximative Lösungsverfahren wird im Kapitel 8 noch näher eingegangen.

Bemerkung

Abschließend sei bemerkt, daß das reine Wartesystem aus dem kombinierten Warte-Verlustsystem durch Grenzübergang $s_1 \rightarrow \infty$ erhalten wird. Aus Stationaritätsgründen muß dann $\rho < 1$ gefordert werden.

Durch den Grenzübergang $s_1 \rightarrow \infty$ können einige Erleichterungen erreicht werden, insbesondere verschwindet der Unterschied zwischen den Disziplinen D3a und D3b.

Das reine Wartesystem ist in der Literatur bereits intensiv untersucht worden, siehe beispielsweise J. RIORDAN [24].

Charakteristische Größen	Angebot A	Abfertigungsdisziplinen			
		D1	D2	D3a	D3b
Verlustwahrscheinl. B	0,8	1,150 ₁₀ -2	1,150 ₁₀ -2	0	1,150 ₁₀ -2
	1,0	7,143 ₁₀ -2	7,143 ₁₀ -2	0	7,143 ₁₀ -2
	1,5	3,345 ₁₀ -1	3,345 ₁₀ -1	0	3,345 ₁₀ -1
Verkehrswert Y	0,8	7,908 ₁₀ -1	7,908 ₁₀ -1	7,908 ₁₀ -1	7,908 ₁₀ -1
	1,0	9,286 ₁₀ -1	9,286 ₁₀ -1	9,286 ₁₀ -1	9,286 ₁₀ -1
	1,5	9,983 ₁₀ -1	9,983 ₁₀ -1	9,983 ₁₀ -1	9,983 ₁₀ -1
Wartewahrscheinl. W	0,8	7,793 ₁₀ -1	7,793 ₁₀ -1	7,908 ₁₀ -1	7,793 ₁₀ -1
	1,0	8,571 ₁₀ -1	8,571 ₁₀ -1	9,286 ₁₀ -1	8,571 ₁₀ -1
	1,5	6,638 ₁₀ -1	6,638 ₁₀ -1	9,983 ₁₀ -1	6,638 ₁₀ -1
Erfolgswartewahrscheinl. W*	0,8	7,793 ₁₀ -1	7,793 ₁₀ -1	7,793 ₁₀ -1	7,793 ₁₀ -1
	1,0	8,571 ₁₀ -1	8,571 ₁₀ -1	8,571 ₁₀ -1	8,571 ₁₀ -1
	1,5	6,638 ₁₀ -1	6,638 ₁₀ -1	6,638 ₁₀ -1	6,638 ₁₀ -1
Misserfolgswahrscheinl. W**	0,8	0	0	1,150 ₁₀ -2	0
	1,0	0	0	7,143 ₁₀ -2	0
	1,5	0	0	3,345 ₁₀ -1	0
Mittlere Warteschlangenlänge Ω	0,8	2,565 ₁₀ +0	2,565 ₁₀ +0	2,565 ₁₀ +0	2,565 ₁₀ +0
	1,0	5,571 ₁₀ +0	5,571 ₁₀ +0	5,571 ₁₀ +0	5,571 ₁₀ +0
	1,5	1,005 ₁₀ +1	1,005 ₁₀ +1	1,005 ₁₀ +1	1,005 ₁₀ +1
1. Moment, Mittlere Wartezeit $\frac{M_1}{W} = t_W$	0,8	4,114 ₁₀ +0	4,114 ₁₀ +0	4,055 ₁₀ +0	4,114 ₁₀ +0
	1,0	6,500 ₁₀ +0	6,500 ₁₀ +0	6,000 ₁₀ +0	6,500 ₁₀ +0
	1,5	1,009 ₁₀ +1	1,009 ₁₀ +1	6,711 ₁₀ +0	1,009 ₁₀ +1
Mittlere erfolgr. Wartezeit $\frac{M_1^*}{W^*} = t_W^*$	0,8	4,114 ₁₀ +0	4,114 ₁₀ +0	3,708 ₁₀ +0	4,114 ₁₀ +0
	1,0	6,500 ₁₀ +0	6,500 ₁₀ +0	4,167 ₁₀ +0	6,500 ₁₀ +0
	1,5	1,009 ₁₀ +1	1,009 ₁₀ +1	1,896 ₁₀ +0	1,009 ₁₀ +1
Mittlere erfolgl. Wartezeit $\frac{M_1^{**}}{W^{**}} = t_W^{**}$	0,8	0	0	2,756 ₁₀ +1	0
	1,0	0	0	2,800 ₁₀ +1	0
	1,5	0	0	1,627 ₁₀ +1	0
2. Moment $\frac{M_2}{W}$	0,8	2,963 ₁₀ +1	4,670 ₁₀ +1	9,390 ₁₀ +1	1,150 ₁₀ +2
	1,0	6,067 ₁₀ +1	1,019 ₁₀ +2	1,820 ₁₀ +2	3,943 ₁₀ +2
	1,5	1,168 ₁₀ +2	2,084 ₁₀ +2	1,281 ₁₀ +2	3,978 ₁₀ +3
2. Moment (erfolgreich) $\frac{M_2^*}{W^*}$	0,8	2,963 ₁₀ +1	4,670 ₁₀ +1	7,994 ₁₀ +1	1,150 ₁₀ +2
	1,0	6,067 ₁₀ +1	1,019 ₁₀ +2	1,050 ₁₀ +2	3,943 ₁₀ +2
	1,5	1,168 ₁₀ +2	2,084 ₁₀ +2	1,856 ₁₀ +1	3,978 ₁₀ +3
2. Moment (erfolglos) $\frac{M_2^{**}}{W^{**}}$	0,8	0	0	1,040 ₁₀ +3	0
	1,0	0	0	1,106 ₁₀ +3	0
	1,5	0	0	3,454 ₁₀ +2	0
3. Moment $\frac{M_3}{W}$	0,8	2,833 ₁₀ +2	9,182 ₁₀ +2	4,193 ₁₀ +3	7,014 ₁₀ +3
	1,0	6,825 ₁₀ +2	2,564 ₁₀ +3	9,282 ₁₀ +3	5,323 ₁₀ +4
	1,5	1,450 ₁₀ +3	6,488 ₁₀ +3	3,416 ₁₀ +3	5,530 ₁₀ +6
3. Moment (erfolgreich) $\frac{M_3^*}{W^*}$	0,8	2,833 ₁₀ +2	9,182 ₁₀ +2	3,483 ₁₀ +3	7,014 ₁₀ +3
	1,0	6,825 ₁₀ +2	2,564 ₁₀ +3	5,107 ₁₀ +3	5,323 ₁₀ +4
	1,5	1,450 ₁₀ +3	6,488 ₁₀ +3	3,915 ₁₀ +2	5,530 ₁₀ +6
3. Moment (erfolglos) $\frac{M_3^{**}}{W^{**}}$	0,8	0	0	5,232 ₁₀ +4	0
	1,0	0	0	5,938 ₁₀ +4	0
	1,5	0	0	9,420 ₁₀ +3	0

Tabelle 1. Charakteristische Größen in Abhängigkeit von Angebot und Abfertigungsdisziplinen
 System: $g = 1, n = 1, s_1 = 12$
 (Wartezeiten und Momente sind mit der mittleren Belegungsdauer normiert)

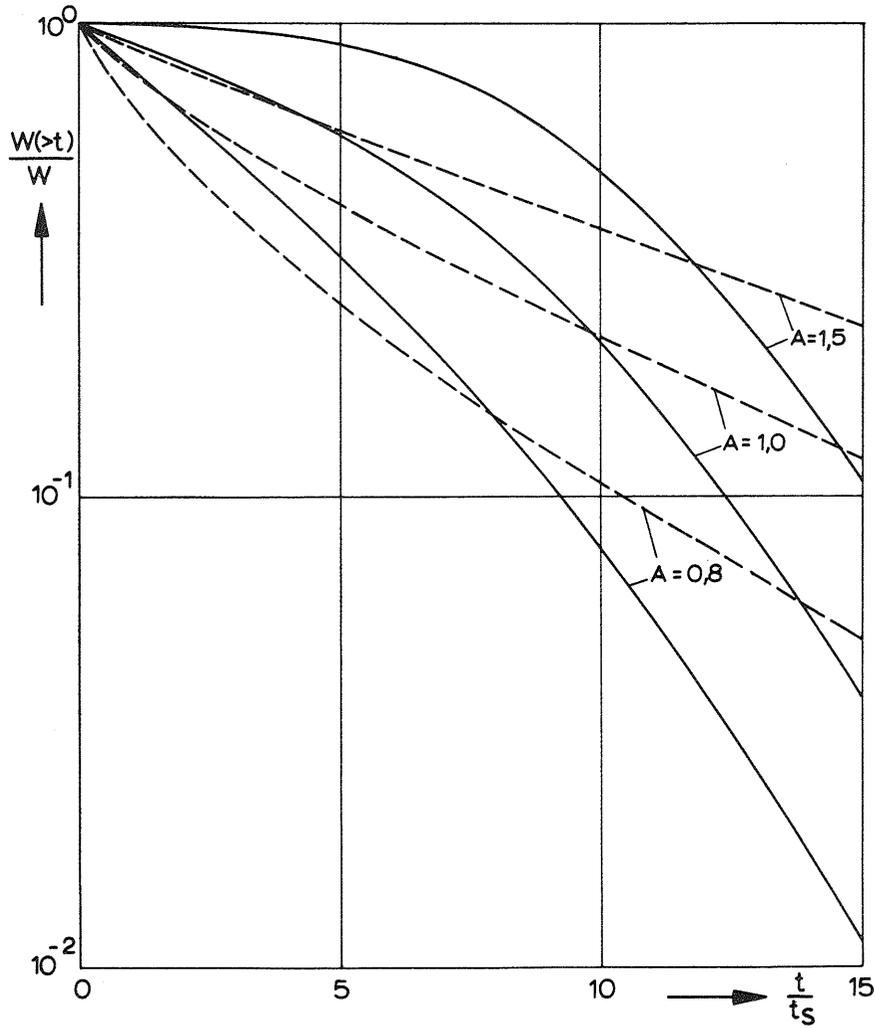


Bild 15a. Wartezeitverteilungsfunktionen für verschiedene Abfertigungsdisziplinen

System: $g = 1, n = 1, s_1 = 12$
 Parameter: Angebot A

- Abfertigung in der Reihenfolge des Eintreffens (D1)
- Abfertigung in zufälliger Reihenfolge (D2)

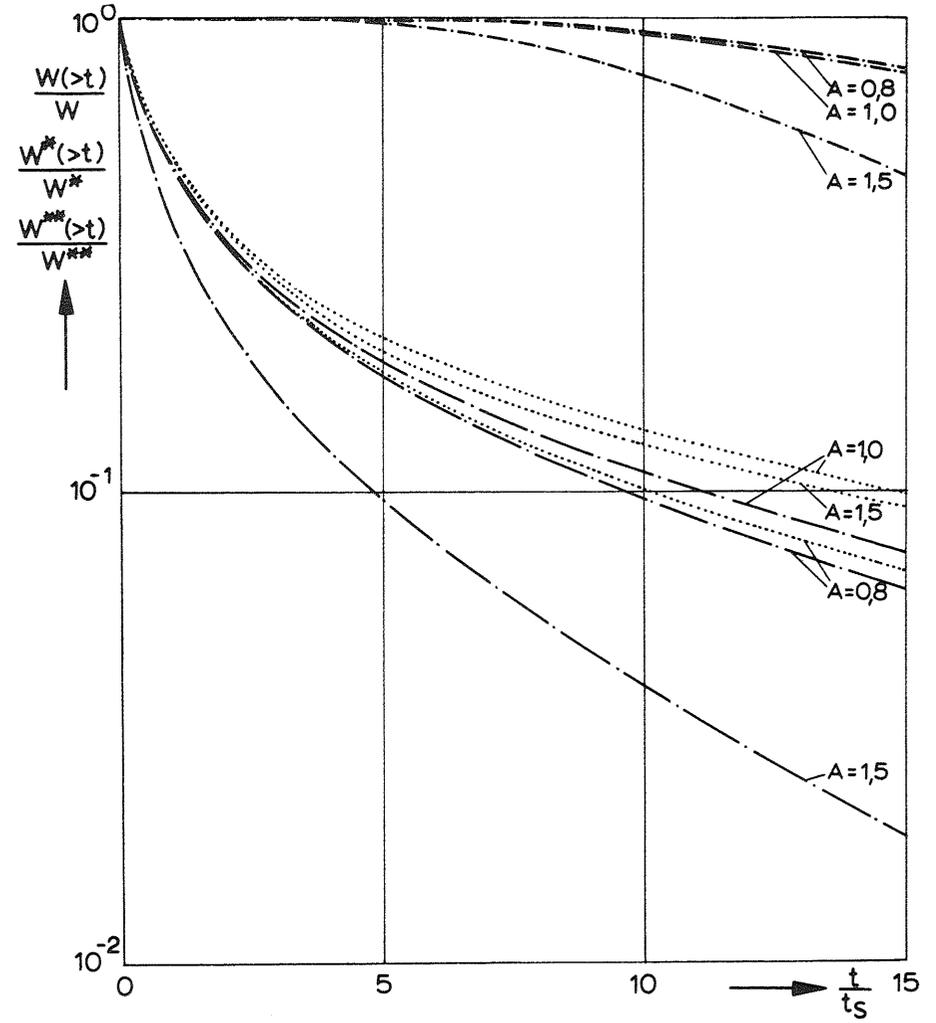


Bild 15b. Wartezeitverteilungsfunktionen für verschiedene Abfertigungsdisziplinen

System: $g = 1, n = 1, s_1 = 12$
 Parameter: A

- Abfertigung in der inversen Reihenfolge des Eintreffens (D3)
- mit verdrängender Priorität (D3a)
- erfolgreich wartende Rufe (*)
- erfolglos wartende Rufe (**)
- mit nichtverdrängender Priorität (D3b)

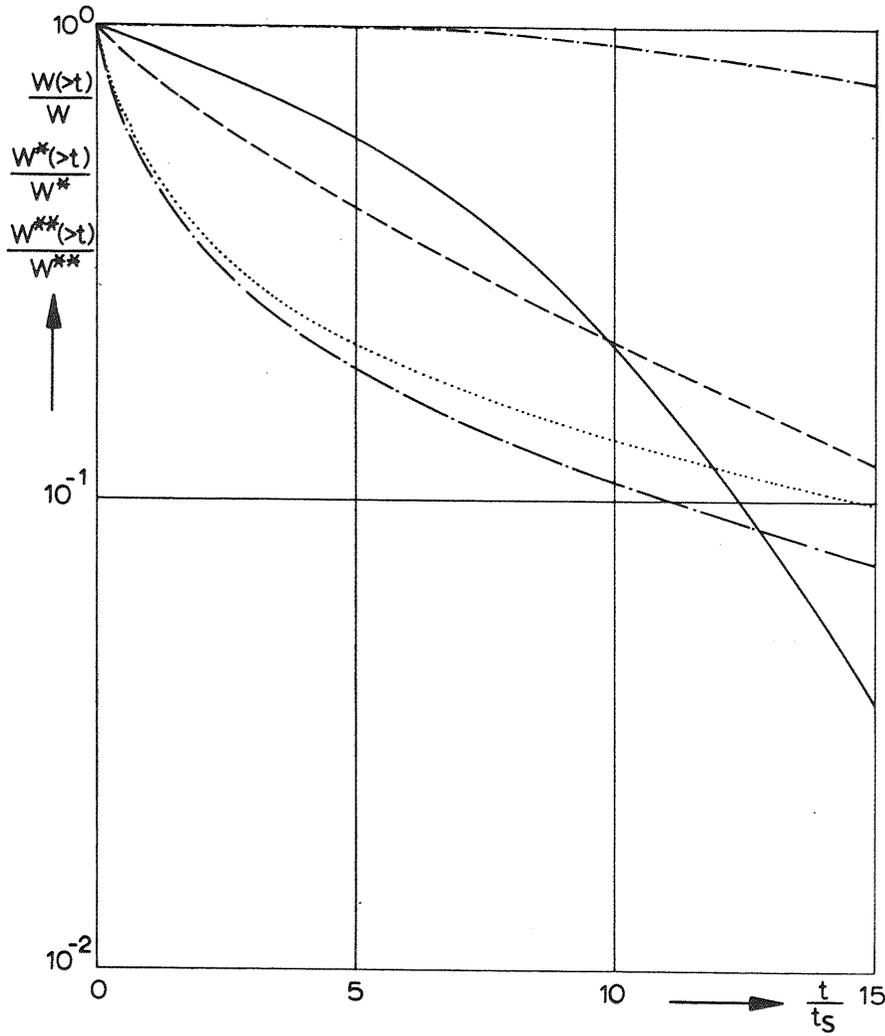


Bild 15c. Wartezeitverteilungsfunktionen für verschiedene Abfertigungsdisziplinen

System: $g = 1, n = 1, s_1 = 12$; Angebot $A = 1,0$

- Abfertigung in der Reihenfolge des Eintreffens (D1)
- - - Abfertigung in zufälliger Reihenfolge (D2)
- Abfertigung in der inversen Reihenfolge des Eintreffens mit verdrängender Priorität (D3a)
- *— erfolgreich wartende Rufe (*)
- - *— erfolglos wartende Rufe (**)
- Abfertigung in der inversen Reihenfolge des Eintreffens mit nichtverdrängender Priorität (D3b)

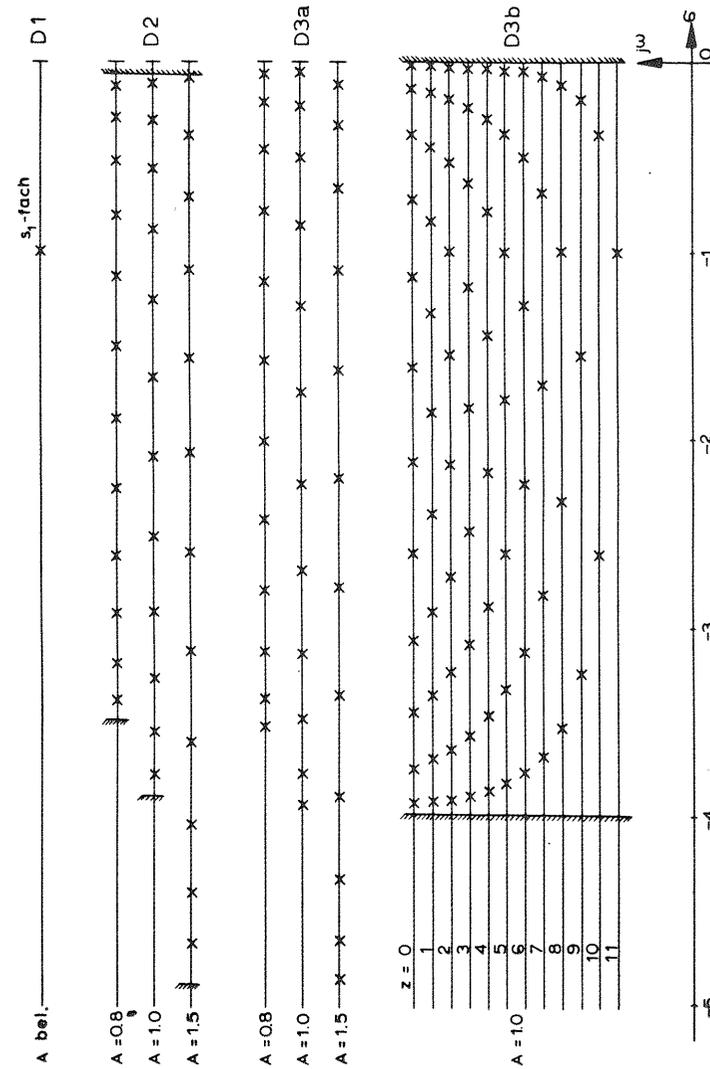


Bild 16. Verteilung der Eigenwerte auf der negativ-reellen Achse
 System: $g = 1, n = 1, s_1 = 12$
 Parameter: Angebot A
 Abfertigungsdisziplinen D1, D2, D3a, D3b

6. BEDIENUNGSSYSTEME MIT VOLLKOMMENER ERREICHBARKEIT UND MEHREREN WARTESCHLANGEN

In diesem Abschnitt werden Bedienungssysteme mit $g \geq 2$ Anrufprozessen bzw. Warteschlangen behandelt. Alle n Leitungen sind vollkommen erreichbar. Dem J -ten Anrufprozeß ist ein Wartespeicher mit s_J Wartepätzen zugeordnet, vergl. Bild 1(b). Es werden verschiedene Abfertigungsdisziplinen zwischen und innerhalb der Warteschlangen zugrunde gelegt. Der übersichtlichen Darstellung wegen werden die Verhältnisse zumeist am Beispiel $g = 2$ dargestellt.

6.1 Zustandswahrscheinlichkeiten

6.1.1 Allgemeine Zustandsgleichungen

Bei vollkommen erreichbaren Leitungen können sich grundsätzlich nur dann Warteschlangen aufbauen, wenn alle Leitungen belegt sind. Zustandsbeschreibung und Zustandsraum bleiben daher unverändert wie in Abschnitt 5.1, solange keine Warteschlangen existieren und wenn

$$\lambda = \sum_{j=1}^g \lambda_j \quad (6.1)$$

als mittlere Summen-Anrufrate eingesetzt wird. Die Zustände der Warteschlangen werden durch einen g -dimensionalen Vektor (z_1, z_2, \dots, z_g) beschrieben, worin $z_j \in [0, s_j]$ die Anzahl Wartender in der J -ten Warteschlange angibt, $J = 1, 2, \dots, g$. Der $\xi(t)$ -Prozeß wird somit durch eine $(n+g)$ -dimensionale Zufallsvariable beschrieben:

$$\xi(t) = (x_1, x_2, \dots, x_n; z_1, z_2, \dots, z_g) \quad (6.2)$$

Der gesamte Zustandsraum zerfällt wieder in 2 Unterräume, einen n -dimensionalen für die Belegungsmuster der Leitungen, und einen g -dimensionalen für die Belegungszustände der Warteschlangen. Beide mehrdimensionalen Unterräume sind - wie bei Systemen mit einer Warteschlange - über einen einzigen Zustand miteinander verknüpft, welcher als beiden Unterräumen zugehörig angesehen werden kann, nämlich dem Zustand $(1, 1, \dots, 1; 0, 0, \dots, 0)$. Die Menge Z der Zustände des $\xi(t)$ -Prozesses ist:

$$2^n - 1 + \prod_{j=1}^g (s_j + 1)$$

In Bild 17 ist der Zustandsraum für das Beispiel $n = 3$ und $g = 2$ für die fünfdimensionale Zustandsvariable $(x_1, x_2, x_3; z_1, z_2)$ angegeben. Es wurde dabei beispielsweise angenommen, daß die Leitungen geordnet abgesucht werden (Für zufallsmäßiges Absuchen vergl. Bild 9).

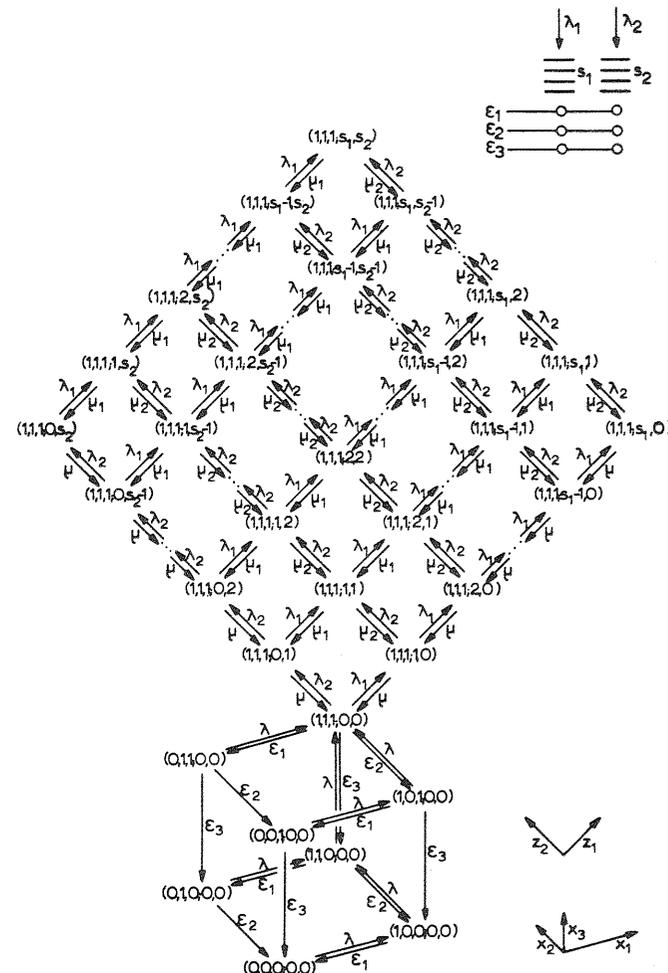


Bild 17. Zustandsraum und Übergangskoeffizienten für das Bedienungssystem vollkommener Erreichbarkeit mit $n = 3$ Leitungen und $g = 2$ Warteschlangen (geordnetes Absuchen der Leitungen)

Zustände und Übergangskoeffizienten des 3(allgemein n)-dimensionalen Unterraumes sind dieselben wie bei Systemen mit einer Warteschlange. Die Zustände der Warteschlangen sind durch einen 2(g)-dimensionalen Unterraum gekennzeichnet. Der Übergangskoeffizient für die Belegung eines weiteren Warteplatzes in der J-ten Warteschlange ist λ_J , $J = 1, 2, \dots, g$. Der Übergangskoeffizient für die Abfertigung eines Wartenden aus der J-ten Warteschlange ist

$$\mu_J = \mu \cdot p_J, \quad J = 1, 2, \dots, g, \quad (6.3)$$

wobei die bedingte Abfertigungswahrscheinlichkeit für die Warteschlange J, p_J , i.a. eine Funktion des momentanen Zustands ist.

Die Zustandsgleichungen werden nach Abschnitt 4.1.5 gewonnen, indem für jeden Zustand die Gleichung für das "statistische Gleichgewicht" angeschrieben wird. Als Beispiel werde hierzu die Zustandsgleichung für den Zustand $(1, 1, 1; z_1, z_2)$ des zweidimensionalen Unterraumes Bild 17 betrachtet:

$$(\lambda + \mu) p(1, 1, 1; z_1, z_2) = \lambda_1 p(1, 1, 1; z_1 - 1, z_2) + \lambda_2 p(1, 1, 1; z_1, z_2 - 1) + \mu_1 p(1, 1, 1; z_1 + 1, z_2) + \mu_2 p(1, 1, 1; z_1, z_2 + 1). \quad (6.4)$$

Zusammen mit der "Normalisierungsbedingung" entsprechend Gl.(4.15) ergibt sich ein lineares Gleichungssystem zur eindeutigen Bestimmung der stationären Zustandswahrscheinlichkeiten $p(x_1, \dots, x_n; z_1, \dots, z_g)$.

Der Rang des linearen Gleichungssystems zur Bestimmung der Zustandswahrscheinlichkeiten läßt sich jedoch - wie beim System mit einer Warteschlange - erniedrigen. Der gesamte "Verkehr" zwischen den beiden Unterräumen verläuft über den "Schlüsselzustand" $(1, 1, \dots, 1; 0, \dots, 0)$. Dieser Zustand muß bei Stationarität bereits mit allen oberen bzw. allen unteren Nachbarzuständen im statistischen Gleichgewicht sein, d.h. er muß von oben (unten) her ebenso häufig entstehen wie er nach oben (unten) hin verschwindet (Diese Folgerung wäre nicht möglich, wenn beide Unterräume über mehrere Zustände miteinander verknüpft wären, wodurch ein sog. "Kreisverkehr" entstehen kann, welcher nicht im Widerspruch zur Stationarität stehen muß).

Das System der linearen Zustandsgleichungen kann dadurch in zwei Teilsysteme des Ranges $2^n - 1$ bzw. $\prod_{j=1}^g (s_j + 1) - 1$ zerlegt werden.

Diese Teilsysteme können unabhängig gelöst werden, wobei jeweils $p(1, 1, \dots, 1; 0, \dots, 0)$ als Parameter in die Lösungen eingeht. $p(1, 1, \dots, 1; 0, \dots, 0)$ wird schließlich mit Hilfe der Normalisierungsbedingung bestimmt.

Bei zufallsmäßigem Absuchen der Leitungen kann insbesondere die Lösung für den n-dimensionalen Unterraum wieder sofort angegeben werden (vergl. Abschnitt 5.1.2).

Haben schließlich alle Leitungen gleiche mittlere Enderaten \mathcal{E} , so kann der n-dimensionale Unterraum wieder auf einen eindimensionalen Zustandsraum reduziert werden, dessen Zustandswahrscheinlichkeiten bekannt sind (vergl. Abschnitt 5.1.3).

6.1.2 Sonderfälle

Die Zustände des 2(g)-dimensionalen Unterraumes zur Beschreibung der Warteschlangen müssen i.a. aus einem linearen Gleichungssystem berechnet werden. Im folgenden werden einige Sonderfälle für die Abfertigungsdisziplin zwischen den Warteschlangen behandelt, welche auf rekursive Lösungsverfahren oder sogar auf eine geschlossene Lösung führen.

a) Fest vorgeschriebene Prioritäten zwischen den Warteschlangen

Es werde im folgenden der praktisch wichtige Fall von $g = 2$ Warteschlangen betrachtet. Besitzt die erste Warteschlange nichtunterbrechende Priorität gegenüber der zweiten Warteschlange, so kann immer nur dann ein Wartender der zweiten Warteschlange abgefertigt werden, wenn der erste Wartespeicher leer ist. Dieser Sonderfall folgt aus dem allgemeinen Modell für die spezielle Form der bedingten Abfertigungswahrscheinlichkeiten p_J

$$\begin{aligned} p_1 &= 1, & p_2 &= 0 & \text{für } z_1 > 0, z_2 \geq 0, \\ p_1 &= 0, & p_2 &= 1 & \text{für } z_1 = 0, z_2 > 0. \end{aligned} \quad (6.5)$$

Im Zustandsraum Bild 17 verschwinden alle Übergänge, welche mit dem Koeffizienten μ_2 bezeichnet sind, d.h.

$$\mu_2 = 0. \quad (6.6)$$

Wegen Gl.(6.6) kann ein Lösungsverfahren zur rekursiven Bestimmung

aller Zustandswahrscheinlichkeiten des 2-dimensionalen Unterraumes angegeben werden.

Die Zustandswahrscheinlichkeit $p(1,1,1;0,0)$ wird als Parameter betrachtet, welcher am Schluß aus der Normalisierungsbedingung bestimmt wird. Nehmen wir vorübergehend $p(1,1,1;0,1)$ als Unbekannte an, so können durch Ansetzen der statistischen Gleichgewichte für die Zustände $(1,1,1;0,0), (1,1,1;1,0), \dots, (1,1,1;s_1-1,0)$ die Zustandswahrscheinlichkeiten $p(1,1,1;1,0), p(1,1,1;2,0), \dots, p(1,1,1;s_1,0)$ als Funktionen der Unbekannten $p(1,1,1;0,1)$ ausgedrückt werden. Das statistische Gleichgewicht für den Zustand $(1,1,1;s_1,0)$ erlaubt die Berechnung der Unbekannten $p(1,1,1;0,1)$. Nun nehmen wir vorübergehend $p(1,1,1;0,2)$ als Unbekannte an und drücken $p(1,1,1;1,1), \dots, p(1,1,1;s_1,1)$ als Funktionen von $p(1,1,1;0,2)$ aus. Das statistische Gleichgewicht für den Zustand $(1,1,1;s_1,1)$ erlaubt die Berechnung der Unbekannten $p(1,1,1;0,2)$ usw.

Auf diese Weise ist es möglich, sämtliche Zustandswahrscheinlichkeiten rekursiv zu berechnen, wobei anstelle eines Gleichungssystems des Ranges $(s_1+1)(s_2+1)-1$ lediglich s_2 Gleichungen für jeweils eine Unbekannte aufzulösen sind.

b) Variable Prioritäten zwischen den Warteschlangen nach momentanen Warteschlangenlängen

Bei dieser Abfertigungsdisziplin erhält die momentan längste Warteschlange nichtunterbrechende Priorität. Gibt es mehrere Warteschlangen der momentan größten Länge, so soll eine beliebige Aufteilungsregel für die Abfertigung zwischen diesen Warteschlangen gelten (z. B. alle gleichwahrscheinlich).

Dieser Fall stellt sich im Zustandsraum Bild 17 wie folgt dar:

$$\mu_1 = \begin{cases} 0 & \text{für } z_2 > z_1 \\ \mu & \text{für } z_1 > z_2 \end{cases} \quad (6.7a)$$

$$\mu_2 = \begin{cases} 0 & \text{für } z_1 > z_2 \\ \mu & \text{für } z_2 > z_1 \end{cases} \quad (6.7b)$$

Es werde wieder $p(1,1,1;0,0)$ als Parameter angenommen. Die Wahrscheinlichkeiten $p(1,1,1;s_1,0), p(1,1,1;0,s_2)$ und $p(1,1,1;1,1)$

erklären wir vorübergehend als Unbekannte. Durch Rekursion können die Wahrscheinlichkeiten $p(1,1,1;s_1-1,0), \dots, p(1,1,1;1,0)$ bzw. $p(1,1,1;0,s_2-1), \dots, p(1,1,1;0,1)$ als Funktionen von $p(1,1,1;s_1,0)$ bzw. $p(1,1,1;0,s_2)$ ausgedrückt werden. Der Ansatz der statistischen Gleichgewichte für die drei Zustände $(1,1,1;0,0), (1,1,1;1,0)$ und $(1,1,1;0,1)$ liefert drei Gleichungen zur Bestimmung der drei Unbekannten $p(1,1,1;s_1,0), p(1,1,1;0,s_2)$ und $p(1,1,1;1,1)$. Dieselbe Prozedur wird nun fortlaufend auf den jeweiligen Restraum ausgeübt.

In diesem Beispiel ist es ebenfalls möglich, die Zustandswahrscheinlichkeiten rekursiv zu berechnen, wobei anstelle eines Gleichungssystems des Ranges $(s_1+1)(s_2+1)-1$ nur $\min(s_1, s_2)$ Gleichungssysteme von jeweils 3. Range aufzulösen sind.

c) Abfertigung nach Wahrscheinlichkeiten proportional zu den momentanen Warteschlangenlängen

Bei dieser Disziplin wird eine Warteschlange mit einer Wahrscheinlichkeit abgefertigt, welche dem Verhältnis aus eigener Länge zur gesamten Länge aller Warteschlangen gleich ist. Diese Disziplin steht gleichzeitig dafür, daß unter a l l e n Wartenden einer zufällig herausgesucht wird.

Die bedingten Abfertigungswahrscheinlichkeiten sind

$$p_J = \frac{z_J}{\sum_{j=1}^g z_j} \quad \text{für } \sum_{j=1}^g z_j > 0 \quad \text{und } J = 1, 2, \dots, g. \quad (6.8)$$

Eine längere Warteschlange wird mit einer größeren Wahrscheinlichkeit abgefertigt als eine kürzere.

Es kann gezeigt werden, daß die Zustandsgleichungen für den g -dimensionalen Unterraum von folgender Lösung befriedigt werden:

$$p(1, \dots, 1, z_1, \dots, z_g) = p(1, \dots, 1, 0, \dots, 0) \cdot \frac{(z_1 + \dots + z_g)!}{z_1! \dots z_g!} \cdot \prod_{j=1}^g \rho_j^{z_j}, \quad z_j \in [0, s_j], \quad J = 1, 2, \dots, g, \quad (6.9)$$

wobei $\rho_j = \lambda_j / \mu$. In diesem Falle ergibt sich außerdem, daß sich schon jeweils zwei benachbarte Zustände im statistischen Gleichgewicht befinden.

6.2 Charakteristische Größen

Die charakteristischen Größen sollen - soweit sie sich nicht auf das ganze System beziehen - für die 1. Warteschlange, d.h. 1-Rufe, definiert werden; die Ergebnisse können durch Vertauschung der Indices auch auf andere Warteschlangen angewendet werden. Die Wahrscheinlichkeiten $p(1, \dots, 1; z_1, \dots, z_g)$ werden kurz mit $p(n; z_1, \dots, z_g)$ bezeichnet.

a) Wartewahrscheinlichkeit W_1

$$W_1 = \sum_{z_1=0}^{s_1-1} \sum_{z_2=0}^{s_2} \dots \sum_{z_g=0}^{s_g} p(n; z_1, z_2, \dots, z_g) \quad \text{für D1, D2, D3b} \quad (6.10a)$$

$$W_1 = \sum_{z_1=0}^{s_1} \sum_{z_2=0}^{s_2} \dots \sum_{z_g=0}^{s_g} p(n; z_1, z_2, \dots, z_g) \quad \text{für D3a} \quad (6.10b)$$

b) Verlustwahrscheinlichkeit B_1

$$B_1 = \sum_{z_2=0}^{s_2} \dots \sum_{z_g=0}^{s_g} p(n; s_1, z_2, \dots, z_g) \quad \text{für D1, D2, D3b} \quad (6.11a)$$

$$B_1 = 0 \quad \text{für D3a} \quad (6.11b)$$

c) Erfolgs-Wartewahrscheinlichkeit W_1^*

$$W_1^* = W_1 \quad \text{für D1, D2, D3b} \quad (6.12a)$$

$$W_1^* = \sum_{z_1=0}^{s_1-1} \sum_{z_2=0}^{s_2} \dots \sum_{z_g=0}^{s_g} p(n; z_1, z_2, \dots, z_g) \quad \text{für D3a} \quad (6.12b)$$

d) Mißerfolgs-Wartewahrscheinlichkeit W_1^{**}

$$W_1^{**} = 0 \quad \text{für D1, D2, D3b} \quad (6.13a)$$

$$W_1^{**} = \sum_{z_2=0}^{s_2} \dots \sum_{z_g=0}^{s_g} p(n; s_1, z_2, \dots, z_g) \quad \text{für D3a} \quad (6.13b)$$

Zwischen W_1 , W_1^* und W_1^{**} gilt, wie bei $g = 1$, die Beziehung

$$W_1 = W_1^* + W_1^{**} \quad (6.14)$$

e) Verkehrswert Y

$$Y = \sum_{x_1=0}^1 \dots \sum_{x_{n-1}=0}^1 \sum_{z_1=0}^{s_1} \dots \sum_{z_g=0}^{s_g} (x_1 + \dots + x_n) p(x_1, \dots, x_n; z_1, \dots, z_g) \quad (6.15)$$

f) Mittlere Warteschlangenlänge Ω_1

$$\Omega_1 = \sum_{z_1=0}^{s_1} \sum_{z_2=0}^{s_2} \dots \sum_{z_g=0}^{s_g} z_1 \cdot p(n; z_1, z_2, \dots, z_g) \quad (6.16)$$

g) Mittlere Wartezeit der Wartenden t_{W_1}

$$t_{W_1} = \frac{\Omega_1}{W_1 \cdot \lambda_1} \quad (6.17)$$

6.3 Wartezeiten

Das Problem der Wartezeitverteilung bei Systemen mit 2 (allgemein g) Warteschlangen und vollkommen erreichbaren Leitungen wird für zwei verschiedene Gruppen von Abfertigungsdisziplinen **z w i s c h e n** den Warteschlangen (A bzw. B, vergl. Abschn. 2.4.3) behandelt. Wie bei Systemen mit einer Warteschlange werden **i n n e r h a l b** der bezüglich eines Testrufs betrachteten Warteschlangen vier verschiedene Abfertigungsdisziplinen zugrunde gelegt. Die Abfertigungsdisziplinen innerhalb der restlichen Warteschlangen dürfen beliebiger Art (D1, ..., D3b) sein.

Bei vollkommen erreichbaren Leitungen kann ein Warteprozess nur dann existieren, wenn alle n Leitungen belegt sind. Diese Leitungsbelegungen haben in jedem Falle einen direkten Einfluß auf die Wartezeit eines Testrufs. Die n Leitungsbelegungen brauchen jedoch nicht explizit in der Zustandsbeschreibung berücksichtigt werden.

Alle Ausführungen beziehen sich aus Übersichtlichkeitsgründen auf Systeme mit $g = 2$ Warteschlangen. Ohne Einschränkung der Allgemeinheit werden die Warteprozesse nur bezüglich Rufen des 1. Anrufprozesses betrachtet.

6.3.1 Abfertigung zwischen den Warteschlangen nach Wahrscheinlichkeiten unabhängig von den momentanen Warteschlangenlängen (A)

Die Abfertigung zwischen den Warteschlangen soll nach Regeln erfolgen, welche nicht von den momentanen Warteschlangenlängen innerhalb der (nichtleeren) Wartespeicher abhängen. Der Spaltenvektor (p_j) der bedingten Abfertigungswahrscheinlichkeiten für die Warteschlange J (vergl. Abschn. 2.4.3) ist also für alle jene Zustände identisch, welche dieselbe Kombination belegter Wartespeicher aufweisen.

6.3.1.1 Abfertigung in der Reihenfolge des Eintreffens (AD1)

a) Der Warteprozess

Wegen der Disziplinen A und D1 haben dem 1-Testruf nachfolgende 1-Rufe keinen Einfluß auf dessen Wartezeit. Während des $\xi_1(t)$ -Prozesses ist der 1-Testruf im ersten Wartespeicher anwesend, d.h. während dieser Zeit ist der erste Wartespeicher belegt und wird deshalb mit der bedingten Wahrscheinlichkeit p_1 abgefertigt. Wartende des zweiten Anrufprozesses haben in jedem Falle einen Einfluß auf die Wartezeit des 1-Testrufs.

Der $\xi_1(t)$ -Prozeß kann durch eine zweidimensionale Zufallsvariable

$$\xi_1(t) = (i_1, i_2), \quad \begin{matrix} i_1 = 0, 1, \dots, s_1 - 1, \\ i_2 = 0, 1, \dots, s_1, \end{matrix} \quad (6.18)$$

beschrieben werden, worin i_1 die Zahl der vor dem 1-Testruf Wartenden in der ersten Warteschlange und i_2 die Zahl der insgesamt Wartenden in der zweiten Warteschlange angibt. Die Antreffwahrscheinlichkeit ist

$$P_j(i_1, i_2) = P\{\xi_1(0) = (i_1, i_2)\} = p(n; i_1, i_2), \quad \begin{matrix} i_1 = 0, 1, \dots, s_1 - 1, \\ i_2 = 0, 1, \dots, s_2. \end{matrix} \quad (6.19)$$

Bild 18 zeigt den Wartezustandsraum und die Übergänge für den $\xi_1(t)$ -Prozeß mit den zugehörigen Übergangskoeffizienten. Die Endkoeffizienten des $\xi_1(t)$ -Prozesses können dem Bild 18 ebenfalls entnommen werden.

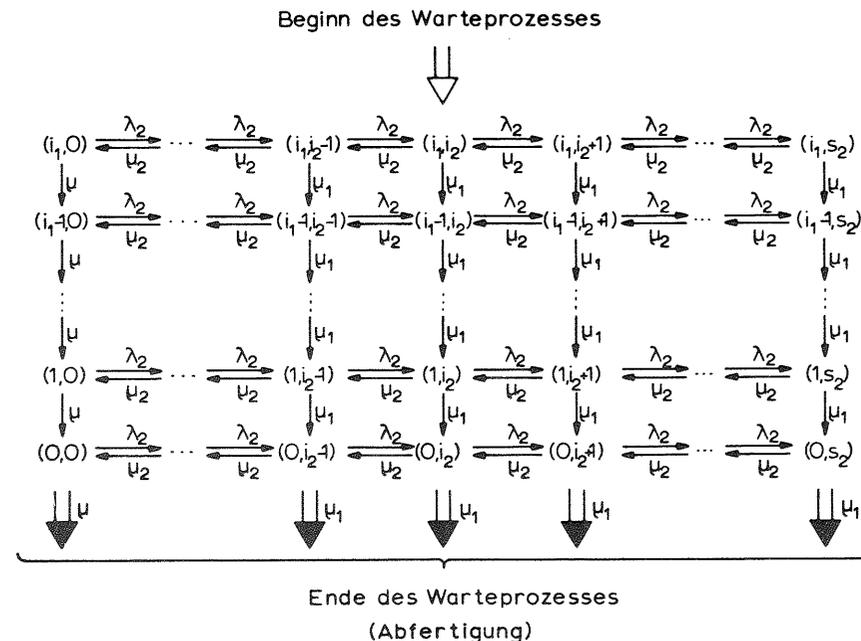


Bild 18. Wartezustandsraum für Warten nach AD1

Entsprechend der Feststellung, daß dem 1-Testruf nachfolgende 1-Rufe keinen Einfluß auf dessen Wartezeit haben, ist der $\xi_1(t)$ -Prozeß von λ_1 unabhängig.

Für den Sonderfall der Abfertigung nach Prioritäten (Warteschlange 1 habe Priorität über Warteschlange 2) ist nach Gl.(6.6) $\mu_2 = 0$ und $\mu_1 = \mu$. Hierdurch ergeben sich weitere Vereinfachungen im Wartezustandsraum, auf die wir später noch näher eingehen werden.

Die Wartezustandsräume für Bedienungssysteme mit mehreren Eingangswarteschlangen sind grundsätzlich mehrdimensional. Die zugehörigen Differentialgleichungssysteme sind daher von hohem Range, so daß es vorteilhaft sein kann, wenn sich der gesamte Warteprozess in unabhängig lösbare Teil-Warteprozesse zerlegen läßt. Hierzu wird die grundlegende Eigenschaft mehrdimensionaler Warteprozesse aus Abschnitt 4.4.3c herangezogen. Sie besagt, daß stets derjenige

Teilraum für sich betrachtet werden kann, auf welchen nur einseitig gleichsinnig gerichtete Übergänge weisen. Diese Eigenschaft läßt sich aus den geometrischen Wartezustandsräumen leicht erkennen.

Angewendet auf den Wartezustandsraum Bild 18 stellt man hiernach fest, daß zunächst der Unterraum $(0, i_2)$, $i_2 = 0, 1, \dots, s_2$, für sich behandelt werden kann. Wenn die Lösungen hierfür bekannt sind, kann nun der Unterraum $(1, i_2)$, $i_2 = 0, 1, \dots, s_2$, für sich behandelt werden. Dabei werden außerdem die Zustände des zuerst betrachteten Unterraumes $(0, i_2)$ auftreten, welche aber bereits aus dem vorausgegangenen Lösungsschritt bekannt sind, usw.

b) Wartezeitverteilungsfunktionen

Es wird eine bedingte Wartezeitverteilungsfunktion $w_1(t|i_1, i_2)$ eingeführt, wobei $\xi_1(0) = (i_1, i_2)$. Da keine akzeptierten 1-Rufe verdrängt werden können (D1), ist $w_1^*(t|i_1, i_2) = w_1(t|i_1, i_2)$, $w_1^*(t|i_1, i_2) = 0$.

Entsprechend Gl.(4.32) gehorchen die bedingten Wartezeitverteilungsfunktionen einem System von linearen Differentialgleichungen, dessen Koeffizienten aus Bild 18 entnommen werden können. Für den allgemeinen Zustand (i_1, i_2) gilt folgende Differentialgleichung:

$$\frac{dw_1(t|i_1, i_2)}{dt} = -(\lambda_2 + \mu)w_1(t|i_1, i_2) + \lambda_2 w_1(t|i_1, i_2+1) + \mu_1 w_1(t|i_1-1, i_2) + \mu_2 w_1(t|i_1, i_2-1). \quad (6.20)$$

(Am rechten, linken und unteren Rand des Wartezustandsraumes Bild 18 vereinfachen sich die Gl.(6.20) entsprechenden Differentialgleichungen).

Die Anfangsbedingungen sind einheitlich $w_1(0|i_1, i_2) = 1$. Dies folgt außerdem aus Gl.(6.20) mit Gl.(4.34), wenn die in Bild 18 angegebenen Endkoeffizienten berücksichtigt werden.

Der Übersichtlichkeit halber wird wieder das lineare Gleichungssystem der Laplace-Transformierten $W_1(s|i_1, i_2)$ der bedingten Wartezeitverteilungsfunktionen $w_1(t|i_1, i_2)$ in Matrixdarstellung angegeben. Die am oberen Rand angeschriebenen Zustände (i_1, i_2) deuten die Zugehörigkeit der in dieser Spalte stehenden Koeffizienten zu den "Unbekannten" $W_1(s|i_1, i_2)$ an.

$$\begin{array}{c|cccccccc|cccc|cccc|c} & (0,0) & (0,1) & (0,2) & \dots & (0,s_2-1) & (0,s_2) & (1,0) & (1,1) & (1,2) & \dots & (1,s_2-1) & (1,s_2) & \dots & (s_2-1,0) & (s_2-1,1) & \dots & (s_2-1,s_2) & \\ \hline (s+\lambda_2+\mu) - \lambda_2 & & & & & & & & & & & & & & & & & & & 1 \\ -\mu_2 (s+\lambda_2+\mu) - \lambda_2 & & & & & & & & & & & & & & & & & & & 1 \\ & & & & & & & & & & & & & & & & & & & 1 \\ & & & & & & & & & & & & & & & & & & & \vdots \\ & & & & & & & & & & & & & & & & & & & 1 \\ & & & & & & & & & & & & & & & & & & & 1 \\ \hline -\mu & & & & & & & (s+\lambda_2+\mu) - \lambda_2 & & & & & & & & & & & & 1 \\ & & & & & & & & & & & & & & & & & & & 1 \\ & & & & & & & & & & & & & & & & & & & \vdots \\ & & & & & & & & & & & & & & & & & & & 1 \\ & & & & & & & & & & & & & & & & & & & 1 \\ \hline & & & & & & & & & & & & & & & & & & & \vdots \\ & & & & & & & & & & & & & & & & & & & 1 \\ & & & & & & & & & & & & & & & & & & & 1 \\ & & & & & & & & & & & & & & & & & & & \vdots \\ & & & & & & & & & & & & & & & & & & & 1 \\ & & & & & & & & & & & & & & & & & & & 1 \end{array} \quad (6.21)$$

Die Gl.(6.21) entsprechende Matrix A_1 hat allgemein den Rang $N = s_1(s_2+1)$. Wegen der speziellen Struktur dieser Matrix ist jedoch eine Sektionalisierung möglich: Das Teilsystem für die Zustände $(0,0), \dots, (0, s_2)$ kann für sich gelöst werden. Mit den bekannten Lösungen des ersten Teilsystems kann das zweite Teilsystem für die Zustände $(1,0), \dots, (1, s_2)$ gelöst werden usw. Die Möglichkeit der Sektionalisierung kann außerdem aus dem Wartezustandsraum erkannt werden, vergl. a). Auf diese Weise sind anstelle eines Systems $s_1(s_2+1)$ -ten Ranges s_1 Systeme von jeweils nur (s_2+1) -tem Range zu lösen.

Über die Eigenwerte der Teilsysteme von Gl.(6.21) können folgende Aussagen gemacht werden:

1. Die Teilsysteme haben jeweils den gleichen Satz von Eigenwerten $e_p, p = 1, 2, \dots, s_2+1$.
2. Die Eigenwerte jedes Teilsystems sind negativ-reell und verschieden, sie liegen sämtlich im Intervall $[-(2\lambda_2+\mu+\mu_2), -\mu_1]$

Die erste Aussage gilt, weil die Systemdeterminanten der durch Sek-
tionalisierung gewonnenen Teilsysteme identisch sind. Der Beweis
der zweiten Aussage erfolgt entsprechend zu den Beweisführungen für
das Bedienungssystem mit einer Warteschlange, vergl. Anhang A1 und
A3 (Die Systemdeterminanten der Teilsysteme sind vom gleichen Typ
wie bei Bedienungssystemen mit einer Warteschlange und der Abfertigungs-
disziplin D3b).

Die Matrix Gl.(6.21) besitzt also $N' = s_2 + 1$ verschiedene Eigenwerte
 e_{ν} mit der Vielfachheit $K_{\nu} = s_1, \nu = 1, 2, \dots, s_2 + 1$.

Löst man sukzessiv die Teilsysteme von Gl.(6.21) auf, so erhält man
als Lösungen Ausdrücke der folgenden Form:

$$W_{\nu}(s | i_1, i_2) = \sum_{\nu=1}^{s_2+1} \sum_{x=1}^{i_1+1} \frac{a_{\nu}(x)}{(s - e_{\nu})^x} \quad (6.22)$$

Die Laplace-Rücktransformation ergibt

$$w_{\nu}(t | i_1, i_2) = \sum_{\nu=1}^{s_2+1} \exp(e_{\nu} t) \sum_{x=1}^{i_1+1} a_{\nu}(x) \cdot \frac{t^{x-1}}{(x-1)!} \quad (6.23)$$

Mit den Antreffwahrscheinlichkeiten $P_1(i_1, i_2)$ nach Gl.(6.19) folgt
entsprechend Gl.(4.39) die Wartezeitverteilungsfunktion bezüglich
1-Rufen

$$W_1(>t) = \sum_{i_1=0}^{s_2-1} \sum_{i_2=0}^{s_2} p(n; i_1, i_2) w_{\nu}(t | i_1, i_2) \quad (6.24)$$

Sonderfälle:

1. Prioritätsfall

Den Sonderfall, bei welchem die erste Warteschlange nichtunterbre-
chende Priorität über die zweite Warteschlange hat, erhält man für
 $\mu_2 = 0, \mu_1 = \mu$. Die Eigenwerte fallen alle zusammen auf

$$e_{\nu} = -\mu \quad (6.25)$$

Beginnend mit dem Zustand $(0, s_2)$ können nacheinander für alle Zu-
stände $(0, i_2), i_2 = s_2 - 1, \dots, 0$, die Lösungen rekursiv berechnet
werden. Man erhält

$$w_1(s | 0, i_2) = \frac{1}{s + \mu}, \quad i_2 = s_2, \dots, 0.$$

Das Ergebnis ist unabhängig von λ_2 , da nachfolgende 2-Rufe keinen
Einfluß auf die Wartezeit des 1-Testrufs haben. Die allgemeine Lö-
sung ist in diesem Sonderfall

$$W_{\nu}(s | i_1, i_2) = \sum_{\nu=1}^{i_1+1} \frac{\mu^{\nu-1}}{(s + \mu)^{\nu}} \quad i_2 = 0, 1, \dots, s_2 \quad (6.26)$$

Die Laplace-Rücktransformation liefert

$$w_{\nu}(t | i_1, i_2) = \sum_{\nu=0}^{i_1} \frac{(\mu t)^{\nu}}{\nu!} \exp(-\mu t), \quad i_2 = 0, 1, \dots, s_2 \quad (6.27)$$

$w(t | i_1, i_2)$ gehorcht also einer ERLANG- $(i_1 + 1)$ -Verteilung wie im Falle
von nur einer Warteschlange und der Abfertigung in der Reihenfolge
des Eintreffens (D1). Dieses Ergebnis ist anschaulich klar, da der
1-Testruf im Prioritätsfall genau $(i_1 + 1)$ negativ-exponentiell ver-
teilte Phasen bis zu seiner Abfertigung warten muß, unabhängig da-
von, wieviele 2-Rufe in dieser Zeit eintreffen.

2. Unbegrenzter Warteraum für 2-Rufe bzw. verschwindendes Angebot
der zweiten Warteschlange

Einen weiteren interessanten Sonderfall erhält man für $s_2 \rightarrow \infty$ bzw.
für sehr kleine λ_2 . In diesen Fällen können die Eigenwerte wieder
explizit angegeben werden, da die einzige Unsymmetrie der System-
determinanten der Teilsysteme - im letzten Hauptdiagonalelement -
verschwindet. In diesen Fällen erhalten wir denselben Matrixtyp
wie bei Bedienungssystemen mit einer Warteschlange und inverser
Abfertigungsreihenfolge D3a, vergl. Abschnitt 5.3.3 sowie Anhang A2.

Die Eigenwerte dieses (symmetrischen) Matrixtyps sind

$$e_{\nu} = 2\sqrt{\lambda_2 \mu_2} \cos \frac{\nu \pi}{s_2 + 2} - (\lambda_2 + \mu) \quad , \quad \nu = 1, 2, \dots, s_2 + 1 \quad (6.28)$$

Die Eigenwerte nach Gl.(6.28) sind exakt für $s_2 \rightarrow \infty$ bzw. $\lambda_2 \rightarrow 0$, sie
können jedoch für $s_2 \gg 1$ und $\lambda_2 \ll \mu$ als Näherungslösung benutzt wer-
den.

In diesen Fällen kann die Wartezeitverteilungsfunktion wieder ex-
plizit angegeben werden, worauf hier aber nicht näher eingegangen
werden soll.

läßt sich insbesondere nicht mehr in Teilsysteme aufteilen. Es muß daher ein Differentialgleichungssystem des Ranges $s_1 \cdot (s_2 + 1)$ aufgelöst werden.

In der Matrix treten von Null verschiedene Elemente nur symmetrisch zur Hauptdiagonale auf. Da i.a. $a_{1k} \neq a_{k1}$ gilt, ist die Matrix nicht-symmetrisch. Ähnlich wie beim Bedienungssystem mit einer Warteschlange und Abfertigung in zufälliger Reihenfolge kann erwartet werden, daß die Matrix nach Gl.(6.31) zur Klasse der symmetrisierbaren Matrizen gehört und daher notwendig reelle Eigenwerte besitzen muß (vergl. Anhang A2).

Der Bereich der Eigenwerte kann am schnellsten mit Hilfe des Theorems von S. GERSCHGORIN (siehe Anhang A1) nachgewiesen werden. Die Zeilenform des Kriteriums liefert folgende Abschätzung des Eigenwertbereichs auf der negativ-reellen Achse:

$$\left[-\max\left(2\lambda + 2\mu_2 + \mu_1 \frac{2s_1 - 3}{s_1 - 1}, 2\lambda_2 + 2\mu_2 + \mu_1 \frac{2s_1 - 1}{s_1}\right), -\frac{\mu_1}{s_1} \right].$$

wegen der allgemeineren Struktur der Matrix kann nicht mehr ausgeschlossen werden, daß alle Eigenwerte verschieden sind. Numerische Ergebnisse haben bestätigt, daß mehrfache Eigenwerte (zumeist zweifache) vorkommen können.

Die Lösungen des Gleichungssystems (6.31) sind i.a. von folgendem Typ:

$$W_{\nu}(s | i_1, i_2) = \sum_{\nu=1}^{N'} \sum_{\nu=1}^{K_{\nu}} \frac{\alpha_{\nu}(\nu)}{(s - e_{\nu})^{\nu}} \quad (6.32)$$

wobei e_{ν} ein K_{ν} -facher Eigenwert der Matrix A_1 ist, $\nu = 1, 2, \dots, N'$, und

$$\sum_{\nu=1}^{N'} K_{\nu} = N = s_1(s_2 + 1).$$

Die bedingten Wartezeitverteilungsfunktionen sind vom Typ

$$w_{\nu}(t | i_1, i_2) = \sum_{\nu=1}^{N'} \exp(e_{\nu} t) \sum_{\nu=1}^{K_{\nu}} \alpha_{\nu}(\nu) \frac{t^{\nu-1}}{(\nu-1)!} \quad (6.33)$$

Für die absolute Wartezeitverteilungsfunktion $W_1(>t)$ gilt weiterhin Gl.(6.24). Die Berechnung der bedingten mittleren Wartezeiten und der bedingten Momente der Wartezeitverteilungsfunktion erfolgt wie in 6.3.1.1 c).

6.3.1.3 Abfertigung in der inversen Reihenfolge des Eintreffens mit verdrängender Priorität (AD3a)

a) Der Warteprozess

Der 1-Testruf ordnet sich bei seiner Ankunft vor allen Wartenden der ersten Warteschlange ein. Die bei Ankunft angetroffenen Wartenden der ersten Warteschlange haben keinen Einfluß auf die Wartezeit des 1-Testrufs, da sie erst - wenn überhaupt - nach dem 1-Testruf abgefertigt werden können (wegen D3a) und ihre Anwesenheit auch keine Beschleunigung in der Abfertigung der ersten Warteschlange bedingt (wegen A). Dagegen konkurrieren Wartende der zweiten Warteschlange in jedem Falle um die Belegung einer frei werdenden Leitung.

Der $\xi_1(t)$ -Prozeß wird durch die zweidimensionale Zufallsvariable

$$\xi_1(t) = (i_1, i_2), \quad \begin{matrix} i_1 = 0, 1, \dots, s_1 - 1, \\ i_2 = 0, 1, \dots, s_2, \end{matrix} \quad (6.34)$$

beschrieben, wobei i_1 die Zahl der vor dem 1-Testruf in der ersten Warteschlange wartenden und i_2 die Zahl der in der zweiten Warteschlange insgesamt wartenden bedeutet. Für die gesamte Wartezeit des Testrufs gilt $\xi_1(0) = (0, i_2)$; für die Antreffwahrscheinlichkeiten gilt

$$P_{\nu}(0, i_2) = \sum_{z_1=0}^{s_1} p(\nu, z_1, i_2) \quad (6.35)$$

In Bild 20 ist der Wartezustandsraum des Warteprozesses $\xi_1(t)$ mit den zugehörigen Übergängen und Übergangskoeffizienten angegeben.

Wie aus Bild 20 ersichtlich ist, kann der $\xi_1(t)$ -Prozeß nur von bestimmten Zuständen aus beendet werden und zwar entweder durch Abfertigung oder Verdrängung aus dem Wartespeicher. Die Endkoeffizienten sind im Wartezustandsraum Bild 20 ebenfalls angegeben.

b) Wartezeitverteilungsfunktion

Wegen der Abfertigungsdisziplin D3a muß zwischen 1-Rufen unterschieden werden, welche nach Ankunft warten ohne Rücksicht auf ihr späteres Schicksal (Abfertigung bzw. Verdrängung) und welche erfolgreich oder erfolglos warten. Die zugehörigen bedingten Wartezeitverteilungsfunktionen sind $w_1(t | i_1, i_2)$, $w_1^*(t | i_1, i_2)$ bzw. $w_1^{**}(t | i_1, i_2)$. Für diese drei Funktionen gilt jeweils dasselbe Differentialgleichungssystem, jedoch mit unterschiedlichen Anfangsbedingungen.

Die Matrix A_1 des Systems Gl.(6.38) hat dieselbe Struktur wie im Falle AD2. Es ist insbesondere keine Aufteilung in Teilsysteme erkennbar. Als Bereich der Eigenwerte auf der negativ-reellen Achse erhält man das Intervall

$$[-2(\lambda + \mu), 0].$$

Da auch hier mehrfache Eigenwerte nicht ausgeschlossen werden können, werden die bedingten Wartezeitverteilungsfunktionen i.a. von der Form Gl.(6.33) sein.

Die absolute Wartezeitverteilungsfunktion bezüglich 1-Rufen stellt sich mit Hilfe der Antreffwahrscheinlichkeiten nach Gl.(6.35) wie folgt dar:

$$W_1(>t) = \sum_{i_2=0}^{s_2} P_1(0, i_2) w_1(t | 0, i_2). \quad (6.39)$$

Aus dem System (6.38) sind also nur die (s_2+1) -ten Größen zu berechnen. Gl.(6.39) gilt entsprechend auch für $W_1^*(>t)$ und $W_1^{**}(>t)$.

Im Sonderfall $\mu_1 = \mu, \mu_2 = 0$ (nichtunterbrechende Priorität von 1-Rufen über 2-Rufe) kann in Gl.(6.38) durch Umordnung eine Aufspaltung in Teilsysteme erreicht werden. Die Teilsysteme haben dieselbe Form wie bei Systemen mit einer Warteschlange und der Disziplin D3a. Die Eigenwerte, und damit die Lösungen des Systems, sind dann angebar.

Die Berechnung der bedingten mittleren Wartezeiten sowie der bedingten Momente erfolgt entsprechend Abschnitt 6.3.1.c).

6.3.1.4 Abfertigung in der inversen Reihenfolge des Eintreffens mit nichtverdrängender Priorität (AD3b)

Ein eintreffender 1-Ruf ordnet sich vor allen Wartenden der ersten Warteschlange ein und verschiebt alle angetroffenen Wartenden dieser Warteschlange um einen Platz nach hinten solange noch mindestens ein Warteplatz im ersten Wartespeicher frei ist. Trifft dagegen der 1-Ruf den ersten Wartespeicher vollbelegt vor, so wird er abgewiesen. Wie bei Bedienungssystemen mit einer Warteschlange und der Disziplin D3b haben die angetroffenen 1-Wartenden einen indirekten, die nach dem 1-Testruf eingetroffenen und akzeptierten 1-Wartenden einen direkten Einfluß auf dessen Wartezeit. Wartende der zweiten Warteschlange konkurrieren in jedem Falle um die Belegung einer frei werdenden Leitung.

Der $\xi_1(t)$ -Prozeß wird durch eine dreidimensionale Zufallsvariable beschrieben:

$$\xi_1(t) = (i_1, z_1, i_2), \quad \begin{aligned} i_1 &= 0, 1, \dots, s_1 - z_1 - 1, \\ z_1 &= 0, 1, \dots, s_1 - 1, \\ i_2 &= 0, 1, \dots, s_2. \end{aligned} \quad (6.40)$$

Es bedeuten i_1 die Zahl der v o r und z_1 die Zahl der h i n t e r dem 1-Testruf wartenden 1-Rufe, i_2 die Zahl der in der zweiten Warteschlange insgesamt wartenden Rufe. Für jede mögliche Zahl z_1 entwickelt sich ein spezieller Prozeß; die Zahl z_1 ändert sich während eines speziell betrachteten Warteprozesses nicht. Für die Antreffwahrscheinlichkeiten gilt:

$$P_1(0, z_1, i_2) = p(n; z_1, i_2). \quad (6.41)$$

In Bild 21 ist der Warte Zustandsraum mit Übergängen des $\xi_1(t)$ -Prozesses für eine allgemeine Zahl z_1 wiedergegeben.

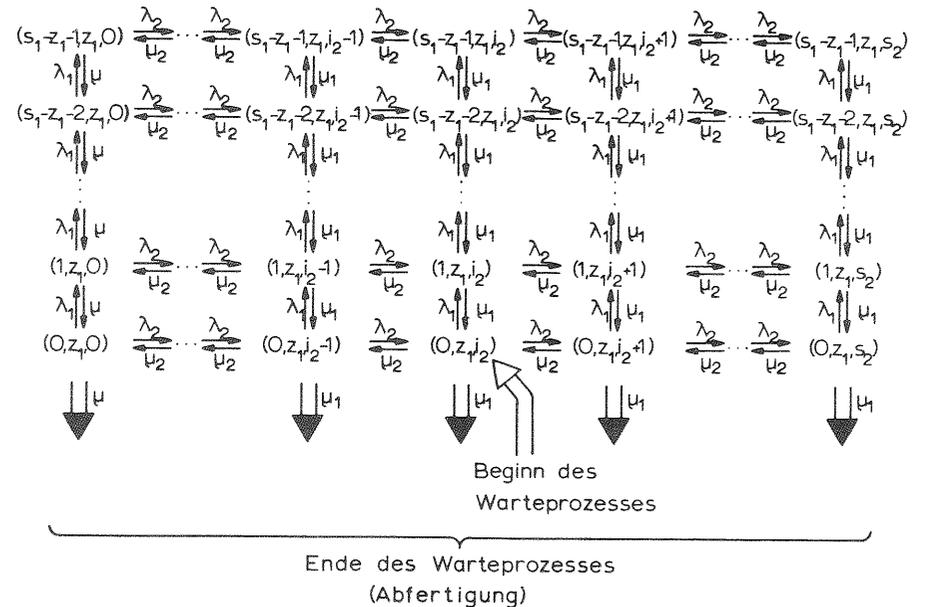


Bild 21. Warte Zustandsraum für Warten nach AD3b

Das lineare Differentialgleichungssystem für die bedingten Wartezeitverteilungsfunktionen $w_1(t | i_1, z_1, i_2)$ kann mit Hilfe des Zustandsdiagrammes Bild 21 leicht angegeben werden, wobei z_1 ein Parameter ist. Die Behandlung des Differentialgleichungssystems, die Berechnung der absoluten Wartezeitverteilungsfunktion, der mittleren Wartezeiten und Momente erfolgen in entsprechender Weise wie in den vorausgegangenen Abschnitten.

6.3.2 Abfertigung zwischen den Warteschlangen nach Wahrscheinlichkeiten abhängig von den momentanen Warteschlangenlängen (B)

Den allgemeinsten Fall der wahrscheinlichkeitmäßigen Abfertigung zwischen den Warteschlangen erhält man, wenn die Aufteilungswahrscheinlichkeiten nicht nur von den nichtleeren Warteschlangen (wie bei A) sondern auch noch von den momentanen Warteschlangenlängen abhängig sind. Der Spaltenvektor (p_j) der bedingten Abfertigungswahrscheinlichkeiten hängt also jeweils vom Zustand des Gesamtsystems ab.

6.3.2.1 Abfertigung in der Reihenfolge des Eintreffens (BD1)

a) Der Warteprozess

wie im Falle AD1 über den 1-Testruf wartende der ersten Warteschlange als auch wartende der zweiten Warteschlange einen direkten Einfluß auf die Wartezeit des 1-Testrufs aus. Dem 1-Testruf nachfolgende 1-Rufe haben wegen Disziplin B einen indirekten Einfluß auf dessen Wartezeit. Werden beispielsweise die Warteschlangen nach Wahrscheinlichkeiten abgefertigt, welche proportional zu den momentanen Warteschlangenlängen sind, so bewirken dem 1-Testruf nachfolgende 1-Rufe eine beschleunigende Abfertigung des des 1-Testrufs, obwohl diese Rufe erst nach dem 1-Testruf abgefertigt werden.

Der $\xi_1(t)$ -Prozeß wird durch die dreidimensionale Zufallsvariable

$$\xi_1(t) = (i_1, z_1, i_2), \quad \begin{matrix} i_1 = 0, 1, \dots, s_1 - 1, \\ z_1 = 0, 1, \dots, s_1 - i_1 - 1, \\ i_2 = 0, 1, \dots, s_2, \end{matrix} \quad (6.42)$$

beschrieben. Dabei bedeuten i_1 die Zahl der 1-Wartenden vor und z_1 die Zahl der 1-Wartenden hinter dem 1-Testruf; i_2 ist die Zahl der insgesamt in der zweiten Warteschlange wartenden. Der $\xi_1(t)$ -Prozeß beginnt stets mit $z_1 = 0$, für $z_1 > 0$ entwickeln sich nur Warteprozesse für partielle Wartezeiten. Als Antreffwahrscheinlichkeiten haben wir

$$P_1(i_1, 0, i_2) = p(n; i_1, i_2). \quad (6.43)$$

Der Wartezustandsraum des Warteprozesses ist nun (echt) dreidimensional. In Bild 22 ist der gesamte Zustandsraum für ein Bedienungssystem mit $s_1 = 3$ und $s_2 = 2$ Wartepunkten angegeben. Es besteht allgemein aus s_1 Ebenen ($z_1 = 0, 1, \dots, s_1 - 1$), welche jeweils von den Komponenten i_1 bzw. i_2 aufgespannt werden ($i_1 = 0, 1, \dots, s_1 - z_1 - 1, i_2 = 0,$

$1, \dots, s_2$). Die Ebenen reduzieren ihre Größe bezüglich i_1 mit der Zunahme von z_1 entsprechend $i_{1max} + z_1 + 1 = s_1$. Die s_1 Ebenen haben über einseitig gerichtete Übergänge Verbindung miteinander. Der Warteprozess kann mit einem beliebigen Zustand der untersten Ebene ($z_1 = 0$) beginnen; er endet mit einem beliebigen Zustand am linken Rande des Zustandsraums ($i_1 = 0$). Als Beispiel wurde in Bild 22 angenommen, daß der $\xi_1(t)$ -Prozeß mit dem Zustand $\xi_1(0) = (1, 0, 1)$ beginnt.

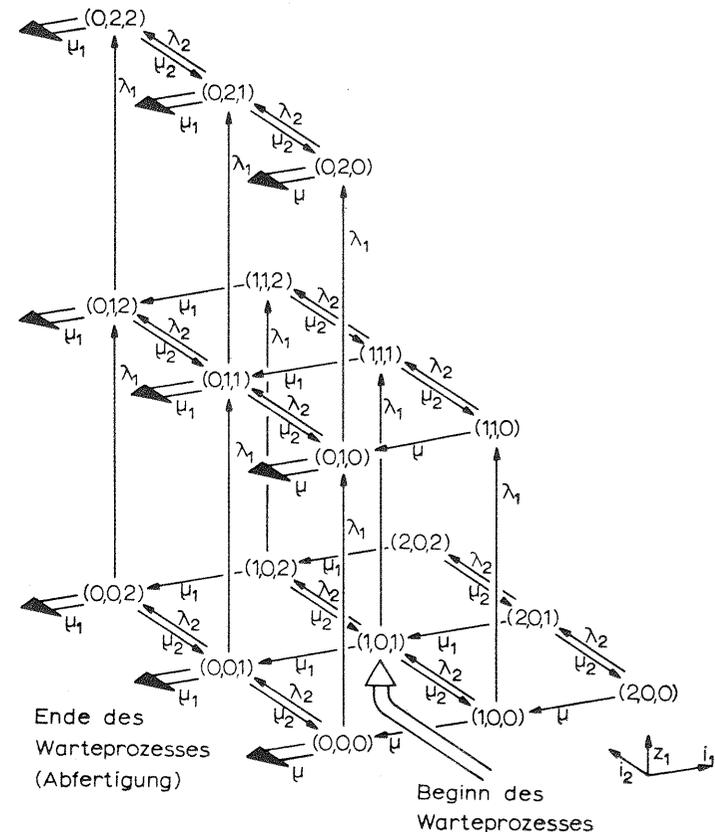


Bild 22. Wartezustandsraum für Warten nach BD1

Beispiel: $s_1 = 3, s_2 = 2, \xi_1(0) = (1, 0, 1)$

Das Gleichungssystem (6.45) hätte man auch aus einer Umordnung der Unbekannten des Systems (6.44) erhalten können; an diesem Beispiel sollte demonstriert werden, wie man auf einfache Weise mit Hilfe des Wartezustandsraumes Aufspaltungen in Teilsysteme erkennen kann.

Wie mit den Überlegungen anhand des Wartezustandsraumes bereits erkannt wurde, läßt die Matrix Gl.(6.45) weitgehende Aufspaltungen in Teilsysteme zu. Zunächst kann ein Teilsystem (s₂+1)-ten Ranges für i₁ = 0 und z₁ = s₁-1 allein gelöst werden, mit Hilfe dieser Lösungen das Teilsystem (s₂+1)-ten Ranges für i₁ = 0 und z₁ = s₁-2 usw. bis schließlich i₁ = 0 und z₁ = 0. Entsprechend verfährt man nun mit den Teilsystemen für i₁ = 1, 2, ..., s₁-1. Auf diese Art sind jeweils nur Systeme des Ranges (s₂+1) aufzulösen.

Es sei darauf hingewiesen, daß Teilsysteme mit formal gleichen Koeffizienten nicht notwendigerweise gleiche Eigenwerte besitzen müssen, da die Koeffizienten μ₁ und μ₂ noch vom jeweiligen Systemzustand abhängen.

Allgemein kann durch Betrachtung der Teilsysteme gezeigt werden, daß die Eigenwerte negativ-reell und im Intervall

$$[-(\lambda + \lambda_2 + \mu + \mu_2), -\mu_1]$$

gelegen sein müssen (für μ₁ und μ₂ müssen dabei die Werte eingesetzt werden, welche das maximal größte Intervall ergeben). Es kann nicht ausgeschlossen werden, daß mehrfache Eigenwerte auftreten können. Die Lösungen für die bedingten Wartezeitverteilungsfunktionen werden deshalb entsprechend Gl.(6.33) aufgebaut sein.

Die gesamte Wartezeitverteilungsfunktion für 1-Rufe ist

$$W_1(t) = \sum_{i_1=0}^{s_1-1} \sum_{i_2=0}^{s_2} p(n; i_1, i_2) w_1(t | i_1, 0, i_2) . \tag{6.46}$$

Die mittleren Wartezeiten und die Momente der Wartezeitverteilungsfunktionen werden in entsprechender Weise aus dem System (6.45) bestimmt wie in vorangegangenen Abschnitten gezeigt wurde.

6.3.2.2 Abfertigung in zufälliger Reihenfolge (BF2)

Alle Wartenden der Warteschlangen sind Konkurrenten des 1-Testruffs. Der Warteprozess wird daher durch eine zweidimensionale Zufallsvariable entsprechend Gl.(6.29) beschrieben. Es gelten sämtliche Überlegungen aus Abschnitt 6.3.1.2 für AD2, wenn anstelle der konstanten Koeffizienten μ₁ und μ₂ die entsprechenden zustandsabhängigen Koeffizienten eingesetzt werden.

6.3.2.3 Abfertigung in der inversen Reihenfolge des Eintreffens mit verdrängender Priorität (BD3a)

a) Der Warteprozess

Ein eintreffender 1-Testruf belegt bei vollbelegtem Leitungsbündel wie bei AD3a den ersten Warteplatz des ersten Wartespeichers und verschiebt alle Wartenden der ersten Warteschlange um einen Platz nach hinten. War der erste Wartespeicher bereits vollbelegt, so geht der am längsten Wartende verloren. Wie bei AD3a haben alle dem 1-Testruf nachfolgenden 1-Rufe sowie alle Wartenden der zweiten Warteschlange einen direkten Einfluß auf dessen Wartezeit. Im Gegensatz zu AD3a üben aber bei BD3a die hinter dem 1-Testruf Wartenden einen indirekten Einfluß auf die Wartezeit des 1-Testruffs aus (wegen B).

Der ξ₁(t)-Prozess läßt sich eindeutig durch die dreidimensionale Zufallsvariable

$$\xi_1(t) = (i_1, z_1, i_2), \quad \begin{aligned} i_1 &= 0, 1, \dots, s_1-1, \\ z_1 &= 0, 1, \dots, s_1-i_1-1, \\ i_2 &= 0, 1, \dots, s_2, \end{aligned} \tag{6.47}$$

beschreiben. Dabei bedeutet i₁ die Zahl der v o r , z₁ die Zahl der h i n t e r dem 1-Testruf wartenden Rufe; i₂ ist die Zahl der insgesamt in der zweiten Warteschlange Wartenden (Die Definitionen des ξ₁(t)-Prozesses stimmen also bei BD1 und BD3a formal überein, obwohl die Warteprozesse völlig unterschiedlich sind).

Der ξ₁(t)-Prozess beginnt stets mit i₁ = 0, für i₁ > 0 entwickeln sich nur Warteprozesse für partielle Wartezeiten. Die Antreffwahrscheinlichkeiten sind

$$P_1(0, z_1, i_2) = p(n; z_1, i_2) . \tag{6.48}$$

In Bild 23 ist der Wartezustandsraum des Warteprozesses für das System mit $s_1 = 3$ und $s_2 = 2$ Wartepätzen angegeben. Als Startzustand wurde beispielsweise der Zustand $(0,1,1)$ gewählt. Wegen der Disziplin D3a kann der Warteprozess entweder durch Abfertigung oder durch Verdrängung des Testrufs enden.

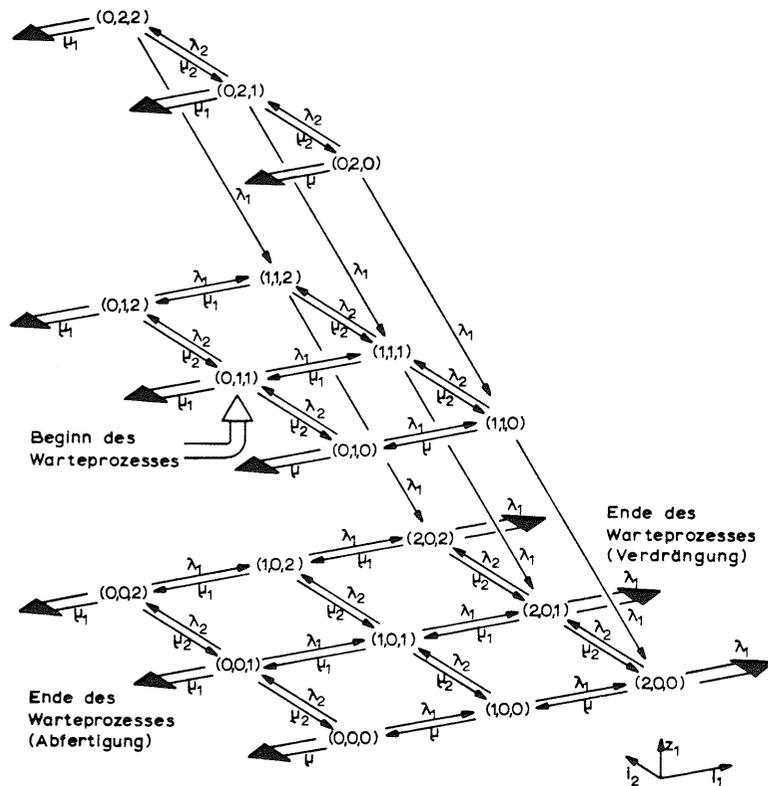


Bild 23. Wartezustandsraum für Warten nach BD3a
 Beispiel: $s_1 = 3, s_2 = 2, \xi_1(0) = (0,1,1)$

Der gesamte Zustandsraum besteht aus gleich vielen Zuständen wie im Falle BD1. Wie man aus Bild 23 leicht entnimmt, kann der gesamte Zustandsraum in Ebenen ($z_1 = 0, 1, \dots, s_1 - 1$) unterteilt werden. Die Zustände der untersten Ebene ($z_1 = 0$) sind unabhängig von allen anderen Zuständen und können für sich durch ein System von Rück-

wärts-Gleichungen beschrieben werden. Mit Hilfe der Lösungen kann dies fortschreitend auch für die Ebenen $z_1 = 1, 2, \dots, s_1 - 1$ durchgeführt werden.

b) Wartezeitverteilungsfunktionen

Für die bedingten Wartezeitverteilungsfunktionen $w_1(t|i_1, z_1, i_2)$ der wartenden 1-Rufe ohne Ansehen, ob sie später abgefertigt oder verdrängt werden, gilt ein System von Kolmogoroff-Rückwärts-Gleichungen. Für die Verteilungsfunktionen erfolgreich bzw. erfolglos wartender 1-Rufe, $w_1^*(t|i_1, z_1, i_2)$ bzw. $w_1^{**}(t|i_1, z_1, i_2)$, gilt dasselbe Differentialgleichungssystem, jedoch mit anderen Anfangsbedingungen. Die Anfangsbedingungen werden allgemein aus demselben System an der Stelle $t = 0+$ bestimmt (siehe Abschnitt 4.3.3), wobei folgende Endekoeffizienten einzusetzen sind:

$$\varepsilon_1(i_1, z_1, i_2) = \begin{cases} \mu & \text{für } i_1 = i_2 = 0 \\ \mu_1 & \text{für } i_1 = 0, i_2 > 0 \\ \lambda_1 & \text{für } i_1 = s_1 - 1, z_1 = 0 \\ 0 & \text{sonst} \end{cases} \quad (6.49a)$$

$$\varepsilon_1^*(i_1, z_1, i_2) = \begin{cases} \mu & \text{für } i_1 = i_2 = 0 \\ \mu_1 & \text{für } i_1 = 0, i_2 > 0 \\ 0 & \text{sonst} \end{cases} \quad (6.49b)$$

$$\varepsilon_1^{**}(i_1, z_1, i_2) = \begin{cases} \lambda_1 & \text{für } i_1 = s_1 - 1, z_1 = 0 \\ 0 & \text{sonst} \end{cases} \quad (6.49c)$$

Nach Gl.(4.33) ist $w_1(0|i_1, z_1, i_2) = 1$ und nach Gl.(4.36) gilt

$$w_1^*(0|i_1, z_1, i_2) + w_1^{**}(0|i_1, z_1, i_2) = 1.$$

Beispielsweise gilt für den Zustand $(1,1,1)$ folgende Differentialgleichung:

$$\frac{dw_1(t|1,1,1)}{dt} = -(\lambda + \mu)w_1(t|1,1,1) + \lambda_1 w_1(t|2,0,1) + \lambda_2 w_1(t|1,1,2) + \mu_1 w_1(t|0,1,1) + \mu_2 w_1(t|1,1,0) \quad (6.50)$$

Für das Beispiel aus Bild 23 gilt folgendes System von Gleichungen für die Laplace-Transformierten $w_1(s|i_1, z_1, i_2)$:

Anruf- bzw. Enderaten wird im Abschnitt 7.4 behandelt). Als Abfertigungsdisziplin innerhalb der einzelnen Warteschlangen wurde D1 zugrunde gelegt. Für die Abfertigung zwischen den Warteschlangen werden drei Disziplinen vorausgesetzt:

- Abfertigung der ersten Warteschlange mit nichtunterbrechender Priorität
- Abfertigung nach Wahrscheinlichkeiten proportional zu den momentanen Warteschlangenlängen
- Gleichwahrscheinliche Abfertigung der Warteschlangen.

In den Bildern 24a,b, S.132 und 133, sind zunächst die charakteristischen Größen W_J , B_J und t_{W_J} , $J = 1,2$, in Abhängigkeit des Angebots $A_1 = A_2$ dargestellt. Aus Bild 24a geht hervor, daß sich die verschiedenen Abfertigungsdisziplinen erst bei größeren Angebotswerten auf Verlust- und wartewahrscheinlichkeiten auswirken. Infolge der begrenzten Speicherkapazitäten und der damit verbundenen Verlustwahrscheinlichkeiten nehmen die Wartewahrscheinlichkeiten nach Überschreiten eines Maximalwerts mit zunehmender Überlastung wieder ab. Aus Bild 24b läßt sich erkennen, daß die Prioritätsdisziplin schon bei kleinen Angebotswerten einen starken Einfluß auf die mittleren Wartezeiten der wartenden ausübt, während die Abfertigung nach Wahrscheinlichkeiten proportional zu den momentanen Warteschlangenlängen den durch die verschiedenen großen Speicherkapazitäten erwarteten Unterschied in den mittleren Wartezeiten nahezu kompensiert.

Die Verläufe der Wartezeitverteilungsfunktionen $W_J(>t)/W_J$, $J = 1,2$, sind in Bild 25, S.134, für die drei obengenannten Abfertigungsdisziplinen zwischen den Warteschlangen angegeben für $A_1 = A_2 = 1,5$. Auch hier bedingt die Prioritätsdisziplin einen starken Unterschied zwischen den Wartezeitverteilungen bezüglich der beiden Warteschlangen, während sich die Abfertigung nach Wahrscheinlichkeiten proportional zu den momentanen Warteschlangenlängen kompensierend auswirkt.

Das lineare Gleichungssystem für die stationären Zustandswahrscheinlichkeiten wurde nach dem Verfahren der "sukzessiven Überrelaxation" aufgelöst. Für die Lösung der linearen Differentialgleichungssysteme wurde das in dieser Arbeit vorgeschlagene Verfahren der "sukzessiven Approximation durch Potenzreihenentwicklungen" (vergl. Abschnitte 8.2.4 und 8.3.2) angewendet. Dieses Verfahren ermöglicht numerisch die Auflösung linearer Differentialgleichungssysteme extrem hohen Ranges mit beliebig vorschreibbarer Genauigkeit.

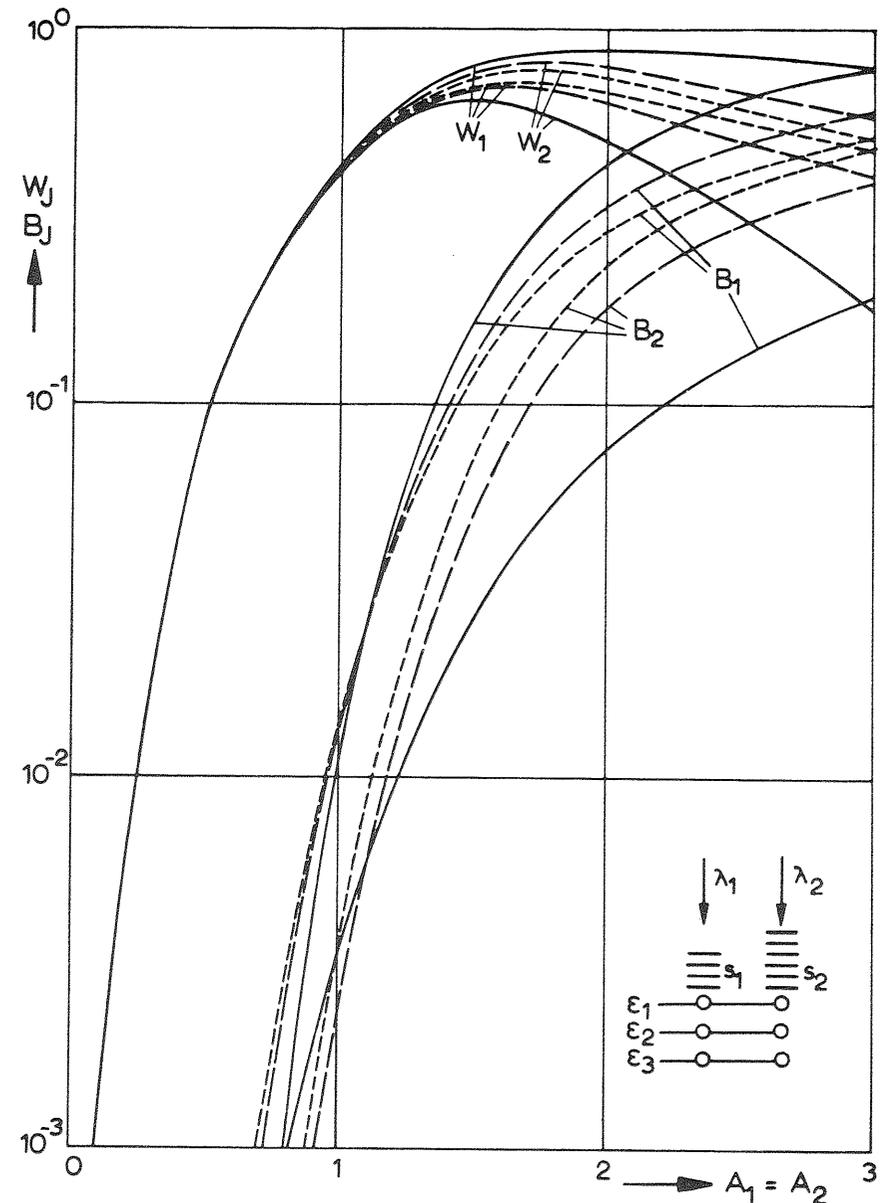


Bild 24a. Warte- und Verlustwahrscheinlichkeiten in Abhängigkeit des Angebots
 System: $n=k=3$, $\rho=2$, $s_1=4$, $s_2=6$ (vollk. Erreichbarkeit)
 ——— Abfertigung der ersten Warteschlange mit nichtunterbrechender Priorität
 - - - - - Abfertigung nach Wahrscheinlichkeiten proportional zu den momentanen Warteschlangenlängen
 ······ Gleichwahrscheinliche Abfertigung der Warteschlangen

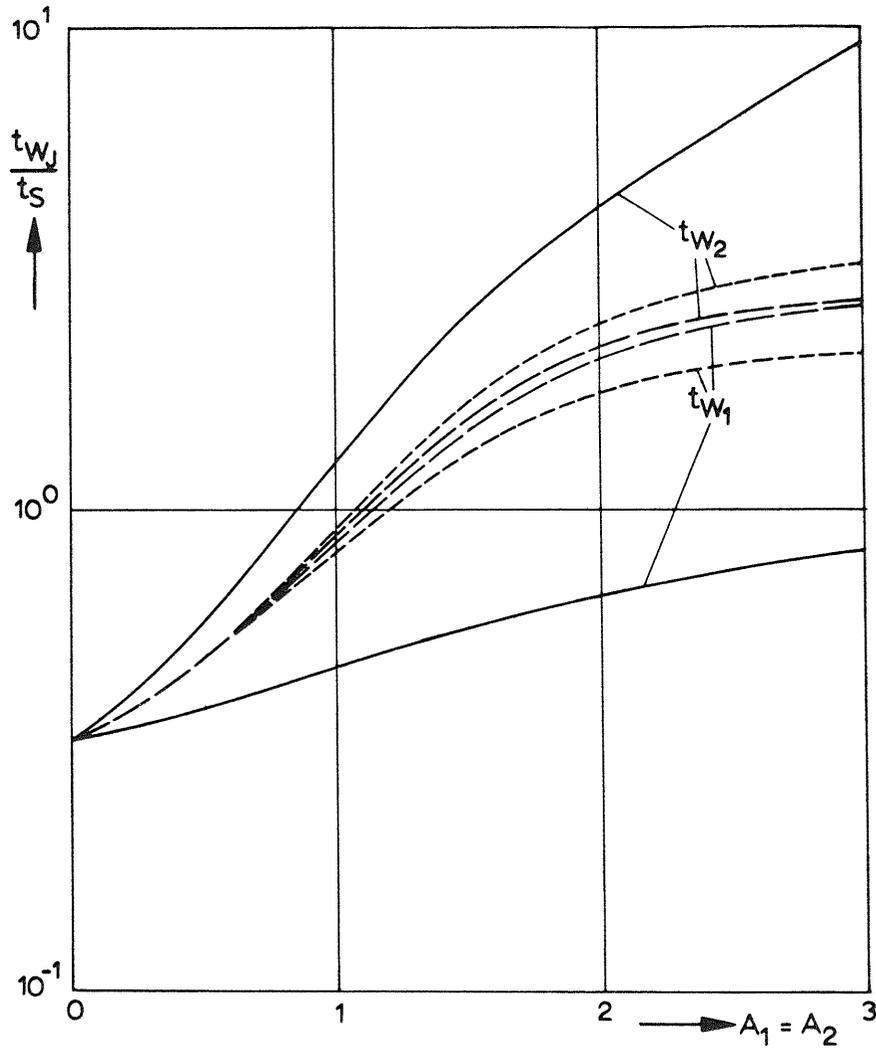


Bild 24b. Mittlere Wartezeiten in Abhängigkeit des Angebots
 System: $n=k=3, g=2, s_1=4, s_2=6$ (vollk. Erreichbarkeit)
 — Abfertigung der ersten Warteschlange mit nichtunterbrechender Priorität
 - - - Abfertigung nach Wahrscheinlichkeiten proportional zu den momentanen Warteschlangenlängen
 - · - Gleichwahrscheinliche Abfertigung der Warteschlangen

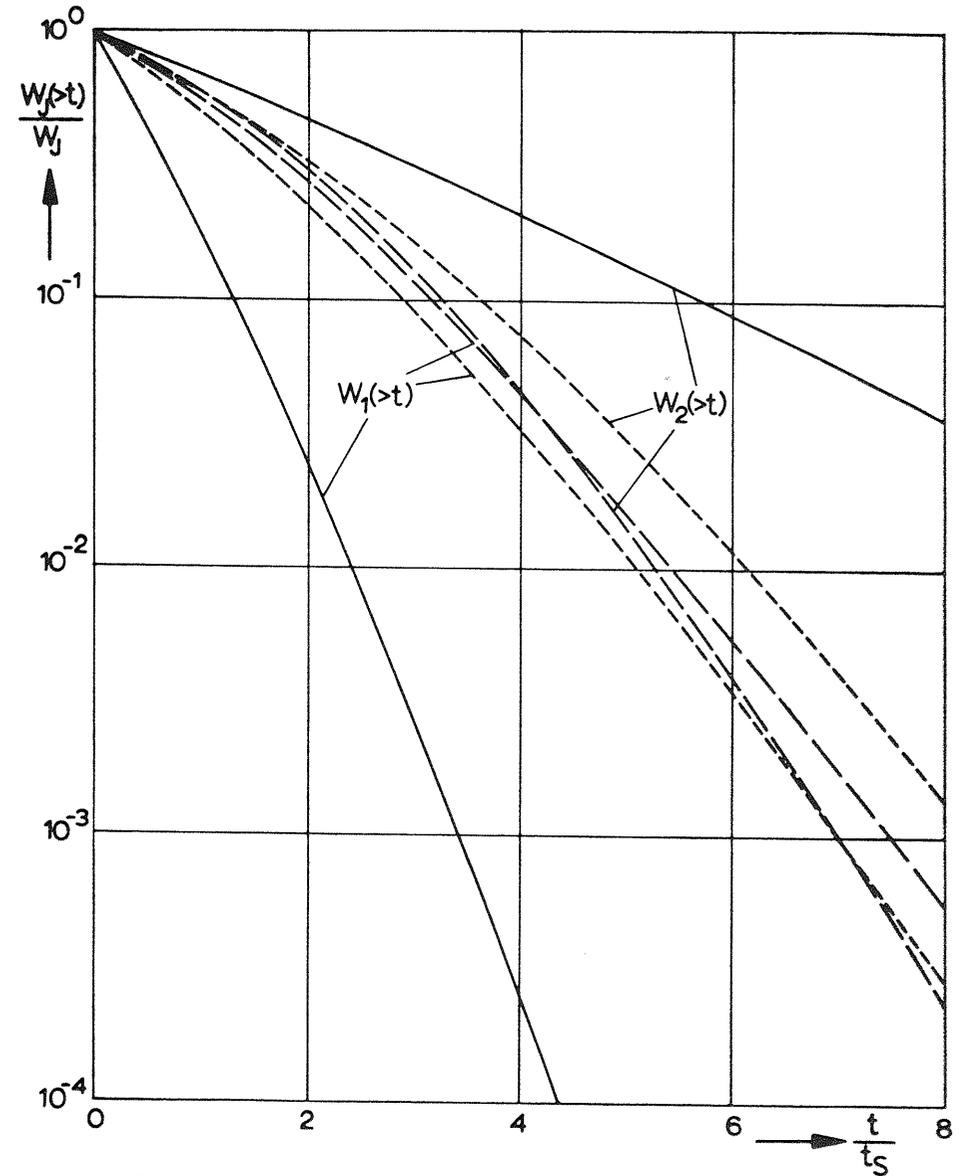


Bild 25. Wartezeitverteilungsfunktionen
 Angebote $A_1 = A_2 = 1,5$
 System: $n=k=3, g=2, s_1=4, s_2=6$ (vollk. Erreichbarkeit)
 — Abfertigung der ersten Warteschlange mit nichtunterbrechender Priorität
 - - - Abfertigung nach Wahrscheinlichkeiten proportional zu den momentanen Warteschlangenlängen
 - · - Gleichwahrscheinliche Abfertigung der Warteschlangen

7. BEDIENTUNGSSYSTEME MIT UNVOLLKOMMENER ERREICHBARKEIT UND MEHREREN WARTESCHLANGEN

In Kapitel 6 wurden für Bedienungssysteme mit vollkommener Erreichbarkeit und $g = 2$ Warteschlangen die Zustandswahrscheinlichkeiten, charakteristischen Größen und Wartezeiten bei unterschiedlichen Abfertigungsdisziplinen innerhalb und zwischen den Warteschlangen behandelt. Diese Überlegungen sollen nun auf Bedienungssysteme mit unvollkommener Erreichbarkeit (vergl. Bild 1(c)) erweitert werden. Aus Übersichtlichkeitsgründen werden die Ausführungen zumeist anhand einer einfachen Mischung für $g = 2$ Teilgruppen und $n = 3$ Leitungen demonstriert. Bei der Behandlung der Wartezeiten werden nur noch die wichtigsten Abfertigungsdisziplinen betrachtet; die Verhältnisse für die restlichen Abfertigungsdisziplinen können mit Hilfe der Überlegungen bei Bedienungssystemen mit vollkommener Erreichbarkeit leicht eingesehen werden.

7.1 Zustandswahrscheinlichkeiten

Es werde zunächst allgemein angenommen, daß das Bedienungssystem g Zubringerteilgruppen, n unvollkommen erreichbare Leitungen (d.h. $k_J < n$) und s_J Warteplätze im J -ten Wartespeicher besitzt, $J = 1, 2, \dots, g$. Der Zufallsprozeß der Systemzustände ($\xi(t)$ -Prozeß) wird durch eine $(n+g)$ -dimensionale Zufallsvariable beschrieben:

$$\xi(t) = (x_1, x_2, \dots, x_n; z_1, z_2, \dots, z_g), \quad (7.1)$$

wobei $x_I \in [0, 1]$, $z_J \in [0, s_J]$, $I = 1, 2, \dots, n$, $J = 1, 2, \dots, g$ (vergl. Abschnitt 4.4.1).

Bei Bedienungssystemen mit vollkommener Erreichbarkeit war grundsätzlich eine Aufteilung des $(n+g)$ -dimensionalen Zustandsraumes in einen n - und einen g -dimensionalen Unterraum möglich, welche beide nur über einen Schlüsselzustand, der zu beiden Unterräumen gehörte, verbunden waren. Dies war nur deshalb möglich, weil sich erst bei vollbelegtem Leitungsbündel Warteschlangen aufbauen können.

Im Falle von unvollkommen erreichbaren Leitungen kann sich eine Warteschlange J bereits dann aufbauen, wenn alle $k_J < n$ von der J -ten Teilgruppe aus erreichbaren Leitungen belegt sind. Aus diesem

Grunde ist i.a. keine Aufspaltung des gesamten Zustandsraumes in Unterräume möglich.

Die maximale Zahl von Zuständen der $(n+g)$ -dimensionalen Zufallsvariablen nach Gl.(7.1) ist

$$2^n \cdot \prod_{J=1}^g (s_J + 1)$$

Bei vollkommener Erreichbarkeit hatten wir nur

$$2^n - 1 + \prod_{J=1}^g (s_J + 1)$$

mögliche Zustände. Bei unvollkommener Erreichbarkeit existiert nicht zu jeder Kombination von belegten Leitungen die volle Kombination von belegten Warteplätzen; die Zahl der möglichen Zustände liegt deshalb zwischen diesen beiden (extremen) Werten.

Wie bei Bedienungssystemen mit vollkommener Erreichbarkeit beeinflusst die Abfertigungsdisziplin innerhalb der Warteschlangen den $\xi(t)$ -Prozeß nicht. Die Abfertigungsdisziplin zwischen den Warteschlangen wird durch die bedingte Wahrscheinlichkeit p_{IJ} beschrieben, welche dafür steht, daß die Warteschlange J abgefertigt wird, vorausgesetzt, daß die Leitung I frei wird. p_{IJ} hängt i.a. vom Systemzustand ab.

Es werden folgende Abkürzungen vereinbart:

$$\lambda = \sum_{J=1}^g \lambda_J, \quad (7.2)$$

$$\epsilon_{IJ} = \epsilon_I \cdot p_{IJ}, \quad \begin{matrix} I = 1, 2, \dots, n, \\ J = 1, 2, \dots, g, \end{matrix} \quad (7.3a)$$

$$\mu = \sum_{I=1}^n \epsilon_I. \quad (7.3b)$$

Ferner sei noch an die aus Gl.(2.12a) und Gl.(7.3a) resultierende Beziehung

$$\sum_{J=1}^g \epsilon_{IJ} = \epsilon_I \quad (7.4)$$

erinnert.

In Bild 26 ist der Zustandsraum für das Beispiel $g = 2$, $n = 3$ (Mischung siehe Bild 26) angegeben. Es wurde dabei beispielsweise angenommen, daß die Leitungen geordnet abgesucht werden.

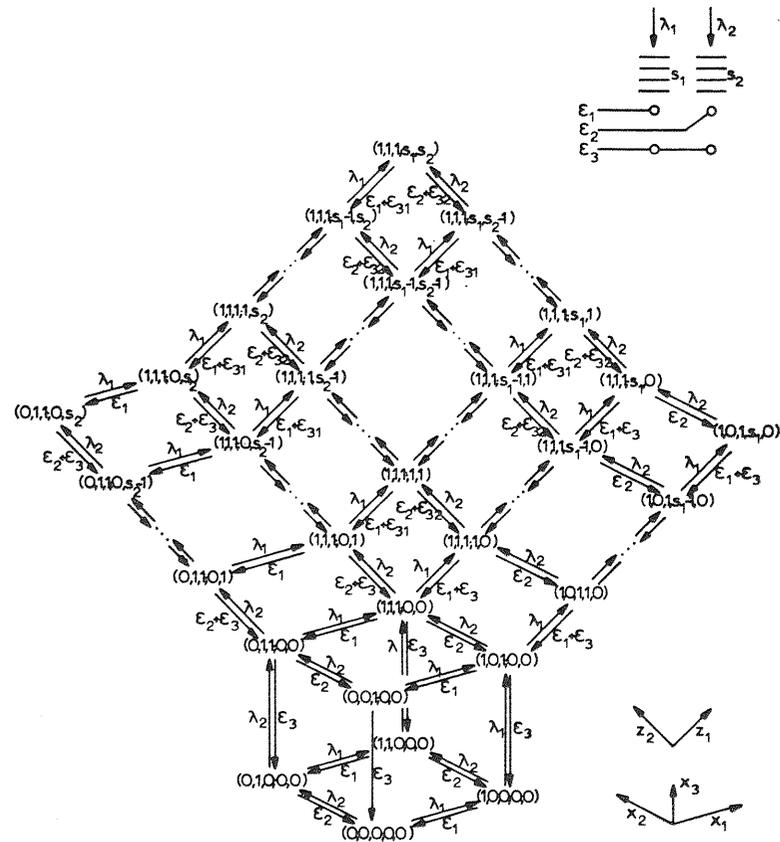


Bild 26. Zustandsraum und Übergangskoeffizienten für ein Bedienungssystem unvollkommener Erreichbarkeit mit $n = 3$ Leitungen und $g = 2$ Warteschlangen (geordnetes Absuchen der Leitungen)

Wie man aus dem Zustandsdiagramm Bild 26 entnimmt, läßt sich keine Aufteilung des gesamten Zustandsraumes in Teilräume durchführen. Die Zustandsgleichungen werden mit Hilfe des Zustandsraumes und der Übergangskoeffizienten entsprechend Abschnitt 4.1.5 gewonnen, indem für jeden Zustand die Kolmogoroff-Vorwärts-Gleichung für den stationären Fall ("statistisches Gleichgewicht") angewendet wird. Auf diese Weise erhält man - zusammen mit der "Normalisierungsbedingung" entsprechend Gl.(4.15) - ein lineares inhomogenes Gleichungssystem

vom Range der Zahl der Zustände. Im Beispiel Bild 26 ist der Rang des Gleichungssystems

$$2^3 - 3 + (s_1 + 1)(s_2 + 1) + (s_1 + 1) + (s_2 + 1).$$

Es sei abschließend noch bemerkt, daß sich - im Gegensatz zu vollkommener Erreichbarkeit - bei Betrachtung der Sonderfälle

- gleiche Enderaten der Leitungen
- feste Prioritäten zwischen den Warteschlangen
- variable Prioritäten zwischen den Warteschlangen entsprechend der momentan längsten Warteschlange
- Abfertigung nach Wahrscheinlichkeiten proportional zu den momentanen Warteschlangenlängen

keine wesentlichen Vereinfachungen im Zustandsraum Bild 26 ergeben.

7.2 Charakteristische Größen

Die charakteristischen Größen sollen der Einfachheit halber für 1-Rufe des in Bild 26 zugrunde gelegten Systems angegeben werden. Die Definitionen können sinngemäß auf beliebige Mischungen erweitert werden.

a) Wartewahrscheinlichkeit W_1

$$W_1 = \sum_{z_1=0}^{s_1-1} \left[p(1,0,1,z_1,0) + \sum_{z_2=0}^{s_2} p(1,1,1,z_1,z_2) \right] \quad \text{für D1, D2, D3b} \quad (7.5a)$$

$$W_1 = \sum_{z_1=0}^{s_1} \left[p(1,0,1,z_1,0) + \sum_{z_2=0}^{s_2} p(1,1,1,z_1,z_2) \right] \quad \text{für D3a.} \quad (7.5b)$$

b) Verlustwahrscheinlichkeit B_1

$$B_1 = p(1,0,1,s_1,0) + \sum_{z_2=0}^{s_2} p(1,1,1,s_1,z_2) \quad \text{für D1, D2, D3b} \quad (7.6a)$$

$$B_1 = 0 \quad \text{für D3a.} \quad (7.6b)$$

c) Erfolgs-Wartewahrscheinlichkeit W_1^*

$$W_1^* = W_1 \quad \text{für D1, D2, D3b} \quad (7.7a)$$

$$W_1^* = \sum_{z_1=0}^{s_1-1} \left[p(1,0,1,z_1,0) + \sum_{z_2=0}^{s_2} p(1,1,1,z_1,z_2) \right] \quad \text{für D3a.} \quad (7.7b)$$

d) Mißerfolgs-Wartewahrscheinlichkeit W_1^{**}

$$W_1^{**} = 0 \quad \text{für D1, D2, D3b} \quad (7.8a)$$

$$W_1^{**} = p(1,0,1; s_1, 0) + \sum_{z_2=0}^{s_2} p(1,1,1; z_1, z_2) \quad \text{für D3a.} \quad (7.8b)$$

Zwischen W_1 , W_1^* und W_1^{**} gilt weiterhin die Beziehung (6.14).

e) Verkehrswert Y

$$Y = \sum_{x_1=0}^1 \sum_{x_2=0}^1 \sum_{x_3=0}^1 \sum_{z_1=0}^{s_1} \sum_{z_2=0}^{s_2} (x_1 + x_2 + x_3) p(x_1, x_2, x_3; z_1, z_2). \quad (7.9)$$

f) Mittlere Warteschlangenlänge Ω_1

$$\Omega_1 = \sum_{z_1=0}^{s_1} z_1 \left[p(1,0,1; z_1, 0) + \sum_{z_2=0}^{s_2} p(1,1,1; z_1, z_2) \right]. \quad (7.10)$$

g) Mittlere Wartezeit der wartenden 1-Rufe t_{W_1}

$$t_{W_1} = \frac{\Omega_1}{W_1 \cdot \lambda_1}. \quad (7.11)$$

7.3 wartezeiten

In Abschnitt 6.3 sind für Systeme mit vollkommener Erreichbarkeit, zwei verschiedenen Disziplinen zwischen den Warteschlangen und jeweils vier verschiedenen Disziplinen innerhalb der Warteschlangen die Warteprozesse erläutert worden. In diesem Abschnitt werden diese Überlegungen auf Systeme mit unvollkommener Erreichbarkeit erweitert. Die Konstruktionen der Warteprozesse werden an jeweils zwei Abfertigungsdisziplinen zwischen und innerhalb der Warteschlangen gezeigt.

Damit ein Warteprozess für einen J-Testruf ($\xi_J(t)$ -Prozess) überhaupt existieren kann, müssen notwendig alle k_J von der Teilgruppe J aus erreichbaren Leitungen belegt sein. Diese k_J Leitungsbelegungen haben in jedem Falle einen direkten Einfluß auf die Wartezeit eines J-Rufes. Bei der Definition des Zustandes ξ_J brauchen diese Leitungsbelegungen jedoch nicht berücksichtigt zu werden. Die Belegungen der restlichen $(n-k_J)$ Leitungen können direkt oder indirekt die Wartezeit des J-Testrufs beeinflussen und müssen aus diesem Grunde in Form von Komponenten in die Beschreibung eingehen, vergl. Abschnitt 4.4.2. Ohne Einschränkung der Allgemeinheit seien wieder Warteprozesse bezüglich Rufen des ersten Anrufprozesses betrachtet.

7.3.1 Abfertigung zwischen den Warteschlangen nach Wahrscheinlichkeiten unabhängig von den momentanen Warteschlangenlängen (A)

Die Abfertigung zwischen den Warteschlangen, welche von einer betrachteten Leitung aus abgefertigt werden können, erfolge nach bedingten Wahrscheinlichkeiten unabhängig von den momentanen Warteschlangenlängen der (nichtleeren) Wartespeicher.

7.3.1.1 Abfertigung in der Reihenfolge des Eintreffens (AD1)

a) Der Warteprozess

Als Beispiel werde zunächst das Bedienungssystem mit $n = 3$ Leitungen und $g = 2$ Teilgruppen nach Bild 26 betrachtet. Damit ein $\xi_1(t)$ -Prozess existieren kann, müssen Leitung 1 und Leitung 3 belegt sein; Leitung 2 kann dagegen belegt oder frei sein. Ist Leitung 2 ebenfalls belegt, so kann sich eine zweite Warteschlange aufbauen. 1-Rufe, welche vor dem 1-Testruf warten, üben einen direkten Einfluß, 1-Rufe, welche hinter dem 1-Testruf warten, üben wegen AD1 keinen Einfluß auf die Wartezeit des 1-Testrufs aus. 2-Rufe beeinflussen über Leitung 2 indirekt und über Leitung 3 direkt die Wartezeit des 1-Testrufs.

Der $\xi_1(t)$ -Prozess wird durch eine dreidimensionale Zufallsvariable

$$\xi_1(t) = (x_2; i_1, i_2), \quad \begin{aligned} x_2 &= 0, 1, \\ i_1 &= 0, 1, \dots, s_1-1, \\ i_2 &= \begin{cases} 0 & \text{für } x_2 = 0 \\ 0, 1, \dots, s_2 & \text{für } x_2 = 1 \end{cases} \end{aligned} \quad (7.12)$$

eindeutig definiert. Es bedeuten hierbei:

- x_2 Zustand der Leitung 2
- i_1 Zahl der vor dem 1-Testruf wartenden 1-Rufe
- i_2 Zahl der insgesamt wartenden 2-Rufe.

Die Antreffwahrscheinlichkeiten sind

$$P_1(x_2; i_1, i_2) = p(1, x_2, 1; i_1, i_2) \quad (7.13)$$

für x_2, i_1, i_2 nach Gl. (7.12).

Bild 27 zeigt den Wartezustandsraum des Warteprozesses. Aus dem Wartezustandsraum entnimmt man, daß wieder eine Aufteilung des Gesamtsystems in Teilsysteme möglich ist, wenn nacheinander die Zustände für $i_1 = 0, 1, \dots, s_1-1$ betrachtet werden.

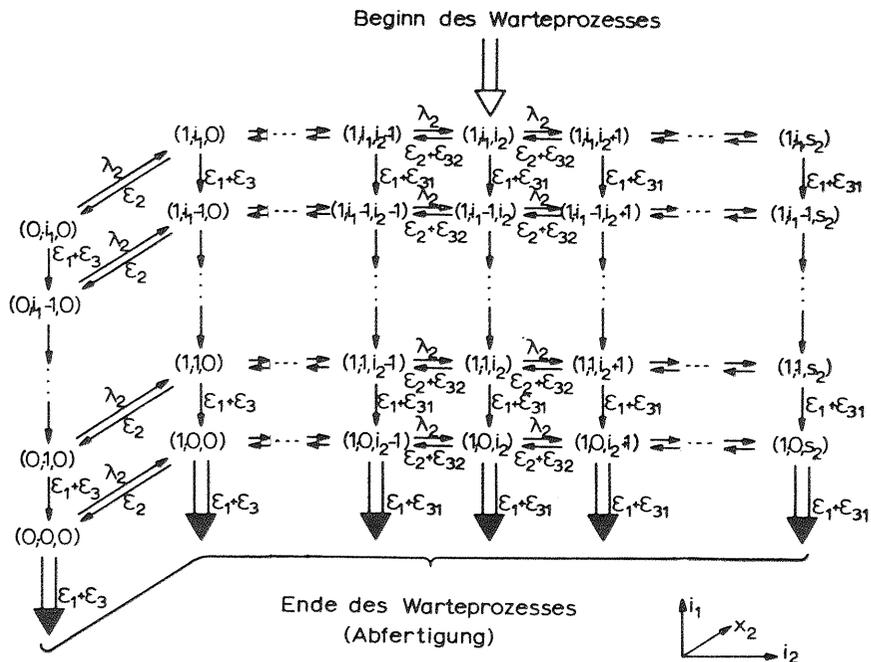


Bild 27. Wartezustandsraum für Warten nach AD1
 Beispiel: $\xi_1(0) = (1; i_1, i_2)$

b) Wartezeitverteilungsfunktionen

Die bedingten Wartezeitverteilungsfunktionen $w_1(t|x_2; i_1, i_2)$ gehorchen einem System von gekoppelten Differentialgleichungen, wobei die Anfangsbedingungen einheitlich $w_1(0|x_2; i_1, i_2) = 1$ sind. Für den Zustand $(1; i_1, 0)$, $i_1 > 0$, erhält man durch Anwendung der Kolmogoroff-Rückwärts-Gleichung beispielsweise

$$\frac{dw_1(t|i_1, i_1, 0)}{dt} = -(\lambda_2 + \mu)w_1(t|i_1, i_1, 0) + \lambda_2 w_1(t|i_1, i_1, 1) + (\epsilon_1 + \epsilon_3) w_1(t|i_1, i_1 - 1, 0) + \epsilon_2 w_1(t|0, i_1, 0) \quad (7.14)$$

Das Gleichungssystem der Laplace-Transformierten der bedingten Wartezeitverteilungsfunktionen hat die folgende Gestalt:

(7.15)

$(0, 0, 0) \quad (1, 0, 0) \quad (1, 0, 1) \quad \dots \quad (1, 0, s_2 - 1) \quad (1, 0, s_2)$	$(0, 1, 0) \quad (1, 1, 0) \quad (1, 1, 1) \quad \dots \quad (1, 1, s_2 - 1) \quad (1, 1, s_2)$	$(0, s_1 - 1, 0) \quad (1, s_1 - 1, 0) \quad \dots \quad (1, s_1 - 1, s_2)$	1
$(s_1 \lambda_2 + \epsilon_1 + \epsilon_3) - \lambda_2$ $-\epsilon_2 (s + \lambda_2 + \mu) - \lambda_2$ $-\dots$ $-(\epsilon_2 + \epsilon_{32})(s + \lambda_2 + \mu) - \lambda_2$ $-\dots$ $-(\epsilon_2 + \epsilon_{32})(s + \mu)$			1
$-(\epsilon_1 + \epsilon_3)$ $-\dots$ $-(\epsilon_1 + \epsilon_{31})$ $-\dots$ $-(\epsilon_1 + \epsilon_{31})$	$(s + \lambda_2 + \epsilon_1 + \epsilon_3) - \lambda_2$ $-\epsilon_2 (s + \lambda_2 + \mu) - \lambda_2$ $-\dots$ $-(\epsilon_2 + \epsilon_{32})(s + \lambda_2 + \mu) - \lambda_2$ $-\dots$ $-(\epsilon_2 + \epsilon_{32})(s + \mu)$		1
	$-(\epsilon_1 + \epsilon_3)$ $-\dots$ $-(\epsilon_1 + \epsilon_{31})$ $-\dots$ $-(\epsilon_1 + \epsilon_{31})$	$(s_1 \lambda_2 + \epsilon_1 + \epsilon_3) - \lambda_2$ $-\epsilon_2 (s + \lambda_2 + \mu) - \lambda_2$ $-\dots$ $-(\epsilon_2 + \epsilon_{32})(s + \mu)$	1

Wie aus Gl. (7.15) hervorgeht, kann allgemein eine Aufteilung des Gesamtsystems - hier des Ranges $s_1(s_2+2)$ - in s_1 Teilsysteme - hier des Ranges (s_2+2) - vorgenommen werden, indem zunächst das unabhängige Teilsystem für $i_1 = 0$ gelöst wird, dann mit dessen Lösungen das Teilsystem für $i_1 = 1$ gelöst wird usw.

Entsprechend zu den Beweisführungen im Abschnitt 6.3.1.1 und im Anhang A1 können über die Eigenwerte der Teilsysteme von (7.15) folgende Aussagen bewiesen werden:

1. Die Teilsysteme haben jeweils den gleichen Satz von Eigenwerten ϵ_p , $p = 1, 2, \dots, s_2 + 2$.
2. Die Eigenwerte jedes Teilsystems sind negativ-reell und verschieden, sie liegen sämtlich im Intervall

$$[-(2\lambda_2 + \mu + \epsilon_2 + \epsilon_{32}), -(\epsilon_1 + \epsilon_{31})].$$

Für kompliziertere Strukturen der Mischung kann i. a. nicht mehr

bewiesen werden, daß die Eigenwerte der Teilsysteme verschieden sind; die Teilsysteme der zugehörigen Matrix weichen dann i.a. von der Tridiagonalform ab, und es können deshalb mehrfache Eigenwerte nicht ausgeschlossen werden (vergl. Beispiel in Teilabschnitt c).

Die Lösungen des Differentialgleichungssystems haben im obigen Beispiel eine Gl.(6.23) entsprechende Form. Die gesamte Wartezeitverteilungsfunktion stellt sich mit den Antreffwahrscheinlichkeiten nach Gl.(7.13) wie folgt dar:

$$W_1(>t) = \sum_{i_1=0}^{s_1-1} \left[p(1,0,1,i_1,0) \omega_1(t|0,i_1,0) + \sum_{i_2=0}^{s_2} p(1,1,1,i_1,i_2) \omega_1(t|i_1,i_1,i_2) \right]. \quad (7.16)$$

Die bedingten mittleren Wartezeiten sowie die Momente der bedingten Wartezeitverteilungsfunktionen können wieder in einer Abschnitt 4.3 entsprechenden Weise aus dem System (7.15) berechnet werden.

c) Beispiel einer Mischung $n = 6, k = 4, g = 3$

Abschließend seien die Verhältnisse an einem in der Vermittlungstechnik häufig auftretenden Mischungstyp mit Verschränkung aufgezeigt. Bild 28 zeigt das Bedienungssystem, bestehend aus $g = 3$ Zubringerteilgruppen, $n = 6$ Leitungen, den Erreichbarkeiten $k_1=k_2=k_3=k_4$ und 3 Wartespeichern.

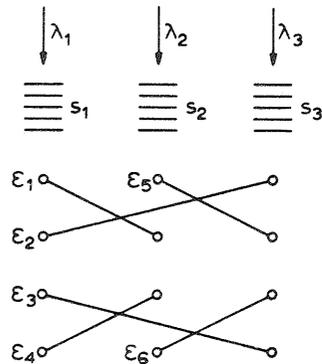


Bild 28. Bedienungssystem mit unvollkommen erreichbaren Leitungen (Homogene zyklische Mischung mit Übergreifen und Verschränken)

Der Zufallsprozeß der Systemzustände ($\xi(t)$ -Prozeß) wird durch die 9-dimensionale Zufallsvariable

$$\xi(t) = (x_1, x_2, x_3, x_4, x_5, x_6; z_1, z_2, z_3) \quad (7.17)$$

beschrieben. Der Warteprozeß bezüglich 1-Rufen ($\zeta_1(t)$ -Prozeß) kann bei der Abfertigungsdisziplin AD1 durch die 5-dimensionale Zufallsvariable

$$\zeta_1(t) = (x_5, x_6; i_1, i_2, i_3) \quad (7.18)$$

eindeutig gekennzeichnet werden.

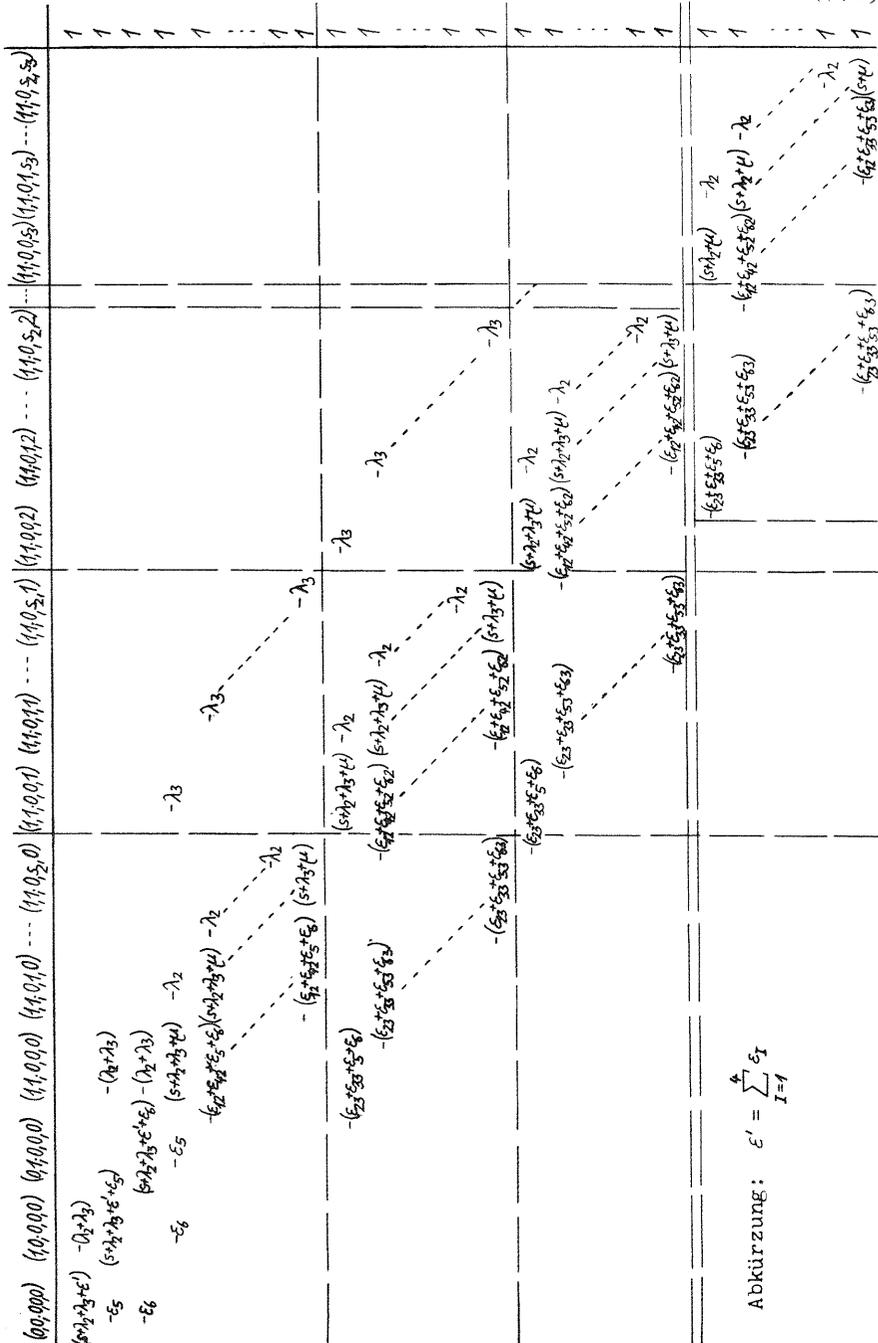
Wie im Abschnitt a) bereits erkannt wurde, läßt sich das Differentialgleichungssystem in Teilsysteme für $i_1 = 0, 1, \dots, s_1-1$ aufteilen. In Gl.(7.19) ist das erste Teilsystem ($i_1 = 0$) für die Laplace-Transformierten der bedingten Wartezeitverteilungsfunktionen $w_1(t|x_5, x_6; 0, i_2, i_3)$ angegeben. Es wurde ferner angenommen, daß die Leitungen geordnet abgesucht werden.

Für das Beispiel Bild 28 haben wir einen allgemeineren Matrixtyp erhalten, für den i.a. keine Aussagen über die Eigenwerte gemacht werden können. Aus physikalischen Gründen ist jedoch zu erwarten, daß sämtliche Eigenwerte negativ-reell sind. Das Beispiel Bild 28 wird in Abschnitt 7.4 numerisch behandelt.

7.3.1.2 Abfertigung in der inversen Reihenfolge des Eintreffens mit verdrängender Priorität (AD3a)

Es werde wieder das Bedienungssystem nach Bild 26 betrachtet. Sind Leitung 1 und Leitung 3 belegt, so erfolgt die Einordnung eines 1-Rufs in die erste Warteschlange wie im Falle vollkommener Erreichbarkeit (vergl. Abschnitt 6.3.1.3). Die zweite Warteschlange kann erst dann aufgebaut werden, wenn die zweite Leitung auch belegt ist. In der ersten Warteschlange *h i n t e r* dem 1-Testruf Wartende haben wegen AD3a keinen Einfluß auf die Wartezeit des 1-Testrufs, 2-Rufe über Leitung 3 einen direkten, über Leitung 2 einen indirekten Einfluß auf die Wartezeit des 1-Testrufs aus.

Der $\zeta_1(t)$ -Prozeß wird formal durch dieselbe Zufallsvariable wie bei AD1 nach Gl.(7.12) beschrieben. Als Antreffwahrscheinlichkeiten



haben wir

$$P_1(x_2; 0, i_2) = \sum_{z_1=0}^{s_1} p(1, x_2, 1; z_1, i_2) \quad (7.20)$$

In Bild 29 ist der Wartezustandsraum für den Warteprozess bezüglich 1-Kufen dargestellt. Der $\xi_1(t)$ -Prozess beginnt stets bei $i_1 = 0$, er endet entweder bei $i_1 = 0$ (Abfertigung) oder bei $i_1 = s_1 - 1$ (Verdrängung).

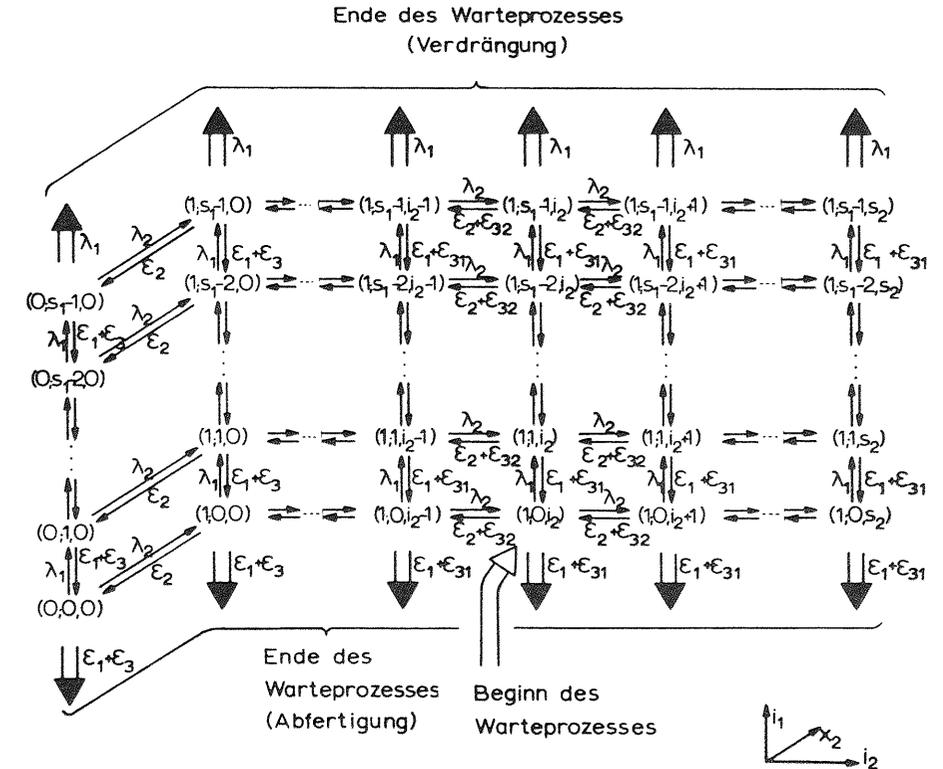


Bild 29. Wartezustandsraum für Warten nach AD3a

Beispiel: $\xi_1(0) = (1; 0, i_2)$

Das Aufstellen und die Lösung des Differentialgleichungssystems für die bedingten wartezeitverteilungsfunktionen $w_1(t|x_2; i_1, i_2)$ erfolgt

in entsprechender Weise wie in den vorausgegangenen Abschnitten gezeigt wurde.

7.3.2 Abfertigung zwischen den Warteschlangen nach Wahrscheinlichkeiten abhängig von den momentanen Warteschlangenlängen (B)

Hängt die Abfertigungsdisziplin zwischen den von einer betrachteten Leitung aus erreichbaren Wartespeichern von den momentanen Warteschlangenlängen ab, so müssen in den Fällen BD1 und BD3a - wie schon gezeigt wurde - zusätzlich die hinter dem 1-Testruf wartenden 1-Rufe berücksichtigt werden. Dadurch erhöht sich in diesen Fällen die Zahl der Komponenten des Vektors ξ_1 um eins.

7.3.2.1 Abfertigung in der Reihenfolge des Eintreffens (BD1)

Einem 1-Testruf nachfolgende und akzeptierte 1-Rufe beeinflussen dessen Wartezeit indirekt. Der Warteprozess kann für das Beispiel Bild 26 durch eine vierdimensionale Zufallsvariable beschrieben werden:

$$\xi_1(t) = (x_2; i_1, z_1, i_2) , \quad x_2 = 0, 1$$

$$i_1 = 0, 1, \dots, s_1 - 1$$

$$z_1 = 0, 1, \dots, s_1 - i_1 - 1, \quad (7.21)$$

$$i_2 = \begin{cases} 0 & \text{für } x_2 = 0 \\ 0, 1, \dots, s_2 & \text{für } x_2 = 1, \end{cases}$$

wobei in Erweiterung zu Gl.(7.12) z_1 die Zahl der hinter dem 1-Testruf wartenden Rufe bedeutet. Der $\xi_1(t)$ -Prozess beginnt stets mit $z_1 = 0$, $i_1 \geq 0$, und endet stets mit $i_1 = 0$, $z_1 \geq 0$ (vergl. auch Abschnitt 6.3.2.1).

7.3.2.2 Abfertigung in der inversen Reihenfolge des Eintreffens mit verdrängender Priorität (BD3a)

Der Warteprozess kann formal nach der Definition (7.21) beschrieben werden. Er beginnt stets bei $i_1 = 0$, $z_1 \geq 0$ und endet entweder bei $i_1 = 0$, $z_1 \geq 0$ (Abfertigung) oder bei $i_1 = s_1 - 1$, $z_1 = 0$ (Verdrängung), vergl. hierzu auch Abschnitt 6.3.2.3.

7.4 Numerische Beispiele

1. Beispiel

Als erstes Beispiel sei das Bedienungssystem mit $g = 2$ Zubringerteilgruppen, $n = 3$ unvollkommen erreichbaren Leitungen, den Erreichbarkeiten $k_1 = k_2 = 2$ und $s_1 = 4$ bzw. $s_2 = 6$ Warteplätzen nach Bild 26 betrachtet. Die Anrufrate λ_1 der ersten Zubringerteilgruppe sei variabel, die Anrufrate der zweiten Zubringerteilgruppe sei konstant $\lambda_2 = 1$. Die Enderaten der drei Leitungen seien unterschiedlich, und zwar $\epsilon_1 = 2$, $\epsilon_2 = 1$ und $\epsilon_3 = 10$ (Anruf- und Enderaten sind auf die Zeiteinheit normiert). Die Abfertigungsdisziplin innerhalb der Warteschlangen sei D1. Die Abfertigung zwischen den Warteschlangen (bezüglich Leitung 3) erfolge nach zwei verschiedenen Disziplinen:

- Abfertigung der ersten Warteschlange mit nichtunterbrechender Priorität
- Abfertigung nach Wahrscheinlichkeiten proportional zu den momentanen Warteschlangenlängen.

Bild 30, Seite 151, zeigt die charakteristischen Größen W_J , B_J und tw_J , $J = 1, 2$, in Abhängigkeit der Anrufrate λ_1 für die zwei genannten Abfertigungsdisziplinen zwischen den Warteschlangen. Während sich für die zwei Disziplinen die charakteristischen Werte bezüglich der 1-Rufe nur wenig unterscheiden, bedingt die Prioritätsdisziplin höhere Verluste und Wartezeiten für 2-Rufe.

Dem Beispiel liegt folgender Anwendungsfall zugrunde: Zwei Rechner (Leitung 1 bzw. 2) bedienen ausschließlich Rufe aus den ihnen zugeordneten Warteschlangen 1 bzw. 2. Aus Sicherheits- und Überlastungsgründen kann ein schnellerer ferner Rechner (Leitung 3) beide Warteschlangen zusätzlich bedienen. 1-Rufe stellen Realzeitprogramme dar, während 2-Rufe weniger dringende Hintergrundprogramme repräsentieren. Mit Hilfe der charakteristischen Größen in Bild 30 können Aussagen über die Bedienungsgüte des Systems, insbesondere über den Einfluß zunehmender Realzeitaufgaben auf die Hintergrundprogramme gemacht werden.

2. Beispiel

Das zweite Beispiel ist der Fernsprechvermittlungstechnik entnommen. Zwei symmetrische Bedienungssysteme mit $g = 3$ Teilgruppen,

$n = 6$ Leitungen, $s_1 = s_2 = s_3 = 4$ Warteplätzen und vollkommen ($k = 6$) bzw. unvollkommen erreichbaren ($k = 4$) Leitungen (Mischung nach Bild 28) sollen für verschiedene Angebotswerte bezüglich charakteristischer Größen und Wartezeitverteilungsfunktionen verglichen werden. Die Abfertigung innerhalb der Warteschlangen erfolge nach D1, nichtleere Warteschlangen werden gleichwahrscheinlich abgefertigt.

In Tabelle 2 sind für die drei Angebotswerte $A_J = 1$ (Normalbelastung), $A_J = 2$ (Grenzbelastung) und $A_J = 3$ (Überbelastung) die charakteristischen Größen B_J , W_J , Y , Ω_J und t_{W_J} angegeben, $J = 1, 2, 3$.

Angebot A_J	Erreichbarkeit k	Verlustwahrsch. B_J	Warte-wahrsch. W_J	Verkehrs-wert Y	Mittl. WS-länge Ω_J	Mittl. Wartezeit t_{W_J}/t_S
1	4	0,000704	0,133923	2,997889	0,052376	0,391093
	6	0,000386	0,098409	2,998841	0,032045	0,325634
2	4	0,090920	0,678501	5,454480	1,152212	0,849087
	6	0,084510	0,681000	5,492940	1,087175	0,798219
3	4	0,337179	0,644810	5,965386	2,552559	1,319540
	6	0,335676	0,652093	5,978913	2,548244	1,302597

Tabelle 2. Charakteristische Größen in Abhängigkeit von Angebot und Erreichbarkeit
 Systeme: $g = 3$, $n = 6$, $s_1 = s_2 = s_3 = 4$,
 $k = 6$ bzw. $k = 4$ (Mischung nach Bild 28)

Bild 31, Seite 152, zeigt die Verläufe der normierten Wartezeitverteilungsfunktionen $W_J(>t)/W_J$ für die drei Angebotswerte.

Aus den Werten der Tabelle 2 und den Kurven des Bildes 31 läßt sich erkennen, daß die relativen Unterschiede zwischen Bedienungssystemen vollkommener und unvollkommener Erreichbarkeit bei kleinen Angebotswerten am größten sind. Für $A_J \rightarrow \infty$ verschwinden die Unterschiede vollends, da sämtliche Einheiten des Systems ohnedies belegt sind; der Vorteil der (aufwendigeren) vollkommener Erreichbarkeit gegenüber unvollkommener Erreichbarkeit, welcher in der besseren Ausnützung der einzelnen Leitungen liegt, verschwindet.

Die linearen Gleichungssysteme für die stationären Zustandswahrscheinlichkeiten wurden mit Hilfe eines allgemeinen Programmes aus den Struktur-, Verkehrs- und Organisationsparametern ermittelt und nach dem Verfahren der "sukzessiven Überrelaxation" [19] aufgelöst. Das Programm wurde für beliebige einstufige Bedienungssysteme mit unsymmetrischer Struktur, unsymmetrischen Anruf- bzw. Enderaten sowie für mehrere Abfertigungsdisziplinen aufgebaut. Bei der Ermittlung des Gleichungssystems müssen aus allen möglichen Zustandskombinationen (im zweiten Beispiel unvollkommener Erreichbarkeit: 8000) die unter den gegebenen Voraussetzungen physikalisch nur auftretenden (in diesem Beispiel: 224) herausgesucht und die Übergangskoeffizienten zu den Nachbarzuständen bestimmt werden.

Die Lösung der linearen Differentialgleichungssysteme für die bedingten Wartezeitverteilungsfunktionen erfolgte nach dem Verfahren der "sukzessiven Approximation durch Potenzreihenentwicklungen" (vergl. Abschnitte 8.2.4 und 8.3.2). Im zweiten Beispiel waren im Falle vollkommener Erreichbarkeit 100, im Falle unvollkommener Erreichbarkeit 112 simultane Differentialgleichungen zu lösen. Bei der Zusammensetzung der bedingten Wartezeitverteilungsfunktionen zur gesamten Wartezeitverteilungsfunktion sind außerdem Untermengen der Zustandswahrscheinlichkeiten (100 bzw. 112) als "Antreffwahrscheinlichkeiten" nötig.

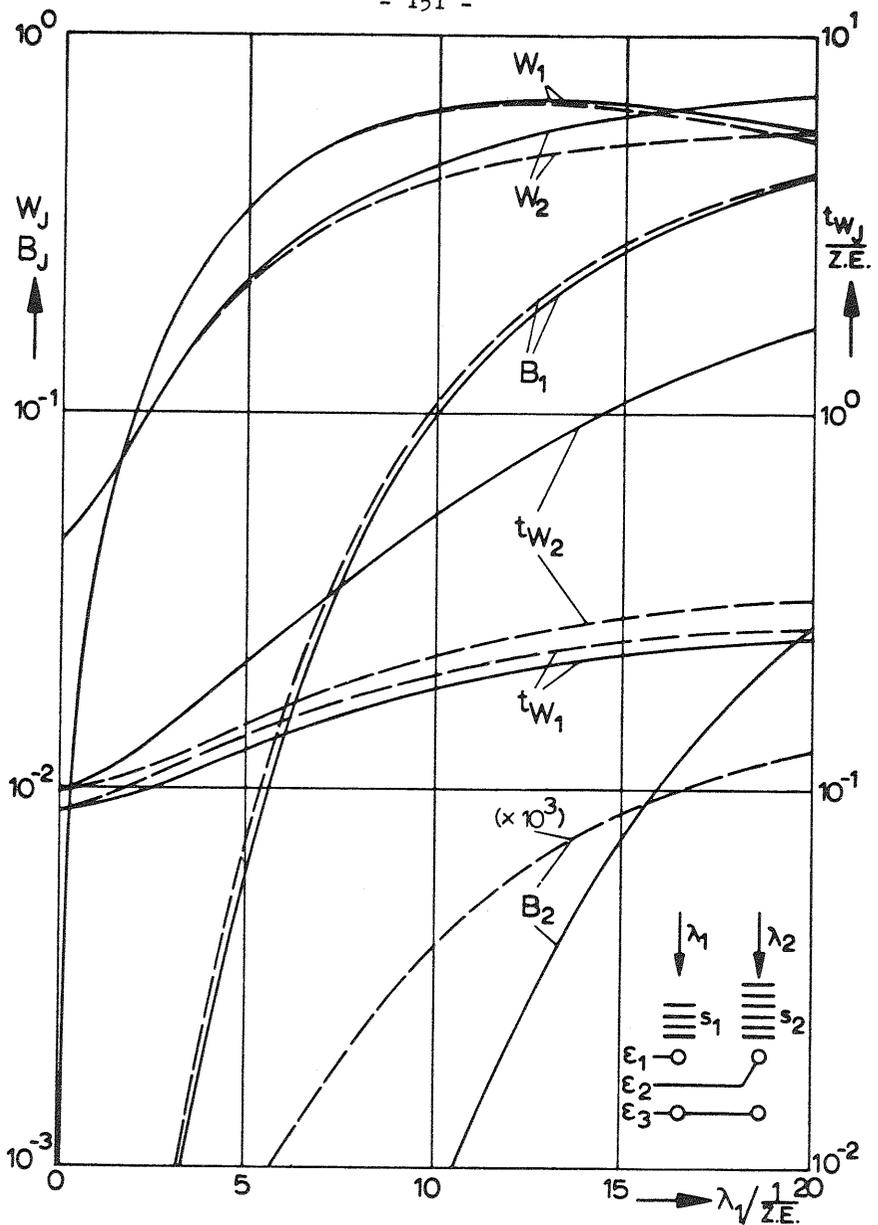


Bild 30. Charakteristische Größen in Abhängigkeit der Anrufrate λ_1
 System: $n=3, k=2, g=2, s_1=4, s_2=6$ (unvollk. Erreichbarkeit)
 $\lambda_2=1, \epsilon_1=2, \epsilon_2=1, \epsilon_3=10$ (in 1/Z.E.)
 ——— Abfertigung der ersten Warteschlange mit nichtunterbrechender Priorität
 - - - - Abfertigung nach Wahrscheinlichkeiten proportional zu den momentanen Warteschlangenlängen

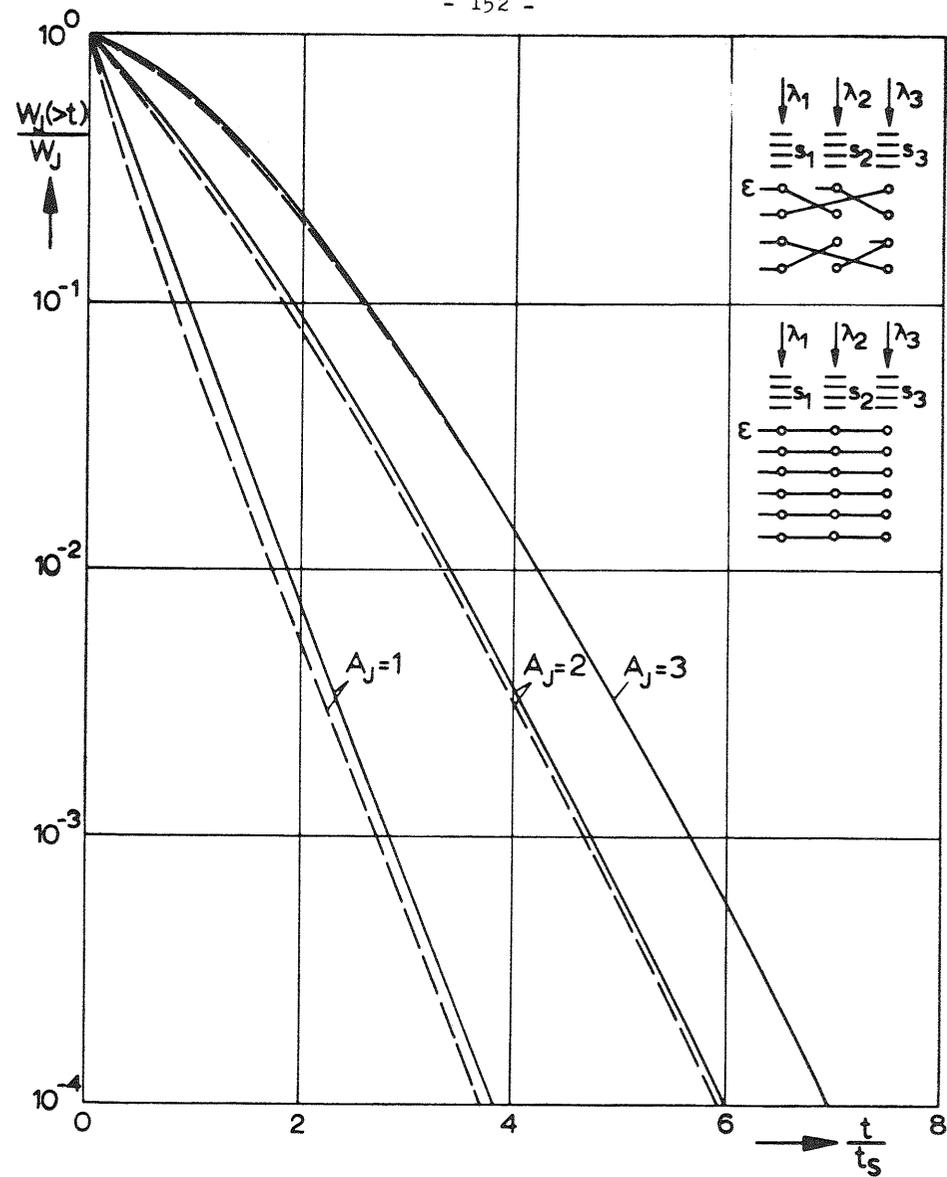


Bild 31. Wartezeitverteilungsfunktion in Abhängigkeit von Angebot und Erreichbarkeit

Systeme: $g = 3, n = 6, s_1 = s_2 = s_3 = 4$
 ——— $k = 4$ (unvollkommene Erreichbarkeit)
 - - - - $k = 6$ (vollkommene Erreichbarkeit)

8. APPROXIMATIVE BERECHNUNGSMETHODEN

Die exakte Berechnung von Markoff'schen Bedienungssystemen läuft auf zwei Grundprobleme hinaus:

1. Die Lösung eines linearen inhomogenen Gleichungssystems für die stationären Zustandswahrscheinlichkeiten.
2. Die Lösung eines linearen homogenen Differentialgleichungssystems erster Ordnung für die bedingten Wartezeitverteilungsfunktionen.

In den meisten Fällen ist der Rang der linearen Gleichungssysteme bzw. der linearen Differentialgleichungssysteme so hoch, daß eine exakte Berechnung mit Hilfe eines Digitalrechners entweder nicht möglich (zu großer Speicherbedarf) oder nicht wirtschaftlich (zu großer Rechenzeitbedarf) ist. Eine häufig angewendete Methode ist die Simulation von Bedienungssystemen, bei der System, Verkehr und Organisation künstlich nachgebildet werden. Neben langen Rechenzeiten hat diese Methode jedoch außerdem den Nachteil, daß man keinerlei analytische Erkenntnisse gewinnt. Aus diesen Gründen ergibt sich die Notwendigkeit approximativer Berechnungsmethoden. Exakte Lösungen bzw. Simulationsergebnisse können zur Beurteilung analytischer Näherungsverfahren herangezogen werden.

8.1 Zustandswahrscheinlichkeiten

8.1.1 Die exakte Lösung

Der Zufallsprozeß der Systemzustände ($\xi(t)$ -Prozeß) wird durch die Zustandswahrscheinlichkeiten $P(t; j)$ beschrieben, wobei (j) i.a. ein mehrdimensionaler Zustand ist, gebildet aus Komponenten der Leitungs- und Speicherbelegungen. Die Zustandswahrscheinlichkeiten gehorchen wegen der Markoff'schen Eigenschaft einem System von linearen homogenen Differentialgleichungen erster Ordnung (Kolmogoroff-Vorwärts-Gleichung), vgl. Gl.(4.12). Im speziellen Fall der Stationarität des Zufallsprozesses sind die Zustandswahrscheinlichkeiten zeitunabhängige Größen $p(j)$; das lineare homogene Differentialgleichungssystem reduziert sich zu einem linearen homogenen Gleichungssystem entsprechend Gl.(4.14)

$$q_j p(j) = \sum_{k \neq j} q_{kj} p(k) \quad (8.1a)$$

Gl.(8.1a) stellt die Gleichung für das "statistische Gleichgewicht" des Zustands (j) dar. Das homogene Gleichungssystem (8.1a) besteht zunächst aus sovielen Gleichungen wie es Zustände gibt. Die Gleichungen sind jedoch nicht linear unabhängig voneinander; es ist stets eine (beliebige) Gleichung linear abhängig von den anderen und kann daher weggelassen werden. Der Rang des gesamten Gleichungssystems (8.1a) ist um eins geringer als die Zahl der Zustände (Rangabfall = 1). Für die Zustandswahrscheinlichkeiten gilt - da die Zustände sich gegenseitig ausschließende Ereignisse sind - außerdem die "Normalisierungsbedingung"

$$\sum_j p(j) = 1. \quad (8.1b)$$

Das um eine (beliebige) Gleichung reduzierte System (8.1a) und Gl. (8.1b) bilden zusammen ein lineares inhomogenes Gleichungssystem mit verschwindendem Rangabfall (d.h. Systemdeterminante $\neq 0$) und kann eindeutig aufgelöst werden.

Sonderfälle:

1. Eindimensionale Zustandsräume

Ist der gesamte Zustandsraum eindimensional oder läßt sich aus dem Zustandsraum ein eindimensionaler Teilraum abspalten, so können die Zustandswahrscheinlichkeiten für den eindimensionalen Raum durch Rekursion bestimmt werden, vgl. Bild 32 und Ausführungen in Abschnitt 5.1.1. Es kann allgemein gezeigt werden, daß in diesem Fall

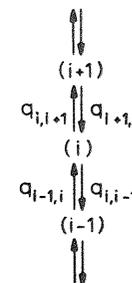


Bild 32. Eindimensionaler Zustandsraum

schon immer zwei Nachbarzustände im statistischen Gleichgewicht sind:

$$q_{i,i-1} \cdot p(i) = q_{i-1,i} \cdot p(i-1) \quad (8.2)$$

Mittels der Rekursionsgleichung (8.2) können alle Zustandswahrscheinlichkeiten eines eindimensionalen Raumes rekursiv berechnet werden, wobei die Zustandswahrscheinlichkeit eines Randzustandes als Parameter einget.

2. Mehrdimensionale Zustandsräume

Werden allgemein die Zustandswahrscheinlichkeiten des aus (n-1) Komponenten gebildeten Teilraumes eines n-dimensionalen Zustandsraumes als Parameter angesehen, so können wieder die restlichen Zustandswahrscheinlichkeiten durch Rekursion bestimmt werden. Bei zweidimensionalen Zustandsräumen können deshalb beispielsweise alle Zustandswahrscheinlichkeiten rekursiv bestimmt werden, wenn eine Randverteilung bekannt ist.

Wir sprechen von "Halbsymmetrie", wenn ein Zustand bereits mit allen seinen "unteren" (bzw. "oberen") Nachbarzuständen im statistischen Gleichgewicht ist. Betrachten wir hierzu z.B. einen zweidimensionalen Zustandsraum Bild 33.

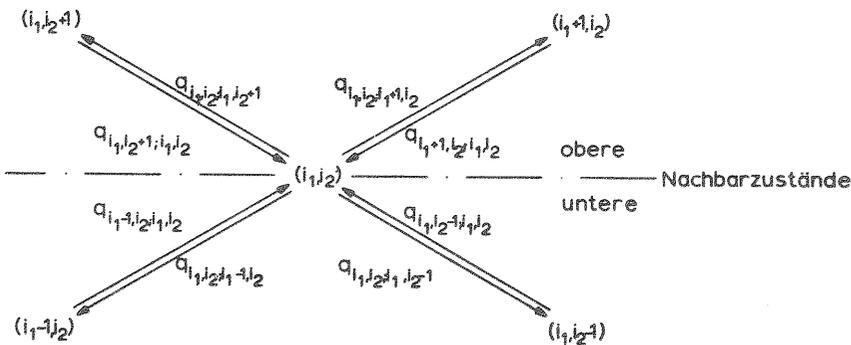


Bild 33. Zweidimensionaler Zustandsraum

Im Falle von "Halbsymmetrie" gelten folgende Gleichungen:

$$[q_{i_1, j_2; i_1-1, j_2} + q_{i_1, j_2; i_1, j_2-1}] p(i_1, j_2) = q_{i_1-1, j_2; i_1, j_2} \cdot p(i_1-1, j_2) + q_{i_1, j_2-1; i_1, j_2} \cdot p(i_1, j_2-1) \quad (8.3a)$$

$$[q_{i_1, j_2; i_1+1, j_2} + q_{i_1, j_2; i_1, j_2+1}] p(i_1, j_2) = q_{i_1+1, j_2; i_1, j_2} \cdot p(i_1+1, j_2) + q_{i_1, j_2+1; i_1, j_2} \cdot p(i_1, j_2+1) \quad (8.3b)$$

Gilt die Eigenschaft der Halbsymmetrie für alle Zustände eines beliebig vieldimensionalen Zustandsraumes, so können alle Zustandswahrscheinlichkeiten wieder rekursiv berechnet werden, wobei die Zustandswahrscheinlichkeit eines Randzustands als Parameter einget.

Sind in einem beliebig vieldimensionalen Zustandsraum schon zwei benachbarte Zustände im statistischen Gleichgewicht, so sprechen wir von "Vollsymmetrie". Für das Beispiel eines zweidimensionalen Zustandsraumes Bild 33 gilt dann zusätzlich zu Gl.(8.3a) beispielsweise

$$q_{i_1-1, j_2; i_1, j_2} \cdot p(i_1-1, j_2) = q_{i_1, j_2; i_1-1, j_2} \cdot p(i_1, j_2) \quad (8.4a)$$

$$q_{i_1, j_2-1; i_1, j_2} \cdot p(i_1, j_2-1) = q_{i_1, j_2; i_1, j_2-1} \cdot p(i_1, j_2) \quad (8.4b)$$

Vollsymmetrie kann höchstens zwischen solchen Zuständen herrschen, welche durch gegenseitig gerichtete Übergänge verbunden sind, vergl. hierzu als Beispiele die Abschnitte 5.1.2 und 6.1.2 c.

Es sei noch bemerkt, daß bei eindimensionalen Zustandsräumen Halbsymmetrie und Vollsymmetrie identisch sind. Bei mehrdimensionalen Zustandsräumen folgt aus der Eigenschaft der Vollsymmetrie auch stets die Eigenschaft der Halbsymmetrie.

Das Ziel approximativer Berechnungsmethoden wird es sein, einmal die geschlossene Auflösung eines großen Gleichungssystems zu umgehen und zum anderen die Zahl der Unbekannten zu verkleinern.

Große lineare Gleichungssysteme können praktisch nur auf iterativer Basis (Überrelaxationsverfahren) [19] gelöst werden und sind daher rechenzeitintensiv. Rekursive Berechnungsverfahren sind i.a. rechenzeitarm. Mit Hilfe der Konzepte der Halb- und Vollsymmetrie sind Näherungsverfahren für rekursive Berechnungen möglich.

Die Zahl der Zustände kann nur dadurch reduziert werden, daß mehrere Mikrozustände zu einem Makrozustand zusammengefaßt werden, wobei

1.a. ein Informationsverlust bei der Beschreibung in Kauf genommen werden muß. Oft liegt schon in der Definition eines Makrozustandes eine Näherung. Eine weitere Näherung kommt hinzu, wenn das Gleichungssystem für die Makrozustände selbst nur näherungsweise gelöst wird.

8.1.2 Das Konzept der Halbsymmetrie

Wie im vorangegangenen Abschnitt erläutert wurde, sind sämtliche Zustandswahrscheinlichkeiten rekursiv berechenbar, wenn sämtliche Zustände die Eigenschaft "Halbsymmetrie" besitzen. Das "Konzept der Halbsymmetrie" beruht auf der Näherungsannahme, daß für einen beliebigen Zustandsraum die Eigenschaft "Halbsymmetrie" gilt.

Die Güte einer solchen Näherung kann je nach Struktur, Organisation und Verkehrsparametern eines Bedienungssystems unterschiedlich sein. Insbesondere sind um so bessere Ergebnisse zu erwarten, je symmetrischer Struktur, Organisation und Verkehrsparameter sind (z.B. gleiche Anrufraten der verschiedenen Teilgruppen, gleiche Enderaten der Leitungen, zufallmäßiges Ausschuchen von Leitungen, gleichgroße Wartespeicher usw.). Ungünstige Ergebnisse sind dann zu erwarten, wenn "Kreisverkehre" stark ausgeprägt sind, d.h. wenn einzelne Übergänge stark von der Vollsymmetrie abweichen oder gar nur einseitig gerichtet sind. Als Beispiel hierfür werde der 3-dimensionale Unterraum aus Bild 8 betrachtet. Der Zustand (0,1,1;0) kann beispielsweise nur von "oben" her entstehen über (1,1,1;0), während er auf drei verschiedene Arten nach "oben" bzw. "unten" verschwinden kann. Ausgehend vom Zustand (1,1,1;0) können mit Hilfe der Annahme der Halbsymmetrie zwar sämtliche Zustandswahrscheinlichkeiten rekursiv berechnet werden; das Beispiel (0,1,1;0) zeigt jedoch, daß wegen des starken Kreisverkehrs ein grober Fehler entsteht.

Im Zusammenhang mit der Berechnung großer, weitgehend symmetrischer Systeme auf der Grundlage der Beschreibung mittels Makrozuständen wurden mit dem Konzept der Halbsymmetrie sehr gute Erfolge erzielt (vergl. auch 8.1.4).

8.1.3 Das Konzept der Vollsymmetrie

Ist der gesamte Zustandsraum ausschließlich durch doppelt gerichtete Übergänge zwischen benachbarten Zuständen gekennzeichnet, so kann

bei besonders symmetrischen Verhältnissen angenommen werden, daß sich schon jeweils zwei benachbarte Zustände im statistischen Gleichgewicht befinden. Dieses "Konzept der Vollsymmetrie" stellt im Vergleich zum "Konzept der Halbsymmetrie" eine gröbere Näherung dar. Der Vorteil des Konzepts der Vollsymmetrie liegt darin, daß die Rekursion mit einem beliebigen Zustand beginnen kann und daß sich einfachere Rekursionsalgorithmen finden lassen. Wie man sich aber bereits anhand eines zweidimensionalen Zustandsraumes verdeutlichen kann, können sich Widersprüchlichkeiten einstellen.

8.1.4 Das Konzept der Makrozustände

Voraussetzung für die Definition von Makrozuständen sind möglichst symmetrische Verhältnisse, wie z.B.

- gleiche Anrufraten $\lambda_J = \frac{\lambda}{g}$, J = 1,2,...,g
- gleiche Enderaten $\epsilon_I = \epsilon$, I = 1,2,...,n
- gleiche Erreichbarkeiten $k_J = k$, J = 1,2,...,g
- gleiche Speichergrößen $s_J = s$, J = 1,2,...,g
- zufälliges Absuchen von Leitungen und Warteplätzen.

Das "Konzept der Makrozustände" werde anhand eines Beispiels verdeutlicht.

Beispiel: Kombiniertes Warte- und Verlustsystem mit unvollkommener Erreichbarkeit

Gegeben sei eine Koppelanordnung unvollkommener Erreichbarkeit mit Wartespeichern entsprechend Bild 2. Es wird folgender Makrozustand definiert:

$$(x; u_1, u_2, \dots, u_g) : \begin{matrix} x \text{ Leitungen belegt,} & x = 0, 1, \dots, n, \\ u_\gamma \text{ } \gamma\text{-te Warteplätze belegt,} & u_\gamma = 0, 1, \dots, g, \\ & \gamma = 1, 2, \dots, s. \end{matrix}$$

Aufgrund dieser Beschreibung durch "Warteplatzreihen", der Zugrundelegung der kombinatorischen Sperrwahrscheinlichkeit $\mathcal{G}(x)$ zur Berechnung der Anzahl blockierter Teilgruppen bei Mischungen [5,25] und des "Konzepts der Halbsymmetrie" ist es möglich, eine einfache Rekursionsformel zur Berechnung aller Zustandswahrscheinlichkeiten abzuleiten [11]:

$$\begin{aligned}
 p(x; u_1, \dots, u_s) &= \frac{A}{x} \cdot [1 - \sigma(x-1)] \cdot p(x-1; u_1, \dots, u_s) \\
 &+ \frac{A}{g \cdot x} \cdot [g\sigma(x) - (u_1-1)] \cdot p(x; u_1-1, \dots, u_s) \\
 &+ \frac{A}{g \cdot x} \cdot \sum_{r=2}^s [u_{r-1} - (u_r-1)] \cdot p(x; \dots, u_{r-1}, \dots) .
 \end{aligned} \tag{8.5}$$

Im Sonderfalle vollkommener Erreichbarkeit führt diese Rekursionsformel sogar auf die exakten Lösungen für die Abfertigung zwischen den Warteschlangen nach Wahrscheinlichkeiten proportional zu den momentanen Warteschlangenlängen.

In Abschnitt 8.3 sind Ergebnisse von Berechnungen und Simulationen angegeben.

8.1.5 Das Konzept vorgegebener Verteilungen mit iteriertem Parameter

Im Vorangehenden wurden, ausgehend vom Zustandsraum, verschiedene Möglichkeiten der approximativen Lösung eines Gleichungssystems diskutiert. Es gibt außer diesen Möglichkeiten noch andere Verfahren der näherungsweise Berechnung.

Ein sehr bekanntes und wirkungsvolles Verfahren besteht darin, eine bekannte Verteilungsfunktion für die Zustandswahrscheinlichkeiten anzunehmen und einen Parameter (z.B. das Angebot A) so lange durch Iteration zu verändern, bis sich eine vorgegebene charakteristische Größe (z.B. Verkehrswert Y) einstellt. Mit Hilfe dieser Verteilung, welche beispielsweise denselben vorgegebenen Verkehrswert Y wie die exakte Verteilung ergibt, können dann weitere Werte wie Verlustwahrscheinlichkeit B und tatsächliches Angebot A berechnet werden (vergl. Ableitung der MPJ-Formel für Mischungen in Verlustsystemen [5] und Berechnung von Linksystemen [26] von A. LOTZE).

8.2 Wartezeitverteilungsfunktionen

8.2.1 Die exakte Lösung

Der Warteprozess eines Testrufs ($\xi(t)$ -Prozess) kann aufgrund der Markoff'schen Eigenschaft durch ein System linearer Differentialgleichungen erster Ordnung (Kolmogoroff-Rückwärts-Gleichung) für die

bedingten Wartezeitverteilungsfunktionen $w(t|i)$ beschrieben werden, wobei $\xi(0) = i$ derjenige Wartezustand ist, mit welchem der spezielle Warteprozess startet. Nach Gl.(4.32) lautet die allgemeine Form der Kolmogoroff-Rückwärts-Gleichung für die bedingten Wartezeitverteilungsfunktionen

$$q(i)w(t|i) - \sum_{\substack{k \neq i \\ k \in H}} q(i,k)w(t|k) = -\frac{d}{dt} w(t|i), \quad i \notin H. \tag{8.6}$$

Endet der Warteprozess grundsätzlich erfolgreich, d.h. durch Abfertigung, so sind die Anfangsbedingungen der bedingten Wartezeitverteilungsfunktionen einheitlich

$$w(0|i) = 1, \quad i \notin H. \tag{8.7}$$

Allgemein können die Anfangsbedingungen $w(0|i)$ aus dem aus Gl.(8.6) entstehenden linearen Gleichungssystem berechnet werden, wenn für

$$-\lim_{t \rightarrow 0^+} \frac{dw(t|i)}{dt} = \mathcal{E}(i), \quad i \notin H, \tag{8.8}$$

eingesetzt wird und wobei $\mathcal{E}(i)$ der Endekoeffizient für das Ende des Warteprozesses vom Zustand i aus ist:

$$q(i)w(0|i) - \sum_{\substack{k \neq i \\ k \in H}} q(i,k)w(0|k) = \mathcal{E}(i), \quad i \notin H. \tag{8.9}$$

Die Anfangsbedingungen müssen nur für die bedingten Wartezeitverteilungsfunktionen erfolgreich bzw. erfolglos wartender Rufe bei verdrängenden Disziplinen berechnet werden.

Aus Gl.(8.6) folgt durch Laplace-Transformation das lineare Gleichungssystem für die Laplace-Transformierten $W(s|i)$ der bedingten Wartezeitverteilungsfunktionen $w(t|i)$:

$$[s + q(i)]W(s|i) - \sum_{\substack{k \neq i \\ k \in H}} q(i,k)W(s|k) = w(0|i), \quad i \notin H. \tag{8.10}$$

Die exakte Lösung des Systems (8.10) erfolgt über die Bestimmung der Eigenwerte; die Lösungen für $W(s|i)$ sind rationale Funktionen in s, welche nach Rücktransformation in den Zeitbereich i.a. mit Polynomen gewichtete Exponentialsummen ergeben, vergl. Gl.(4.48).

Die K-ten Momente $M_K(i)$ der bedingten Wartezeitverteilungsfunktionen $w(t|i)$ sind aus folgendem linearem inhomogenen Gleichungssystem zu berechnen:

$$q(i)M_K(i) - \sum_{\substack{k \neq i \\ k \in H}} q(i,k)M_K(k) = KM_{K-1}(i), \quad i \in H, \quad (8.11)$$

$$K = 1, 2, \dots$$

wobei $M_0(i) = w(0|i)$. Die Systeme lassen sich bezüglich K rekursiv auflösen, wobei der K-fache Lösungsvektor der (K-1)-ten Momente die "rechte Seite" des Gleichungssystems für die K-ten Momente bildet. Für den Sonderfall $K = 1$ erhält man wegen $M_1(i) = t_w(i)$ gerade die bedingten mittleren Wartezeiten.

Wir bemerken, daß die Systeme (8.6), (8.9), (8.10) für $s = 0$, sowie (8.11) die identische Koeffizientenmatrix besitzen. Das Problem ist daher schon immer dann eindeutig vorgegeben, wenn diese Koeffizientenmatrix vorliegt.

Die gesamte Wartezeitverteilungsfunktion $W(>t)$ stellt sich mit Hilfe der Antreffwahrscheinlichkeiten $P(i)$ wie folgt dar:

$$W(>t) = \sum_{i \in H} P(i)w(t|i) \quad (8.12)$$

Sonderfälle:

1. Eindimensionale Zustandsräume

Ist ein Teil des Zustandsraumes oder der gesamte Zustandsraum eindimensional, wobei die Übergänge zwischen den Zuständen jeweils nur einseitig gerichtet sind, so ist eine rekursive Auflösung der zugehörigen Laplace-Transformierten der bedingten Wartezeitverteilungsfunktionen möglich (vergl. 5.3.1 sowie 6.3.1.1, erster Sonderfall).

2. Mehrdimensionale Zustandsräume

Läßt sich allgemein ein Teil eines mehrdimensionalen Zustandsraumes finden mit der Eigenschaft, daß von den Zuständen dieses Teilraumes keine Übergänge nach Zuständen außerhalb dieses Teilraumes möglich sind, so kann der zu diesen Zuständen gehörige Teil der Gleichungssysteme (8.6), (8.9), (8.10) und (8.11) selbständig gelöst werden (vergl. Abschnitte 6.3.1.1, 6.3.2.1, 6.3.2.3, 7.3.1.1).

3. Spezielle Matrixstrukturen

Zu einem speziellen Typ der tridiagonalen Koeffizientenmatrix ("Differenzenmatrix") sind die Eigenwerte explizit angebar, womit auch die Wartezeitverteilungsfunktion explizit ausgedrückt werden kann (vergl. 5.3.3 und Anhang A2 sowie 6.3.1.1, zweiter Sonderfall).

Das Ziel approximativer Berechnungsmethoden wird es sein, die geschlossene Auflösung eines großen linearen Differentialgleichungssystems zu umgehen. Ähnlich wie beim Problem der Zustandswahrscheinlichkeiten werden im folgenden Näherungsverfahren angegeben, welche rechnerisch einfacher zu handhaben sind.

8.2.2 Das Konzept der Halbsymmetrie

Die Halbsymmetrie bei den Zustandswahrscheinlichkeiten wurde anschaulich eingeführt, indem eine Symmetrieebene durch einen Zustand des mehrdimensionalen Zustandsraums gelegt und weiterhin angenommen wurde, daß der jeweilige Zentralzustand sich bereits mit allen seinen "unteren" (bzw. "oberen") Nachbarzuständen im statistischen Gleichgewicht befindet. In der Matrix des zugehörigen Gleichungssystems läßt sich die Halbsymmetrie als charakteristische "Dreiecksstruktur" erkennen:

(1)	(2)	(3)	(4)	(N)	
a_{21}	a_{22}				0
a_{31}	a_{32}	a_{33}			0
a_{41}	a_{42}	a_{43}	a_{44}		0
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
a_{N1}	a_{N2}	a_{N3}	a_{N4}	\dots	a_{NN}
1	1	1	1	1	1

(8.13)

Es sind höchstens nur diejenigen Elemente a_{ik} von Null verschieden, für welche $k \leq i$ gilt. Die Wahrscheinlichkeiten der Zustände (2), (3), ..., (N) können rekursiv berechnet werden, wobei die Wahrscheinlichkeit des Zustands (1) zunächst als bekannt angenommen wird. Die letzte Zeile bedeutet die "Normalisierungsbedingung", aus ihr wird die Wahrscheinlichkeit des Zustands (1) bestimmt.

Das Konzept der Halbsymmetrie für ein Differentialgleichungssystem

bedingter Wartezeitverteilungsfunktionen werde nun entsprechend zu Gl.(8.13) eingeführt. Werden alle die bedingten Wartezeitverteilungsfunktionen für jene Startzustände, welche "obere" Nachbarstartzustände eines betrachteten Startzustandes sind, mit der bedingten Wartezeitverteilungsfunktion des betrachteten Startzustandes gleichgesetzt, so ergibt sich eine Gleichungsstruktur entsprechend zu Gl. (8.13). Die bedingten Wartezeitverteilungsfunktionen können dann rekursiv gefunden werden.

Im Differentialgleichungssystem wird die rechte Seite durch den Vektor der negativen Differentialquotienten der bedingten Wartezeitverteilungsfunktionen gebildet; im System der Laplace-Transformierten der bedingten Wartezeitverteilungsfunktionen wird die rechte Seite durch den Vektor der Anfangsbedingungen gebildet (vergl. Gl.(8.6) bzw. (8.10)).

Als Beispiel werde die allgemeine Differentialgleichung (5.41a) für Systeme mit einer Warteschlange und zufallsmäßigem Abfertigen (D2) betrachtet:

$$-\frac{i}{i+1}\mu w(t|i-1) + (\lambda + \mu)w(t|i) - \lambda w(t|i+1) = -\frac{dw(t|i)}{dt} \quad (8.14a)$$

Das Konzept der Halbsymmetrie erzwingt

$$w(t|i+1) = w(t|i) \quad (8.14b)$$

Gl.(8.14b) in Gl.(8.14a) eingesetzt, ergibt

$$-\frac{i}{i+1}\mu w(t|i-1) + \mu w(t|i) = -\frac{dw(t|i)}{dt} \quad (8.14c)$$

Die Lösung des Differentialgleichungssystems (8.14c) für $i = 0, 1, \dots, s_1-1$ läßt sich rekursiv durchführen. Als Lösung erhält man

$$w(t|i) = \sum_{\nu=0}^i \frac{i+1-\nu}{i+1} \cdot \frac{(\mu t)^\nu}{\nu!} \cdot \exp(-\mu t), \quad i = 0, 1, \dots, s_1-1 \quad (8.15)$$

Im Vergleich zur exakten Lösung liefert diese Näherung zu günstige Ergebnisse; für große Zahlen i wird sie jedoch zunehmend genauer, weil dann der Unterschied zwischen $w(t|i)$ und $w(t|i+1)$ geringer wird.

8.2.3 Das Konzept der Vollsymmetrie

Wird das Konzept der Halbsymmetrie auch auf die zu einem betrachteten Startzustand benachbarten "unteren" Nachbarstartzustände erweitert, so artet die zugehörige Koeffizientenmatrix zur Diagonalmatrix aus, und es kann jede Gleichung unabhängig gelöst werden. Dieses "Konzept der Vollsymmetrie" soll beispielsweise auf Gl.(8.14a) angewendet werden, wobei

$$w(t|i-1) = w(t|i) = w(t|i+1) \quad (8.16a)$$

Wird Gl.(8.16a) auf Gl.(8.14a) angewendet, so erhalten wir

$$\frac{1}{i+1}\mu w(t|i) = -\frac{dw(t|i)}{dt} \quad (8.16b)$$

Die Lösung von Gl.(8.16b) ist

$$w(t|i) = \exp\left(-\frac{\mu}{i+1}t\right) \quad (8.16c)$$

Die Annahme (8.16a) wurde speziell für die Berechnung der Wartezeitverteilungsfunktion beim reinen Wartesystem und bei zufallsmäßiger Abfertigung von J.W.MELLOR [28,29] bereits im Jahre 1942 getroffen. Die Lösung Gl.(8.16c) ist in die Literatur unter dem Namen "Mellor-Approximation" eingegangen.

Wo liegt nun die Grenze der Anwendbarkeit des Konzepts der Vollsymmetrie? Wird dieses Konzept auf die allgemeine Gl.(8.6) unter Beachtung von Gl.(4.28c) angewendet, erhält man

$$\mathcal{E}(i)w(t|i) = -\frac{dw(t|i)}{dt}, \quad i \notin H \quad (8.17a)$$

Die Lösung von Gl.(8.17a) ist

$$w(t|i) = \exp[-\mathcal{E}(i)t], \quad i \notin H \quad (8.17b)$$

Aus Gl.(8.17a,b) folgt, daß unter Annahme der Vollsymmetrie sinnvolle Lösungen nur für Zustände (i) erhalten werden, von welchen aus der Warte-prozeß mit der Endedichte $\mathcal{E}(i) \neq 0$ direkt beendet werden kann.

Gl.(8.17b) stellt die allgemeine Lösung des Differentialgleichungssystems (8.6) nach dem Konzept der Vollsymmetrie dar ("Verallgemeinerte MELLOR-Approximation"). Die Lösung gilt nur für Zustände mit $\mathcal{E}(i) \neq 0$ und wird i.a. immer bei zufallsmäßigen Abfertigungsdisziplinen anwendbar sein. Die verallgemeinerte MELLOR-Approximation bedeutet außerdem eine praktische Anwendung der zunächst nur formal

eingeführten Endedichten $\mathcal{E}(i)$ des $\zeta(t)$ -Prozesses.

Für jene Anwendungsfälle, in denen das Konzept der Vollsymmetrie nicht durchweg brauchbar ist, kann ein Mischverfahren aus Voll- und Halbsymmetrie angewendet werden.

8.2.4 Das Konzept der sukzessiven Approximation durch Potenzreihenentwicklungen

Da die bedingten Wartezeitverteilungsfunktionen monoton abnehmende Funktionen sind, besteht die Aussicht, das Differentialgleichungssystem (8.6) für einen genügend großen Zeitbereich mit Hilfe von Reihenentwicklungen zu lösen. Der Ansatz erfolgt nach einer TAYLOR-Entwicklung der Ordnung K_{max} :

$$w(t|i) = \sum_{x=0}^{K_{max}} w^{(x)}(0|i) \cdot \frac{t^x}{x!}, \quad i \notin H. \quad (8.18)$$

Bei bekannten Anfangsbedingungen $w(0|i) = 1$ sind hierfür außerdem die Ableitungen der bedingten Wartezeitverteilungsfunktionen an der Stelle $t = 0+$ zu bestimmen. Aus Gl.(8.6) gewinnt man durch sukzessive Differentiation:

$$w^{(x+1)}(t|i) = -q(i)w^{(x)}(t|i) + \sum_{\substack{k \neq i \\ k \notin H}} q(i,k)w^{(x)}(t|k), \quad i \notin H. \quad (8.19)$$

Für $x=0,1,\dots,K_{max}+1$ können aus Gl.(8.19) nacheinander sämtliche höheren Ableitungen im Entwicklungsursprung rekursiv ermittelt werden (die $(K_{max}+1)$ -te Ableitung wird zur Restgliedabschätzung benutzt). Die Koeffizienten der Reihenentwicklung des Differentialgleichungssystems werden in einem Feld der Größe $N \cdot (K_{max}+2)$ abgespeichert ($N = \text{Rang des Differentialgleichungssystems}$).

Durch fortgesetzte Anwendung der TAYLOR-Entwicklung in den Punkten $t = t_0, t_1, t_2, \dots$, wobei $t_0=0 < t_1 < t_2 < \dots$, ist es möglich, den Verlauf der Verteilungsfunktionen über den gesamten interessierenden Bereich zu berechnen. Um die Genauigkeit eines derartigen sukzessiven Approximationsprozesses beliebig einstellen zu können, ist es zweckmäßig, die Länge eines Approximationsintervalls aus einem vorgeschriebenen maximalen Fehler, welcher mit Hilfe der Restgliedformel ermittelt wird, zu berechnen.

Im einzelnen bietet dieses Verfahren folgende Vorteile:

- Rekursivität

Sämtliche Ableitungen der bedingten Wartezeitverteilungsfunktionen sind im Entwicklungsursprung t_n , $n = 0,1,\dots$, rekursiv berechenbar, vergl. Gl.(8.19). Das Verfahren kann fortgesetzt angewendet werden, indem der Endpunkt t_n des n -ten Intervalls als Entwicklungsursprung des $(n+1)$ -ten Intervalls angesehen wird, $n = 1,2,\dots$. Das Feld der Größe $N \cdot (K_{max}+2)$ kann für jedes neue Approximationsintervall zur Aufnahme der Ableitungen im Entwicklungsursprung überschrieben werden.

- Vorschreibbare Genauigkeit

Die Länge des n -ten Approximationsintervalls wird dadurch bestimmt, daß der maximale Fehler der normierten Wartezeitverteilungsfunktion $W(>t)/W$ am Intervallende, R_n , vorgeschrieben wird (z.B. 10^{-7}). Aus der Abschätzung des Fehlers mittels der Restgliedformel nach TAYLOR für die gesamte Wartezeitverteilungsfunktion erhält man die Länge des n -ten Approximationsintervalls:

$$t_n - t_{n-1} \leq \sqrt[K_{max}+1]{\frac{R_n \cdot (K_{max}+1)!}{\frac{1}{W} \sum_{i \notin H} P(i) |w^{(K_{max}+1)}(t_n|i)|}} \quad (8.20)$$

Die Werte der bedingten Wartezeitverteilungsfunktionen am Ende des n -ten Intervalls sind gleichzeitig die Anfangswerte für den Entwicklungsursprung t_n des $(n+1)$ -ten Intervalls.

- Austauschbarkeit zwischen Speicherplatz und Rechenzeit

Wird der Wert K_{max} verkleinert, so reduziert sich - bei vorgeschriebenem Fehler R_n - die Länge des Approximationsintervalls. Die Lösung des Differentialgleichungssystems für einen vorgeschriebenen Zeitbereich benötigt dann mehr Schritte, d.h. grössere Rechenzeit.

Für den Minimalwert $K_{max} = 1$ können auf einer Rechanlage somit Differentialgleichungssysteme bis zum Range von einem Drittel der verfügbaren Arbeitsspeicherkapazität behandelt werden. Umgekehrt kann für kleinere Differentialgleichungssysteme die Ordnung K_{max} zur Erzielung großer Intervalle - und damit kleiner Rechenzeiten - optimal groß gewählt werden.

Es sei abschließend bemerkt, daß sich das Konzept der sukzessiven Approximation durch Potenzreihenentwicklungen ebenso auf die Behandlung von Differentialgleichungssystemen für die nichtstationären Zustandswahrscheinlichkeiten anwenden läßt (Berechnung von Einschwingvorgängen).

8.2.5 Das Konzept der Approximation durch Exponentialsummen (Momentenmethode)

Alle im vorangehenden vorgeschlagenen Approximationsverfahren sind dadurch charakterisiert, daß die Genauigkeit entweder von den Startzuständen oder der Zeit abhängt. In diesem Abschnitt werden Verfahren behandelt, welche eine definierte Approximationsgüte unabhängig von den Startzuständen und der Zeit gewährleisten.

Als natürliches Approximationsmaß erscheinen die Momente der Wartezeitverteilungsfunktionen, da eine Verteilungsfunktion durch sämtliche Momente eindeutig festgelegt ist. Andererseits ist aus den theoretischen Überlegungen bereits bekannt, welche mathematische Struktur die exakte Lösung für die Wartezeitverteilungsfunktionen besitzt: Sind die Eigenwerte des zugehörigen Eigenwertproblems sämtlich verschieden, so hat die Lösung eine reine Exponentialsummenstruktur; gibt es dagegen auch teilweise mehrfache Eigenwerte, so hat die Lösung eine mit Polynomen in t erweiterte Exponentialsummenstruktur, vergl. Gl.(4.48). Durch Ansatz einer der exakten Lösung ähnlichen Exponentialsumme für eine Verteilungsfunktion können aus der Forderung, daß exakte Lösung und approximative Lösung in den K ersten Momenten übereinstimmen, K Parameter des Exponentialsummenansatzes ermittelt werden. Als entscheidender Rechenaufwand müssen bei dieser "Momentenmethode" K lineare Gleichungssysteme für die K ersten Momente gelöst werden (das 0-te Moment ist dabei identisch mit der Anfangsbedingung und bei nichtverdrängenden Abfertigungsdisziplinen bekannt).

Die Momentenmethode ist bisher auf die gesamte Wartezeitverteilungsfunktion von reinen Wartesystemen angewendet worden, und zwar von J. RIORDAN für zufallsmäßige Abfertigung [29] und Abfertigung in der inversen Ankunftsreihenfolge [24], sowie von W. WAGNER für nichtunterbrechende Prioritätsdisziplinen [6]. Als Ansatz sind dabei reine Exponentialsummen für die gesamte Wartezeitverteilungsfunktion verwendet worden.

In Erweiterung zu diesem Vorschlag werden zwei Verallgemeinerungen vorgenommen:

1. Die Momentenmethode wird nicht nur auf die gesamte, sondern bereits auf jede bedingte Wartezeitverteilungsfunktion angewendet.

Aus der Forderung, daß nicht nur die globale, sondern zusätzlich auch schon jede bedingte Wartezeitverteilungsfunktion in den ersten K Momenten mit der jeweiligen exakten Lösung übereinstimmt, ergibt sich eine bessere Approximation in der "Feinstruktur" der Wartezeitverteilungsfunktion.

2. Es wird der Fall teilweise mehrfacher Eigenwerte mit berücksichtigt.

Mehrfache Eigenwerte haben insbesondere bei Überlastung einen starken Einfluß auf den Verlauf der Wartezeitverteilungsfunktion, und zwar dergestalt, daß ein reiner Exponentialsummenansatz immer ungenauer wird, vergl. z.B. Gl.(5.81).

Der allgemeine Ansatz für die approximative bedingte Wartezeitverteilungsfunktion lautet:

$$w(t|i) \sim \sum_{\nu=1}^N \exp(-\frac{t}{x_{\nu}}) \sum_{x=0}^K a_{\nu x} t^x . \tag{8.21}$$

a) Zwei-Momenten-Methode

Aus Gl.(8.21) folgt die einfachste Näherung, welche in den zwei ersten Momenten mit der exakten Lösung übereinstimmt, nämlich dem Anfangswert $M_0(i) = w(0|i)$ und der mittleren bedingten Wartezeit $M_1(i) = t_w(i)$:

$$w(t|i) \sim a_{10} \exp(-\frac{t}{x_1}) . \tag{8.22}$$

Die zwei ersten Momente berechnen sich zu

$$M_0(i) = a_{10} \tag{8.23a}$$

$$M_1(i) = a_{10} x_1 . \tag{8.23b}$$

Die Auflösung dieses Gleichungssystems nach den unbekanntem Parametern a_{10} und x_1 ergibt

$$a_{10} = M_0(i) \tag{8.24a}$$

$$x_1 = \frac{M_1(i)}{M_0(i)} . \tag{8.24b}$$

Bei dieser (einfachsten) Näherung ergibt sich kein Unterschied zwischen den Fällen einfacher und mehrfacher Eigenwerte.

b) Drei-Momenten-Methode

Aus Gl.(8.21) folgt die dreiparametrische Näherungsverteilung, welche in den ersten drei Momenten $M_0(i)$, $M_1(i)$ und $M_2(i)$ mit der exakten Lösung übereinstimmt und insbesondere bei mehrfachen Eigenwerten geeignet ist:

$$w(t|i) \sim (a_{10} + a_{11}t) \cdot \exp\left(-\frac{t}{x_1}\right). \quad (8.25)$$

Die drei ersten Momente dieser Verteilung sind

$$M_0(i) = a_{10} \quad (8.26a)$$

$$M_1(i) = a_{10}x_1 + a_{11}x_1^2 \quad (8.26b)$$

$$M_2(i) = 2a_{10}x_1^2 + 4a_{11}x_1^3. \quad (8.26c)$$

Werden Gl.(8.26a-c) nach x_1 , a_{10} und a_{11} aufgelöst, erhält man

$$x_1 = \frac{1}{M_0(i)} \left[M_1(i) \pm \sqrt{A} \right] \quad (8.27a)$$

$$a_{10} = M_0(i) \quad (8.27b)$$

$$a_{11} = \frac{\pm \sqrt{A}}{x_1^2}, \quad (8.27c)$$

wobei

$$A = M_1^2(i) - \frac{1}{2}M_0(i)M_2(i). \quad (8.27d)$$

Diese Approximation hat zwei bemerkenswerte Eigenschaften:

1. Die Approximation nach Gl.(8.25) gilt nur für $A \geq 0$, d.h.

$$M_1^2(i) \geq \frac{1}{2}M_0(i)M_2(i). \quad (8.28a)$$

Im Sonderfalle $M_1^2(i) = \frac{1}{2}M_0(i)M_2(i)$ ergibt sich $a_{11} = 0$; die Approximation stimmt dann formal mit der Approximation nach der Zwei-Momenten-Methode überein.

2. Die Approximation ist zunächst nicht eindeutig; es gibt zwei verschiedene Lösungen, welche gleichzeitig die Gl.(8.26a-c) erfüllen. Aus der notwendigen Bedingung $w'(t|i) \leq 0$ folgt jedoch eine weitere Einschränkung:

$$a_{10} + a_{11}t \geq a_{11}x_1, \quad t \geq 0.$$

Diese Relation kann wegen $a_{10} > 0$ und $x_1 > 0$ jedoch nur für $a_{11} > 0$ erfüllt werden, d.h. in den Gl.(8.27a-c) gilt nur das untere Vorzeichen der Wurzel. Die größte Einschränkung besteht dabei für $t = 0$, was auf eine zweite Bedingung für die drei Momente führt:

$$M_1^2(i) \leq \frac{2}{3}M_0(i)M_2(i). \quad (8.28b)$$

Zusammenfassend kann also festgestellt werden, daß die Drei-Momenten-Methode unter der folgenden Bedingung anwendbar ist:

$$\frac{1}{2}M_0(i)M_2(i) \leq M_1^2(i) \leq \frac{2}{3}M_0(i)M_2(i). \quad (8.28c)$$

Gilt das linke Gleichheitszeichen, reduziert sich die Approximation auf eine Exponentialfunktion; gilt das rechte Gleichheitszeichen, so beginnt die Approximation mit horizontaler Tangente.

c) Vier-Momenten-Methode

Der einfachste Fall einer vierparametrischen Verteilung ist

$$w(t|i) \sim a_{10} \exp\left(-\frac{t}{x_1}\right) + a_{20} \exp\left(-\frac{t}{x_2}\right). \quad (8.29)$$

Die ersten vier Momente berechnen sich zu

$$M_0(i) = a_{10} + a_{20} \quad (8.30a)$$

$$M_1(i) = a_{10}x_1 + a_{20}x_2 \quad (8.30b)$$

$$\frac{M_2(i)}{2} = a_{10}x_1^2 + a_{20}x_2^2 \quad (8.30c)$$

$$\frac{M_3(i)}{6} = a_{10}x_1^3 + a_{20}x_2^3. \quad (8.30d)$$

Wird a_{20} aus jeweils zwei aufeinanderfolgenden Gleichungen eliminiert und werden a_{10} und x_1 aus den erhaltenen drei Gleichungen durch Division jeweils zweier Gleichungen eliminiert, entsteht eine quadratische Gleichung für x_2 . Durch Auflösen dieser Gleichung erhält man mit den anderen Zwischenergebnissen schließlich

$$x_1 = A + \sqrt{A^2 - B} \quad (8.31a)$$

$$x_2 = A - \sqrt{A^2 - B} \quad (8.31b)$$

$$a_{10} = \frac{1}{2(A^2 - B)} \left[m_0(A^2 - B) + (m_1 - m_0A)\sqrt{A^2 - B} \right] \quad (8.31c)$$

$$a_{20} = \frac{1}{2(A^2 - B)} \left[m_0(A^2 - B) - (m_1 - m_0A)\sqrt{A^2 - B} \right], \quad (8.31d)$$

wobei

$$A = \frac{m_1 m_2 - m_0 m_3}{m_1^2 - m_0 m_2}, B = \frac{m_2^2 - m_1 m_3}{m_1^2 - m_0 m_2}, m_K = \frac{M_K(i)}{K!}, K=0,1,2,3. \quad (8.31e)$$

Aus der Forderung, daß x_1 und x_2 positiv-reell sein müssen, resultieren folgende notwendige einschränkende Bedingungen:

$$A > 0 \quad (8.32a)$$

$$B > 0 \quad (8.32b)$$

$$A^2 - B \geq 0 \quad (8.32c)$$

Im Sonderfalle

$$M_1^2(i) = \frac{1}{2} M_0(i) M_2(i) \quad (8.33)$$

reduziert sich die Approximation wieder auf eine einzige Exponentialfunktion.

d) 2n-Momenten-Methode

Im allgemeinsten Falle der 2n-Momenten-Methode und reinem Exponentialsummenansatz haben wir

$$w(t|i) \sim \sum_{\nu=1}^n a_{\nu 0} \exp\left(-\frac{t}{x_{\nu}}\right). \quad (8.34)$$

Die Momente dieser Verteilung sind

$$M_K(i) = K! \cdot \sum_{\nu=1}^n a_{\nu 0} x_{\nu}^K, K=0,1,\dots,2n-1. \quad (8.35)$$

Die Auflösung des Gleichungssystems (8.35) nach $a_{\nu 0}$ und x_{ν} , $\nu = 1, 2, \dots, n$, führt i.a. über eine algebraische Gleichung n-ten Grades und kann deshalb nicht in geschlossener Form angegeben werden.

Über die Genauigkeit der in den Teilabschnitten 8.24 und 8.25 vorgeschlagenen Verfahren wird im folgenden Abschnitt 8.3 berichtet.

8.3 Numerische Beispiele

8.3.1 Zustandswahrscheinlichkeiten

a) Konzept von Voll- und Halbsymmetrie

Betrachtet werde das Bedienungssystem mit $g = 2$ Warteschlangen, $n = 3$ vollkommen erreichbaren Leitungen und $s_1 = 4$ bzw. $s_2 = 6$ Wartepplätze aus Abschnitt 6.4. Wie in Abschnitt 6.1.2c mit Hilfe der Lösung (6.9) gezeigt werden kann, gilt im Falle der Abfertigung nach

Wahrscheinlichkeiten proportional zu den momentanen Warteschlangenlängen die Vollsymmetrie exakt, damit gilt auch die Halbsymmetrie. Die Ergebnisse für obengenannte Disziplin sind also identisch mit den Ergebnissen, welche man aufgrund des Konzepts der Voll- bzw. Halbsymmetrie erhält.

Aus den Bildern 24a,b, S.132 und 133, kann somit beispielsweise der Fehler beurteilt werden, welcher bei näherungsweise Berechnung nach dem Konzept der Vollsymmetrie im Falle der gleichwahrscheinlichen Abfertigung der nichtleeren Warteschpeicher entsteht. Wie aus den Bildern 24a,b zu entnehmen ist, ergeben sich erst für größere Angebotswerte nennenswerte Unterschiede zwischen exakter und approximativer Lösung.

b) Konzept der Makrozustände in Verbindung mit dem Konzept der Halbsymmetrie

Zu der in Abschnitt 8.1.4 skizzierten Methode sollen zwei Beispiele angegeben werden.

Im ersten Beispiel werde eine sog. "ideale Erlang-Mischung" [5] mit $n = 9$ Leitungen, der Erreichbarkeit $k = 6$ und $g = \binom{n}{k} = 84$ Teilgruppen mit jeweils $s = 2$ Wartepätzen zugrunde gelegt. Für ideale Erlang-Mischungen liegt die einzige Näherungsannahme in der "Halbsymmetrie". In Bild 34, S.173, sind die charakteristischen Größen W_J , B_J und t_{WJ} in Abhängigkeit des Angebots A aufgetragen. Die eingetragenen Werte der Verkehrssimulation (mit 95%-Vertrauensintervallen) bestätigen die Wirksamkeit der Methode.

Im zweiten Beispiel soll eine reale Mischung (vereinfachte Normmischung der DBP) mit $n = 60$ Leitungen, Erreichbarkeit $k = 10$, $g = 12$ Teilgruppen und $s = 4$ Wartepätzen je Teilgruppe betrachtet werden. Im Gegensatz zu idealen Mischungen ist für reale Mischungen der Erwartungswert der Anzahl blockierter Teilgruppen (d.h. die mittlere Zahl der möglichen Warteschlangen) im Zustand "x Leitungen belegt" ($k < x < n$) i.a. keine ganze Zahl und muß auf- oder abgerundet werden. Die Ergebnisse in Bild 35, S.174, werden aber auch hier von der Simulation bestätigt. Der Verlauf der Verlustwahrscheinlichkeit des zugehörigen Verlustsystems ($s = 0$) zeigt den Unterschied in der Ausnutzung der Leitungen im Vergleich zu Systemen mit Wartemöglichkeit. Für eine ausführliche Behandlung von Wartezeitproblemen aus der Vermittlungstechnik sei ferner auf [27] verwiesen.

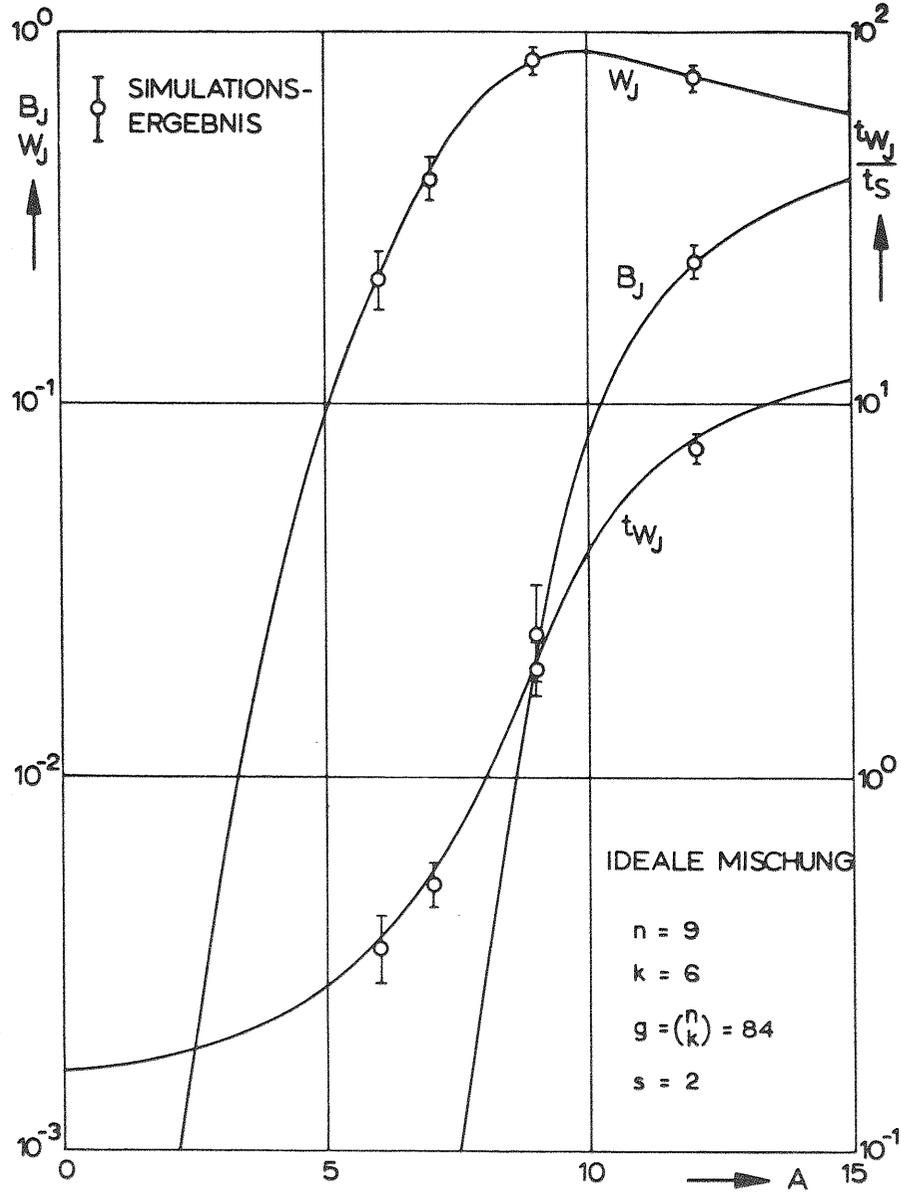


Bild 34. Charakteristische Größen in Abhängigkeit des Angebots A (Lösung durch Makrozustände in Verbindung mit dem Konzept der Halbsymmetrie)
 System: $n=9$, $k=6$, $g=84$, $s=2$ (Ideale Erlang-Mischung)

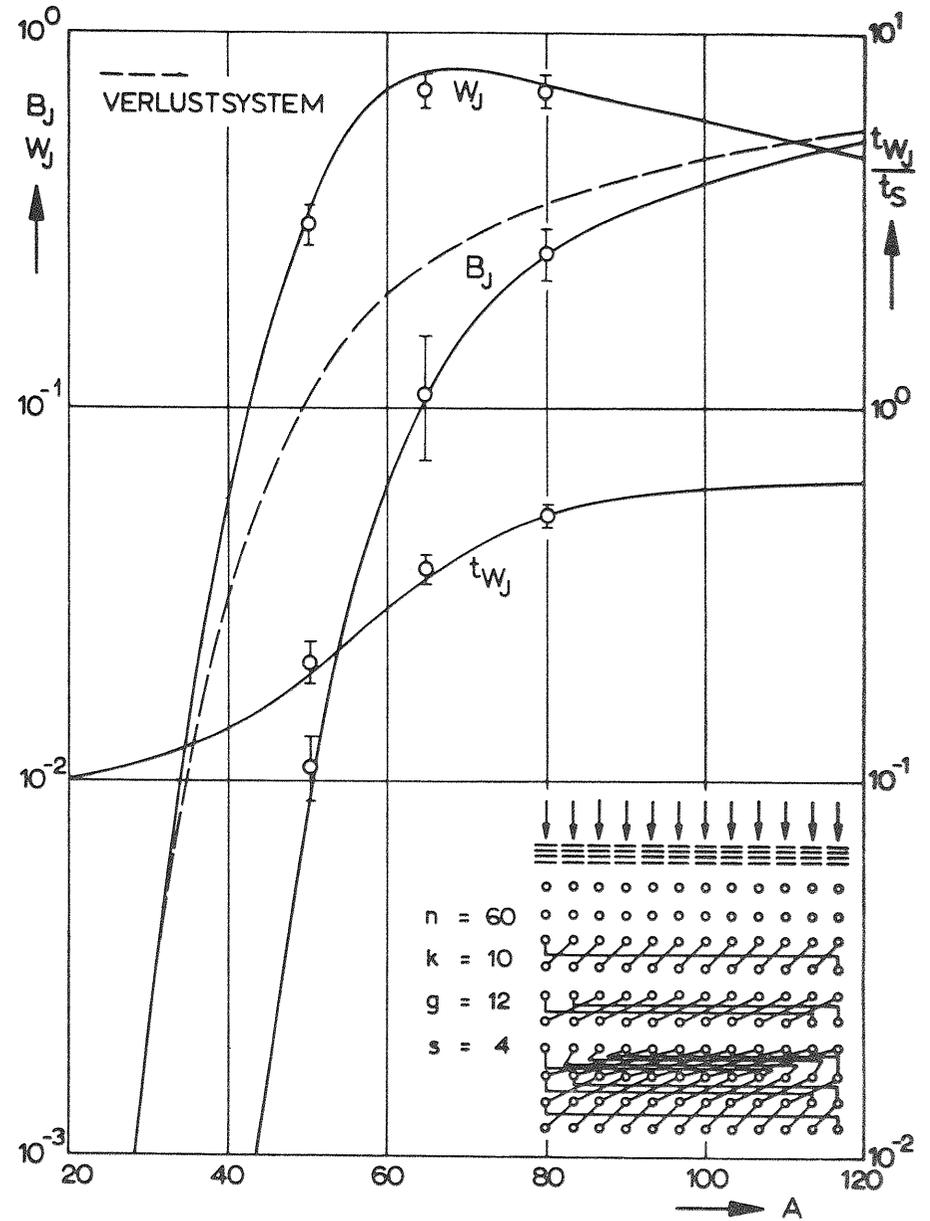


Bild 35. Charakteristische Größen in Abhängigkeit des Angebots A (Lösung durch Makrozustände in Verbindung mit dem Konzept der Halbsymmetrie)
 System: $n=60$, $k=10$, $g=12$, $s=4$ (Vereinfachte Normmischung)

8.3.2 Wartezeitverteilungsfunktionen

a) Momentenmethoden

Zur Demonstration der Güte verschiedener Approximationen mit Exponentialsummen nach der Momentenmethode werde das System mit $n = 1$ Leitung, $s_1 = 12$ Warteplätzen und der Abfertigungsdisziplin D2 herangezogen. Es werden die exakten Ergebnisse nach Abschnitt 5.5 mit Ergebnissen von vier verschiedenen Momentenmethoden verglichen, und zwar

1. Approximation der gesamten Wartezeitverteilungsfunktion mit Hilfe der Gesamtmomente nach
 - a) Zwei-Momenten-Methode
 - b) Vier-Momenten-Methode
2. Approximation der bedingten Wartezeitverteilungsfunktionen mit Hilfe der bedingten Momente nach
 - a) Zwei-Momenten-Methode
 - b) Vier-Momenten-Methode.

Insbesondere soll die Verbesserung in der "Feinstruktur" der Wartezeitverteilungsfunktion mit Hilfe der bedingten Momente gegenüber der Approximation mit Hilfe der Gesamtmomente demonstriert werden. Tabelle 3, S. 176, zeigt die Ergebnisse in Abhängigkeit der Zeit sowie dreier typischer Angebotswerte.

Der Vergleich der verschiedenen Näherungsergebnisse (1a und 2a bzw. 1b und 2b) mit den exakten Ergebnissen ergibt, daß die Approximation der bedingten Wartezeitverteilungsfunktionen genauere Resultate liefert. Dieses Ergebnis war auch zu erwarten, da z.B. bei der Zwei-Momenten-Methode im ersten Falle (1a) exakte und näherungsweise Lösung nur im Mittelwert der gesamten Wartezeitverteilungsfunktion, im zweiten Falle (2a) jedoch außerdem in sämtlichen Mittelwerten der bedingten Wartezeitverteilungsfunktionen übereinstimmen.

Zur Berechnung bei der Vier-Momenten-Methode (2b) ist noch anzumerken, daß bei der Approximation der bedingten Wartezeitverteilungsfunktionen durch zwei Exponentialfunktionen im Falle der Verletzung einer der Bedingungen (8.32) die Zwei-Momenten-Methode angewendet wurde.

Angebot	Zeit t/t _S	Wartezeitverteilungsfunktion W(>t)/W					
		Exakte Berechnung	Absolute Abweichungen bei approximativer Berechnung (Momentenmethoden)				
			mit Gesamtmomenten 2 Momente		mit bedingten Momenten 4 Momente		X
0,8	0	1,000000 ₁₀ ⁺⁰	0	0	0	0	
	2	5,183882 ₁₀ ⁻¹	+0,96	+0,04	+0,33	-0,024	10 ⁻¹
	4	3,224001 ₁₀ ⁻¹	+0,56	-0,09	+0,10	-0,037	10 ⁻¹
	6	2,157855 ₁₀ ⁻¹	+0,17	-0,07	-0,02	-0,010	10 ⁻¹
	8	1,504248 ₁₀ ⁻¹	-0,07	-0,03	-0,07	+0,006	10 ⁻¹
	10	1,075993 ₁₀ ⁻¹	-0,19	-0,003	-0,08	+0,012	10 ⁻¹
	12	7,831954 ₁₀ ⁻²	-2,42	+0,11	-0,79	+0,113	10 ⁻²
	14	5,770725 ₁₀ ⁻²	-2,45	+0,16	-0,70	+0,086	10 ⁻²
	16	4,289136 ₁₀ ⁻²	-2,24	+0,16	-0,59	+0,055	10 ⁻²
	18	3,207969 ₁₀ ⁻²	-1,94	+0,13	-0,48	+0,031	10 ⁻²
	20	2,410230 ₁₀ ⁻²	-1,64	+0,09	-0,39	+0,012	10 ⁻²
	1,0	0	1,000000 ₁₀ ⁺⁰	0	0	0	0
2		6,630156 ₁₀ ⁻¹	+0,72	+0,03	+0,30	-0,009	10 ⁻¹
4		4,812784 ₁₀ ⁻¹	+0,59	-0,04	+0,18	-0,023	10 ⁻¹
6		3,617579 ₁₀ ⁻¹	+0,36	-0,04	+0,07	-0,011	10 ⁻¹
8		2,770914 ₁₀ ⁻¹	+0,15	-0,04	-0,01	-0,0002	10 ⁻¹
10		2,146944 ₁₀ ⁻¹	-0,0002	-0,01	-0,04	+0,005	10 ⁻¹
12		1,675850 ₁₀ ⁻¹	-0,10	+0,001	-0,06	+0,006	10 ⁻¹
14		1,314557 ₁₀ ⁻¹	-0,15	+0,01	-0,07	+0,005	10 ⁻¹
16		1,034555 ₁₀ ⁻¹	-0,18	+0,01	-0,07	+0,004	10 ⁻¹
18		8,160088 ₁₀ ⁻²	-1,89	+0,07	-0,66	+0,027	10 ⁻²
20		6,446038 ₁₀ ⁻²	-1,83	+0,07	-0,57	+0,016	10 ⁻²
1,5		0	1,000000 ₁₀ ⁺⁰	0	0	0	0
	2	8,096418 ₁₀ ⁻¹	+0,11	+0,0017	+0,08	-0,0003	10 ⁻¹
	4	6,620001 ₁₀ ⁻¹	+0,10	-0,0013	+0,07	-0,0010	10 ⁻¹
	6	5,435592 ₁₀ ⁻¹	+0,08	-0,0021	+0,05	-0,0003	10 ⁻¹
	8	4,472714 ₁₀ ⁻¹	+0,05	-0,0012	+0,02	+0,00005	10 ⁻¹
	10	3,684694 ₁₀ ⁻¹	+0,03	-0,0004	+0,005	+0,0002	10 ⁻¹
	12	3,037337 ₁₀ ⁻¹	+0,01	+0,0001	-0,007	+0,0002	10 ⁻¹
	14	2,504600 ₁₀ ⁻¹	-0,006	+0,0003	-0,01	+0,0001	10 ⁻¹
	16	2,065690 ₁₀ ⁻¹	-0,02	+0,0010	-0,02	+0,00007	10 ⁻¹
	18	1,703870 ₁₀ ⁻¹	-0,23	+0,0003	-0,20	+0,00003	10 ⁻¹
	20	1,405505 ₁₀ ⁻¹	-0,27	+0,0002	-0,21	+0,00002	10 ⁻¹

Tabelle 3. Vergleich exakter und approximativ (Momentenmethoden) gewonnener Werte der Wartezeitverteilungsfunktion
System: $n = 1, s_1 = 12$
Abfertigungsdisziplin : zufallsmäßiges Abfertigen (D2)

b) Sukzessive Approximation durch Potenzreihenentwicklungen

Für größere Differentialgleichungssysteme (Rang >100) ist das exakte Verfahren (Abschnitt 5.5) aus Rechenzeit-, Speicherplatz- und Konvergenzgründen ungeeignet. Bei Anwendung der Vier-Momenten-Methode sind in diesem Falle 4 (3) lineare inhomogene Gleichungssysteme entsprechenden Ranges auf iterative Weise aufzulösen; da die Lösungen des K-ten Gleichungssystems die rechten Seiten des (K+1)-ten Gleichungssystems bilden und derartige Gleichungssysteme sehr empfindlich bezüglich der Konvergenz sind, werden große Zykluszahlen und damit lange Rechenzeiten benötigt.

Aus obengenannten Gründen wurde das Verfahren der "sukzessiven Approximation durch Potenzreihenentwicklungen" entwickelt, vgl. Abschnitt 8.2.4.

Die Beispiele in den Abschnitten 6.4 und 7.4 wurden mit Hilfe dieses Verfahrens berechnet. Die Genauigkeit dieses Verfahrens liegt in der Größenordnung des "exakten Verfahrens" nach Abschnitt 5.5, bei welchem die Bestimmung der Eigenwerte auf iterativer Basis vorgenommen und daher ebenfalls nach Erreichen einer vorgeschriebenen Genauigkeit abgebrochen wird.

Zur Austauschbarkeit zwischen Speicherplatz und Rechenzeit werde als Beispiel das Bedienungssystem mit $n = 1$ Leitung und $s_1 = 12$ Warteplätzen und der Abfertigungsdisziplin D2 betrachtet ($A = 0,8$). Die Lösung für $W(>t)/W$ soll sich über den Zeitbereich $t/t_S = 10$ erstrecken. Tabelle 4 zeigt die Werte für K_{max} und n_{min} bei gleichem maximalem Schrittfehler R_n .

Ordnung der Taylor-Entw.	Zahl der Schritte	Maximaler Schrittfehler	Durchschnittl. Länge eines Intervalls	Speicherbedarf
K_{max}	n_{min}	R_n	$\Delta t/t_S$	$s_1(K_{max}+2)$
3	80	10^{-7}	0,125	60
5	24	10^{-7}	0,417	84
8	11	10^{-7}	0,908	120
10	9	10^{-7}	1,111	144

Tabelle 4. Zur Austauschbarkeit zwischen Speicherplatz und Rechenzeit bei der "Sukzessiven Approximation durch Potenzreihenentwicklungen"

Für größere Schrittzahlen muß die Möglichkeit der Fehlerfortpflanzung in Betracht gezogen werden. Als grobe Abschätzung kann hierfür der Wert $\sum_n R_n$ genommen werden. Vergleiche mit exakten Berechnungen haben gezeigt, daß dieser Wert unterschritten wird. Vorteilhaft wirkt sich hierbei aus, daß sämtliche Lösungen monoton abnehmende Funktionen sind und den asymptotischen Wert Null besitzen. Das Problem der numerischen Instabilität ist deshalb auf ein Minimum beschränkt.

Bei der Berechnung eines Beispiels von 600 simultanen Differentialgleichungen wurden für 20 Entwicklungsschritte ($K_{max} = 10, R_n = 10^{-6}$) ca. 10 min Rechenzeit (TR4) benötigt.

ZUSAMMENFASSUNG

In der vorliegenden Arbeit wurden einstufige Bedienungssysteme mit vollkommen bzw. unvollkommen erreichbaren Abfertigungsorganen, $g \geq 1$ Warteschlangen und verschiedenen Abfertigungsdisziplinen unter Markoffschen Voraussetzungen analytisch untersucht.

Neben der Berechnung der stationären Zustandswahrscheinlichkeiten wurde insbesondere auf das Problem der Wartezeiten eingegangen. Die Einführung eines aus dem Prozeß der Systemzustände konstruierten "Warteprozesses" ermöglichte eine methodisch einheitliche Formulierung des Wartezeitproblems für verschiedene Abfertigungsarten sowie für verschiedenartige Systemstrukturen, hier erstmals für Systeme mit mehreren Warteschlangen und unvollkommen erreichbaren Abfertigungsorganen. Durch die Untersuchung der zu den Warteprozessen zugehörigen Eigenwertprobleme konnte ein weitreichender Einblick in die Struktur der Wartezeitverteilungsfunktionen gewonnen werden.

Für die linearen Gleichungssysteme der stationären Zustandswahrscheinlichkeiten konnten teilweise rekursive Lösungsalgorithmen, z.T. sogar explizite Lösungen gefunden werden. Es wurden numerische Ergebnisse für die "charakteristischen Größen" von Bedienungssystemen angegeben, welche den Einfluß von Systemstruktur und Abfertigungsdisziplin zeigen.

Die linearen Differentialgleichungssysteme für die bedingten Wartezeitverteilungsfunktionen wurden unter Zuhilfenahme der Eigenwerttheorie untersucht. Hierdurch war es möglich, Aussagen über die Eigenwertverteilungen und die Struktur der Wartezeitverteilungsfunktionen zu machen. Insbesondere konnte die Wartezeitverteilungsfunktion für $g = 1$ Warteschlange und inverse Abfertigungsreihenfolge explizit abgeleitet werden. Ergebnisse in Form von Kurven und Tabellen demonstrieren die Einflüsse von Abfertigungsdisziplinen innerhalb und zwischen den Warteschlangen sowie der Struktur von Bedienungssystemen (vollkommene bzw. unvollkommene Erreichbarkeit) auf die Wartezeitcharakteristik.

Im abschließenden Kapitel wurden approximative Lösungsverfahren für lineare Gleichungs- und Differentialgleichungssysteme behandelt. Es wurde gezeigt, wie durch Einführung von Symmetrieannahmen rekursive

Lösungsalgorithmen für lineare Gleichungs- und Differentialgleichungssysteme gefunden werden können. Die Approximation der Wartezeitverteilungsfunktion durch Exponentialsummen wurde derart erweitert, daß auch schon die bedingten Wartezeitverteilungsfunktionen mit den bedingten Momenten bis zu einer vorgeschriebenen Ordnung übereinstimmen, womit eine bessere Genauigkeit gegenüber der Approximation der gesamten Wartezeitverteilungsfunktion erzielt werden konnte. Für die numerische Auflösung extrem großer linearer Differentialgleichungssysteme wurde ein Verfahren vorgeschlagen, welches die Lösung mit Hilfe von Potenzreihenentwicklungen auf rekursive Weise mit beliebig vorschreibbarer Genauigkeit erhält.

$(A + sI)$ betrachtet. Durch rekursive Entwicklung dieser Matrix folgt

$$D_i(s) = (s + \lambda + \mu) D_{i-1}(s) - \frac{i-1}{i} \lambda \mu D_{i-2}(s), \quad i = 1, 2, \dots, s-1, \quad (A1.7a)$$

$$D_s(s) = (s + \mu) D_{s-1}(s) - \frac{s-1}{s} \lambda \mu D_{s-2}(s), \quad (A1.7b)$$

wobei $D_0 = 1, D_{-1} = 0$.

Durch Vergleich von (A1.3b,c) mit (A1.7a,b) an der Stelle $s = 0$ folgt mit Gl.(A1.5)

$$D_i(0) > 0, \quad i = 0, 1, \dots, s. \quad (A1.8a)$$

Die höchste Potenz von $D_i(s)$ ist nach Gl.(A1.7a,b) $1 \cdot s^i$. Sie bestimmt das Verhalten von $D_i(\sigma)$ für $\sigma \rightarrow -\infty$ ($s = \sigma + j\omega, \omega = 0$):

$$\lim_{\sigma \rightarrow -\infty} D_i(\sigma) = (-1)^i \cdot \infty, \quad i = 1, 2, \dots, s. \quad (A1.8b)$$

In Bild A.1. ist der prinzipielle Verlauf der Funktionen $D_i(\sigma)$, $i = 0, 1, 2, \dots$ für $\sigma \leq 0$ angegeben.

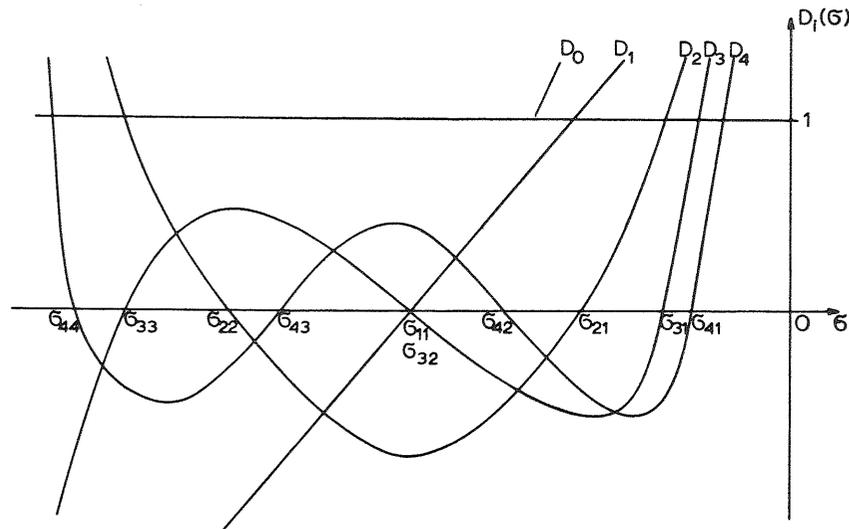


Bild A.1. Verlauf der Folge von Abschnittdeterminanten $D_i(\sigma)$

$D_0 = 1$ ist eine Konstante und besitzt daher keine Nullstelle.

$D_1(\sigma) = (\sigma + \lambda + \mu) \cdot D_0$ ist eine lineare Funktion und besitzt eine Nullstelle

$$\sigma_{11} = -(\lambda + \mu) < 0.$$

$D_2(\sigma) = (\sigma + \lambda + \mu) \cdot D_1(\sigma) - \frac{1}{2} \lambda \mu \cdot D_0$ werde an der Nullstelle von $D_1(\sigma)$ betrachtet: $D_2(\sigma_{11}) = -\frac{1}{2} \lambda \mu D_0 < 0$. Da jedoch $D_2(\sigma)$ sowohl an der Stelle $\sigma = 0$ positiv als auch für $\sigma \rightarrow -\infty$ positiv ausfällt, gibt es zwei Nullstellen σ_{21}, σ_{22} mit der Eigenschaft

$$\sigma_{22} < \sigma_{11} < \sigma_{21} < 0.$$

$D_3(\sigma) = (\sigma + \lambda + \mu) D_2(\sigma) - \frac{2}{3} \lambda \mu D_1(\sigma)$ werde an den Nullstellen von $D_2(\sigma)$ betrachtet. Aus den bisher nachgewiesenen Eigenschaften folgt: $D_3(\sigma_{21}) = -\frac{2}{3} \lambda \mu D_1(\sigma_{21}) < 0, D_3(\sigma_{22}) = -\frac{2}{3} \lambda \mu D_1(\sigma_{22}) > 0$. Es muß also zwischen σ_{21} und σ_{22} eine Nullstelle von $D_3(\sigma)$ liegen. Da $D_3(0) > 0$ und $D_3(\sigma) \rightarrow -\infty$ für $\sigma \rightarrow -\infty$ gilt, muß zwischen σ_{21} und 0 als auch links von σ_{22} jeweils eine weitere Nullstelle von $D_3(\sigma)$ gelegen sein. Insgesamt ergibt sich also folgendes Bild:

$$\sigma_{33} < \sigma_{22} < \sigma_{32} < \sigma_{21} < \sigma_{31} < 0.$$

Dieses Verfahren läßt sich beliebig fortsetzen. Im allgemeinen Falle wird $D_i(\sigma)$ an den Nullstellen von $D_{i-1}(\sigma)$ betrachtet und es gilt $D_i(\sigma_{i-1,j}) = -\frac{i-1}{i} \lambda \mu D_{i-2}(\sigma_{i-1,j}), j = 1, 2, \dots, i-1$. Wegen der alternierenden Vorzeichen von $D_{i-2}(\sigma_{i-1,j})$ und Gl.(A1.8a,b) gilt allgemein:

$$\sigma_{ii} < \sigma_{i-1,i-1} < \sigma_{i,i-1} < \sigma_{i-1,i-2} < \dots < \sigma_{i-1,2} < \sigma_{i,2} < \sigma_{i-1,1} < \sigma_{i,1} < 0. \quad (A1.9)$$

Aus Ungleichung (A1.9) folgt, daß alle Nullstellen von $D_i(\sigma)$ einfach sind und daß zwischen je zwei benachbarten Nullstellen von $D_i(\sigma)$ eine Nullstelle von $D_{i-1}(\sigma)$ liegt. Ungleichung (A1.9) gilt auch für den Grenzfall $i = s_1$, womit gezeigt ist, daß die Eigenwerte der Matrix A nach Gl.(5.42) sämtlich verschieden sind.

Satz 3: Die Eigenwerte der Matrix A nach Gl.(5.42) liegen alle im Intervall

$$\left[-\max\left(2\lambda + \frac{2s_1-3}{s_1-1} \mu, \frac{2s_1-1}{s_1} \mu\right), -\frac{\mu}{s_1} \right]$$

Für die Sonderfälle $s_1 = 1$ und $s_1 = 2$ sind die Eigenwerte explizit angebar:

$$e_1 = -\mu, \quad s_1 = 1$$

$$e_{1,2} = -\frac{1}{2} \left[(\lambda + 2\mu) \mp \sqrt{\lambda^2 + 2\lambda\mu} \right], \quad s_1 = 2.$$

Beweis von Satz 3

Ein Theorem von S. GERSCHGORIN [18] erlaubt die Abschätzung der Eigenwertbeträge direkt aus den Elementen der Matrix $A = (a_{ik})$:

Die Eigenwerte von A liegen innerhalb oder auf dem Rande des Gebietes $G(G')$, welches aus den N Kreisen K_i bzw. K'_i mit den Mittelpunkten $-a_{ii}$ und den Radien

$$r_i = \sum_{k \neq i} |a_{ik}| \quad \text{bzw.} \quad r'_i = \sum_{k \neq i} |a_{ki}|$$

gebildet wird. Die Eigenwerte liegen somit im Durchschnitt der Gebiete G und G' , d.h. in dem beiden Gebieten gemeinsamen Teilgebiet $G \cap G'$.

Wenden wir dieses Theorem auf Matrix A nach Gl.(5.42) an, so finden wir:

$$G : \left[-\max \left(2\lambda + \frac{2s_1-3}{s_1-1} \mu, \frac{2s_1-1}{s_1} \mu \right), \right]$$

$$G' : \left[-\left(2\lambda + \frac{2s_1-1}{s_1} \mu \right), -\min \left(\frac{\mu}{s_1}, \mu - \lambda \right) \right].$$

Der Durchschnitt von G und G' ergibt ebenfalls das Intervall von G .

Für die zwei Sonderfälle $s_1 = 1$ bzw. $s_1 = 2$ führt $\det(A + sI) = 0$ auf eine Gleichung 1. bzw. 2. Grades in s , deren Nullstellen explizit angebar sind.

A2 Wartezeitverteilungsfunktion bei einer Warteschlange und Abfertigung in der inversen Reihenfolge des Eintreffens mit verdrängender Priorität (D3a)

Satz 1: Die Matrix A nach Gl.(5.57) hat s_1 verschiedene negativ-reelle Eigenwerte

$$e_{\nu} = 2\sqrt{\lambda\mu} \cos \frac{\nu\pi}{s_1+1} - (\lambda+\mu), \quad \nu = 1, 2, \dots, s_1. \quad (5.58)$$

Beweis von Satz 1

Die Matrix A werde der nichtsingulären linearen Transformation

$$B = T^{-1}AT \quad (A2.1)$$

unterworfen, wobei

$$T = \begin{pmatrix} 1 & & & \\ & \sqrt{\frac{\mu}{\lambda}} & & \\ & & \ddots & \\ & & & \sqrt{\frac{\mu}{\lambda}}^{s_1-1} \end{pmatrix}. \quad (A2.2)$$

Als Ergebnis erhält man

$$B = \sqrt{\lambda\mu} \begin{pmatrix} \frac{\lambda+\mu}{\sqrt{\lambda\mu}} & -1 & & \\ -1 & \frac{\lambda+\mu}{\sqrt{\lambda\mu}} & -1 & \\ & & \ddots & \\ & & & -1 & \frac{\lambda+\mu}{\sqrt{\lambda\mu}} & -1 \\ & & & & -1 & \frac{\lambda+\mu}{\sqrt{\lambda\mu}} \end{pmatrix}. \quad (A2.3)$$

Die Eigenwerte einer Matrix sind invariant in Bezug auf eine nichtsinguläre lineare Transformation nach Gl.(A2.1) [18]. Die erhaltene Matrix B ist die bekannte Differenzenmatrix, von der die Eigenwerte explizit angegeben werden können [18], vergl.(5.58).

Satz 2: Die Wartezeitverteilungsfunktion bei D3a ist durch folgenden Ausdruck gegeben:

$$W(>t) = W \cdot \sum_{\nu=1}^{s_1} \left(\frac{\sum_{i=1}^{s_1} w(0|i-1) \lambda^{i-1} D_{s-i}(e_\nu)}{\prod_{i \neq \nu} (e_\nu - e_i)} \right) \cdot \exp(e_\nu t), \quad (5.59)$$

wobei

$$D_k(e_\nu) = \sqrt{\lambda}^k \cdot \frac{\sin(k+1) \frac{\nu\pi}{s_1+1}}{\sin \frac{\nu\pi}{s_1+1}}, \quad \nu = 1, 2, \dots, s_1, \\ k = 0, 1, \dots, s_1-1.$$

Beweis von Satz 2

w(t|0) ist die VF der Gesamt-wartezeit eines Testrufs. Verbunden mit der Antreffwahrscheinlichkeit P(i) nach Gl.(5.52) folgt für die Wartezeitverteilungsfunktion bezüglich aller Rufe

$$W(>t) = W \cdot w(t|0). \quad (A2.4)$$

Aus dem System Gl.(5.54a-c) ist also nur die Größe w(t|0) zu berechnen. Nach Einführung der Laplace-Transformation setzen wir an:

$$W(s|0) = \frac{Z(s|0)}{D_{s_1}(s)}, \quad (A2.5)$$

wobei

$$D_{s_1}(s) = \det(\mathbf{A} + s\mathbf{I}) = \prod_{i=1}^{s_1} (s - e_i). \quad (A2.6)$$

Das Zählerpolynom Z(s|0) erhält man aus det(A + sI), indem der erste Spaltenvektor durch den Spaltenvektor der Anfangsbedingungen nach Gl.(5.56) ersetzt wird. Durch sukzessive Entwicklung der entstandenen Determinante nach den Elementen der ersten Zeile (oder ersten Spalte) findet man

$$Z(s|0) = \sum_{i=1}^{s_1} w(0|i-1) \lambda^{i-1} \cdot D_{s-i}(s), \quad (A2.7)$$

wobei D_k(s) die k-te Hauptabschnittsdeterminante von (A + sI) bedeutet. Mit Gl.(A2.5), (A2.6) und (A2.7) ergibt sich

$$W(s|0) = \frac{\sum_{i=1}^{s_1} w(0|i-1) \lambda^{i-1} D_{s-i}(s)}{\prod_{i=1}^{s_1} (s - e_i)} = \sum_{\nu=1}^{s_1} \frac{a_\nu}{s - e_\nu}, \quad (A2.8a)$$

wobei

$$a_\nu = \frac{\sum_{i=1}^{s_1} w(0|i-1) \lambda^{i-1} D_{s-i}(e_\nu)}{\prod_{i \neq \nu} (e_\nu - e_i)}, \quad \nu = 1, 2, \dots, s_1. \quad (A2.8b)$$

In Gl.(A2.8b) muß als letzte Größe noch D_k(e_ν), k = 0, 1, ..., s₁-1, berechnet werden. Wegen

$$\det(\mathbf{A} + s\mathbf{I}) = \det[\mathbf{T}^{-1}(\mathbf{A} + s\mathbf{I})\mathbf{T}] \quad (A2.9)$$

und der besonderen Lösung Gl.(5.58) für die Eigenwerte folgt

$$D_k(e_\nu) = \sqrt{\lambda}^k \cdot \begin{vmatrix} 2 \cos \frac{\nu\pi}{s_1+1} & -1 & & & \\ -1 & 2 \cos \frac{\nu\pi}{s_1+1} & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 \cos \frac{\nu\pi}{s_1+1} & -1 \\ & & & -1 & 2 \cos \frac{\nu\pi}{s_1+1} \end{vmatrix} \quad (A2.10)$$

Die Substitution von $2 \cos \frac{\nu\pi}{s_1+1} = x + x^{-1}$, wobei $x = \exp(jx_\nu)$ und $x_\nu = \frac{\nu\pi}{s_1+1}$, $\nu = 1, 2, \dots, s_1$, erlaubt die Berechnung dieser Determinante

$$D_k(e_\nu) = \sqrt{\lambda}^k \cdot \sum_{\mathcal{L}=0}^k x^{k-2\mathcal{L}} = \sqrt{\lambda}^k \sum_{\mathcal{L}=0}^k \exp\left[j \frac{\nu\pi}{s_1+1} (k-2\mathcal{L})\right] \quad (A2.11) \\ = \sqrt{\lambda}^k \cdot \frac{\sin(k+1) \frac{\nu\pi}{s_1+1}}{\sin \frac{\nu\pi}{s_1+1}}, \quad \nu = 1, 2, \dots, s_1, \\ k = 0, 1, \dots, s_1-1.$$

Die Laplace-Rücktransformation von w(s|0) nach Gl.(A2.8a) liefert

$$w(t|0) = \sum_{\nu=1}^{s_1} a_\nu \exp(e_\nu t). \quad (A2.12)$$

Gl.(A2.12) in Gl.(A2.4) eingesetzt, ergibt mit a_ν nach Gl.(A2.8b) und D_k(e_ν) nach Gl.(A2.11) den Ausdruck für die Wartezeitverteilungsfunktion Gl.(5.59).

