

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

**39. Bericht über verkehrstheoretische Arbeiten**

Verkehrsanalyse von Warteschlangennetze  
mit Prioritäten

von  
Wolfgang Schmitt

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

**39th Report on Studies in Congestion Theory**

Traffic Analysis of Priority  
Queueing Networks

by  
Wolfgang Schmitt

1984

© 1986 Institut für Nachrichtenvermittlung und Datenverarbeitung Universität Stuttgart

Druck: E. Kurz & Co., Stuttgart

ISBN 3-922403-49-2

## TRAFFIC ANALYSIS OF PRIORITY QUEUEING NETWORKS

### ABSTRACT

Modelling of data communication networks and computer systems for purposes of performance evaluation, often leads to multi-class queueing networks with priority service disciplines. Whereas a variety of analytical results is available for single stage priority queueing models, the theory of queueing networks in its present form provides very few results for the analysis of networks with priority service disciplines.

This thesis deals in its main parts with the analysis of the stationary processes of system states in Markovian priority queueing networks. In general, those networks do not possess a product-form solution and an exact analysis is only possible for models with state-transition-rate diagrams of moderate complexity. Hence, in the main parts of this thesis I concentrate on the presentation of a new approximation technique which is applicable to an extensive class of open and closed priority queueing networks. My method is based on a decomposition of the multi-class priority queues into single-class queueing models without priority disciplines. I designate these models as "Ersatzmodelle" or "Virtual Server" models and we notice that its application to queueing networks with priorities leads to queueing networks with a product form solution.

The most remarkable thing about these Virtual Server models is a state dependent reduction of its service rates to take the priority mechanism into account. It is proved that this feature provides an accurate description of the processes of system states at least for the single stage queueing models M/M/1/PRE and M/M/1/NONPRE and for the cyclic queueing model  $\cdot/M/1/PRE \rightarrow \cdot/M/1/PRE$ . Compared to decomposition methods known up to now this means an essential improvement.

Subsequently, an outline of the thesis is given.

### CHAPTER 1: Introduction

This chapter describes the objectives of the thesis and contains a summary of its contents.

### CHAPTER 2: Queueing Models

After giving a brief review on Markov chains and its properties the second chapter presents notions and definitions for single- and multistage queueing models. These notions and definitions are indispensable for understanding the report. In the last part of Chapter 2 a queueing model of a local area network and a so called 'Central Server Model' are explained. Both models are analysed in Chapter 6.

### CHAPTER 3: Analysis of Queueing Models

In this chapter some known results on the analysis of single- and multistage queueing models are quoted. The first part contains results for the queueing models GI/G/1 (Refs. [Kraemer, Langenbach-Belz 76], [Marshall 68]), M/M/1/PRE (Ref. [White, Christie 58]) and M/M/1/NONPRE (Refs. [Cobham 54], [Miller 81]). Part two deals with queueing networks and presents first of all the well known BCMP-Theorem [Baskett, Chandy, Muntz, Palacios 75] and the generalised Norton Theorem [Kritzinger, Wyk, Krzesinski 82]. Subsequently, methods for the analysis of nonseparable queueing networks without and with priority service disciplines are briefly discussed (Refs. [Shum 76], [Chandy, Herzog, Woo 75], [Sauer, Chandy 75], [Courtois 77], [Herfort, Müller-Clostermann 83], [Kobayashi 74], [Kühn 79] and [Spirn 76], [Morris 81], [Wong, Sauvé 82], [Chandy, Lakshmi 83], [Bryant, Krzesinski, Teunissen 83], [Sauer, Chandy 75], [Chow, Yu 83], [Reiser 76], [Sevcik 77], [Kaufman 82], [Uspenski, Dowling 79] respectively).

### CHAPTER 4: Single-Stage Queueing Models with Priorities

The intention of the preceding chapters was to supply some fundamentals and to summarize the state of affairs in queueing network analysis. Henceforth we are concerned with new results related to single- and multistage queueing models with priority service disciplines.

Subject of Chapter 4 is the development of new methods to decompose the two-class priority queueing models M/M/1/PRE and M/M/1/NONPRE into single-class Virtual Server models. These Virtual Server models are called "Ersatzmodelle" (German) and described in the sections 4.3 and 4.4.

The Virtual Server models presented in Section 4.3 (see Fig. 4.1) are characterized by exponential service time distributions with state-dependent reduced service rates and they are valid for preemptive and nonpreemptive service disciplines. It is proved in Subsection 4.1.1 for preemptive and in Subsection 4.2.1 for nonpreemptive priorities that in the case of Poisson arrival processes these Virtual Server models provide an exact description of the marginal probabilities  $p_1(m)$  (high priority class) and  $p_2(n)$  (low priority class) of the queueing models M/M/1/PRE and M/M/1/NONPRE. The results lead also to very efficient algorithms to calculate these marginal probabilities (see Appendix 1 (ANHANG 1)) and the related blocking probabilities  $B_1$  and  $B_2$  (see Subsection 4.1.2 and Subsection 4.2.2). This decomposition method is called the "state dependent reduced occupancy approximation" and abbreviated to s d r o a.

The Virtual Server models presented in Section 4.4 (see Fig. 4.2) are only valid for preemptive priorities and they provide a moment approximation of the virtual service time  $T_{H2}^*$  of low priority customers. Hence they are appropriate for the queuing network analysis method described in [Kuehn 79] which requires the characterization of the virtual service time of low priority customers by its mean  $h_2^*$  and its coefficient of variation  $c_{H2}^*$ . Fig. 4.4 reflects two different approaches for the approximation of these quantities. I have called the approaches the "completion time approximation" (c t a) and the "flow time approximation" (f t a). Both approximation methods are derived in Subsection 4.1.3. The c t a assumes the virtual service time  $T_{H2}^*$  to be equal to the completion time  $T_{C2}$  of low priority customers. It follows from this

assumption, that the mean waiting time of a second class customer would be given by Equation (4.1,21). However, comparing this equation with the exact formula (see Eq. 4.1,9) makes clear that only in case of a saturated server ( $\rho_1 + \rho_2 = 1$ ) the c t a provides the right result for the mean waiting time  $w_2$  of a second class customer. The f t a avoids this lack of the c t a by determining the quantities  $h_2^*$  and  $c_{H2}^*$  in such a manner that the mean flow time of second class customers is exactly reproduced.

Both, the c t a and the f t a provide a well working approximation for the coefficient of variation  $c_{D2}$  of the output intervals  $T_{D2}$  of the M/M/1/PRE queue by setting the formulas (4.1,20) and (4.1,24) respectively into Equation (3.1,5). Some results are given in Table 4.1 (Tabelle 4.1) and validated by simulations. Table 4.2 contains the corresponding simulation results for the different output processes (first class:  $c_{D1}$ , second class:  $c_{D2}$ , both classes:  $c_D$ ) in the M/M/1/NONPRE queue. For this model, it turned out to be very difficult to find a moment approximation like the c t a or the f t a which provides good results over a wide parameter range.

#### CHAPTER 5: Priority Queueing Networks

This chapter contains new results for the traffic analysis of Markovian queueing networks with priority service disciplines. It is divided into the sections 5.1 and 5.2. The first of these sections deals with open priority queueing networks and the second one with closed priority queueing networks.

In Subsection 5.1.1 an outline of the analysis of the tandem queueing model M/M/1/NONPRE + M/M/1/NONPRE (see Fig. 5.1) by setting up the global balance equations and solving them numerically is given. Appendix 2 is pertinent to this problem and describes the classification of system states. Some numerical results for the mean flow times  $f_r$  of class r customers ( $r = 1,2$ ), dependent on class r arrival rates  $\lambda_r$  and mean service times  $h_{i_r}$

at station  $i$  ( $i = 1, 2$ ) are shown in Table 5.1. These quasi exact results can be compared there with those obtained by assuming the arrival processes at station two to be Poisson. We see that the maximum amount of the relative error between these values comes to 4.4 per cent.

In Subsection 5.1.2 the same tandem model is analysed by means of the  $c t a$  and the  $f t a$  method. Some results are given in Fig. 5.4.

Subsection 5.1.3 deals with the approximation of open priority queueing networks by separable queueing networks. This method is motivated by the decomposition of priority queues into Virtual Server models with state-dependent reduced service rates and it comprises the following steps:

Step 1: Compute the throughputs  $\lambda_{ir}$  of each queueing station.

Step 2: Analyse each queueing station separated from the network and subjected to Poisson arrivals with the rates  $\lambda_{ir}$ , obtained from Step 1.

Concerning the second step I refer to the benefit derived from the new algorithms to calculate the marginal state probabilities (see Appendix 1).

The Figs. 5.6 and 5.7 ... 5.9 show some results for the mean flow times  $f_r$  in the tandem queueing models  $M/M/1/PRE \rightarrow \cdot/M/1/PRE$  and  $M/M/1/NONPRE \rightarrow \cdot/M/1/NONPRE$  respectively. The choice of a tandem structure for the test-bed network is justified because this structure provides the worst case with regard to the interdependencies between the arrival-processes inside the network.

Section 5.2 begins with a study of the cyclic queueing model  $\cdot/M/1/PRE \rightarrow \cdot/M/1/PRE$ . It is proved in Subsection 5.2.1 that the queue-length-equivalence property (see Subsection 4.1.1 and 4.1.2) Subsection 4.1.2) of the decomposed priority stations holds even

for this cyclic queueing model where the arrival processes are in general not Poisson.

In Subsection 5.2.2 it is proved that the generalised Norton Theorem for queueing networks [Kritzinger, Wyk, Krzesinski 82] is also valid in case of the presence of  $\cdot/M/1/FIFO$  queueing stations with state-dependent service rates. This result will be needed for the analysis of closed priority queueing networks. The proof of the assertion is based on a lemma (Hilfssatz 5.1) which is proved in Appendix 3.

In Subsection 5.2.3 a method for the approximate analysis of closed Markovian queueing networks with priority service disciplines is presented. The general procedure is to decompose each priority station into Virtual Server models with state-dependent reduced service rates. That means to approximate a nonseparable priority queueing network by a separable queueing network. This approach is motivated in particular through the result of Subsection 5.2.1. The main problem is that the reduced service rates are a-priori unknown and hence, they must be approximated. Concerning this approximation, two different methods are presented. The first one is applicable to networks consisting of  $\cdot/M/1/PRE$ -stations exclusively and the second one allows the presence of priority stations (PRE and/or NONPRE) and BCMP-stations. Once the virtual service rates are known, it remains to solve a BCMP-network. The whole procedure is called the "modified state dependent reduced occupancy approximation" (m s d r o a).

#### CHAPTER 6: Examples of Application

The last chapter contains some numerical results for the network examples explained already in Chapter 2. Section 6.1 deals with the analysis of the queueing model of a local area network called RDC-ring (see Fig. 2.8 and Fig. 6.1) and Section 6.2 deals with the analysis of a Central Server Model (see Fig. 6.6). All results are validated by simulations with 95% confidence intervals. For comparison, the diagrams for the Central Server Model (see Figs. 6.7 ... 6.18) contain also results achieved by means of the reduced occupancy approximation (r o a) method

(see Refs. [Kaufman 82] and [Sevcic 77]) and by means of the composite centre approximation (c c a) method (see Refs. [Sauer, Chandy 75] and [Chow, Yu 83]).

INHALTSVERZEICHNIS

LITERATUR

BEZEICHNUNGEN

Seite

1.	<u>EINFÜHRUNG</u>	1
1.1	Entwicklung der Kommunikations- und Rechner-technik	1
1.2	Bedeutung der Verkehrsanalyse für die Planung und Entwicklung von Kommunikations- und Rechnersystemen	2
1.3	Methoden zur Verkehrsanalyse	5
1.4	Überblick über die Arbeit	8
2.	<u>WARTESCHLANGENMODELLE</u>	11
2.1	<u>Einiges über stochastische Prozesse</u>	11
2.1.1	Definition und Eigenschaften stochastischer Prozesse	11
2.1.2	Markoff-Ketten und ihre Eigenschaften	12
2.1.3	Der Poisson-Prozeß	19
2.2	<u>Warteschlangenmodelle mit einstufiger Bedienung</u>	22
2.2.1	Grundstruktur	22
2.2.2	Beschreibung der Last	23
2.2.3	Abfertigungs- und Prioritätsstrategien	25
2.2.4	Zustands- und Ausgangsprozesse	27
2.2.5	Kurzbezeichnungen	29
2.3	<u>Warteschlangenmodelle mit mehrstufiger Bedienung</u>	31
2.3.1	Offene Warteschlangennetze	31
2.3.2	Geschlossene Warteschlangennetze	32
2.3.3	Mehrklassige Warteschlangennetze	33

2.4	<u>Kenngrößen zur Leistungsbewertung</u>	35
2.5	<u>Beispiele für die Modellierung von Kommunikations- und Rechnersystemen</u>	38
2.5.1	Einige Gesichtspunkte zur Modellbildung	38
2.5.2	Die Unabhängigkeitsannahme von KLEINROCK	39
2.5.3	Erstes Beispiel: Kommunikationssystem mit Ringstruktur	40
2.5.4	Zweites Beispiel: "Multiprogramming" in Rechnern mit Speicherhierarchie	43
3.	<u>ANALYSE VON WARTESCHLANGENMODELLEN</u>	47
3.1	<u>Warteschlangenmodelle mit einstufiger Bedienung</u>	48
3.1.1	Warteschlangenmodell GI/G/1	48
3.1.2	Warteschlangenmodell M/M/1/PRE	50
3.1.3	Warteschlangenmodell M/M/1/NONPRE	52
3.2	<u>Warteschlangenmodelle mit mehrstufiger Bedienung</u>	54
3.2.1	Separable Warteschlangennetze	54
3.2.2	Nichtseparable Warteschlangennetze	60
3.2.3	Warteschlangennetze mit Prioritäten	62
4.	<u>EINSTUFIGE WARTESCHLANGENMODELLE MIT PRIORITÄTEN</u>	68
4.1	<u>Warteschlangenmodell M/M/1/PRE</u>	68
4.1.1	Zustandsverteilungen	68
4.1.2	Blockierungswahrscheinlichkeiten	71
4.1.3	Ausgangsprozesse	73
4.2	<u>Warteschlangenmodell M/M/1/NONPRE</u>	76
4.2.1	Zustandsverteilungen	76

4.2.2 Blockierungswahrscheinlichkeiten 81

4.2.3 Ausgangsprozesse 82

4.3 Einklassige Ersatzmodelle mit zustandsabhängiger Bedienstungsrate 84

4.4 Einklassige Ersatzmodelle mit konstanter Bedienstungsrate 87

5. WARTESCHLANGENNETZE MIT PRIORITÄTEN 89

5.1 Offene Warteschlangennetze 89

5.1.1 Zustandsanalyse eines seriellen Warteschlangenmodells mit nichtunterbrechenden Prioritäten 89

5.1.2 Ein approximatives Momentenverfahren 94

5.1.3 Approximation durch separable Warteschlangennetze 98

5.2 Geschlossene Warteschlangennetze 107

5.2.1 Zustandsanalyse eines zyklischen Warteschlangenmodells mit unterbrechenden Prioritäten 107

5.2.2 Erweiterung des Gültigkeitsbereiches für das "Norton Theorem" 110

5.2.3 Approximation durch separable Warteschlangennetze 112

6. AUSGEWÄHLTE SYSTEMSTUDIEN 127

6.1 Kommunikationssystem mit Ringstruktur 127

6.2 Multiprogramming in Rechnern mit Speicherhierarchie 134

ANHANG 1 151

ANHANG 2 156

ANHANG 3 158

LITERATURVERZEICHNIS

Avi-Itzak, D. P. Approximate Queueing Models for Multiprogramming Computer Systems. Op. Res. 21 (1973), 1212-1230.

Heyman, D. P.

Baskett, F. Open, Closed and Mixed Networks of Queues with Different Classes of Customers. Journal of the ACM 22 (1975), 248-260.

Chandy, K. M.

Muntz, R. R.

Palacios, G. F.

Bolch, G. Analyse von Rechnersystemen. Teubner Verlag, Stuttgart (1982).

Akyildiz, I. F.

Bryant, R. M. The MVA Pre-empt resume Priority Approximation. ACM Sigmetrics Conference (1983).

Krzesinski, A. E.

Teunissen, P.

Burke, P. J. The Output of a Queueing System. Op. Res. 4 (1956), 699-704.

Buzen, J. P. Queueing Network Models of Multiprogramming. PhD Thesis, Div. Eng. and Applied Physics, Harvard Univ., Cambridge, Mass. (1971).

Chandy, K. M. Parametric Analysis of Queueing Network Models. IBM Journal Res. Dev. 19 (1975), 36-42.

Herzog, U.

Woo, L.

Chandy, K. An Approximation Technique for Queueing Networks with Preemptive Priority Queues. Dept. of Computer Sciences, The University of Texas at Austin, Texas 78712 (1983).

Lakhshmi, M. S.

Chow, W. M. M/M/1 Priority Queues with State Dependent Arrival and Service Rates. IBM Res. Report RC8570 (1980).

Chow, W. M. An Approximation Technique for Central Server Queueing Models with a Priority Dispatching Rule. Performance Evaluation 3 (1983), 55-62.

Yu, P. S.

Chung, K. L. Markov Chains with Stationary Transition Probabilities. Springer Verlag, New York Inc. (1967).

Cobham, A. Priority Assignment in Waiting Line Problems. Op. Res. 2 (1954), 70-76.

Cohen, J. W. Boundary Value Problems in Queueing System Analysis. North Holland, Amsterdam/New York/Oxford (1983).

Boxma, O. J.

Conway, R. W. Theory of Scheduling. Addison-Wesley Publ. Comp., Reading/Massachusetts (1967).

Maxwell, W. L.

Miller, L. W.



Cooper, R. B. Introduction to Queueing Theory. North Holland, New York/Oxford (1981).

Courtois, P. J. Decomposability - Queueing and Computer System Applications. Academic Press, London/New York/Toronto/Sydney/San Francisco (1977).

Fahrmaier, L. Kaufmann, H. Ost, F. Stochastische Prozesse - Eine Einführung in Theorie und Anwendungen. Hanser Verlag, München/Wien (1981).

Forys, L. J. Performance Analysis of a New Overload Strategy. 10th International Teletraffic Congress, Montreal (1983).

Gaede, K. W. Zuverlässigkeit - Mathematische Modelle. Hanser Verlag, München/Wien (1977).

Gnedenko, B. W. Kowalenko, I. N. Einführung in die Bedienungstheorie. Oldenbourg Verlag, München/Wien (1971).

Gordon, W. J. Newell, G. F. Closed Queueing Systems with Exponential Servers. Op. Res. 15 (1967), 254-265.

Gross, D. Harris, C. Fundamentals of Queueing Theory. John Wiley and Sons, Inc., New York/London/Sydney/Toronto (1974).

Heger, D. Zur Klassifizierung, theoretischen Beschreibung und Bewertung lokaler Datensammelleitungen. Fortschritt-Berichte der VDI Zeitschriften, Reihe 10, Nr. 24.

Herfort, J. Müller-Clostermann, B. Strukturierung Markovscher Zustandsräume zur effizienten numerischen Modellierung. 2. GI/NTG-Fachtagung: Messung, Modellierung und Bewertung von Rechensystemen, Stuttgart (1983).

Isaacson, D. L. Madson, R. W. Markov Chains - Theory and Applications. John Wiley and Sons, New York/London/Sydney/Toronto (1976).

Jackson, J. R. Jobshop-Like Queueing Systems. Management Science 10 (1963), 131-142.

Jaiswal, N. K. Priority Queues. Academic Press, New York/London (1968).

Jordan, C. Calculus of Finite Differences. Chelsea Publishing Company, New York (1975).

Kameda, H. A Finite Source Queue with Different Customers. Journal ACM 29 (1982), 478-491.

Kaufman, J. S. Approximate Analysis of Priority Scheduling Disciplines in Queueing Network Models of Computer Systems. 6th International Conference on Computer Communication, London (1982), 955-961.

Kelly, F. Reversibility and Stochastic Networks. John Wiley & Sons, Chichester/New York/Brisbane/Toronto (1979).

Kendall, D. G. Stochastic Processes Occuring in the Theory of Queues and Their Analysis by the Method of Imbedded Markov Chains. Ann. Math. Statist. 24 (1953), 338-354.

Kleinrock, L. Communication Nets - Stochastic Message Flow and Delay. McGraw Hill, New York/San Francisco/Toronto/London (1964).

Kleinrock, L. Queueing Systems. Vol. I: Theory, Vol. II: Applications. John Wiley and Sons, Inc., New York/London/Sydney/Toronto (1975 and 1976).

Kobayashi, H. Application of the Diffusion Approximation to Queueing Networks. Parts I and II. Journal ACM 21 (1974), 316-328 und 459-469.

Kobayashi, H. Modeling and Analysis. An Introduction to System Performance Evaluation. Addison-Wesley Publ. Comp., Reading/Massachusetts (1978).

Krämer, W. Langenbach-Belz, M. Approximate Formulae for the Delay in the Queueing System GI/G/1. 8th International Teletraffic Congress, Melbourne (1976).

Kritzinger, P. S. van Wyk, S. Krzesinski, A. E. A Generalisation of Norton's Theorem for Multiclass Queueing Networks. Performance Evaluation 2 (1982), 98-107.

Kühn, P. J. Approximate Analysis of General Queueing Networks by Decomposition. IEEE Transactions on Communications 27 (1979), 113-126.

Kühn, P. J. Analyse zufallsabhängiger Prozesse in Systemen zur Nachrichtenvermittlung und Nachrichtenverarbeitung. Habilitationsschrift, Universität Stuttgart (1981).

- Kühn, P. J. Analysis of Busy Periods and Response Times in Queueing Networks by the Method of First Passage Times. International Symposium on Computer Performance Modeling, Measurement and Evaluation, College Park, Md., USA (1983).
- Lehnert, R. Ein modularer Prozessor zur Simulation von Zufallsnetzwerken. Dissertation RWTA, Aachen (1979).
- Lindley, D. W. The Theory of Queues with a Single Server. Proc. Cambridge Philosophical Society 48 (1952), 277-289.
- Little, J. D. C. A Proof of the Queueing Formula  $L = \lambda W$ . Op. Res. 9 (1961), 383-387.
- Marks, B. I. State Probabilities of M/M/1 Priority Queues. Op. Res. 21 (1973), 974-987.
- Marshall, K. T. Some Inequalities in Queueing. Op. Res. 16 (1968), 651-665.
- Miller, D. G. Computation of Steady-State Probabilities for M/M/1 Priority Queues. Op. Res. (1981), 945-958.
- Morris, R. J. T. Priority Queueing Networks. Bell Syst. Techn. Journal 60 (1981), 1745-1763.
- Muntz, R. R. Poisson Departure Processes and Queuing Networks. IBM Res. Rep. RC-4145 (1972).
- Neuts, M. F. Matrix-Geometric Solutions in Stochastic Models. The John Hopkins University Press, Baltimore/London (1981).
- Pollaczek, F. Problèmes Stochastiques Posés par le Phénomène de Formation d'une Queue d'Attente à un Guichet et par de Phénomènes Apparentes. Gauthiers Villars (Paris), (1957).
- Reiser, M. Interactive Modeling of Computer Systems. IBM Systems Journal 15 (1976), 309-327.
- Reiser, M. Mean-Value Analysis of Closed Multichain Queueing Networks. Journal ACM 27 (1980), 312-322.
- Sauer, C. H. Approximate Analysis of Central Server Models. IBM Journal Res. Dev. 19 (1975), 301-313.

- Sauer, C. H. Computer Systems Performance Modeling. Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1981).
- Chandy, K. M.
- Schassberger, R. Warteschlangen. Springer-Verlag, Wien/New York (1973).
- Schmitt, W. Analytische Untersuchung eines offenen Tandem-Wartesystems mit nichtunterbrechenden Prioritäten. Studienarbeit Nr. 29, FG Nachrichtentechnik, Universität Siegen (1982).
- Schneider, M.
- Schmitt, W. Untersuchung von Ausgangsprozessen in Wartesystemen mit Prioritäten. Studienarbeit Nr. 40, FG Nachrichtentechnik, Universität Siegen (1983).
- Carl, U.
- Schmitt, W. Approximate Analysis of Markovian Queueing Networks with Priorities. 10th International Teletraffic Congress, Montreal (1983).
- Schmitt, W. Brockert, K. Berechnung von Zustandswahrscheinlichkeiten im Wartesystem M/M/1-NONPRE. Studienarbeit Nr. 41, FG Nachrichtentechnik, Universität Siegen (1983).
- Schmitt, W. Gotthart, C. Wirksamkeit von Prioritäten in Rechnern mit "Multiprogramming". Diplomarbeit Nr. 42, FG Nachrichtentechnik, Universität Siegen (1983).
- Schneemann, J. Numerische Aspekte dreigliedriger linearer Rekursionen. Kolloquium des Fachgebiets Theoretische Elektrotechnik, Technische Hochschule Darmstadt (1975).
- Sevcik, K. C. Priority Scheduling Disciplines in Queueing Network Models of Computer Systems. Information Processing 77 (1977), 565-570.
- Shum, A. W. Queueing Models for Computer Systems with General Service Time Distributions. PhD Thesis, Div. Eng. and Applied Physics, Harvard Univ., Cambridge, Mass. (1976).
- Spirn, J. R. Multi-Queue Scheduling of Two Tasks. International Symposium on Computer Performance Modeling, Measurement, and Evaluation; Harvard Univ., Cambridge, Mass. (1976), 102-106.
- Stein, M. D. The Memory Organization of Future Large Processors. Quaestiones Informaticae 1 (1980).

Stephan, F. F. Two Queues under Preemptive Priority with Poisson Arrival and Service Rates. Op. Res. 6 (1958), 399-418.

Takahashi, Y. Asymptotic Exponentiality of the Tail of the Waiting-Time Distribution in a PH/PH/c Queue. Adv. Appl. Prob. 13 (1981), 619-630.

Törnig, W. Numerische Mathematik für Ingenieure und Physiker, Band 1. Springer Verlag, Berlin/Heidelberg/New York (1979).

Uspenski, A. Approximation Method for Closed Loop Model with Priority Discipline at CPU. 13th Asilomar Conference Circuits, Systems Computers, Pacific Grove, CA (1979), 574-578.

Van As, H. R. QSIMLIB, A Modular FORTRAN IV Package Supporting Development of Stationary and Transient Simulation Programs for Complex Queueing Systems. Interner Bericht der Fachgruppe Nachrichtentechnik der Universität GH Siegen (1984).

Walke, B. Realzeitrechner-Modelle, Theorie und Anwendung. R. Oldenbourg Verlag, München/Wien (1978).

White, H. Queueing with Preemptive Priorities or with Breakdown. Op. Res. (1958), 79-95.

Christie, L. S.

Wojtkowiak, H. H. Effects of a Reduced Memory Residence Set on System Performance. IBM Res. Report RC8607 (1980).

Wong, J. W. A Study of Fairness in Packet-Switching Networks. IEEE Transactions on Communications 30 (1982), 346-353.

Sauvé, J. P.

BEZEICHNUNGEN

BE	Bedienungseinheit
BS	Bedienungsstation
CPU	Central Processing Unit
CSM	Central Server Model
M-K	Markoff Kette
VF	Verteilungsfunktion
WS	Wartespeicher
WSN	Warteschlangennetz
ZV	Zufallsvariable
i.Z.	"in Zeichen"
E[X]	Erwartungswert der ZVen X
VAR[X]	Varianz der ZVen X
o(h)	Landau-Symbol (siehe Seite 14)
$\delta_{ik}$	Kronecker-Symbol (siehe Seite 14)
$\underline{A}$	Matrix A
$\underline{A}^T$	zu $\underline{A}$ transponierte Matrix
$\mathbb{N}$	Menge der natürlichen Zahlen
$\mathbb{N}_0$	Menge der nichtnegativen ganzen Zahlen
$\mathbb{R}$	Menge der reellen Zahlen
$\mathbb{R}^+$	Menge der nichtnegativen reellen Zahlen
$x := A$	die Größe x ist durch den Ausdruck A definiert
$x \approx y$	die Größe x ist ungefähr gleich der Größe y
$x \stackrel{!}{=} y$	die Größe x ist gleich der Größe y gesetzt
$x \downarrow x_0$	die Größe x strebt gegen den Wert $x_0$ und $x > x_0$
$\Delta$	"für alle"

## 1. EINFÜHRUNG

### 1.1 Entwicklung der Kommunikations- und Rechnertechnik

Betrachtet man die Entwicklungen auf dem Gebiet der vermittelnden Kommunikationstechnik von ihrem Beginn im letzten Jahrhundert an bis heute, so läßt sich ein ständig wachsender Automatisierungsgrad in den Systemen und, vor allem in neuerer Zeit, ein steigendes Angebot an Diensten für den Kunden feststellen. Parallel zu diesen Fortschritten ist eine rapide zunehmende Komplexität der nachrichtentechnischen Einrichtungen zu beobachten. Als Beispiel hierfür sei der Werdegang der Fernsprechtechnik von der Erfindung des Telefons im Jahre 1861 an bis hin zur nahezu vollständigen Automatisierung der internationalen Fernwahl in der Bundesrepublik ab 1970 genannt.

Für die Öffentlichkeit noch aufsehenerregender stellen sich sicherlich die Entwicklungen in der Rechnertechnik mit all ihren Konsequenzen für unsere Arbeits- und Freizeitwelt dar. Vor allem die Fortschritte in der Hardware/Software-Technologie und auf dem Gebiet der Rechnerarchitektur haben zur Leistungsfähigkeit heutiger Rechner beigetragen und ihnen als neues Anwendungsgebiet die Informationsverarbeitung zur Lösung sehr komplexer Steuerungsprobleme erschlossen. Diese Möglichkeit, komplizierteste Steuerungsaufgaben mit Hilfe von Groß- und/oder Kleinstrechnern (Mikrocomputer, Mikroprozessoren) wirtschaftlich zu bewältigen, wird zusammen mit der Digitalisierung im Bereich der Übertragungs- und Vermittlungstechnik sowie der Bereitstellung hoher Bandbreiten durch optische Wellenleiter die Zukunft der Kommunikationstechnik hinsichtlich der Erschließung neuer Dienste wesentlich mitbestimmen. Als Beispiele für solche neuen Dienste seien Teletex, Telefax, Bildschirmtext, Datex, Bildfernsprechen und Bildkonferenz genannt. Ein wichtiger Trend in der Entwicklung von Rechnersystemen zeichnet sich in der Kopplung von Einzelsystemen zu Rechnernetzen ab. Berührungspunkte von Nachrichten- und Rechnertechnik zeigen sich somit einerseits beim Einsatz von Rechnern zur Bewältigung von Steuerungsaufgaben in Kommunikationsnetzen und andererseits bei der Bereitstellung von Datennetzen für Über-

mittlungsaufgaben in Rechnerverbundsystemen. Aus diesem Grunde wird in der vorliegenden Arbeit sowohl auf Kommunikations- als auch auf Rechnersysteme eingegangen werden.

## 1.2 Bedeutung der Verkehrsanalyse für die Planung und Entwicklung von Kommunikations- und Rechnersystemen

Das Ablaufgeschehen in modernen Kommunikations- und Rechnersystemen ist durch einen regen Verkehr charakterisiert, welcher durch den Austausch von Nutz- und Steuerinformationen sowohl zwischen Benutzer und System als auch zwischen Systemteilen untereinander hervorgerufen wird. Dieser Informationsaustausch ist mit der Bearbeitung vielfältiger Aufgaben durch das System verbunden, welches hierfür eine Anzahl geeigneter Betriebsmittel bereitstellt. So müssen beispielsweise im Falle einer Mensch-Computer-Kommunikation über ein Paketvermittlungnetz Nutz- und Steuerdaten übermittelt werden, wofür Paketvermittlungsstellen und Datenübertragungskanäle zur Verfügung stehen. Allgemeiner gesprochen, können Betriebsmittel in vielen Fällen als Bedienungseinheiten, bestehend aus einer oder mehreren Bedienungseinheiten (z.B. Übertragungskanälen) mit vorgegebener Bedienungsgeschwindigkeit und einem oder mehreren Wartespeichern mit bestimmter Kapazität, angesehen werden. Die Aufgaben stellen im allgemeineren Sinne Aufträge oder Anforderungen dar, welche die Bedienungseinheiten mit einer gewissen Last belegen. Während die für das Verkehrsgeschehen maßgebenden Kenngrößen der Bedienungseinheiten (Anzahl der Bedienungseinheiten, Bedienungsgeschwindigkeit, Anzahl der Speicherplätze) rein deterministischer Natur sind, können die verkehrsbestimmenden Eigenschaften der Last oft nur in statistischem Sinne beschrieben werden. So ist etwa die Länge eines Datenpakets eine Zufallsgröße, und die Folge der Eintreffzeitpunkte von Datenpaketen an einer Paketvermittlungsstelle stellt einen stochastischen Prozeß dar. Dementsprechend beschreibt man die Eigenschaften der Systemlast durch statistische Kenngrößen wie Verteilungsfunktionen, Erwartungswerte, Variationskoeffizienten u.a.m. . Die zufallsbedingten Eigenschaften der Systemlast haben natürlich auch einen zufallsbedingten Ablauf des

allgemeinen Verkehrsgeschehens im System zur Folge.

Für die Planung und Entwicklung moderner Kommunikations- und Rechnersysteme ergibt sich mit deren zunehmender Komplexität immer mehr die Notwendigkeit einer verkehrsgerechten Auslegung der Einrichtungen. Es sei hierzu das Problem genannt, die Anzahl der bereitzustellenden Betriebsmittel günstig festzulegen und auf die sinnvolle Vergabe und Organisation dieser Betriebsmittel nach Gesichtspunkten der Wirtschaftlichkeit und Leistungsfähigkeit zu achten. Zur Lösung dieser Aufgaben stellt die Verkehrstheorie geeignete Verfahren zur Verfügung. Bevor diese ausführlicher diskutiert werden, seien noch einige Bemerkungen zu den Begriffen "Leistung" und "Betriebsorganisation" vorangestellt.

Im Sinne der Verkehrstheorie kann man zwischen benutzerorientierten Leistungsmerkmalen und systemorientierten Leistungsmerkmalen unterscheiden. Zur ersten Kategorie zählen z.B. Zyklus- und Durchlaufzeiten (Aufenthaltszeiten) in Systemen mit Stapelbetrieb sowie Warte- und Antwortzeiten in interaktiven Systemen. Bezeichnend für diese Größen ist, daß sie wiederum im allgemeinen zufallsabhängig sind. Typische Vertreter der Klasse systemorientierter Leistungsmerkmale sind Systemdurchsatz (kurz: Durchsatz) und Systemauslastung (kurz: Auslastung) sowie Blockierungswahrscheinlichkeiten. Es ist zu bemerken, daß benutzer- und systemorientierte Leistungsgrößen nicht voneinander unabhängig sind. So kann die Ursache für eine hohe Antwortzeit ein Betriebsmittelengpaß sein, welcher sich durch einen sehr hohen Ausnutzungsgrad der entsprechenden Bedienungseinheit bemerkbar macht. Da solche Betriebsmittelengpässe gegebenenfalls zu Blockierungen infolge von Speicherüberlauf führen können, ist es für den Entwicklungsingenieur sehr wesentlich, diese schon während der Entwicklungsphase des Systems zu erkennen, um geeignete Gegenmaßnahmen treffen zu können. Es stellt sich hier das Problem der Leistungsvorhersage. Die erwähnten Gegenmaßnahmen können zum einen in der Erweiterung von Speicher- und Bedienungskapazität und zum anderen in einer Verbesserung der Betriebsorganisation bestehen, wobei aus Gründen der Wirtschaftlichkeit das letztgenannte Mittel oft vorzuziehen ist.

Hinsichtlich der Betriebsorganisation unterscheidet man zwischen den Gesichtspunkten der Betriebsmittelzuteilung und -verwaltung sowie der Verkehrslenkung. Zu letzterer ist neben den derzeit häufig implementierten Strategien der festen sowie alternativen Verkehrslenkung die adaptive Verkehrslenkung zu nennen, welche mit der Einführung separater Signalisierungsnetze zunehmende Bedeutung erlangt. Bei der adaptiven Verkehrslenkung erfolgt die Wegesuche durch ein Kommunikationsnetz nach Kriterien der momentan kürzestmöglichen Übertragungszeit.

Bezüglich der Betriebsmittelzuteilung und -verwaltung seien zunächst die Warteschlangen- und Belegungsorganisation genannt. Die Warteschlangenorganisation legt die Reihenfolge des Ein- und Ausspeicherns von Anforderungen fest, während durch die Belegungsorganisation die Zuteilung von eintreffenden oder wartenden Anforderungen zu freien Bedienungseinheiten geregelt wird. Weitere Einzelheiten hierzu findet man z.B. in [Conway, Maxwell, Miller 67] oder [Kühn 81].

Eine in vielen Anwendungsfällen sehr wichtige Art der Betriebsmittelzuteilung bietet sich an, wenn die Anforderungen unterschiedliche Dringlichkeit aufweisen. Dies kann z.B. bei Realzeitanforderungen gegenüber zeitunkritischen Anforderungen der Fall sein. Es ergibt sich dann auf natürliche Weise eine Prioritätsorganisation, indem die Anforderungen bestimmten Prioritätsklassen zugeteilt werden, welche gewisse Unterbrechungseigenschaften aufweisen. Innerhalb jeder Klasse erfolgt die Bearbeitung nach einer festen Strategie wie z.B. "FIFO", d.h. Abarbeitung in der Reihenfolge des Eintreffens. Im Hinblick auf die Unterbrechungseigenschaften werden unterbrechende und nichtunterbrechende Prioritäten unterschieden. Bei der ersten Art wird die Bearbeitung einer Anforderung beim Eintreffen einer Anforderung höherer Priorität sofort unterbrochen, während bei der zweiten Art die sich gerade in Bedienung befindende Anforderung erst zu Ende bearbeitet wird. Es sei erwähnt, daß auch Mischformen dieser Unterbrechungsarten angewendet werden. Eine sehr ausführliche Übersicht über die verschiedenen Abfertigungsstrategien findet man in [Walke 78].

Für den Systemplaner und -entwickler ist es wiederum sehr wesentlich, aufgrund einer Verkehrsanalyse über die Wirksamkeit der genannten Verfahren zur Betriebsmittelvergabe und -verwaltung sowie zur Verkehrslenkung Bescheid zu wissen. Auf die wichtigsten Methoden zur Durchführung einer Verkehrsanalyse soll im nächsten Abschnitt eingegangen werden.

### 1.3 Methoden zur Verkehrsanalyse

Eine Möglichkeit zur Ermittlung von Verkehrsdaten besteht in der Durchführung von Messungen am realen System, wobei man zwischen Hardware und Software-Messungen unterscheidet. Hardware-Messungen werden mit sogenannten Hardware-Monitoren durchgeführt. Sie gestatten die Ermittlung von Verkehrsgrößen durch das Anschalten von Sensoren am Betriebsmittel und einer geeigneten quantitativen Auswertung der Meßgrößen. Das Ablaufgeschehen im Betriebsmittel wird hierbei nicht gestört.

Software-Messungen werden mit Hilfe von Software-Monitoren durchgeführt und sind vor allem geeignet, um programmgesteuerte Ereignisse sowie Systemzustände zu erfassen. Hierzu werden entweder sogenannte Meßprogramme zur Registrierung von signifikanten Ereignissen zu natürlichen Unterbrechungspunkten im Ablaufgeschehen eingeschleift (event-tracing monitor), oder es erfolgt eine künstliche Unterbrechung in konstanten Zeitabständen, an denen der Software-Monitor "Schnappschüsse" des Systemzustandes aufnimmt (sampling monitor). Der Einsatz von Software Monitoren ist somit immer mit einem Eingriff in das normale Ablaufgeschehen im Betriebsmittel verbunden.

Die Anwendungsgebiete von Meßmethoden erstrecken sich vorwiegend auf Überwachungsfunktionen (z.B. Erkennen von Überlastsituationen) und auf die Ermittlung von Daten für die Systemabstimmung. Darüber hinaus dienen sie der Beschaffung von zuverlässigen Eingabeparametern für die als nächstes besprochene Methode der Verkehrsanalyse mit Hilfe von Verkehrsmodellen sowie der Verifikation der mit dieser Methode erzielten Ergeb-

nisse. Die wesentlichen Nachteile von Meßmethoden liegen darin, daß sie zum einen in der Planungs- und Entwicklungsphase nicht anwendbar sind (keine Leistungsvorhersage möglich) und zum anderen bei vertretbarem Aufwand nur einen Überblick über das lokale Verkehrsgeschehen geben.

Die erwähnten Nachteile von Verkehrsmessungen vermeidet die Verkehrsanalyse mit Hilfe von Verkehrsmodellen. Bei diesem Verfahren wird das reale System zunächst auf ein Verkehrsmodell abgebildet, welches die verkehrsbestimmenden System- und Lastmerkmale möglichst gut wiedergibt. Entsprechend berücksichtigt ein Verkehrsmodell folgende Komponenten:

1. Struktur des Systems
2. Beschreibung der Last
3. Betriebsorganisation des Systems.

Unter einer Vielzahl von Modellen zur Beschreibung des realen Verkehrsgeschehens nehmen die Warteschlangenmodelle eine besonders wichtige Stellung ein. Sie gestatten die Modellierung von Systemen, welche als strukturbezogene Komponenten im wesentlichen Verkehrsquellen, Speicher und Bedienungseinheiten enthalten. Im einfachsten Falle besteht ein Warteschlangenmodell aus einer Verkehrsquelle, einem Wartespeicher und einer Bedienungseinheit. Man spricht dann von einem Warteschlangenmodell mit einstufiger Bedienung oder kürzer einem einstufigen Warteschlangenmodell. Durchläuft eine Anforderung auf ihrem Weg mehrere Bedienstationen, die nach Gesichtspunkten der Verkehrslenkung ausgewählt werden, so führt dies auf Warteschlangenmodelle mit mehrstufiger Bedienung, welche auch als Warteschlangennetze bezeichnet werden. Warteschlangennetze treten z.B. bei der Modellierung von Kommunikations- und Rechnersystemen sowie von Vermittlungssteuerungen auf. Da solche Einrichtungen im Hinblick auf ihre Betriebsorganisation im allgemeinen mehrere Prioritätsebenen aufweisen, gelangt man bei der Modellbildung zu Warteschlangennetzen mit Prioritäten.

Ist die Phase der Systemmodellierung abgeschlossen, so folgt im zweiten Schritt die Modellanalyse. Diese kann entweder auf dem Wege der Simulation oder mit Hilfe von mathematischen Verfahren erfolgen.

Bei der Simulation wird das Ablaufgeschehen im Modell entweder mittels eines Hardware-Simulators [Lehnert 79] oder mit Hilfe eines Rechnerprogramms nachgebildet [Kobayashi 79, Kühn 81]. Unter verschiedenen, für den Digitalrechner geeigneten, Simulationsmethoden hat das Verfahren nach dem Prinzip der zeit-treuen Simulation die größte praktische Bedeutung. Dieses Verfahren gestattet einerseits eine große Modellierungstiefe (Wirklichkeitstreue des Verkehrsmodells) und erweist sich andererseits bei entsprechender Modularität des verwendeten Simulationsprogramms als recht flexibel im Hinblick auf Modellvariationen. Der Nachteil dieser Methode liegt darin, daß sie einmal wegen ihres hohen Rechenzeitbedarfs sehr kostenaufwendig ist, und daß sie außerdem keinen Einblick in analytische Zusammenhänge liefert. Aus letzterem Grunde sind auch Parameterabhängigkeiten, aus denen sich gegebenenfalls Entscheidungs- und Optimierungskriterien ergeben, nur mit größerem Aufwand zu erkennen.

Unter den mathematischen Verfahren zur Modellanalyse nehmen die Methoden der Warteschlangentheorie [Schassberger 73, Gross, Harris 74, Kleinrock 75, Cooper 81] die wichtigste Stellung ein. Diese Theorie stellt unter Anwendung von wahrscheinlichkeitstheoretischen Hilfsmitteln sowohl approximative als auch exakte analytische Lösungsverfahren für eine Vielzahl von Warteschlangenmodellen bereit, welche jedoch in ihrer Allgemeinheit und Komplexität gegenüber den für die Simulation zugänglichen Modellen oft Einschränkungen aufweisen. Da jedoch der erforderliche Rechenzeitaufwand für analytische Verfahren gegenüber dem der Simulation fast immer klein ist, ergänzen sich die beiden Methoden in sehr guter Weise. Dies wird auch dadurch deutlich, daß sowohl Simulationsergebnisse zur Validierung von analytischen Resultaten als auch analytische Resultate zur Validierung von Simulationsergebnissen verwendet werden. In manchen Fällen findet auch die Technik der Hybrid-

simulation Anwendung, welche eine Kombination aus Simulation und analytischen Methoden darstellt.

Abschließend sei bemerkt, daß in diesem Abschnitt lediglich eine Skizzierung der wichtigsten Methoden zur Verkehrsanalyse vorgenommen werden konnte. Für eine ausführlichere Darstellung zu diesem Problemkreis sei auf [Kobayashi 79] und [Kühn 81] verwiesen.

#### 1.4 Überblick über die Arbeit

Die Entwicklung von Verkehrsmodellen für Kommunikations- und Rechnersysteme führt in vielen Fällen auf Warteschlangennetze mit Prioritäten. Während Prioritätsmechanismen in Einzelstufen in der Vergangenheit weitgehend untersucht worden sind [Jaiswal 68, Schassberger 73], steht die Theorie für die Analyse von Netzen erst am Anfang. In der vorliegenden Arbeit werden approximative Verfahren zur mathematischen Analyse von Prioritätsnetzen mit Markoff-Eigenschaften bereitgestellt. Die hierzu entwickelte Theorie liefert Einblicke in einige grundsätzliche Zusammenhänge in Warteschlangenmodellen mit Prioritäten.

Das Kapitel 2 beginnt mit einer Zusammenstellung von Ergebnissen aus der Theorie stochastischer Prozesse, welche zum Verständnis der Arbeit erforderlich sind. Im Anschluß daran erfolgt eine Beschreibung von ein- und mehrstufigen Warteschlangenmodellen sowie die Definition der darin ablaufenden Zustands- und Ausgangsprozesse. Nach der Erklärung einiger Kenngrößen zur Leistungsbewertung findet anhand zweier anwendungsorientierter Beispiele abschließend eine kurze Einführung in die Methodik der Modellbildung statt.

Im dritten Kapitel sind bekannte Verfahren und Ergebnisse zur Analyse von ein- und mehrstufigen Warteschlangenmodellen zusammengestellt. Insbesondere wird im letzten Abschnitt dieses Kapitels ein kurzer Überblick über den derzeitigen Kenntnisstand zur mathematischen Analyse von Warteschlangennetzen mit

Prioritäten gegeben. Dabei liegt ein Hauptaugenmerk auf der systematischen Eingliederung der teilweise recht unterschiedlichen Lösungsansätze. Die Zusammenstellung geschieht vorwiegend zu Referenzzwecken in den folgenden Kapiteln.

Hauptgegenstand des vierten Kapitels ist die Bereitstellung eigener, neuer Methoden zur "Zerlegung" der zweiklassigen Prioritätsmodelle M/M/1/PRE und M/M/1/NONPRE in einklassige Warteschlangenmodelle ohne Prioritätsabfertigung. Letztere werden kurz als "Ersatzmodelle" bezeichnet. Von besonderer Bedeutung sind die neu entwickelten Ersatzmodelle mit zustandsabhängig reduzierten Bedienungsraten. Diese Modelle garantieren die Äquivalenz ihrer Zustandswahrscheinlichkeiten zu denen der Originalmodelle und sind daher den bislang bekannten Ersatzmodellen entscheidend überlegen. Über diese Zerlegungsmethoden hinaus findet man noch einige Ergebnisse für Ausgangsprozesse sowie ein neues Verfahren zur direkten Berechnung von Randverteilungen und Blockierungswahrscheinlichkeiten in Prioritätsmodellen.

Auf den Ergebnissen des vierten Kapitels aufbauend wurden neue Verfahren für die approximative Analyse von offenen und geschlossenen Warteschlangennetzen mit Prioritäten entwickelt. Die Erläuterung dieser Verfahren bildet den Schwerpunkt des fünften Kapitels. Nach einer einleitenden Betrachtung zur Problematik der exakten Analyse von Warteschlangennetzen mit Prioritäten am Beispiel eines seriellen Warteschlangenmodells wird zunächst ein approximatives Momentenverfahren zur Analyse offener Netze diskutiert. Die im Anschluß daran erläuterte Methode der Approximation von Prioritätsnetzen durch separable Warteschlangennetze ist diesem Verfahren jedoch aus verschiedenen Gründen überlegen und nimmt daher im wesentlichen den verbleibenden Teil des fünften Kapitels ein. Alle approximativen Methoden wurden anhand geeigneter Testmodelle erprobt und die Resultate zum Zwecke des Genauigkeitsnachweises mit exakten und/oder Simulationsergebnissen verglichen. Außerdem stehen Ergebnisse zur Verfügung, die mit einem der in Kapitel 3 skizzierten Näherungsverfahren berechnet worden sind.



Das abschließende sechste Kapitel enthält zwei anwendungsorientierte Systemstudien. Es handelt sich dabei um das schon im Kapitel 2 eingeführte Modell für ein Kommunikationssystem mit Ringstruktur sowie um das dort ebenfalls schon erläuterte "Central Server Model". Letzteres gestattet die Untersuchung des Verkehrsgeschehens in Rechnersystemen mit "Multiprogramming". Die in diesem Kapitel angegebenen numerischen Ergebnisse bestätigen nochmals die Tauglichkeit der auf Ersatzmodellen mit zustandsabhängig reduzierten Bedienungsdaten beruhenden Approximationsmethode.

Zum Abschluß sei noch bemerkt, daß im folgenden nicht mehr auf die Technik der zeitreuen Simulation, welche bei der Validierung von Approximationsergebnissen eine grundlegende Rolle spielt, eingegangen werden wird. Es sei diesbezüglich auf die Arbeit [Van As 84] verwiesen. Alle in der vorliegenden Arbeit angegebenen Simulationsergebnisse wurden mit FORTAN-Programmen erzielt, die auf dem dort beschriebenen Programmpaket QSIMLIB aufbauen.

## 2. WARTESCHLANGENMODELLE

### 2.1 Einiges über stochastische Prozesse

Für die mathematische Beschreibung und Analyse von Warteschlangenmodellen werden einige grundlegende Definitionen und Sätze aus der Theorie stochastischer Prozesse benötigt, welche, soweit dies zum Verständnis der vorliegenden Arbeit erforderlich ist, in den nachfolgenden drei Abschnitten zusammengestellt sind. Dabei ist die Kenntnis von Grundbegriffen und Grundergebnissen aus dem Gebiete der Wahrscheinlichkeitsrechnung vorausgesetzt. Um Verwechslungen vorzubeugen, sei noch bemerkt, daß in diesem Unterkapitel der Buchstabe  $h$  ausnahmsweise zur Bezeichnung von Zeitspannen verwendet wird. Später ist dieses Zeichen ausschließlich für die Bezeichnung der mittleren Bedienungszeit vorgesehen.

#### 2.1.1 Definition und Eigenschaften stochastischer Prozesse

Die mathematische Beschreibung der Ergebnisse von zufallsabhängigen Vorgängen geht vom Begriff des Wahrscheinlichkeitsraumes aus, worunter ein Tripel  $(\Omega, \alpha, P)$  verstanden wird, in welchem  $\Omega$  die Ergebnismenge,  $\alpha$  eine  $\sigma$ -Algebra von Teilmengen von  $\Omega$  (Ereignisse) und die Abbildung  $P$  von  $\alpha$  in die Menge der reellen Zahlen  $\mathbb{R}$  (i. Z.  $P: \alpha \rightarrow \mathbb{R}$ ) ein Wahrscheinlichkeitsmaß auf  $\alpha$  darstellt. Eine Abbildung  $X$  von  $\Omega$  in den  $n$ -dimensionalen Euklidischen Raum  $\mathbb{R}^n$  wird als ( $n$ -dimensionale) Zufallsvariable (ZV) bezeichnet. Die Bereitstellung dieser Begriffe diente zur Vorbereitung der folgenden

#### Definition 2.1: Stochastischer Prozeß

Sei  $T \subseteq \mathbb{R}$ . Dann heißt eine Familie  $\{X_t: t \in T\}$  von ZVn auf einem Wahrscheinlichkeitsraum  $(\Omega, \alpha, P)$  ein stochastischer Prozeß.

Die Menge  $T$  wird auch als Parameterraum bezeichnet und die Vereinigung  $E$  aller Wertemengen von  $X_t$  heißt Zustandsraum

des Prozesses. Die Elemente  $x_t$  aus  $E$  heißen Zustände des Prozesses. In den späteren Anwendungen wird der Parameter  $t$  stets als Zeit gedeutet, während der Zustandsraum  $E$  entweder aus der Menge  $\mathbb{R}^+$  der nichtnegativen reellen Zahlen (zur Repräsentation von Zeiten) oder aber aus einer Teilmenge des  $n$ -dimensionalen kartesischen Produktes  $\mathbb{N}_0^n$  der Menge der nichtnegativen ganzen Zahlen (zur Repräsentation von Modellzuständen) bestehen wird. Ordnet man jeder ZVn  $X_t$  einen ihrer Werte  $x_t$  zu, so heißt die dadurch auf  $T$  definierte Funktion eine Realisation oder Trajektorie des stochastischen Prozesses. Um später die Verwendung von Staffelindices zu vermeiden, wird im folgenden die Schreibweise  $X(t)$  (bzw.  $x(t)$ ) anstelle von  $X_t$  (bzw.  $x_t$ ) verwendet.

Zur Vorbereitung späterer Betrachtungen seien noch zwei Eigenschaften stochastischer Prozesse zusammengestellt, welche sich aus den Verteilungsfunktionen (VFen) der ZVn  $X(t)$  ergeben.

Sind die ZVn  $X(t)$  voneinander unabhängig und identisch verteilt, so heißt der stochastische Prozeß  $\{X(t): t \in T\}$  ein unabhängiger Prozeß. Wenn für  $\{t_1, t_2, \dots, t_n\} \subset T$  und für  $\{t_1+h, t_2+h, \dots, t_n+h\} \subset T$ , ( $n \in \mathbb{N}$ ,  $h \in \mathbb{R}$ ), die Verteilungen der  $n$ -dimensionalen ZVn  $(X(t_1), \dots, X(t_n))$  und  $(X(t_1+h), \dots, X(t_n+h))$  gleich sind, so heißt der stochastische Prozeß stationär.

### 2.1.2 Markoff-Ketten und ihre Eigenschaften

Dieser Abschnitt enthält einige wichtige Definitionen und Sätze (ohne Beweis) über Markoff-Ketten mit kontinuierlichem Parameter.

#### Definition 2.2: Markoff-Kette

Ein stochastischer Prozeß  $\{X(t): t \in T\}$  mit dem abzählbaren Zustandsraum  $E$  heißt eine Markoff-Kette (M-K) mit kontinuierlichem Parameter, wenn für jede Parameterfolge  $0 \leq t_1 < t_2 < \dots < t_n$  und jede Zustandsfolge  $x_1, x_2, \dots, x_n$ , ( $x_i \in E$ ), mit  $n > 2$  die Beziehung (Markoff-Eigenschaft)

$$\begin{aligned} P\{X(t_n)=x_n | X(t_{n-1})=x_{n-1}, \dots, X(t_1)=x_1\} \\ = P\{X(t_n)=x_n | X(t_{n-1})=x_{n-1}\} \end{aligned} \quad (2.1,1)$$

gilt. Ist  $E$  endlich, so heißt auch die M-K endlich.

Der Zusatz "mit kontinuierlichem Parameter" wird im folgenden weggelassen, da in dieser Arbeit nur solche M-Kn auftreten werden.

Aufgrund der letzten Gleichung werden M-Kn auch oft als Prozesse ohne Gedächtnis bezeichnet. Die Markoff-Eigenschaft findet man in vielen praktischen Anwendungen vor.

Hängen die Übergangswahrscheinlichkeiten  $P\{X(t_n)=x_n | X(t_{n-1})=x_{n-1}\}$  nur von der Differenz  $h=t_n-t_{n-1}$  ab, so heißt die M-K homogen oder auch eine M-K mit stationären Übergängen. Wird  $E=\mathbb{N}$  angenommen, was wegen der vorausgesetzten Abzählbarkeit von  $E$  keine Einschränkung der Allgemeinheit bedeutet, so genügt es, für  $i, j \in E$  die Übergangswahrscheinlichkeiten durch

$$p_{ij}(h) := P\{X(t_n)=j | X(t_{n-1})=i\} \quad (2.1,2)$$

zu bezeichnen. Die Größen  $p_{ij}(h)$  genügen den Beziehungen [Gaede 77]:

$$p_{ij}(h) \geq 0, \quad (i, j \in E; h \geq 0), \quad (2.1,3a)$$

$$\sum_{j \in E} p_{ij}(h) = 1, \quad (i, j \in E; h \geq 0), \quad (2.1,3b)$$

$$p_{ij}(h+s) = \sum_{k \in E} p_{ik}(h) \cdot p_{kj}(s), \quad (i, j \in E; h, s \geq 0). \quad (2.1,3c)$$

Das Gleichungssystem (2.1,3c) nennt man auch die Gleichungen von Chapman und Kolmogoroff. Fortan werden ausschließlich M-Kn mit stationären Übergängen betrachtet.

Zur Herleitung eines Differentialgleichungssystems zur Bestimmung der Übergangswahrscheinlichkeiten  $p_{ij}(h)$  wird die folgen-

de Definition der Übergangsraten  $q_{ij}$ , ( $i, j \in E$ ), benötigt:

$$q_{ij} := \lim_{h \rightarrow 0} \frac{p_{ij}(h) - \delta_{ij}}{h} \quad (2.1,4a)$$

mit

$$q_{ij} \in [0, \infty) \quad \text{für } j \neq i \quad (2.1,4b)$$

und

$$q_{ii} \in [-\infty, 0] \quad \text{für } j = i, \quad (2.1,4c)$$

welche für

$$\lim_{h \rightarrow 0} p_{ij}(h) = \delta_{ij} \quad (2.1,5)$$

stets existieren [Isaacson, Madson 79]. Mit  $\delta_{ij}$  ist das durch

$$\delta_{ij} := \begin{cases} 1 & \text{für } i=j \\ 0 & \text{sonst} \end{cases} \quad (2.1,6)$$

definierte Kronecker Symbol bezeichnet. Anstelle von (2.1,4a) kann man für kleine  $h$  auch fordern:

$$p_{ij}(h) = \delta_{ij} + q_{ij}h + o(h), \quad h \rightarrow 0. \quad (2.1,7)$$

wobei  $o(h)$  (Landausches Symbol) eine Funktion mit der Eigenschaft

$$\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0 \quad (2.1,8)$$

symbolisiert. Die Matrix  $Q := (q_{ij})$  wird als Intensitätsmatrix der M-K bezeichnet. Ist die Folge  $\{|q_{ii}|\}$  beschränkt, so heißt  $Q$  beschränkt, und es gilt [Schassberger 73]

$$\sum_{j \in E} q_{ij} = 0. \quad (2.1,9)$$

Damit kann der folgende Satz angegeben werden:

Satz 2.1: [Schassberger 73]

Die Intensitätsmatrix der M-K  $\{X(t): t \in [0, \infty)\}$  sei beschränkt. Dann sind die Übergangswahrscheinlichkeiten  $p_{ij}(h)$ , ( $i, j \in E$ ), im Intervall  $[0, \infty)$  stetig differenzierbar und genügen dem Differentialgleichungssystem (Kolmogoroffsche Vorwärtsgleichungen)

$$\frac{d}{dh} p_{ij}(h) = \sum_{k \in E} p_{ik}(h) \cdot q_{kj}, \quad (i, j \in E; h \geq 0), \quad (2.1,10a)$$

mit den Anfangsbedingungen

$$p_{ij}(0) = \delta_{ij}, \quad (i, j \in E). \quad (2.1,10b)$$

Es sei erwähnt, daß auch ein als "Kolmogoroffsche Rückwärtsgleichungen" bekanntes Differentialgleichungssystem hergeleitet werden kann, dessen Lösungen mit denen der Vorwärtsgleichungen übereinstimmen.

Ein weiteres charakteristisches Merkmal zur Kennzeichnung des Verhaltens einer M-K bilden die Zustandswahrscheinlichkeiten (Zustandsverteilung)

$$p(j;t) := P\{X(t)=j\}, \quad (j \in E, t \in [0, \infty)). \quad (2.1,11)$$

Sie geben die Wahrscheinlichkeit an, mit welcher sich der Prozeß zum Zeitpunkt  $t$  im Zustand  $j$  befindet. Vor der Angabe eines sehr grundlegenden Satzes zur Bestimmung der stationären (von  $t$  unabhängigen) Zustandsverteilung  $p(j)$ , werden noch einige Definitionen vorangestellt, welche eine Klassifizierung der Zustände im Hinblick auf ihre Erreichbarkeit und ihr Rückkehrverhalten ermöglichen. Als sehr nützliche Anschauungshilfe erweist sich dabei der sogenannte Übergangsgraph (Bild 2.1), in welchem alle möglichen Zustände der M-K durch Knoten (dargestellt durch Kreise) mit der zugehörigen Zustandsnummer und alle möglichen Übergänge durch Pfeile mit der zugehörigen Übergangsrate symbolisiert sind.

Führt im Übergangsgraph ein Pfeil von  $i$  nach  $j$  oder gibt es

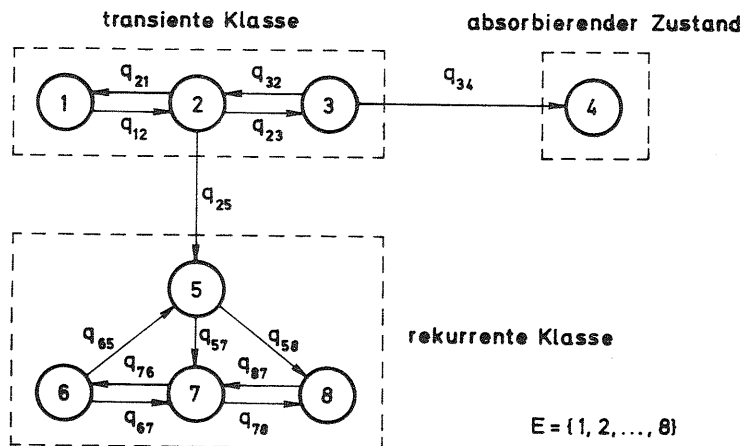


Bild 2.1: Übergangsgraph einer Markoff-Kette

eine Folge von Pfeilen, die (über eine Folge von Zuständen) von  $i$  nach  $j$  führen, so heißt der Zustand  $j$  vom Zustand  $i$  aus erreichbar (i.Z.:  $i \sim j$ ). Gilt außerdem  $j \sim i$  so heißen  $i$  und  $j$  verbunden (i.Z.:  $i \sim j$ ). Vereinbart man noch, daß jeder Zustand als mit sich selbst verbunden ( $i \sim i$ ) gilt, so bildet  $\sim$  eine Äquivalenzrelation und induziert eine Klasseneinteilung des Zustandsraumes  $E$  derart, daß all diejenigen Zustände in einer Klasse zusammengefaßt werden, welche miteinander verbunden sind (Bild 2.1). Im Falle, daß alle Zustände miteinander verbunden sind, existiert nur eine Klasse und die M-K heißt irreduzibel. Ist von keinem der Zustände einer Klasse ein Zustand einer anderen Klasse erreichbar, so nennt man die Klasse abgeschlossen. Besteht eine abgeschlossene Klasse nur aus einem Zustand, so wird dieser als absorbierend bezeichnet.

Die folgende Erklärung betrifft das Rückkehrverhalten der Zustände. Ein Zustand  $i \in E$  heißt rekurrent, wenn die M-K bei Start in  $i$  mit Wahrscheinlichkeit 1 nach einer gewissen Rückkehrzeit nach  $i$  zurückkehrt, oder wenn  $i$  ein absorbierender Zustand ist. Nichtrekurrente Zustände werden als transiente Zustände bezeichnet. Man kann zeigen [Gaede 77], daß in einer homogenen M-K, deren Zustandsraum endlich ist und deren durch

(2.1,4) definierte Übergangsraten existieren, ein Zustand genau dann rekurrent ist, wenn er einer abgeschlossenen Klasse angehört. Sind alle Zustände einer Klasse rekurrent (transient), so heißt die Klasse rekurrent (transient).

Als nächstes wird ein fundamentaler Satz zur Bestimmung der stationären Zustandsverteilung angegeben:

Satz 2.2 [Gaede 77]

Ist  $\{X(t): t \in [0, \infty)\}$  eine endliche M-K mit beschränkter Intensitätsmatrix, so gilt:

- i) Die reellen Zahlen  $p(j)$ , ( $j \in E$ ), bilden genau dann eine stationäre Zustandsverteilung der M-K, wenn die globalen Gleichgewichtsgleichungen

$$p(i) \sum_{j \in E} q_{ij} = \sum_{j \in E} p(j) \cdot q_{ji}, \quad (i \in E), \quad (2.1,12a)$$

mit der Normierungsbedingung

$$\sum_{i \in E} p(i) = 1 \quad (2.1,12b)$$

erfüllt sind.

- ii) Ist die M-K irreduzibel, so hat (2.1,12) nur eine einzige Lösung. Dieses ist die (vom Startzustand des Prozesses unabhängige) Grenzverteilung  $\pi_j := \lim_{h \rightarrow \infty} p_{ij}(h)$ , ( $j \in E$ ).

Die Übergangswahrscheinlichkeiten  $p_{ij}(h)$  sind Lösungen von (2.1,10a-b).

Bemerkungen

- i) Satz 2.2 behält unter verschärften Annahmen, welche in der Praxis meist erfüllt sind und im wesentlichen beinhalten,

daß der Prozeß in einer endlichen Zeitspanne nicht unendlich viele Zustände durchlaufen kann (vgl. [Chung 67] sowie [Fahrmeir, Kaufmann, Ost 81]), seine Gültigkeit auch für abzählbares E.

ii) Die Terme  $p(i) \cdot q_{ij}$  können für jedes Zustandspaar  $i, j \in E$ , ( $j \neq i$ ), als "Wahrscheinlichkeitsfluß" vom Zustand  $i$  zum Zustand  $j$  interpretiert werden und korrespondieren im Übergangsgraphen mit dem entsprechenden Pfeil von  $i$  nach  $j$ . Demzufolge können die Gleichungen (2.1,12a) direkt aus dem Übergangsgraphen der M-K aufgestellt werden.

Zum Ende dieses Abschnitts sei noch auf eine Entstehungsmöglichkeit von endlichen, homogenen M-Kn eingegangen, über welche der folgende Satz Auskunft gibt.

Satz 2.3: [Gaede 77]

Es sei  $\{X(t): t \in [0, \infty)\}$  ein stochastischer Prozeß mit endlichem Zustandsraum E. Ferner sei  $t_0=0$  oder ein Zeitpunkt, in dem ein Zustandswechsel von  $X(t)$  erfolgt, und es gelte  $X(t_0)=i$ .

Für jeden Zustand  $i$  sei eine Teilmenge  $E_i$  der Zustandsmenge E festgelegt, welche  $i$  nicht enthält. Der nächste Übergang aus dem Zustand  $i$  erfolge mit Sicherheit nicht in einen Zustand  $j$  aus der Differenzmenge  $E \setminus E_i$ . Für alle  $k \in E_i$  dagegen seien vom Verlauf des Prozesses vor dem Zeitpunkt  $t_0$  und voneinander unabhängige  $\alpha_{ik}$ -exponentialverteilte Zufallsvariable  $T_{ik}$  gegeben. Der nächste Zustandswechsel von  $X(t)$  nach dem Zeitpunkt  $t_0$  erfolge zum Zeitpunkt  $t_0 + V_0$  mit

$$V_0 := T_{i, \min} = \min \{T_{ik}: k \in E_i\}$$

und zwar in den Zustand  $j$ , wenn  $T_{ij} = T_{i, \min}$  ist. Dann gilt:  $\{X(t): t \in [0, \infty)\}$  ist eine endliche homogene M-K mit den Übergangsraten:

$$q_{ij} := \begin{cases} \alpha_{ij} & \text{für } j \in E_i \\ 0 & \text{sonst} \end{cases}, \quad (i \neq j; i, j \in E). \quad (2.1,13)$$

Bemerkungen

i) Eine ZV T wird als  $\alpha$ -exponentialverteilt bezeichnet, falls für deren VF  $F(t) = P\{T \leq t\}$  die Beziehung

$$F(t) = 1 - \exp(-\alpha t), \quad (t \geq 0), \quad (2.1,14a)$$

gilt. Für das k-te Moment  $E[T^k]$  von T folgt dann

$$E[T^k] = k! \cdot \alpha^{-k} \quad (2.1,14b)$$

Die Exponentialverteilung ist die einzige "gedächtnislose" Verteilung, d.h. es gilt

$$P\{T \leq t+h | T > t\} = P\{T \leq h\}. \quad (2.1,14c)$$

ii) Bezüglich der Gültigkeit des Satzes 2.3 für abzählbares E gilt das in der Bemerkung i) zum Satz 2.2 Gesagte entsprechend.

Auf diesen grundlegenden Satz wird im folgenden stets dann stillschweigend zurückgegriffen, wenn in den Anwendungen direkt der Übergangsgraph oder die globalen Gleichgewichtsgleichungen für den gerade betrachteten Prozeß angegeben werden.

2.1.3 Der Poisson-Prozeß

Für die Beschreibung der Last in einem Wartemodell spielen die sogenannten Zählprozesse eine wesentliche Rolle. Darunter sind stochastische Prozesse  $\{N(t): t \in [0, \infty)\}$  zu verstehen, in denen  $N(t)$  die Anzahl von Ereignissen angibt, die bis zum Zeitpunkt t einschließlich eingetreten sind. Es wird  $N(0)=0$  vereinbart. Ein für die Praxis sehr wichtiger Zählprozeß ist der Poisson-Prozeß, welcher mit Hilfe der folgenden Forderungen definiert werden kann:

i)  $N(t)$  hat unabhängige Zuwächse. D.h., für jede Folge  $0 \leq t_0 < t_1 \dots < t_n$  ist  $\{N(t_i) - N(t_{i-1}): i=1, 2, \dots, n\}$  eine unabhängige Familie von ZVn.

ii)  $N(t)$  hat stationäre Zuwächse. D.h., für beliebige Zeitpunkte  $t_1, t_2$  mit der Eigenschaft  $0 \leq t_1 < t_2$  sind für ein  $h > 0$  die Zuwächse  $N(t_2) - N(t_1)$  und  $N(t_2+h) - N(t_1+h)$  identisch verteilt.

iii)  $P\{(N(t+h) - N(t)) > 1\} = o(h), \quad h > 0. \quad (2.1,15a)$

Man kann zeigen [Gnedenko, Kowalenko 71], daß aus diesen drei Forderungen die Existenz einer Zahl  $\lambda > 0$  folgt, so daß

iv)  $P\{N(t+h) - N(t) = 1\} = \lambda h + o(h), \quad h > 0 \quad (2.1,15b)$

gilt.

Diese axiomatische Definition des Poisson-Prozesses hat den Vorteil, daß sie einen Einblick in dessen Entstehungsgesetz gibt, was sowohl bei der Modellierung von realen Prozessen, als auch bei der Approximation von anderen, mathematisch definierten Prozessen durch Poisson-Prozesse von Vorteil sein kann. Im folgenden wird nun eine Anzahl von Eigenschaften des Poisson-Prozesses angegeben, welche sich aus den Forderungen i) bis iv) ergeben. Die zugehörigen Beweise findet man z.B. in [Fahrmeir, Kaufmann, Ost 81].

Zunächst ist festzustellen, daß das Verteilungsgesetz des Poisson-Prozesses durch die Beziehung (Poisson-Verteilung)

$$P\{N(t)=n\} = P\{N(t+h)-N(h)=n\} = \exp(-\lambda t) \frac{(\lambda t)^n}{n!}, \quad (n \in \mathbb{N}_0), \quad (2.1,16a)$$

gegeben ist. Für den Erwartungswert und die Varianz von  $N(t)$  folgt dann

$$E[N(t)] = \text{VAR}[N(t)] = \lambda t. \quad (2.1,16b)$$

Deutet man im Hinblick auf die späteren Anwendungen  $N(t)$  als die Anzahl von eintreffenden Anforderungen (kurz: Ankünften) in einem Wartesystem im Zeitintervall  $[0, t]$ , so ist  $\lambda t$  die mittlere Anzahl von Ankünften in diesem Zeitintervall. Wegen

$E[N(t)/t] = \lambda$  und  $\text{VAR}[N(t)/t] = \lambda/t \rightarrow 0$  für  $t \rightarrow \infty$ , gilt die Beziehung

$$\lim_{t \rightarrow \infty} \frac{N(t)}{t} = \lambda, \quad (2.1,16c)$$

woraus sich die Identifizierung von  $\lambda$  als Ankunftsrate ergibt. Wird mit  $T_{Ai}$ , ( $i \in \mathbb{N}$ ), die Zeitspanne zwischen dem  $(i-1)$ -ten und dem  $i$ -ten Ereignis (Zwischenzeit, Ankunftsabstand) bezeichnet, so liefert der folgende Satz eine Charakterisierung des Poisson-Prozesses durch seine Zwischenzeiten.

Satz 2.4: [Fahrmeir, Kaufmann, Ost 81]

Seien  $T_{Ai}$ , ( $i \in \mathbb{N}$ ), die Zwischenzeiten eines Zählprozesses  $\{N(t): t \in [0, \infty)\}$ , dann handelt es sich genau dann um einen Poisson-Prozeß mit dem Parameter  $\lambda$ , falls die  $T_{Ai}$  voneinander unabhängig und identisch  $\lambda$ -exponentialverteilt sind.

Aus diesem Theorem folgt, daß der Poisson-Prozeß eine M-K ist.

Abschließend wird noch auf die für die Analyse von Warteschlangennetzen wichtige Überlagerung und Zerlegung von Poisson-Prozessen eingegangen. Anstelle einer formalen Definition dieser Begriffe, seien für eine plausible Erklärung die Bilder 2.2a und 2.2b herangezogen. Demnach werden bei der Überlagerung von  $m$  gegenseitig unabhängigen Poisson-Prozessen  $\{N_i(t): t \in [0, \infty)\}$ , ( $i=1, \dots, m$ ), mit den Parametern  $\lambda_1, \dots, \lambda_m$  alle Ereignisströme "zusammengefaßt" und bilden einen neuen Prozeß  $\{N(t): t \in [0, \infty)\}$ , welcher die Ereignisse aller Prozesse in der Reihenfolge ihres Auftretens zählt. Man kann durch Verifizierung der Forderungen i) bis iv) zeigen, daß der so konstruierte Zählprozeß ebenfalls ein Poisson-Prozeß mit dem Parameter  $\lambda = \lambda_1 + \lambda_2 + \dots + \lambda_n$  ist.

Bei der Zerlegung eines Poisson-Prozesses  $N := \{N(t): t \in [0, \infty)\}$  in  $m$  Teilprozesse  $N_i := \{N_i(t): t \in [0, \infty)\}$ , ( $i=1, \dots, m$ ), wird jedes in  $N$  auftretende Ereignis aufgrund eines unabhängigen Zufallsexperimentes mit der Wahrscheinlichkeit  $p_i$  dem neuen Zählprozeß  $N_i$  zugeordnet. Auch hier kann man zeigen, daß die

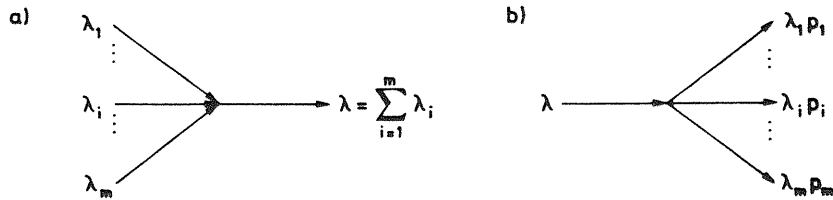


Bild 2.2: a) Überlagerung von Poisson-Prozessen  
b) Zerlegung eines Poisson-Prozesses

Prozesse  $N_i$  voneinander unabhängige Poisson-Prozesse mit den Parametern  $\lambda_i = \lambda p_i$  sind, wobei natürlich  $\sum_i p_i = 1$  gelten muß.

## 2.2 Warteschlangenmodelle mit einstufiger Bedienung

### 2.2.1 Grundstruktur

Von der Struktur her gesehen, besteht ein Warteschlangenmodell mit einstufiger Bedienung im einfachsten Fall aus einem Wartespeicher (WS) mit  $s$  Speicherplätzen und einer Bedienungseinheit (BE), deren verkehrsbestimmender Parameter die Bedienungsgeschwindigkeit  $C$  ist. Unter der Bedienungsgeschwindigkeit ist dabei die Anzahl von Bedienstücken zu verstehen, welche von der BE in einer bestimmten Zeitspanne bearbeitet werden kann. Sie wird z.B. in kbit/sek bei einem Übertragungskanal oder in Instruktionen/sek bei einer Zentraleinheit (CPU) angegeben. WS und BE bilden zusammen eine Bedienstation (BS),

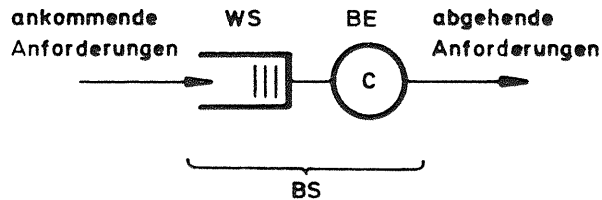


Bild 2.3: Grundstruktur eines einstufigen Warteschlangenmodells

die von eintreffenden Anforderungen "betreten" und nach Abschluß der Bedienung wieder verlassen wird. Bild 2.3 zeigt die symbolische Darstellung eines solchen Modells.

Genauer betrachtet, erfährt eine Anforderung in Abhängigkeit vom Zustand der BS eine der folgenden Behandlungen:

- i) Die BE ist nicht belegt.  
Dann wird die Anforderung sofort bedient.
- ii) Die BE ist belegt, es stehen jedoch noch freie Warteplätze zur Verfügung.  
In diesem Falle wird die Anforderung bis zur späteren Bedienung in die Warteschlange eingereiht.
- iii) Die BE und alle Warteplätze sind belegt.  
Tritt dieses ein, so ist die BS blockiert, und die Anforderung wird abgewiesen.

Für die BE soll gelten, daß sie tätig wird, sobald eine Anforderung auf ihre Bedienung wartet.

In vielen Anwendungsfällen wird die Anzahl der Speicherplätze als unbegrenzt angenommen (d.h.  $s \rightarrow \infty$ ), so daß der Fall iii) entfällt. Diese Annahme ist immer dann gerechtfertigt, wenn der Speicherplatz im realen System so groß ist, daß eine Blockierung praktisch ausgeschlossen werden kann. In der vorliegenden Arbeit wird, soweit nicht anders vermerkt, stets  $s \rightarrow \infty$  vorausgesetzt. Außerdem werden ausschließlich BSen mit nur einer BE betrachtet.

### 2.2.2 Beschreibung der Last

Ein mathematisches Modell der von der Verkehrsquelle an die BS gerichtete Last muß die beiden folgenden Größen berücksichtigen:

- i) Anzahl  $N_A(T)$  der in einem Zeitintervall  $T$  in der BS ein-

treffenden Anforderungen unter Miteinbeziehung der VF der Ankunftsabstände  $T_A(t_i)$ ,  $t_i \in T$ , (Bild 2.4).

- ii) Anzahl  $N_H$  der Bedienstungsquanten, welche von jeder Anforderung zur Erfüllung ihres Bedienungswunsches benötigt wird.

Zum zweiten Punkt ist hinzuzufügen, daß anstelle der Größe  $N_H$  meistens die Bedienstungszeit  $T_H$  betrachtet wird, welche durch die Beziehung

$$T_H = \frac{N_H}{C} \quad (2.2,1)$$

gegeben ist.

Im allgemeinen sind  $N_A(t)$ ,  $T_A(t_i)$  und  $N_H$  bzw.  $T_H$  Zufallsgrößen. Während  $N_A(t)$  von  $T_A(t_i)$  abhängig ist, können  $N_A(t)$  und  $N_H$  meist als voneinander unabhängig angesehen werden.

Eine Beschreibung der Last kann demzufolge durch eine Charakterisierung des stochastischen Prozesses  $\{N_A(t) : t \in [0, \infty)\}$  (Ankunftsprozeß) und durch Vorgabe der VF von  $T_H$  erfolgen. In vielen Fällen wird angenommen, daß der Ankunftsprozeß ein Poisson-Prozeß mit Ankunftsrate  $\lambda$  ist und daß die Bedienstungszeit  $\mu$ -exponentialverteilt ist. Für die mittlere Bedienstungszeit  $h$  gilt dann wegen (2.1,14b)

$$h := E[T_H] = \mu^{-1}, \quad (2.2,2a)$$

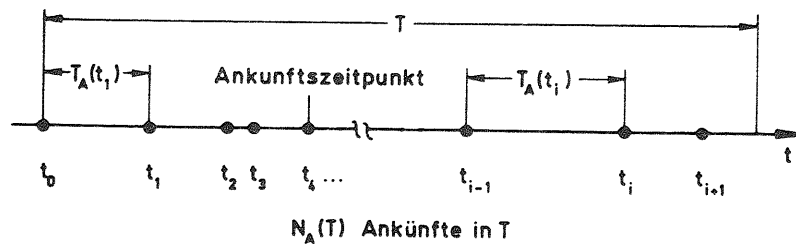


Bild 2.4: Zur Definition des Ankunftsprozesses

und für den mittleren Ankunftsabstand  $a$  gilt:

$$a := E[T_A] = \lambda^{-1}. \quad (2.2,2b)$$

$\mu$  heißt Bedienstungsrate.

### 2.2.3 Abfertigungs- und Prioritätsstrategien

Wie schon in der Einleitung erwähnt, ist neben Struktur und Last die Betriebsorganisation die dritte verkehrsbestimmende Komponente eines realen Systems. Sie muß bei der Modellierung berücksichtigt und in die Analyse mit einbezogen werden. In diesem Abschnitt sind daher einige Abfertigungs- und Prioritätsstrategien für Warteschlangenmodelle mit einer BE kurz beschrieben. Auf das im Zusammenhang mit Warteschlangennetzen auftretende Problem der Verkehrslenkung wird im Abschnitt 2.3 eingegangen werden.

Die Abfertigungsstrategie legt die Reihenfolge fest, in welcher die in einer Warteschlange stehenden Anforderungen abgearbeitet werden. Beispiele sind:

- i) FIFO (first in, first out): Die Anforderungen werden in der Reihenfolge ihres Eintreffens bedient. Diese Warteschlangenorganisation ist mit wenig Verwaltungsaufwand zu implementieren und wird wahrscheinlich am häufigsten angewendet.
- ii) LIFO (last in, first out): Die Anforderungen werden in der umgekehrten Reihenfolge ihres Eintreffens bedient. Dieses Verfahren findet z.B. in rechnergesteuerten Vermittlungssystemen als Abwehrmaßnahme in Überlastsituationen Anwendung [Forys 83].
- iii) LIFO preemptive resume (kurz: LIFO-PRE): Hierbei wird eine neu eintreffende Anforderung sofort zur Bedienung zugelassen. D.h., eine sich eventuell in Bedienung befindende Anforderung wird augenblicklich unterbrochen



und jeweils an den Anfang der Warteschlange eingeordnet. Ihre Bearbeitung wird später an der Stelle fortgesetzt, an welcher die Unterbrechung stattgefunden hat. Anwendungen für diese Strategie sind nicht bekannt. Sie ist hier nur im Hinblick auf das später zu besprechende BCMP-Theorem aufgeführt.

- iv) RR (round robin): Bei dieser Strategie wird allen wartenden Anforderungen die BE in zyklischer Reihenfolge für die Dauer einer vorgegebenen Zeitspanne (Zeitscheibe) so oft zur Verfügung gestellt, bis der jeweilige Bedienungswunsch erfüllt ist. Neu eintreffende Anforderungen werden am Ende der Warteschlange eingereiht. "round robin" und das nachfolgend beschriebene Verfahren "processor sharing" spielen eine gewichtige Rolle in Rechnersystemen mit "time sharing" Betrieb, wo eine Quantelung der zur Verfügung stehenden CPU-Zeit vorgenommen wird.
- v) PS (processor sharing): Dieses Verfahren geht aus dem RR hervor, indem die Zeitscheibe verschwindend klein gemacht wird (mathematisches Modell).

Prioritätsstrategien ermöglichen die bevorzugte Behandlung von Anforderungen, indem diese auf Grund ihrer Wichtigkeit, Dringlichkeit, Herkunft usw. bestimmten Prioritätsklassen zugeordnet werden, welche gewisse Unterbrechungseigenschaften aufweisen. Für die Klassenbezeichnung hat sich eine fortlaufende Nummerierung mit einem Klassenindex  $r$  ( $r=1,2,\dots,R$ ) eingebürgert. Der Index 1 kennzeichnet die Klasse höchster Priorität.  $R$  ist die Anzahl der Prioritätsklassen. Bezüglich der Unterbrechungseigenschaften kann man die folgenden Möglichkeiten unterscheiden:

- i) NONPRE (nonpreemptive priorities): Bei nichtunterbrechenden Prioritäten wird eine einmal begonnene Bedienung ohne Unterbrechung zu Ende geführt. Die Implementierung ist relativ einfach, da nach jedem Bedienungsende lediglich die Wartespeicher von Klasse eins aus beginnend nach der nächsten zu bedienenden Anforderung abgesucht werden müs-

sen. Letzteres läßt sich meist als Software-Maßnahme realisieren.

- ii) PRE (preemptive priorities): Bei unterbrechenden Prioritäten wird die Bearbeitung einer Anforderung beim Eintreffen einer Anforderung höherer Priorität sofort unterbrochen und erst dann weitergeführt, wenn keine Anforderungen höherer Priorität mehr im System sind. Je nachdem ob die Bearbeitung einer unterbrochenen Anforderung an der Unterbrechungsstelle fortgesetzt oder wieder von neuem begonnen wird, unterscheidet man die Varianten "PRE resume" oder "PRE restart". Bei der letzten Form differenziert man nochmals zwischen "PRE restart without resampling", was eine Erhaltung der Bedienungszeit bedeutet und zwischen "PRE restart with resampling", was bedeutet, daß die Bedienungszeit ihrer VF entsprechend durch ein Zufallsexperiment neu festgelegt wird. Die Implementierung von unterbrechenden Prioritäten ist meist aufwendiger als jene von nichtunterbrechenden Prioritäten und erfolgt i.a. durch Hardware-Maßnahmen.

Die Unterbrechungsstrategie kann u.U. klassenweise verschieden sein (Distanzprioritäten), was in dieser Arbeit jedoch ausgeschlossen wird. Abschließend sei bemerkt, daß die Warteschlangenorganisation innerhalb einer jeden Klasse nach einer festen Strategie, wie z.B. FIFO oder RR erfolgt.

#### 2.2.4 Zustands- und Ausgangsprozesse

Eine Beurteilung des Verkehrsverhaltens von Warteschlangenmodellen ergibt sich aus geeigneten Informationen über die in einem solchen Modell erzeugten stochastischen Prozesse. Dieses sind vor allem die Wartezeitprozesse, die Zustandsprozesse und die Ausgangsprozesse. Während für Wartezeitprozesse in einstufigen Warteschlangenmodellen eine Vielfalt von Ergebnissen bekannt ist, steht die Entwicklung von entsprechenden Analyseverfahren für Warteschlangennetze, selbst wenn keine Prioritätsklassen vorhanden sind, erst am Anfang. Aus diesem Grunde werden die Wartezeitprozesse in dieser Arbeit nicht behandelt. Sie

beschränkt sich vielmehr in wesentlichen Teilen auf die Betrachtung der nachfolgend erklärten Zustandsprozesse bei stationärem Verhalten.

Bezeichnet fortan die ZV  $X_r$ , ( $r=1, \dots, R$ ), stets die Anzahl von Anforderungen der Klasse  $r$  in einem Warteschlangenmodell mit  $R$  Klassen und  $t$  die Zeit, dann heißt der stochastische Prozeß  $\{X_r(t): t \in [0, \infty)\}$  mit dem Zustandsraum  $N_0$  der Zustandsprozeß der Klasse  $r$ . In Bild 2.5 ist eine Trajektorie eines solchen Prozesses wiedergegeben. Von praktischem Interesse sind vor allem die den jeweiligen Prozeß charakterisierenden stationären, zeitunabhängigen Zustandsverteilungen

$$P_r(x_r) := P\{X_r=x_r\}, \quad (r=1, \dots, R; x_r \in N_0), \quad (2.2,3a)$$

sowie die Verbundverteilung

$$P(x_1, \dots, x_R) := P\{X_1=x_1, \dots, X_R=x_R\}, \quad (r=1, \dots, R; x_r \in N_0), \quad (2.2,3b)$$

aus welchen die in Abschnitt 2.4 definierten Kenngrößen zur Leistungsbewertung berechnet werden können.

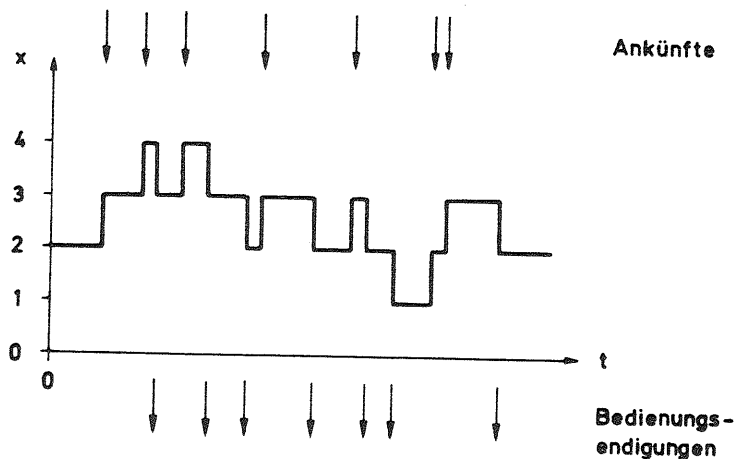


Bild 2.5: Trajektorie eines Zustandsprozesses

Als letztes seien die Ausgangsprozesse erklärt, deren Betrachtung weniger für die Beurteilung des Verkehrsverhaltens von einstufigen Warteschlangenmodellen als für die Analyse von Warteschlangennetzen eine Rolle spielt. Zur Vorbereitung der Definition dieser Prozesse werde die Folge der Zeitpunkte  $0 \leq t_{r1} < t_{r2} < \dots$ , an denen ein Bedienungsende der Klasse  $r$  stattfindet, betrachtet. Setzt man nun  $T_{Dr}(t_{r1}) := t_{r1}$  und  $T_{Dr}(t_{ri}) := t_{ri} - t_{r,i-1}$ , ( $i \in N, i \geq 2$ ), dann heißt der stochastische Prozeß  $\{T_{Dr}(t_{ri}): t_{ri} \in [0, \infty)\}$ , ( $r=1, \dots, R$ ), der Ausgangsprozeß der Klasse  $r$ . Bezieht man sich auf alle Ausgangsabstände  $T_D$  ohne Berücksichtigung der Klassenzugehörigkeit, so ist durch  $\{T_D(t_i): t_i \in [0, \infty)\}$  der allgemeine Ausgangsprozeß definiert.

Es sei zum Abschluß noch bemerkt, daß natürlich auch Ankunftsprozesse sowohl klassenweise als auch allgemein in entsprechender Weise erklärt sind. Der Bezug auf ein einklassiges System in Abschnitt 2.2.2 erfolgte aus Gründen der Übersichtlichkeit.

### 2.2.5 Kurzbezeichnungen

Um eine kurze und übersichtliche Bezeichnungsweise für die Vielfalt der einstufigen Warteschlangenmodelle zu haben, wird heute im allgemeinen das folgende von Kendall [Kendall 53] eingeführte und später erweiterte Notationssystem akzeptiert:

$$A/H/n-s/Y,Z, \dots$$

Darin bedeuten:

- A Kurzbezeichnung für die VF der Ankunftsabstände, z.B.:
  - M Exponentiell verteilte Ankunftsabstände
  - GI Allgemein verteilte und voneinander unabhängige Ankunftsabstände
- H Kurzbezeichnung für die Bedienungszeit-VF, z.B.:
  - M Exponentiell verteilte Bedienungszeiten
  - G Allgemein verteilte Bedienungszeiten

- n Anzahl der Bedienungseinheiten
- s Anzahl der Warteplätze. Diese wird nur angegeben, falls s beschränkt ist.
- X,Y,... Weitere Angaben zum Modell, wie z.B. die Abfertigungsdisziplin

Darüber hinaus sei vereinbart, daß für die VF der Ankunftsabstände ein  $\cdot$  geschrieben wird, wenn diese nicht bekannt ist. Bei mehrklassigen Modellen gilt allgemein die Übereinkunft, daß sich die jeweiligen Symbole auf alle Klassen beziehen, falls dies nicht ausdrücklich anders vermerkt ist. Insbesondere seien die Ankunftsprozesse voneinander unabhängig. Auf weitere Besonderheiten wird an gegebener Stelle hingewiesen. Schließlich sei bemerkt, daß die oben vereinbarten Abkürzungen auch bei der zeichnerischen Darstellung von Warteschlangenmodellen verwendet werden, wie dies in Bild 2.6 am Beispiel des zweiklassigen Prioritätsmodells M/M/1/NONPRE mit Ankunfts-raten  $\lambda_r$  und Bedienungsraten  $\mu_r$ , ( $r=1,2$ ), veranschaulicht ist. Bei einer FIFO-Warteschlangenorganisation wird diese Bezeichnung im weiteren nicht mehr explizit angegeben.

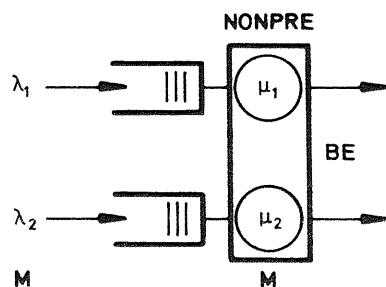


Bild 2.6: Zweiklassiges Warteschlangenmodell M/M/1/NONPRE

### 2.3 Warteschlangenmodelle mit mehrstufiger Bedienung

#### 2.3.1 Offene Warteschlangennetze

In der Praxis kommt es häufig vor, daß zur Erfüllung eines Bedienungswunsches mehrere Stationen durchlaufen werden. Dies ist z.B. bei einem Datenpaket der Fall, welches, um vom Sender zum Empfänger zu gelangen, über mehrere Leitungsabschnitte übermittlelt wird. Jeden Leitungsabschnitt kann man dann als BS auffassen, und auf diese Weise entsteht durch die Verknüpfung von mehreren einstufigen Warteschlangenmodellen ein Warteschlangennetz (WSN).

Besitzt ein WSN mindestens einen Eingang (Verkehrsquelle) und mindestens einen Ausgang (Verkehrssenke), so spricht man von einem offenen WSN. Die durch äußere Verkehrsquellen erzeugten Ankunftsprozesse werden als externe Ankunftsprozesse oder als Quellankunftsprozesse bezeichnet und müssen für die Analyse des WSNe ausreichend charakterisiert sein. Bei Kennzeichnung der einzelnen BSen mit einem Index  $i=1,2,\dots,N$ , bietet es sich an, die Struktur eines WSNe durch einen Graphen zu beschreiben, in welchem die BSen durch Knoten und die gerichteten Verbindungen zwischen den BSen durch Pfeile symbolisiert sind (Bild 2.7a). Jede einzelne BS wird wiederum durch Angaben über ihre Bedienungzeit-VF sowie die jeweils angewendeten Abfertigungs- und Prioritätsstrategien charakterisiert. Schließlich ist noch auf die Beschreibung der Verkehrslenkung (Verzweigungsstrategie) einzugehen, welche in diesem und dem folgenden Abschnitt zunächst nur für einklassige WSNe behandelt wird. Sie kann in vielen praktisch vorkommenden Fällen durch die Angabe einer Verzweigungsmatrix P erfolgen, deren Elemente durch die folgenden, voneinander unabhängigen Verzweigungswahrscheinlichkeiten definiert sind

$$p_{ij} := P\{\text{Anforderung geht nach dem Verlassen von BS } i \text{ nach BS } j\}, \quad (2.3,1a)$$

mit

$$0 \leq p_{ij} \leq 1, \quad (i, j=0, \dots, N), \quad (2.3,1b)$$

und

$$\sum_{j=0}^N p_{ij} = 1, \quad (i=0, \dots, N). \quad (2.3,1c)$$

Dabei sind aus formalen Gründen Verkehrsquellen und -senken einem Knoten "0" (Außenwelt) zugeordnet. Für den Fall, daß alle Knoten im Sinne der Definition aus Abschnitt 2.1.2 miteinander verbunden sind, definiert  $\underline{p}$  eine irreduzible, homogene M-K mit dem Zustandsraum  $E=\{0, \dots, N\}$ .

Ein wichtiger Sonderfall der offenen WSNe sind die seriellen Warteschlangenmodelle, bei welchen  $p_{i,i+1}=1$ , für  $i=0, \dots, N-1$ , und  $p_{N0}=1$  gilt.

### 2.3.2 Geschlossene Warteschlangennetze

Geschlossene WSNe unterscheiden sich von den offenen dadurch, daß sie weder einen Ein- noch einen Ausgang besitzen (Bild 2.7b). Ferner ist die Anzahl  $M$  (Population) der in ihnen zirkulierenden Anforderungen stets konstant. Für ihre Beschrei-

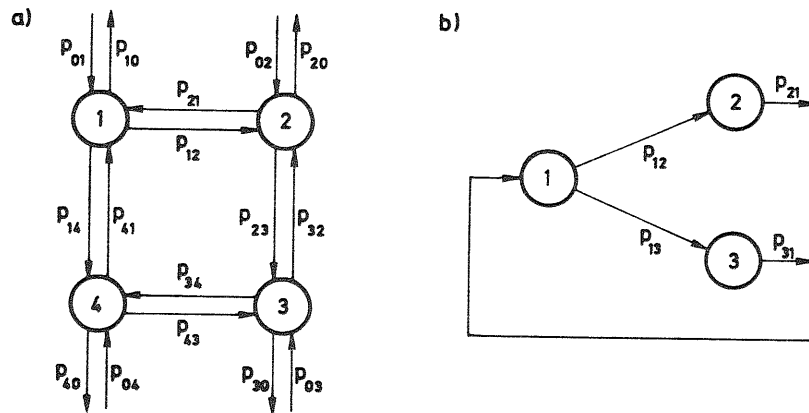


Bild 2.7: a) Graph eines offenen Warteschlangennetzes  
b) Graph eines geschlossenen Warteschlangennetzes

bung gilt das im vorherigen Abschnitt Gesagte entsprechend. wenn man diejenigen Aussagen ausschließt, welche Verkehrsquellen und Verkehrssenken betreffen. Damit erübrigen sich insbesondere Verzweigungswahrscheinlichkeiten des Typs  $p_{i0}$  bzw.  $p_{0j}$ , was bei Aussagen entsprechend den Beziehungen (2.3,1a-c) zu berücksichtigen ist. Den seriellen, offenen Warteschlangenmodellen entsprechen zyklische, geschlossene Warteschlangenmodelle, bei welchen  $p_{i,i+1}=1$ , für  $i=1, \dots, N-1$ , und  $p_{N1}=1$  gilt. Geschlossene WSNe treten häufig bei der Modellierung von Rechensystemen auf.

### 2.3.3 Mehrklassige Warteschlangennetze

Das Klassenkonzept in Warteschlangennetzen verfolgt zwei unterschiedliche Ziele:

- i) Die Modellierung der vom zugrunde liegenden realen System induzierten Klasseneinteilung aufgrund unterschiedlicher Eigenschaften der Anforderungen. Diese können z.B. durch verschiedene Bedienungszeit-VFen und/oder durch die jeweilige Zugehörigkeit zu einer Prioritätsklasse gegeben sein.
- ii) Eine möglichst genaue Modellierung der Verkehrslenkung in einem System, wie z.B. das n-malige Durchlaufen einer bestimmten Schleife, durch ausschließlich im Modell vorgesehene, jedoch nicht im realen System auftretende Klassenwechsel. Näheres hierzu findet man z.B. in [Sauer, Chandy 81].

In dieser Arbeit steht im Hinblick auf die Prioritätsklasseneinteilung zwar das erste Ziel im Vordergrund, dennoch erlauben die später entwickelten Analyseverfahren auch die Berücksichtigung des zweiten Ziels.

Bei mehrklassigen WSNe werden die zur vollständigen Beschreibung erforderlichen Angaben, soweit dieses notwendig ist, klassenabhängig gemacht. Dabei haben sich im Bereich der Verzwei-

ungsstrategie noch einige Begriffe eingebürgert, welche über die Erklärungen von Abschnitt 2.3.1 hinausgehen und im folgenden erläutert werden.

Die Struktur des WSNes sei durch einen Graphen mit  $N$  Knoten gegeben, und die Anzahl der Auftragsklassen sei  $R$ . Verkehrsquellen und -senken seien durch den Knoten "0" repräsentiert. Zur Beschreibung der Verkehrslenkung wird eine, durch die voneinander unabhängigen Wahrscheinlichkeiten

$$P_{ir,js} := P\{\text{Anforderung wechselt von BS } i \text{ nach BS } j; \text{ dabei gleichzeitiger Wechsel von Klasse } r \text{ nach Klasse } s\},$$

$$(i, j=0, \dots, N; r, s=1, \dots, R), \quad (2.3, 2a)$$

definierte, 4-dimensionale Verzweigungsmatrix  $\underline{P}$  benutzt, in welcher  $p_{0r,js}=0$  für  $r \neq s$  gesetzt ist. Es gilt wiederum

$$0 \leq p_{ir,js} \leq 1, \quad (i, j=0, \dots, N; r, s=1, \dots, R), \quad (2.3, 2b)$$

sowie

$$\sum_{j=0}^N \sum_{s=1}^R p_{ir,js} = 1, \quad (i=0, \dots, N; r=1, \dots, R). \quad (2.3, 2c)$$

Für die weiteren Betrachtungen sind zwei Fälle zu unterscheiden.

i) Es finden keine Klassenwechsel statt.

Dann gilt  $p_{ir,js}=0$  für  $r \neq s$ , und  $\underline{P}$  kann in  $R$  Untermatrizen zerlegt werden, von denen jede eine irreduzible, homogene M-K mit dem Zustandsraum  $E_r \subseteq \{(0,r), (1,r), \dots, (N,r)\}$ , ( $r=1, \dots, R$ ), definiert, sofern die jeweils involvierten Zustände  $(i,r)$  miteinander verbunden sind. In geschlossenen WSNen ohne Klassenwechsel kann für jede Klasse eine konstante Population  $M_r$  angegeben werden.

ii) Es finden Klassenwechsel statt.

In diesem Falle wird durch  $\underline{P}$  eine homogene M-K mit den Zuständen  $(i,r)$ , ( $i=0, \dots, N; r=1, \dots, R$ ), definiert. Deren Zustandsraum  $E$  kann entsprechend Abschnitt 2.1.2 in Äquivalenzklassen zerlegt werden, indem man jeweils alle miteinander verbundenen Zustände in einer Teilmenge  $E_k \subseteq E$ , ( $k=1, \dots, K; 1 \leq k \leq R$ ), zusammenfaßt.

Die Äquivalenzklassen  $E_k$  werden in der Literatur als Teilketten oder kürzer als Ketten bezeichnet. Enthält eine Kette Zustände des Typs  $(0,r)$ , so spricht man von einer offenen Kette, anderenfalls von einer geschlossenen Kette. Die Anzahl  $M_k$  der zu einer geschlossenen Kette gehörenden Anforderungen ist konstant und wird als Population der Kette bezeichnet. Ein offenes WSN enthält nur offene, ein geschlossenes nur geschlossene Ketten. WSNs mit offenen und geschlossenen Ketten werden als gemischte WSNs bezeichnet. Durch Definition geeigneter Ketten kann für die Anforderungen einer bestimmten Klasse der Weg durch ein WSN vorgeschrieben werden. Das Kettenprinzip ist natürlich auch auf die Netztypen ohne Klassenwechsel anwendbar.

Abschließend sei bemerkt, daß man aufgrund von Regelmäßigkeiten im betrachteten WSN oft mit weniger als vier Indices zur Kennzeichnung der Verzweigungswahrscheinlichkeiten auskommt.

#### 2.4 Kenngrößen zur Leistungsbewertung

In diesem Abschnitt sollen die in der Einleitung erwähnten Kenngrößen zur Leistungsbewertung definiert werden, welche für die realen Systeme in gleicher Weise wie für deren Modelle erklärt sind.

a) Einstufige Wartemodelle

Wartezeit  $T_w$ :

Die zufällige Zeitspanne zwischen der Ankunft in der Warte-

schlange bis zum (erstmaligen) Bedienungsbeginn wird als Wartezeit  $T_W$  bezeichnet. Mittelwert  $w := E[T_W]$ .

Verweilzeit  $T_F$ :

Hierunter versteht man die zufällige Zeitspanne zwischen der Ankunft in der Warteschlange und dem Bedienungsende. Mittelwert  $f := E[T_F]$ .

Anzahl von Anforderungen im System X:

Bedarf keiner weiteren Erklärung. Mittelwert  $\bar{x} := E[X]$ .

Blockierungswahrscheinlichkeit B:

Dieses ist die Wahrscheinlichkeit dafür, daß eine eintreffende Anforderung keinen freien Platz mehr im WS vorfindet und zurückgewiesen wird.

Auslastung  $\rho$ :

Die Auslastung einer BE ist durch die Beziehung

$$\rho := \frac{\text{mittlere Bedienungszeit}}{\text{mittleren Ankunftsabstand}} = \frac{h}{a} = \frac{\lambda}{\mu} \quad (2.4,1)$$

erklärt und gibt für die Modelle mit einer BE an wie lange diese während einer genügend langen Beobachtungszeit belegt war. Zwischen der Auslastung  $\rho$  und der Wahrscheinlichkeit  $P\{X=0\}=p(0)$  besteht der Zusammenhang (siehe z.B. [Kleinrock 75]):

$$\rho = 1-p(0). \quad (2.4,2)$$

Während man die Größen B und  $\bar{x}$  direkt aus den Zustandswahrscheinlichkeiten bestimmen kann, ist dies für w und f nicht möglich. Zwischen der mittleren Anzahl  $\bar{x}$  von Anforderungen in einer BS und der mittleren Verweilzeit f einer beliebigen Anforderung in der BS besteht jedoch nach [Little 61] der folgende einfache Zusammenhang

$$\bar{x} = \lambda \cdot f. \quad (2.4,3)$$

Diese vom Typ des Ankunftsprozesses und der VF für die Bedienungszeiten unabhängige Beziehung gilt nur für die Mittelwerte der betroffenen ZVen.

b) Offene Warteschlangennetze

Bei einem offenen WSN können die obigen Definitionen sowohl für jede einzelne BS als auch für das gesamte WSN angewendet werden. Im letzteren Sinne ist vor allem die mittlere Verweilzeit einer Anforderung im Netz von Interesse.

c) Geschlossene Warteschlangennetze

Im Falle geschlossener WSNe sind die oben gegebenen Definitionen ebenfalls für jede BS sinnvoll anwendbar. Eine Ausnahme bildet nur die Blockierungswahrscheinlichkeit B, welche wegen der festen Population im Netz durch ausreichende Dimensionierung der WS immer zu Null gemacht werden kann. Als zusätzliche Leistungsgröße interessiert man sich in geschlossenen WSNe oft für den

Durchsatz  $\lambda$  in einer BS:

Hierunter ist die mittlere Anzahl von Anforderungen zu verstehen, welche in einem Zeitintervall fertig bedient werden, geteilt durch die Länge dieses Zeitintervalls.

Der Durchsatz ist eine von der Netzpopulation abhängige Größe und entspricht der in offenen Warteschlangenmodellen definierten Ankunftsrate  $\lambda$ , woraus sich auch die gleiche Bezeichnung für beide Größen erklärt. Die Beziehungen (2.4,1) bis (2.4,3) gelten in geschlossenen WSNe entsprechend.

Bezüglich der Leistungsgrößen, die das gesamte WSN betreffen, ist die Zykluszeit  $T_F$  zu nennen, welche der mittleren Verweilzeit  $T_F$  in offenen WSNe entspricht.

Zykluszeit  $T_F$ :

Diese ist wie folgt erklärt: Betrachtet man eine beliebige

Stelle innerhalb des WSNes (z.B. den Eingang einer BS) und mißt die Zeitspanne, welche vergeht bis eine ausgewählte Anforderung (Testanforderung) nach dem Verlassen dieser Stelle wieder dort eintrifft, so heißt diese zufällige Zeitspanne die Zykluszeit  $T_F$  der Anforderung bezüglich dieser Stelle. Mittelwert  $f := E[T_F]$ .

Liegen Warteschlangenmodelle mit mehreren Auftragsklassen vor, so sind die aufgeführten Kenngrößen zur Leistungsbewertung in gleicher Art klassenweise erklärt.

## 2.5 Beispiele für die Modellierung von Kommunikations- und Rechnersystemen

### 2.5.1 Einige Gesichtspunkte zur Modellbildung

Bei der Erstellung von Warteschlangenmodellen zum Zwecke der Verkehrsanalyse von realen Systemen sind verschiedene Ziele in Einklang zu bringen, von denen einige, soweit sie hier von Interesse sind, kurz diskutiert werden sollen.

Als wichtigstes Ziel mag in vielen Fällen eine möglichst genaue Wiedergabe der realen Verhältnisse (Wirklichkeitstreue) durch eine weitestgehend exakte Modellierung von Systemstruktur, Last und Betriebsorganisation erscheinen. Solche detaillierten Modelle können jedoch mit analytischen Methoden, besonders im Falle von Warteschlangennetzen, kaum oder überhaupt nicht gelöst werden, und man muß auf das meist kostenaufwendige und umständliche Hilfsmittel der Simulation zurückgreifen. Aus dieser Feststellung resultiert ein zweites, wichtiges Kriterium für ein Verkehrsmodell, nämlich dessen Zugänglichkeit für analytische Lösungen. Zur Erfüllung dieses Kriteriums bietet es sich zum einen an, gewisse Abstriche an der Wirklichkeitstreue des Modells zuzugestehen, indem man z.B. die statistischen Eigenschaften des realen Systems durch mathematisch möglichst einfach zu behandelnde VFn nur näherungsweise beschreibt oder gewisse idealisierende Annahmen, wie etwa vernachlässigbar kleine Verwaltungszeiten, macht. Zum anderen erweist es sich oft als sinnvoll, auf eine zu große Modellie-

rungstiefe zu verzichten. D.h., man bildet nicht die in den einzelnen Teilsystemen vorhandenen Feinstrukturen detailliert ab, sondern beschreibt vielmehr ganze Systemblöcke durch ein Warteschlangenmodell. Umgekehrt ist es auch möglich, Teilsysteme bis in Einzelheiten hinein zu modellieren und die an den Schnittstellen wirksame "Umwelt" global zu beschreiben.

In vielen Fällen liefern analytische Lösungen, auch wenn sie unter stark vereinfachenden Annahmen erzielt werden, wertvolle Einblicke in grundsätzliche Zusammenhänge zwischen Ein- und Ausgabeparametern und decken Phänomene auf, welche bei Simulationsstudien verborgen bleiben. Weiterhin reichen sie oft aus, um beim Vorliegen alternativer Vorschläge in der Phase eines Systementwurfs deren Vor- und Nachteile für das Verkehrsverhalten des Systems abzuwägen und damit eine Entscheidungshilfe zu geben.

Mit einer bei der Modellierung von Systemen mit mehrstufiger Bedienung vielfach anzutreffenden Annahme, welche die analytische Behandlung entsprechender WSNe wesentlich erleichtert oder im allgemeinen sogar erst möglich macht, beschäftigt sich der nächste Abschnitt.

### 2.5.2 Die Unabhängigkeitsannahme von KLEINROCK

In technischen Systemen mit mehrstufiger Bedienung ist es meistens der Fall, daß sich die zur Erfüllung eines Bedienungswunsches erforderliche Bedienungszeit bei der Generierung des Auftrages in der Verkehrsquelle einmal zufällig ergibt und dann auf dem Wege durch die einzelnen Bedienstufen erhalten bleibt. So liegt z.B. die Länge eines Datenpaketes nach der Erzeugung im Sender fest und ändert sich bei der Übermittlung zum Empfänger über ein Datenpaketvermittlungnetz nicht mehr. Bei dem Versuch, ein entsprechendes WSN, welches diesen Sachverhalt exakt berücksichtigt, analytisch zu untersuchen, stößt man im allgemeinen auf sehr große Schwierigkeiten. Das liegt daran, daß sowohl die Bedienungszeiten in aufeinanderfolgenden Stationen als auch die Ankunftsabstände und Bedienungszeiten in ein und derselben Station kei-

ne voneinander unabhängigen ZVn mehr sind. Einen Ausweg aus dieser Situation liefert die "Unabhängigkeitsannahme von Kleinrock". Diese ist für Nachrichtennetze mit "genügend großer Vermaschung", Poisson-Verkehrsquellen und  $\mu$ -exponential-verteilten Nachrichtenlängen in [Kleinrock 64] wie folgt erklärt:

"Zu jedem Ankunftszeitpunkt einer Nachricht in einem Netzknoten wird für diese Nachricht eine neue Länge  $v$  entsprechend der Verteilungsdichtefunktion  $p(v)=\mu \cdot \exp(-\mu v)$  ermittelt."

Die Rechtfertigung dieser Annahme ist in dem genannten Buch begründet und dort durch ausführliche Simulationsstudien bestätigt. Ihre Anwendung bei der Modellbildung von mehrstufigen Bedienungssystemen zum Zwecke der Zustandsanalyse führt auf WSNe mit einer sogenannten Produktlösungsform, welche später noch ausführlich besprochen werden. Die Voraussetzung einer "genügend großen Vermaschung" ist in [Kleinrock 75] dahingehend präzisiert, daß die meisten BSen mehr als nur einen Zu- und einen Abgang haben sollten. D.h., es sind die rein seriellen Systeme ausgeschlossen.

2.5.3 Erstes Beispiel: Kommunikationssystem mit Ringstruktur

Das erste hier vorgestellte Beispiel zur Modellbildung ist aus [Heger 83] entnommen. Dort wird u.a. die Leistungsfähigkeit eines als RDC-Ring bezeichneten Kommunikationssystems mit Hilfe von Simulationsstudien bewertet. Dieses System wurde am Fraunhofer-Institut für Informations- und Datenverarbeitung (IITB) in Karlsruhe entwickelt und dient dem Informationsaustausch in einem Echtzeitrechnersystem mit verteilten Mikroprozessoren. Über den prinzipiellen Aufbau des RDC-Ringes gibt Bild 2.8a Auskunft.

Die durch optische Wellenleiter ringförmig miteinander verbundenen Teilnehmerstationen  $T_i$ , ( $i=1, \dots, N$ ), sind alle nach gleichem Muster aufgebaut und enthalten jeweils eine Sende- und Empfangseinrichtung  $S_i$  bzw.  $E_i$ , einen nach dem FIFO-Prin-

zip organisierten Pufferspeicher  $P_i$  sowie einen Teilnehmerrechner  $TR_i$ . Ein Nachrichtenaustausch zwischen zwei nicht benachbarten Stationen kann nur indirekt über die dazwischen liegenden Stationen erfolgen. Jede Station befindet sich demnach in einem der Zustände "Warten", "Weiterleiten einer fremden Nachricht" oder "Senden einer eigenen Nachricht". Weiterzuleitende Nachrichten werden nach dem Empfang stets in  $P_i$  zwischengespeichert bevor sie über  $S_i$  die Station verlassen. Sie haben immer nichtunterbrechende Priorität gegenüber den Sendewünschen von  $TR_i$ , womit u.a. der möglichst zügigen Abwicklung einer einmal begonnenen Übertragung Rechnung getragen wird.

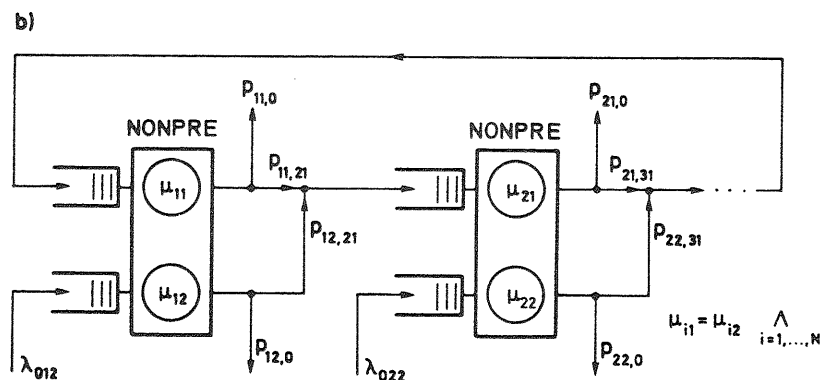
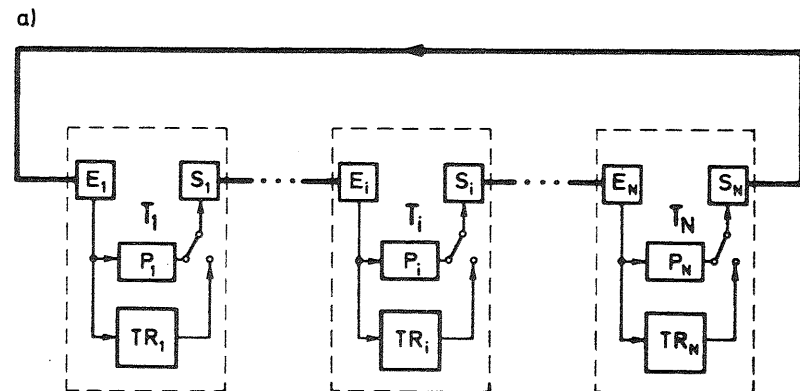


Bild 2.8: a) Schematische Darstellung des RDC-Ringes  
b) Warteschlangenmodell für den RDC-Ring



Für das beschriebene System ist in [Heger 83] das in Bild 2.8b dargestellte WSN angegeben. Darin wird jede Station  $T_i$  durch ein zweiklassiges Warteschlangenmodell mit nichtunterbrechenden Prioritäten repräsentiert, in welchem Sendeeinrichtung und Übertragungsmedium als BE, der Pufferspeicher  $P_i$  als WS für die erste Klasse und der im Teilnehmerrechner  $TR_i$  enthaltene Speicher für die Sendewünsche als WS für die zweite Klasse mit jeweiliger FIFO-Abfertigungsstrategie modelliert sind. Etwaige in den Stationen auftretende Verzögerungszeiten, wie sie z.B. durch die Auswertung der Zielinformation u.a.m. entstehen, können ebenso wie die Kanallaufzeiten als klein gegenüber den Übertragungszeiten angesehen werden und sind im Modell vernachlässigt. Die Teilnehmerrechner stellen als Erzeuger und Abnehmer von Nachrichten sowohl Verkehrsquellen als auch Verkehrssenken dar. Alle BSen sind der Ringstruktur des Systems entsprechend miteinander verbunden. Die Verkehrslenkung wird durch voneinander unabhängige Verzweigungswahrscheinlichkeiten  $p_{ir,i+1}$ ,  $p_{Nr,0s}$  und  $p_{ir,0}$ , ( $i=1, \dots, N-1$ ), beschrieben. Letzteres gibt die Wahrscheinlichkeit für das Verlassen des Ringes an. Die Klassenindices  $r, s \in \{1, 2\}$  sind zur Beschreibung der Klassenwechsel notwendig. Für die Modellierung der Last ist zum einen angenommen, daß die von den verschiedenen Teilnehmerrechnern herrührenden Folgen der Sendewünsche als Poisson-Prozesse mit Parametern  $\lambda_{0i2}$  angesehen werden können und zum anderen, daß die Länge der Nachrichten, bei überall gleicher Übertragungskapazität der Leitungsabschnitte,  $\mu_{ir}$ -exponentialverteilten Übertragungszeiten entspricht. Durch Übernahme der im Abschnitt 2.5.2 besprochenen Unabhängigkeitsannahme ergibt sich dann für jede BS ebenfalls eine  $\mu_{ir}$ -exponentialverteilte Bedienungszeit-VF mit  $\mu_{i1} = \mu_{i2}$ , ( $i=1, \dots, N$ ).

Bei dem besprochenen Modell handelt es sich um ein WSN mit einer zweiklassigen offenen Kette. Durch Bestimmung der stationären Zustandsverteilung und den daraus resultierenden Mittelwerten für die Verweilzeiten in den BSen bzw. im WSN kann eine Beurteilung des RDC-Ringes bezüglich seiner verkehrsmäßigen Leistungsfähigkeit erfolgen.

#### 2.5.4 Zweites Beispiel: "Multiprogramming" in Rechnern mit Speicherhierarchie

Das zweite hier zu besprechende Beispiel für ein Warteschlangenmodell ist allgemein unter der Bezeichnung "Central Server Model" (CSM) bekannt und wurde in seiner ursprünglichen Form von Buzen [Buzen 71] entwickelt. Es beschreibt den Mechanismus des "Multiprogrammings" (Mehrprogrammbetrieb) in Rechnern mit Speicherhierarchie durch ein geschlossenes WSN. Zum besseren Verständnis des Modells sei kurz auf die Konzepte der Speicherhierarchie und des Multiprogrammings eingegangen.

Die Einführung hierarchischer Speicherstrukturen in Rechnern ist vor allem eine Folge der Diskrepanz zwischen den ständig wachsenden Arbeitsgeschwindigkeiten der Prozessoren und den vergleichsweise stagnierenden Zugriffszeiten für Massenspeicher. Sie beinhaltet die Implementierung und Verwaltung von Speichertypen unterschiedlicher Kapazität und Zugriffszeit derart, daß bei der Ausführung von Programmen die aktuellen Programmteile und Datensätze (kurz: Daten) nach Möglichkeit in den unteren, schnelleren Speicherebenen verfügbar sind. Im wesentlichen unterscheidet man:

- i) Schnellspeicher (SSP) mit niedrigen Zugriffszeiten und kleiner Kapazität (Größenordnungen: nsek, kByte).
- ii) Arbeitsspeicher (ASP) mit erhöhten Zugriffszeiten und mittelgroßer Kapazität (Größenordnungen: nsek, MByte).
- iii) Hintergrundspeicher (HSP) mit hohen Zugriffszeiten und sehr großer Kapazität (Größenordnungen: msek, GByte).

Der SSP bildet zusammen mit einem Befehls- und Rechenwerk den Zentralprozessor (CPU), welcher direkten Zugriff auf den ASP hat. Demgegenüber erfolgt der Datentransfer zwischen ASP und HSP, auf Anstoß durch das Betriebssystem, unter der Regie eines separaten Ein-/Ausgabe-Prozessors (EAP). CPU, ASP, EAP sowie die für den Datenaustausch notwendigen Übertragungskanäle sind die wesentlichen Bestandteile der Zentraleinheit (ZE).

Jeder der genannten Speichertypen ist in kleine Bereiche fester Größe, sogenannte Seitenrahmen unterteilt, deren Inhalt bei Bedarf zwischen den Ebenen transferiert werden kann. Dieses ist z.B. notwendig, wenn bei der Ausführung eines Programms ein Seitenfehler auftritt, womit gemeint ist, daß sich eine im ASP benötigte Datenmenge im HSP befindet. Da die für einen solchen Datentransfer in Anspruch genommene Zeitspanne vergleichsweise groß ist, bietet es sich beim Auftreten eines Seitenfehlers an, mit der Bearbeitung anderer, im ASP residenter Programme durch die CPU fortzufahren. Man bezeichnet diese Form der Parallelarbeit als Multiprogramming und die Anzahl  $M_r$ , ( $r=1, \dots, R$ ), der sich gleichzeitig in Bearbeitung befindlichen Programme der Klasse  $r$ , als Multiprogramminggrad der Klasse  $r$ . Fragestellungen, die sich sowohl bei der Entwicklung neuer als auch bei der Abstimmung bestehender Rechensysteme ergeben, betreffen u.a. die Wahl von  $M_r$ , um bei vorgegebenen Randbedingungen (Architektur, mittleren Bedienungszeiten usw.) einerseits bestimmte Antwortzeiten einzuhalten und andererseits eine möglichst hohe Ausnutzung der CPU zu gewährleisten. Eine weitgehende Beantwortung solcher Fragen ergibt sich aus der Analyse des nachfolgend beschriebenen CSMs.

Die hier betrachtete Variante des CSMs ist in Bild 2.9 dargestellt und beschreibt als Besonderheit eine Vergabe der CPU nach Prioritäten (PRE oder NONPRE). Dadurch wird einer bevorzugten Ausführung von zeitkritischen Programmen (1. Klasse) gegenüber solchen, die nicht die Einhaltung einer bestimmten Reaktionszeit erfordern (2. Klasse), Rechnung getragen. Bei dem Modell handelt es sich um ein geschlossenes WSN mit zwei einklassigen Ketten der Population  $M_r$ , ( $r=1,2$ ). Im Gegensatz zum vorherigen Beispiel, wo eine direkte gerätetechnische Interpretation der Komponenten BS und WS ebenso leicht möglich war wie eine übereinstimmende Lokalisierung von Anforderungen in Modell und System, findet die Modellbildung in diesem Beispiel auf einer abstrakteren Ebene statt, welche solche Betrachtungen nur noch bedingt zuläßt.

Die BS 1 repräsentiert die CPU sowie die vom Betriebssystem erstellten Warteschlangen, welche die zur Ausführung klassen-

weise eingereichten Programme (Anforderungen) enthalten. In die zugehörigen Bedienungszeiten  $T_{H1r}$  sind die Bearbeitungszeiten für die Programme mit den auftretenden Verwaltungszeiten zusammengefaßt. Durch die BSen 2, ..., N werden die Seitentransfers zwischen ASP und HSP sowie die Ein-/Ausgabe von Massendaten modelliert. Eine Prioritätsstrategie ist in diesen BSen nicht vorgesehen, was meistens den praktischen Gegebenheiten entspricht. Die Bedienungszeiten  $T_{Hir}$ , ( $i=2, \dots, N$ ;  $r=1,2$ ), berücksichtigen im wesentlichen die Zugriffs- und die Übertragungszeiten, und es ist angenommen, daß ihre Mittelwerte klassenunabhängig sind. Mit Hilfe der Verzweigungswahrscheinlichkeiten  $P_{ir}$ , ( $i=1, \dots, N$ ;  $r=1,2$ ) sind die relativen Häufigkeiten ausgedrückt, mit denen für einen Seitentransfer oder eine Ein-/Ausgabeoperation das betreffende Betriebsmittel belegt wird, bzw. mit der die Ausführung eines Programmes beendet wird. Zu letzterem ist zu bemerken, daß wegen der geschlossenen Netzstruktur von der sofortigen Ersetzung des abgearbeiteten Programms ausgegangen wird (konstanter Multiprogramminggrad). Schließlich seien alle Warteschlangen nach dem FIFO-Prinzip

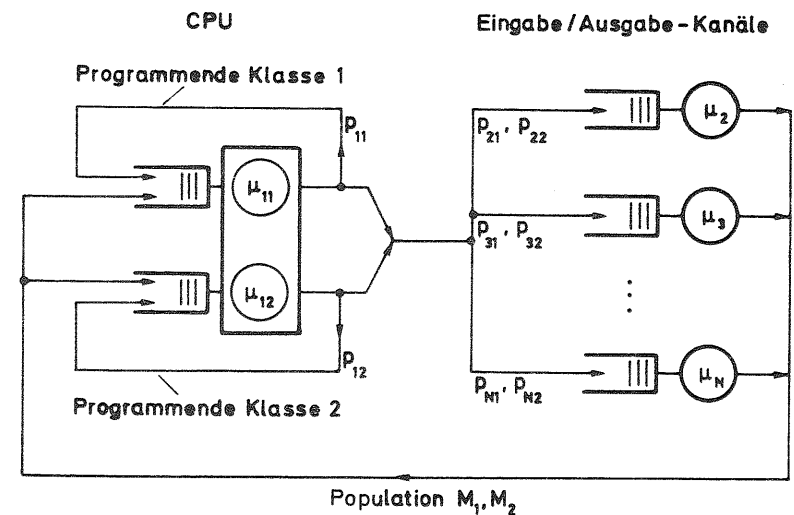


Bild 2.9: "Central Server Model" mit Prioritätsabfertigung in der CPU

organisiert, und die Bedienungszeiten seien sämtlich exponentiell verteilt. Diese Annahmen machen zusammen mit jener der klassenunabhängigen mittleren Bedienungszeiten  $h_1$  in den BSen  $2, \dots, N$  das WSN einer mathematischen Analyse zugänglich.

Abschließend sei für detailliertere Beschreibungen der Konzepte "Speicherhierarchie" und "Multiprogramming" ebenso wie für eingehendere Diskussionen des CSMS auf die Literaturstellen [Buzen 71, Kobayashi 79, Stein 80, Wojtkowiak 80, Kühn 81] verwiesen.

### 3. ANALYSE VON WARTESCHLANGENMODELLEN

Für die mathematische Analyse von Warteschlangenmodellen steht eine Anzahl unterschiedlicher Methoden zur Verfügung, welche in wesentlichen Teilen auf Ergebnissen der Wahrscheinlichkeitstheorie basieren. Unter diesen Methoden hat sich vor allem das Absuchen des zu analysierenden Modells auf geeignete Markoff- bzw. Semi-Markoff-Prozesse [Kleinrock 75] als sehr erfolgreich erwiesen. Insbesondere ergeben sich die stationären Zustandswahrscheinlichkeiten für Warteschlangenmodelle mit Poisson-Ankunftsprozessen und exponentialverteilten Bedienungszeiten durch Lösen der Gleichgewichtsgleichungen (2.1, 12a-b), da die betreffenden Zustandsprozesse homogene M-Kn sind (Satz 2.3). Man spricht dann auch von "Markoffschen Warteschlangenmodellen". Zieht man noch in Betracht, daß man jede auf  $[0, \infty)$  konzentrierte VF für praktische Zwecke beliebig genau durch eine VF approximieren kann, welche aus "exponentiellen Phasen" zusammengesetzt ist [Schassberger 73], so kann die Klasse der mit Hilfe von M-Kn analysierbaren Warteschlangenmodelle beträchtlich erweitert werden. In der Praxis stößt dieses als "Phasenmethode" bekannte Verfahren jedoch dann auf Grenzen, wenn die Zahl der zu berücksichtigenden Zustände so groß ist, daß die Gleichgewichtsgleichungen auch mit numerischen Verfahren nicht mehr gelöst oder unter Umständen erst gar nicht mehr aufgeschrieben werden können. Dieses zentrale Problem des Auffindens geschlossener Lösungen der Gleichgewichtsgleichungen bzw. die Bereitstellung geeigneter numerischer und/oder approximativer Verfahren wird in der Warteschlangentheorie in vielfältiger Weise angegangen und bildet auch den Kernpunkt der vorliegenden Arbeit. Neben den in Lehrbüchern bevorzugten Transformationsmethoden spielen die Anwendung von Ergebnissen aus der Theorie der Differenzgleichungen ebenso wie die Anwendung numerischer Verfahren zur Lösung linearer Gleichungssysteme die größte Rolle bei der Lösung der Gleichgewichtsgleichungen. Neuere Methoden und Ergebnisse findet man in [Kelly 79, Neuts 81, Cohen, Boxma 83]. Von besonderer Bedeutung für diese Arbeit ist die in Abschnitt 3.2.1 angegebene Produktlösung für separable WSNe.

Außer den skizzierten Analysemethoden existiert noch eine Reihe anderer Verfahren, auf die hier nicht eingegangen wird. Die nachfolgenden Abschnitte dieses Kapitels enthalten einige für spätere Betrachtungen benötigte Ergebnisse für ein- und mehrstufige Warteschlangenmodelle.

### 3.1 Warteschlangenmodelle mit einstufiger Bedienung

#### 3.1.1 Warteschlangenmodell GI/G/1

Das Warteschlangenmodell GI/G/1 kann als ein klassisches Objekt der Warteschlangenlehre angesehen werden, welches in der Vergangenheit schon auf verschiedene Weisen mathematisch untersucht worden ist. Es seien diesbezüglich die Arbeiten [Lindley 52, Pollaczek 57, Schassberger 73] genannt. Zur Vorbereitung auf die weiter unten besprochenen Untersuchungen der Ausgangsprozesse in Prioritätsmodellen sind nachfolgend zwei Formeln für die Berechnung der mittleren Wartezeit  $w$  sowie des Variationskoeffizienten  $c_D$  der Ausgangsabstände  $T_D$  angegeben. Der Variationskoeffizient  $c$  einer beliebigen ZVn  $X$  ist durch die Beziehung

$$c^2 := \frac{E[X^2]}{(E[X])^2} - 1 \quad (3.1,1)$$

definiert und wird weiterhin je nach Bezugnahme auf  $T_A$ ,  $T_H$  oder  $T_D$  mit einem der Buchstaben A, H oder D indiziert. Für eine exponentialverteilte ZV folgt aus (2.1,14b) sofort, daß  $c = 1$  gilt.

Als erstes sei aus [Krämer, Langenbach-Belz 76] eine Formel zitiert, welche die approximative Berechnung von  $w$  in Abhängigkeit der beiden ersten Momente von  $T_A$  und  $T_H$  gestattet:

$$w \approx h \cdot \frac{\rho}{2(1-\rho)} \cdot (c_A^2 + c_H^2) \cdot g(\rho, c_A^2, c_H^2) \quad (3.1,2a)$$

mit

$$g(\rho, c_A^2, c_H^2) = \begin{cases} \exp \left[ -\frac{2(1-\rho)}{3\rho} \cdot \frac{(1-c_A^2)^2}{c_A^2 + c_H^2} \right], & c_A < 1 \\ \exp \left[ -(1-\rho) \cdot \frac{c_A^2 - 1}{c_A^2 + 4c_H^2} \right], & c_A \geq 1. \end{cases} \quad (3.1,2b)$$

Für die mittlere Verweilzeit  $f$  gilt:

$$f = h + w. \quad (3.1,3)$$

Die zweite Formel ist [Marshall 68] entnommen und erlaubt die exakte Berechnung von  $c_D$  gemäß

$$c_D^2 = c_A^2 + 2\rho^2 c_H^2 - 2\rho(1-\rho) \cdot \frac{w}{h}. \quad (3.1,4)$$

Durch Einsetzen von (3.1,2a) in (3.1,3) folgt

$$c_D^2 \approx c_A^2 + 2\rho^2 c_H^2 - \rho^2 (c_A^2 + c_H^2) g(\rho, c_A^2, c_H^2). \quad (3.1,5)$$

Zieht man in Betracht, daß im Modell GI/G/1 die Beziehung

$$E[T_A] = E[T_D] = \lambda^{-1} \quad (3.1,6)$$

gilt (keine Quellen und Senken im Innern), so liefert (3.1,5) eine Charakterisierung des Ausgangsprozesses durch die als bekannt vorauszusetzenden beiden ersten Momente von  $T_A$  und  $T_H$ . Es bleibt noch anzumerken, daß die Gleichungen für  $w$  und  $c_D$  nur dann sinnvolle Ergebnisse liefern, wenn die Stabilitätsbedingung

$$\rho < 1 \quad (3.1,7)$$

erfüllt ist. Diese Forderung besagt, daß über einen längeren Zeitraum gemessen im Mittel nicht mehr Anforderungen eintreffen dürfen als bedient werden können.

3.1.2 Warteschlangenmodell M/M/1/PRE

Die ersten Ergebnisse für das Warteschlangenmodell M/M/1/PRE mit zwei Klassen wurden für die Unterbrechungsstrategien "PRE-resume" und "PRE repeat with resampling"<sup>1)</sup> in [White, Christie 58] veröffentlicht und beruhen auf der Analyse der nachfolgend angegebenen Gleichgewichtsgleichungen für den stationären Zustandsprozeß<sup>2)</sup>:

$$0 = -(\lambda + \mu_1)p(m, n) + \lambda_1 p(m-1, n) + \mu_1 p(m+1, n) + \lambda_2 p(m, n-1), \quad (m, n \geq 1), \quad (3.1,8a)$$

$$0 = -\lambda p(0, 0) + \mu_1 p(1, 0) + \mu_2 p(0, 1) \quad (3.1,8b)$$

$$0 = -(\lambda + \mu_1)p(m, 0) + \lambda_1 p(m-1, 0) + \mu_1 p(m+1, 0), \quad (m \geq 1), \quad (3.1,8c)$$

$$0 = -(\lambda + \mu_2)p(0, n) + \lambda_2 p(0, n-1) + \mu_2 p(0, n+1) + \mu_1 p(1, n), \quad (n \geq 1), \quad (3.1,8d)$$

mit

$$\lambda := \lambda_1 + \lambda_2 \quad (3.1,8e)$$

und

$$p(m, n) := P\{X_1=m, X_2=n\}, \quad (m, n \in \mathbb{N}_0). \quad (3.1,8f)$$

Bild 3.1 zeigt den zugehörigen Übergangsgraphen.

Rekursive Lösungen der partiellen Differenzengleichung (3.1,8a) mit den Randbedingungen (3.1,8b-d) sind außer in der o.g. Literaturstelle noch in [Stephan 58, Marks 73, Miller 81] zu finden. Hier interessiert zunächst nur die mittlere Wartezeit  $w_r$  für eine Anforderung der Klasse r auf ihre erste Zulassung zur Bedienung:

1) In diesen Fällen erfolgt ein Zustandswechsel nach voneinander unabhängigen  $\mu_r$ - bzw.  $\lambda_r$ -exponentialverteilten Zeitspannen, ( $r=1,2$ ).

2) Wegen platzsparender Schreibweise werden, wo keine Unklarheiten zu befürchten sind, die Symbole m und n anstelle von  $x_1$  und  $x_2$  verwendet.

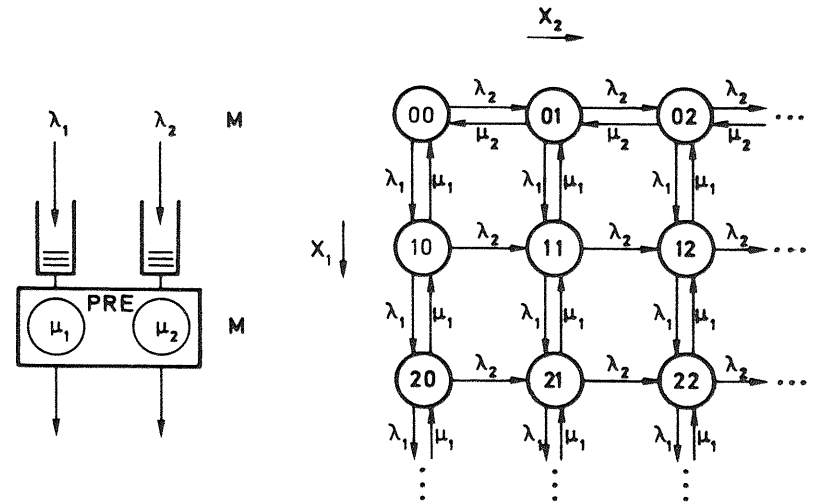


Bild 3.1: Übergangsgraph für den Zustandsprozeß im Warteschlangenmodell M/M/1/PRE

$$w_r = \frac{\sum_{i=1}^r h_i \rho_i}{(1-\rho_a) \cdot (1-\rho_a - \rho_r)}, \quad (r=1, \dots, R). \quad (3.1,9)$$

Darin wurde abgekürzt:

$$\rho_a := \sum_{i=1}^{r-1} \lambda_i h_i = \sum_{i=1}^{r-1} \rho_i, \quad \rho_i = \lambda_i h_i. \quad (3.1,10)$$

Die Beziehung (3.1,9) gilt auch für  $R > 2$  und liefert im Falle  $r=1$  das bekannte Ergebnis für das Modell M/M/1. Für einen stationären Zustandsprozeß muß die Bedingung

$$\rho = \sum_{r=1}^R \rho_r < 1 \quad (3.1,11)$$

erfüllt sein. Der Gleichung (3.1,6) des vorhergehenden Abschnitts entsprechend gelten in Prioritätsmodellen die

Beziehungen

$$E[T_{Ar}] = E[T_{Dr}] = \lambda_r^{-1}, \quad (r=1, \dots, R), \quad (3.1,12a)$$

$$E[T_A] = E[T_D] = (\lambda_1 + \lambda_2 + \dots + \lambda_R)^{-1}. \quad (3.1,12b)$$

Abschließend sei noch auf die Komplettierungszeit  $T_{Cr}$  einer Anforderung der Klasse  $r$  eingegangen. Diese ist für die Strategie "PRE resume" als die Zeitspanne zwischen dem ersten Bedienungsbeginn und dem endgültigen Bedienungsende einer solchen Anforderung erklärt. Sie enthält demnach außer der eigenen Bedienungszeit noch die Summe der Bedienungszeiten von unterbrechenden Anforderungen. Für den Erwartungswert  $E[T_{Cr}]$  und den Variationskoeffizienten  $c_{Cr}$  errechnet man [Conway, Maxwell, Miller 67]:

$$E[T_{Cr}] = \frac{h_r}{1 - \rho_a}, \quad (r=1, \dots, R) \quad (3.1,13a)$$

$$c_{Cr}^2 = 1 + \frac{2 \sum_{i=1}^{r-1} h_i \rho_i}{h_r (1 - \rho_a)}, \quad (r=1, \dots, R). \quad (3.1,13b)$$

Zusammen mit  $w_r$  folgt dann für die mittlere Verweilzeit:

$$f_r = w_r + E[T_{Cr}], \quad (r=1, \dots, R). \quad (3.1,14)$$

Wegen der "Gedächtnislosigkeit"  $\mu_r$ -exponentialverteilter Zeitphasen gelten die letzten Beziehungen ebenso wie die Gleichgewichtsgleichungen auch im Falle der Unterbrechungsstrategie "PRE restart with resampling".

### 3.1.3 Warteschlangenmodell M/M/1/NONPRE

Die Analyse des Prioritätsmodells M/M/1/NONPRE geht auf Cobham zurück, der durch Mittelwertbetrachtungen die Gleichung

$$w_r = \frac{\sum_{i=1}^R h_i \rho_i}{(1 - \rho_a) \cdot (1 - \rho_a - \rho_r)}, \quad (r=1, \dots, R), \quad (3.1,15)$$

für die mittlere Wartezeit einer Anforderung der Klasse  $r$  herleitete [Cobham 54]. Durch Addition von  $h_r$  folgt für die mittlere Verweilzeit  $f_r$ :

$$f_r = h_r + w_r, \quad (r=1, \dots, R). \quad (3.1,16)$$

Zur Beschreibung des Zustandsprozesses wird außer den ZVn  $X_1(t)$  und  $X_2(t)$  noch eine ZV  $S(t)$  benötigt, welche angibt, ob die BE durch eine Anforderung hoher ( $S=1$ ) bzw. niedriger Priorität ( $S=2$ ) belegt ist oder ob keine Anforderung zur Bedienung ansteht ( $S=0$ ). Mit der Bezeichnung

$$p(m, n, s) := P\{X_1=m, X_2=n, S=s\}, \quad (m, n \in \mathbb{N}_0; s \in \{0, 1, 2\}), \quad (3.1,17)$$

lauten die Gleichgewichtsgleichungen für den stationären Zustandsprozeß:

$$0 = -(\lambda + \mu_1)p(m, n, 1) + \lambda_1 p(m-1, n, 1) + \lambda_2 p(m, n-1, 1) + \mu_1 p(m+1, n, 1) + \mu_2 p(m, n+1, 2), \quad (m \geq 2, n \geq 1), \quad (3.1,18a)$$

$$0 = -(\lambda + \mu_2)p(m, n, 2) + \lambda_1 p(m-1, n, 2) + \lambda_2 p(m, n-1, 2), \quad (m \geq 1, n \geq 2), \quad (3.1,18b)$$

$$0 = -\lambda p(0, 0, 0) + \mu_1 p(1, 0, 1) + \mu_2 p(0, 1, 2) \quad (3.1,18c)$$

$$0 = -(\lambda + \mu_2)p(0, 1, 2) + \lambda_2 p(0, 0, 0) + \mu_1 p(1, 1, 1) + \mu_2 p(0, 2, 2) \quad (3.1,18d)$$

$$0 = -(\lambda + \mu_2)p(0, n, 2) + \lambda_2 p(0, n-1, 2) + \mu_1 p(1, n, 1) + \mu_2 p(0, n+1, 2), \quad (n \geq 2), \quad (3.1,18e)$$

$$0 = -(\lambda + \mu_1)p(1, 0, 1) + \lambda_1 p(0, 0, 0) + \mu_1 p(2, 0, 1) + \mu_2 p(1, 1, 2), \quad (3.1,18f)$$

$$0 = -(\lambda + \mu_1)p(m, 0, 1) + \lambda_1 p(m-1, 0, 1) + \mu_1 p(m+1, 0, 1) + \mu_2 p(m, 1, 2), \quad (m \geq 2), \quad (3.1,18g)$$

$$0 = -(\lambda + \mu_1)p(1, n, 1) + \lambda_2 p(1, n-1, 1) + \mu_1 p(2, n, 1) + \mu_2 p(1, n+1, 2), \quad (n \geq 1), \quad (3.1,18h)$$

$$0 = -(\lambda + \mu_2)p(m, 1, 2) + \lambda_1 p(m-1, 1, 2), \quad (m \geq 1). \quad (3.1,18i)$$

Rekursive Lösungen der gekoppelten partiellen Differenzengleichungen (3.1,18a-b) findet man in [Marks 73, Miller 81]. Die letztere Veröffentlichung enthält noch eine Formel für die stationäre Zustandsverteilung  $p_1(m)$  der ersten Klasse:

$$p_1(m) = \rho_1^m(1-\rho) + \frac{\lambda_2}{\lambda_1+\mu_2-\mu_1} \left[ \rho_1^m - \frac{\mu_1 \lambda_1^m}{(\lambda_1+\mu_2)^{m+1}} \right], \quad (m \in \mathbb{N}_0). \quad (3.1,19)$$

Als Voraussetzung für einen stationären Zustandsprozeß muß (3.1,11) erfüllt sein.

### 3.2 Warteschlangenmodelle mit mehrstufiger Bedienung

#### 3.2.1 Separable Warteschlangennetze

Ein großer Teil der bisher erzielten Ergebnisse über Zustandsverteilungen in WSNen beruht, bei Ersetzung nichtexponentieller VFN durch solche vom Phasentyp, auf dem Studium der zugehörigen Gleichgewichtsgleichungen. Wie jedoch schon eingangs erwähnt, sind dieser Methode durch die Anzahl der zu berücksichtigenden Zustände i.a. Grenzen gesetzt, was durch die Tatsache verdeutlicht wird, daß man zur Beschreibung des Zustandsprozesses in einem WSN mit N BSen und R Klassen eine  $N \times R$ -dimensionale ZV

$$\underline{X} := (\underline{X}_1, \dots, \underline{X}_i, \dots, \underline{X}_N)^T, \quad (T \hat{=} \text{Transponierung}) \quad (3.2,1a)$$

$$\underline{X}_i := (X_{i1}, \dots, X_{iR}, \dots, X_{iR}). \quad (3.2,1b)$$

benötigt, in der  $X_{ir}$  die Anzahl von Anforderungen der Klasse  $r$  im Knoten  $i$  bezeichnet. Diese Zustandsbeschreibung enthält schon keine Information mehr über die Bedienungsphasen einer Anforderung sowie die Reihenfolge der Anforderungen in der Warteschlange unter Berücksichtigung der Klassenzugehörigkeit. Die stationäre Zustandsverteilung sei mit

$$p(\underline{x}) := P\{\underline{X}=\underline{x}\}, \quad (3.2,2a)$$

$$\underline{x} := (\underline{x}_1, \dots, \underline{x}_N)^T, \quad (3.2,2b)$$

$$\underline{x}_i := (x_{i1}, \dots, x_{iR}), \quad (x_{ir} \in \mathbb{N}_0; i=1, \dots, N; r=1, \dots, R), \quad (3.2,2c)$$

bezeichnet, und es sei in Verallgemeinerung der bisherigen Notation

$$x_i := \sum_{r=1}^R x_{ir} \quad (3.2,3)$$

die Anzahl von Anforderungen in der  $i$ -ten BS schlechthin.

Als ein grundlegendes Ergebnis der bisherigen Forschungen auf dem Gebiete der Warteschlangentheorie hat sich herausgestellt, daß für eine Klasse von WSNen die zugehörigen Gleichgewichtsgleichungen eine Lösung mit Produktform besitzen. Bekannteste Vertreter dieser als separabel bezeichneten WSNen sind die sogenannten BCMP-Netze, deren Name auf die Initialen der Autoren von [Baskett, Chandy, Muntz, Palacios 75] zurückgeht. Hauptgegenstand dieser Veröffentlichung ist ein als BCMP-Theorem bekannter Satz, welcher sich auf die Analyse von WSNen mit BSen einer der folgenden vier Typen bezieht:

Typ 1:  $\cdot/M/1/FIFO$  mit zustandsabhängigen Bedienungsraten  $\mu_i(x_i)$ , die innerhalb einer Station nicht klassenabhängig sein dürfen.

Typ 2:  $\cdot/G/1/PS$ . Die Laplace-Transformierte der Bedienungzeit-VF muß eine rationale Funktion sein.

Typ 3:  $\cdot/G/\infty$ . Bei dieser BS wird angenommen, daß jede eintreffende Anforderung sofort bedient wird (unendlich viele BEn). Die Laplace-Transformierte der Bedienungzeit-VF muß eine rationale Funktion sein.

Typ 4:  $\cdot/G/1/LIFO-PRE$ . Die Laplace-Transformierte der Bedienungzeit-VF muß eine rationale Funktion sein.

Handelt es sich um ein offenes WSN, so sind für die externen Ankunftsströme zwei Fälle zu unterscheiden:

i) Alle Bedienungswünsche entstammen einer Quelle, welche einen inhomogenen Poisson-Prozeß erzeugen darf. Darunter ist ein Poisson-Prozeß zu verstehen, dessen Ankunftsrate  $\lambda_0(M(\underline{x}))$  von der momentanen Population  $M(\underline{x})$  im WSN abhängen kann. Die Anforderungen werden gemäß den unabhängigen Wahrscheinlichkeiten  $p_{0r,jr}$  den BSen und Klassen zugeordnet, wobei (2.3,2c) gilt.

ii) Alle Anforderungen entstammen aus K voneinander unabhängigen Quellen, von denen jede zu einer der K im WSN vorhandenen Ketten gehört. Diese Quellen können inhomogene Poisson-Prozesse mit Ankunftsrate  $\lambda_{0k}(M_k(\underline{x}))$  erzeugen, die jeweils von der momentanen Population  $M_k(\underline{x})$  der k-ten Kette, ( $k=1, \dots, K$ ), abhängen dürfen. Ein Auftrag der k-ten Quelle tritt mit der Wahrscheinlichkeit  $p_{0r,jr}$  in die BS j ein und gehört dort der Klasse r an. Es gilt dabei

$$\sum_{(j,r) \in E_k} p_{0r,jr} = 1, \quad (k=1, \dots, K). \quad (3.2,4)$$

Die Bereitstellung dieser Erklärungen erlaubt die Formulierung des folgenden Satzes 3.1, welcher aus dem o.g. BCMP-Theorem abgeleitet werden kann und eine auf Anwendungen zugeschnittene Form dieses Theorems ist.

Satz 3.1: [Baskett, Chandy, Muntz, Palacios 75]

Gegeben sei ein offenes, geschlossenes oder gemischtes WSN mit N BSen der Typen 1-4. Die Anzahl der Klassen sei R. Sind externe Ankünfte vorhanden, so sollen diese einer der o.g. Bedingungen i) oder ii) genügen. Die Verkehrslenkung sei durch die voneinander unabhängigen Verzweigungswahrscheinlichkeiten  $p_{ir,jr}$ , ( $i,j=0, \dots, N$ ;  $r,s=1, \dots, R$ ), bestimmt. Dann berechnen sich die stationären Zustandswahrscheinlichkeiten  $p(\underline{x})$  gemäß

$$p(\underline{x}) = G^{-1} \cdot d(\underline{x}) \cdot g_1(\underline{x}_1) \cdot g_2(\underline{x}_2) \cdot \dots \cdot g_N(\underline{x}_N), \quad (3.2,5a)$$

wobei:

a) G eine Konstante ist, die aus der Normierungsbedingung

$$\sum_{\underline{x} \in \mathcal{J}} p(\underline{x}) = 1 \quad (3.2,5b)$$

zu bestimmen ist.  $\mathcal{J}$  bezeichne die Menge aller Zustände.

b) 
$$d(\underline{x}) = \prod_{m=0}^{M(\underline{x})-1} \lambda_0(m) \quad (3.2,5c)$$

falls i) und

$$d(\underline{x}) = \prod_{k=1}^K \left[ \prod_{m=0}^{M_k(\underline{x})-1} \lambda_{0k}(m) \right] \quad (3.2,5d)$$

falls ii) gilt sowie  $d(\underline{x}) \equiv 1$  für geschlossene WSNe.

c)  $g_i(\underline{x}_i)$  Funktionen sind, welche vom Typ der jeweiligen BS abhängen:

Typ 1: 
$$g_i(\underline{x}_i) = \left[ x_i! \prod_{r=1}^R (e_{ir}/x_{ir}!) \right] \cdot \prod_{m=1}^{x_i} (\mu_i(m))^{-1} \quad (3.2,5e)$$

Typen 2, 4: 
$$g_i(\underline{x}_i) = x_i! \left[ \prod_{r=1}^R ((e_{ir}/\mu_{ir})^{x_{ir}}/x_{ir}!) \right] \quad (3.2,5f)$$

Typ 3: 
$$g_i(\underline{x}_i) = \prod_{r=1}^R ((e_{ir}/\mu_{ir})^{x_{ir}}/x_{ir}!). \quad (3.2,5g)$$

Die Größen  $e_{ir}$  bilden eine Lösung des linearen Gleichungssystems

$$e_{js} = p_{0s,jr} + \sum_{i=1}^N \sum_{r=1}^R e_{ir} \cdot p_{ir,jr}, \quad (3.2,5h)$$

das beim Vorhandensein von K Ketten in K Systeme des Typs

$$e_{js} = p_{0s,jr} + \sum_{(i,r) \in E_k} e_{ir} \cdot p_{ir,jr}, \quad ((j,s) \in E_k; k=1, \dots, K) \quad (3.2,5i)$$

zerfällt.



Bemerkungen

i) Für die i.a. aufwendige Bestimmung der Normierungskonstanten G existiert eine Vielzahl problemorientierter Algorithmen, von denen einige in [Bolch, Akyildiz 82] zusammengestellt sind. Diese Algorithmen liefern meist auch Berechnungsmöglichkeiten für Größen wie Durchsatz und Ausnutzung, ohne explizit auf die Zustandswahrscheinlichkeiten zurückzugreifen. Zu diesem Punkt ist insbesondere die sogenannte Mittelwertanalyse [Reiser, Lavenberg 80] zu nennen.

ii) Im Falle eines offenen WSNe mit konstanten Ankunftsrate folgt aus Satz 3.1 die Beziehung

$$p(x_1, \dots, x_N) = p_1(x_1) \cdot \dots \cdot p_N(x_N), \quad (3.2,6)$$

in welcher die Zustandswahrscheinlichkeiten  $p_i(x_i)$  für jeden Knoten i unabhängig so zu berechnen sind als sei der zugehörige Ankunftsprozeß ein Poisson-Prozeß. Dessen Gesamtankunftsrate  $\lambda_i$  bestimmt sich aus einer der Gleichungen (3.2,7a) oder (3.2,7b) durch Summation über den Klassenindex s. Die Ergebnisse aus [Jackson 63, Gordon, Newell 67] sind somit als Sonderfälle in Satz 3.1 enthalten.

iii) Nach [Muntz 72] erzeugt jede isoliert betrachtete BS, die einem der vier o.g. Typen entspricht, bei einem Poisson-Ankunftsprozeß auch wieder einen Poisson-Ausgangsprozeß ("M=>M property").

iv) Die Größe  $e_{ir}$  ist der Ankunftsrate  $\lambda_{ir}$  der Klasse r in der BS i proportional. Im Falle offener WSNe mit konstanten externen Ankunftsrate  $\lambda_{0ir} := \lambda \cdot p_{0r,ir}$  sind (3.2,5h-i) den Flußgleichungen

$$\lambda_{js} = \lambda_{0js} + \sum_{i=1}^N \sum_{r=1}^R \lambda_{ir} \cdot p_{ir,js}, \quad (3.2,7a)$$

$$\lambda_{js} = \lambda_{0js} + \sum_{(i,r) \in E_k} \lambda_{ir} \cdot p_{ir,js}, \quad ((j,s) \in E_k; k=1, \dots, K) \quad (3.2,7b)$$

äquivalent, welche eine eindeutige Lösung besitzen. Für geschlossene WSNe ist  $p_{0r,ir} \equiv 0$  (bzw.  $\lambda_{0ir} \equiv 0$ ), so daß die Größen  $e_{ir}$  (bzw.  $\lambda_{ir}$ ) nur bis auf einen konstanten Faktor bestimmt sind, der in die Normierungskonstante eingeht.

Der nächste Satz wird als "Nortons Theorem für WSNe mit Produktlösung" bezeichnet und sagt über die Möglichkeit aus, Teile eines separablen WSNe durch eine einzige Stufe mit zustandsäquivalentem Verhalten zu ersetzen. Das Theorem wurde für einklassige WSNe erstmals in [Chandy, Herzog, Woo 75] veröffentlicht und später von Kritzingner, Wyk und Krzesinski verallgemeinert.

Satz 3.2: [Kritzingner, Wyk, Krzesinski 82]

Gegeben sei ein geschlossenes WSN mit N BSen der Typen 1-4 und R Klassen. Für BSen des Typs 1 gelte die Einschränkung, daß die Bedienungsrate konstant sein muß. Ein Klassenwechsel soll nicht stattfinden.  $M_r$  sei die Population der Kette r, ( $r=1, \dots, R$ ). Die Stationsnummern der zu aggregierenden BSen mögen die Menge  $\bar{\sigma} = \{1, \dots, n\}$  und diejenigen des restlichen Netzes die Menge  $\sigma = \{n+1, \dots, N\}$  bilden.

Man ersetze das  $\bar{\sigma}$  zugeordnete Teilnetz durch eine BS (Index c) mit zustandsabhängigen Bedienungsrate  $\mu_{cr}(x_c)$ , wobei

$$x_c := (x_{n1}^*, \dots, x_{nR}^*), \quad x_{nr}^* := \sum_{i=1}^n x_{ir}, \quad (r=1, \dots, R). \quad (3.2,8)$$

Die Bedienungsrate sind dem Durchsatz  $\lambda_r(x_c)$  der Klasse r durch das  $\bar{\sigma}$  zugeordnete Teilnetz gleichzusetzen, wenn dieses sich im Zustand  $x_c$  befindet und die Bedienungszeiten in den zu  $\sigma$  gehörenden BSen gleich Null gesetzt sind. Dann gilt:

Die stationäre Randverteilung  $p_{\sigma}^i(x_{n+1}, \dots, x_N)$  des neuen äquivalenten WSNe ist die gleiche wie die Randverteilung  $p_{\sigma}(x_{n+1}, \dots, x_N)$  des gegebenen WSNe.

Im Vorgriff auf den späteren Abschnitt 5.2.2 sei schon hier erwähnt, daß man bei BSen des Typs 1 die Voraussetzung zustandsunabhängiger Bedienungsraten wegfällen lassen kann.

Abschließend ist zu bemerken, daß die Klasse der separablen WSNe über solche vom BCMP-Typ hinaus geringfügig erweitert wurde. Diese Erweiterungen umfassen jedoch mit Ausnahme der im übernächsten Abschnitt zu besprechenden "homogenen WSNe" keine Prioritätsstrategien.

### 3.2.2 Nichtseparable Warteschlangennetze

Trotz der großen Bedeutung der BCMP-Netze für die Modellierung von Kommunikations- und Rechnersystemen weisen sie einige Einschränkungen auf, welche den praktischen Anwendungen mitunter entgegenstehen. Hierzu sind sowohl die Forderung nach Markoffschen Eigenschaften der Last als auch der Ausschluß von Prioritätsstrategien zu nennen. Während den Prioritätsnetzen der weitere Teil dieser Arbeit gewidmet sein wird, sei in diesem Abschnitt noch kurz auf einige approximative Analyseverfahren für WSNe eingegangen, welche vor allem auch nichtexponentiell verteilte Bedienungszeiten für BSen vom Typ 1 zulassen. Da für solche WSNe keine Lösungen mit Produktform bekannt sind, werden sie als nichtseparabel bezeichnet. Nachfolgend eine Aufstellung der wichtigsten Verfahren.

#### a) Erweiterte Produktformlösung [Shum 76]

Wie schon der Name andeutet, geht diese Methode von der näherungsweise Gültigkeit von (3.2,5a) für WSNe mit BSen vom Typ  $\cdot/G/1$ -FIFO aus, wobei die Funktionen  $g_i(\underline{x}_i)$  für diese BSen geeignet zu wählen sind.

#### b) Äquivalenzverfahren [Chandy, Herzog, Woo 75]

Ähnlich wie unter a) wird hier die Gültigkeit des Norton Theorems für allgemeinere Voraussetzungen angenommen. Dieses führt unter Beachtung der Bedingungen i) Konstanz der Population in

geschlossenen WSNe und ii) Gültigkeit der Flußgleichungen (3.2,7a-b) auch für nichtseparable WSNe zu einem Iterationsalgorithmus.

Eine Variante dieses Verfahrens ist in [Sauer, Chandy 75] vorgestellt. Sie beinhaltet eine Miteinbeziehung der beiden ersten Momente für die Bedienungszeiten und führt auf ein aggregiertes WSN mit VFn vom Phasentyp. Außerdem wird dort für ein CSM mit klassenabhängigen Bedienungsraten in der CPU ein Analyseverfahren vorgeschlagen, welches im wesentlichen auf der Transformation des R-klassigen Modells ( $R > 2$ ) in ein zweiklassiges Modell (Zusammenfassen von Klassen) und anschließender Anwendung der zuvor erläuterten Äquivalenzmethode beruht.

#### c) Zerlegungsverfahren [Courtois 77]

Diese Verfahren haben zum Ziel, die Komplexität des Analyseproblems dadurch zu reduzieren, daß das WSN geeigneten Kriterien entsprechend in kleinere Einheiten zerlegt wird, welche entweder unabhängig voneinander (vollständige Zerlegung) oder unter Beachtung zusätzlicher Zustandsbedingungen (nahezu vollständige Zerlegung) untersucht werden.

#### d) Numerische Analyse [Herfort, Müller-Clostermann 83]

Dieses Verfahren beinhaltet als ersten Schritt die automatische Generierung der Intensitätsmatrix  $Q$  für eine reichhaltige Modellklasse, in welcher u.a. auch unterbrechende Prioritätsstrategien zugelassen sind. Der nächste Schritt umfaßt die numerische Lösung der Gleichgewichtsgleichungen. Dazu wird durch Definition einer geeigneten Äquivalenzrelation eine Strukturierung des Zustandsraumes mit Blick auf zerlegungsorientierte Verfahren vorgenommen.

#### e) Diffusionsapproximation [Kobayashi 74]

Diese Methode beruht auf der approximativen Beschreibung des diskreten Zustandsprozesses in einstufigen Warteschlangenmodellen durch einen Wiener-Prozeß, wobei für die Rechtferti-

gung dieser Approximation eine voll ausgelastete BE voraussetzen ist. Durch Ermittlung der (kontinuierlichen) Dichtefunktion für die Zustände des Wiener-Prozesses und anschließender Diskretisierung ergeben sich die Zustandswahrscheinlichkeiten in Abhängigkeit von den beiden ersten Momenten der Ankunftsabstände und der Bedienungszeiten. Verallgemeinerung dieser Prinzipien auf WSNe führt zu Lösungen mit Produktform, welche im Falle eines genügend großen Verkehrsvolumens gute Ergebnisse liefern.

#### f) Momentenverfahren [Kühn 79]

Auf den im Abschnitt 3.1.1 angegebenen Ergebnissen für das Wartemodell GI/G/1 aufbauend, beruht dieses für offene WSNe konzipierte Verfahren auf der isolierten Betrachtung von Einzelstufen (knotenweise Zerlegung). Die jeweiligen Ankunftsprozesse werden durch Erneuerungsprozesse approximiert, deren Charakterisierung durch die beiden ersten Momente der Ankunftsabstände erfolgt. Das Verfahren liefert demzufolge auch keine Zustandsverteilungen sondern Mittelwerte für die Verweilzeiten in den einzelnen Stufen bzw. im Netz.

### 3.2.3 Warteschlangennetze mit Prioritäten

Der letzte Abschnitt in diesem Kapitel gibt einen Überblick über den derzeitigen Entwicklungsstand von Analysemethoden für WSNe mit Prioritätsklassen. Hierbei wird zwischen exakten und approximativen Verfahren unterschieden.

#### a) Exakte Verfahren

Für die exakte Analyse von WSNe mit Prioritäten bietet es sich (gegebenenfalls unter Anwendung der Phasenmethode) prinzipiell an, nach Lösungen der zugehörigen Gleichgewichtsgleichungen zu suchen. Obwohl man diesen Weg in der Vergangenheit für verschiedene geschlossene Modelle geringer Komplexität erfolgreich beschritten hat [Spirn 76, Morris 81, Sauer, Chandy

81, Kameda 82, Kaufman 82] ist er eher von theoretischem Interesse, da die Form der Gleichgewichtsgleichungen bei umfangreicheren Modellen zu unübersichtlich bzw. die Zahl der zu berücksichtigenden Zustände zu groß wird. Diese schon zu Beginn des Kapitels erwähnte Tatsache verdeutlicht sich im Abschnitt 5.1.1 anhand eines seriellen Warteschlangenmodells in sehr eindrucksvoller Weise.

Der Aspekt des theoretischen Interesses trifft auch auf die sogenannten homogenen WSNe mit Markoffschen Eigenschaften zu, für welche im Falle unterbrechender Prioritäten in [Morris 81] eine der Mittelwertanalyse ähnelnde Lösungsform angegeben ist. In dieser Arbeit wird ein WSN als homogen bezeichnet, wenn in jeder BS die Bedienungsraten und Verzweigungswahrscheinlichkeiten für alle Anforderungen gleich sind, was für praktische Zwecke eine starke Einschränkung bedeutet. Das Resultat von Morris stellt eine Erweiterung der in [Avi-Itzhak, Heyman 73] präsentierten Lösung für ein CSM mit den genannten Eigenschaften dar.

Aus Vollständigkeitsgründen sei noch ein Ergebnis aus [Jaiswal 68] erwähnt, wo eine exakte Analyse von Prioritätsmodellen mit endlicher Quellenzahl durchgeführt ist. Diese Modelle sind jeweils einem zyklischen WSN mit einer BS vom Typ  $^*/M/\infty$  und einer BS vom Typ  $^*/G/1/PRE$  (bzw. NONPRE) äquivalent.

#### b) Approximative Verfahren

Die bisher bekannten approximativen Verfahren für WSNe mit Prioritäten stützen sich weitgehend auf die in den Abschnitten 3.2.1 und 3.2.2 besprochenen Verfahren und Ergebnisse, wobei die dort skizzierte "Numerische Analyse" eine Berücksichtigung von unterbrechenden Prioritäten von vornherein vorsieht. Nachfolgend eine Aufstellung verschiedener Methoden.

#### bl) Knotenweise Zerlegung

Im Falle offener WSNe mit Markoff-Eigenschaften nimmt man an, daß die Ankunftsprozesse in den einzelnen Stufen Poisson-

prozesse sind und somit eine knotenweise Analyse mit Hilfe der in den Abschnitten 3.1.2-3 zusammengestellten Ergebnisse erfolgen kann. Dieses Verfahren wird u.a. in [Wong, Sauvé 82] angewendet.

b2) Erweiterte Mittelwertanalyse [Chandy, Lakshmi 83, Bryant, Krzesinski, Teunissen 83]

Dieses ist ein entsprechendes Verfahren für geschlossene WSNe. Es beruht auf der Mittelwertanalyse [Reiser, Lavenberg 80] und berücksichtigt Prioritätsdisziplinen durch Berechnung der mittleren Verweilzeiten in den einzelnen Prioritätsstufen auf der Basis der Mittelwertformeln aus den Abschnitten 3.1.2-3. Demzufolge sind auch hier indirekt die Ankunftsprozesse im WSN als Poisson-Prozesse angenommen.

b3) Aggregierungsverfahren [Sauer, Chandy 75, Chow, Yu 83]

Die in den o.g. Arbeiten vorgeschlagenen Aggregierungsverfahren sind vorwiegend für geschlossene WSNe mit einer Prioritätsstufe und exponentiell verteilten Bedienungszeiten geeignet. Mit ihnen wird eine Reduzierung des Zustandsraumes durch Klassenzusammenfassung und anschließender Anwendung des Norton-Theorems erreicht, so daß die Gleichgewichtsgleichungen gelöst werden können.

b4) Approximation durch separable WSNe [Reiser 76, Sevcik 77, Kaufman 82]

Grundidee dieser Methode ist die Ersetzung einer jeden R-klassigen Prioritätsstufe im WSN, welche vom Typ  $\cdot/M/1/PRE$  oder  $\cdot/M/1/NONPRE$  sei, durch jeweils R einklassige BSen vom Typ  $\cdot/M/1$  (Bild 3.2). Im Falle, daß das gegebene WSN ansonsten höchstens BSen der Typen 1-4 enthält, entsteht auf diese Weise ein separables WSN. Zur Berücksichtigung der Auswirkungen, welche die Prioritätsabfertigung in den BSen auf die Verweilzeiten der betroffenen Anforderungen hat, sind die Bedienungs-raten  $\mu_{ir}^*$  der Ersatzmodelle gegenüber denen im gegebenen WSN um einen konstanten Faktor  $c_{ir}$  gemäß

$$\mu_{ir}^* = \mu_{ir} \cdot c_{ir}, \quad (0 \leq c_{ir} \leq 1; i=1, \dots, N; r=1, \dots, R), \quad (3.2,9)$$

reduziert. Die o.g. Arbeiten unterscheiden sich in der Wahl des Faktors  $c_{ir}$ .

In [Reiser 76, Sevcik 77] ist für den Fall  $R=2$  und unterbrechende Prioritäten

$$c_{i2} = (1 - \rho_{i1}), \quad (i=1, \dots, N), \quad (3.2,10)$$

gesetzt. Als Motivation für diese Wahl kann man anführen, daß

- i)  $1 - \rho_{i1}$  dem für die Bedienung von Anforderungen der Klasse 2 zur Verfügung stehenden Anteil an Bedienungskapazität entspricht,
- ii)  $h_{ir} / (1 - \rho_{ir})$  die mittlere Komplettierungszeit für Anforderungen zweiter Klasse ist und damit eine Art mittlerer Bedienungszeit zu sein scheint.

Gerade das letzte Argument wird jedoch in [Kaufman 82] widerlegt, wo für unterbrechende Prioritäten

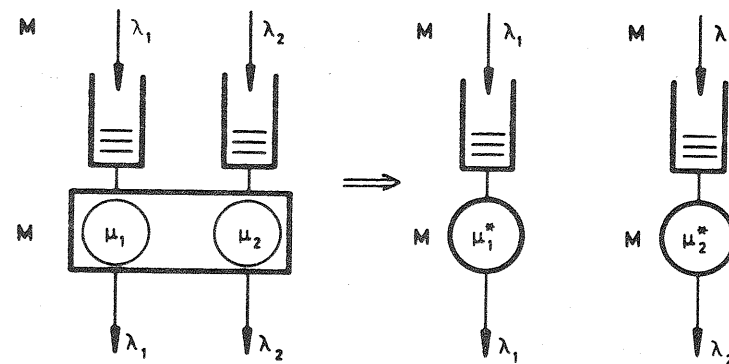


Bild 3.2: "Zerlegung" eines zweiklassigen Warteschlangenmodells mit Prioritätsabfertigung

$$c_{i2} = \frac{P\{X_{i1}=0, X_{i2}>0\}}{P\{X_{i2}>0\}}, \quad (i=1, \dots, N), \quad (3.2,11)$$

und für nichtunterbrechende Prioritäten

$$c_{i1} = 1 - \frac{P\{X_{i1}>0, X_{i2}>0, S=2\}}{P\{X_{i1}>0\}}, \quad (i=1, \dots, N), \quad (3.2,12a)$$

$$c_{i2} = 1 - \frac{P\{X_{i1}>0, X_{i2}>0, S=1\}}{P\{X_{i2}>0\}}, \quad (i=1, \dots, N), \quad (3.2,12b)$$

gesetzt ist. Diese Formeln wurden aus Betrachtungen über die Modelle M/M/1/PRE und M/M/1/NONPRE gewonnen und haben wegen der à priori nicht bekannten Wahrscheinlichkeiten zunächst nur theoretische Bedeutung. In seinem Tagungsbeitrag weist Kaufman jedoch auf die bevorstehende Veröffentlichung einer Approximationstechnik zur Bestimmung der Reduktionskoeffizienten hin. Ein ähnliches Problem besteht auch bei einer Wahl der  $c_{ir}$  nach (3.2,10), da bei geschlossenen WSnen die Ausnutzungen  $\rho_{i1}$  nur dann separat berechnet werden können, wenn alle BSen vom Typ M/1/PRE sind. In [Sevcik 77] ist zur Lösung dieses Problems eine Iterationsmethode vorgeschlagen worden. Um bei späteren Diskussionen auf die beiden Verfahren leicht Bezug nehmen zu können, wird in Übernahme der Notation aus [Kaufman 82] von der "reduced occupancy approximation" (r o a) gesprochen, wenn eine Wahl der  $c_{ir}$  nach (3.2,10) stattfindet und von der "modified reduced occupancy approximation" (m-r o a), wenn die Gleichungen (3.2,11) bzw. (3.2,12a-b) angewendet werden. Unter dieselbe Kategorie fällt ein in Kapitel 5 später behandeltes, neu vorgeschlagenes Verfahren, mit dem eine deutliche Verbesserung der approximativen Ergebnisse erzielt werden kann.

Abschließend sei noch erwähnt, daß in [Uspenski, Dowling 79] eine Analysemethode für ein CSM mit unterbrechenden Prioritäten in der CPU beschrieben ist, welche sich nicht in die hier vorgenommene Klassifizierung einordnen läßt. Das Verfahren besteht aus zwei Folgen von Einzelschritten, die jeweils empirisch getroffene Festlegungen enthalten. Im ersten Teil wird

ohne Berücksichtigung der Prioritäten ein Schätzwert für die mittlere Zykluszeit errechnet, welcher dann durch wiederholte Korrekturen eine Verbesserung erfährt.

4. EINSTUFIGE WARTESCHLANGENMODELLE MIT PRIORITÄTEN

Nachdem der Schwerpunkt in den vorangegangenen Kapiteln auf der Erläuterung von Begriffen und der Zusammenstellung von Ergebnissen aus der Warteschlangentheorie gelegen hat, werden im nun folgenden Teil der Arbeit neue Verfahren zur Analyse von Markoffschen WSNen mit Prioritäten vorgestellt. Die Grundlage für diese Verfahren bilden verschiedene neu entwickelte einklassige Ersatzmodelle für mehrklassige Prioritätsmodelle. Durch Anwendung der Ersatzmodelle kann man das zu untersuchende WSN entweder in ein BCMP-Netz überführen oder in ein WSN, das mittels des o.g. Momentenverfahrens von Kühn analysierbar ist. Ziel dieser Vorgehensweise ist, das komplexere Problem durch Zerlegungsverfahren auf einfachere, beherrschbare Teilprobleme zurückzuführen.

Zur Rechtfertigung der in den beiden letzten Abschnitten dieses Kapitels vorgestellten Ersatzmodelle seien einige Betrachtungen über die Verteilungen der Zustände und der Ausgangsabstände in einstufigen Prioritätsmodellen vorausgestellt. Dabei wird stets von zwei Prioritätsklassen ausgegangen.

4.1 Warteschlangenmodell M/M/1/PRE

4.1.1 Zustandsverteilungen

Das Hauptziel der Betrachtungen in diesem Abschnitt ist, durch das Studium der Randverteilungen

$$p_1(m) := P\{X_1=m\} = \sum_{n=0}^{\infty} p(m,n), \quad (m \in \mathbb{N}_0), \quad (4.1,1a)$$

und

$$p_2(n) := P\{X_2=n\} = \sum_{m=0}^{\infty} p(m,n), \quad (n \in \mathbb{N}_0), \quad (4.1,1b)$$

geeignete Hinweise zur Konstruktion von einklassigen Ersatzmodellen zu gewinnen.

Summiert man in den Gleichgewichtsgleichungen (3.1,8a-d) über alle n und berücksichtigt die Normierungsbedingung

$$\sum_{m=0}^{\infty} \sum_{n=0}^{\infty} p(m,n) = 1, \quad (4.1,2)$$

so folgt nach einigen Umformungen:

$$0 = \lambda_1 p_1(m-1) - (\lambda_1 + \mu_1) p_1(m) + \mu_1 p_1(m+1), \quad (m \geq 1), \quad (4.1,3a)$$

$$0 = -\lambda_1 p_1(0) + \mu_1 p_1(1). \quad (4.1,3b)$$

Die gewöhnliche Differenzgleichung (4.1,3a) mit der Randbedingung (4.1,3b) hat die rekursive Lösung

$$p_1(m) = \rho_1 p_1(m-1), \quad (m \geq 1), \quad (4.1,4a)$$

$$p_1(0) = 1 - \rho_1 \quad (4.1,4b)$$

für welche man auch

$$p_1(m) = \rho_1^m (1 - \rho_1), \quad (m \geq 0), \quad (4.1,5)$$

schreiben kann. Dieses Ergebnis überrascht nicht, da bei unterbrechenden Prioritäten die Verhältnisse für die erste Klasse mit denen in einem Warteschlangenmodell M/M/1 mit Ankunftsrate  $\lambda_1$  und Bedienungsrate  $\mu_1$  übereinstimmen.

Wird in den Gleichgewichtsgleichungen (3.1,8a-d) über alle m summiert, so erhält man in gleicher Weise:

$$\lambda_2 p_2(n-1) - \lambda_2 p_2(n) = \mu_2 p(0,n) - \mu_2 p(0,n+1), \quad (n \geq 1), \quad (4.1,6a)$$

$$\lambda_2 p_2(0) = \mu_2 p(0,1). \quad (4.1,6b)$$

Aus diesen Gleichungen folgt die Beziehung

$$p_2(n) = \rho_2^{-1} p(0,n+1), \quad (n \geq 0), \quad (4.1,7)$$

deren Gültigkeit man durch Einsetzen in (4.1,6a-b) sofort bestätigen kann. Damit ist das Problem, eine geschlossene Lösung für die Randverteilung  $p_2(n)$  zu finden, darauf zurückgeführt, die Verbundwahrscheinlichkeiten  $p(0,n)$  zu bestimmen.

Im Anhang 1 ist hierfür ein Verfahren angegeben, welches der üblichen Methode, zuerst die Verbundverteilung  $p(m,n)$  auszurechnen und dann über  $m$  zu summieren, hinsichtlich des Bedarfs an Rechenzeit und Speicherplatz überlegen ist.

Ein sehr viel interessanterer Aspekt ergibt sich jedoch, wenn man in (4.1,6a-b) die folgenden Umformungen vornimmt

$$0 = \lambda_2 p_2(n-1) - \left[ \lambda_2 + \mu_2 \frac{p(0,n)}{p_2(n)} \right] p_2(n) + \mu_2 \frac{p(0,n+1)}{p_2(n+1)} p_2(n+1), \quad (n \geq 1),$$

$$0 = -\lambda_2 p_2(0) + \mu_2 \frac{p(0,1)}{p_2(1)} p_2(1).$$

Definiert man nun

$$\mu_2^*(n) := c_2(n) \cdot \mu_2, \quad (n \geq 1), \quad (4.1,8a)$$

$$c_2(n) := \frac{p(0,n)}{p_2(n)}, \quad (n \geq 1), \quad (4.1,8b)$$

so folgt

$$0 = \lambda_2 p_2(n-1) - (\lambda_2 + \mu_2^*(n)) p_2(n) + \mu_2^*(n+1) p_2(n+1), \quad (n \geq 1), \quad (4.1,9a)$$

$$0 = -\lambda_2 p_2(0) + \mu_2^*(1) p_2(1). \quad (4.1,9b)$$

Diese Gleichungen beschreiben die Zustandsverteilung in einem Warteschlangenmodell M/M/1 mit Ankunftsrate  $\lambda_2$  und zustandsabhängiger Bedienungsrate  $\mu_2^*(n)$ . Es gilt dann

$$p_2(n) = \frac{\lambda_2}{c_2(n) \mu_2} p_2(n-1), \quad (n \geq 1). \quad (4.1,10)$$

Nachfolgend einige

Eigenschaften der Koeffizienten  $c_2(n)$ :

i) Aus der Definition bedingter Wahrscheinlichkeiten (Gleichung (4.1,8b)) folgt sofort

$$c_2(n) = P\{X_1=0 | X_2=n\}, \quad (n \geq 1). \quad (4.1,11)$$

Hierzu ist die Formulierung

$$c_2(n) = P\{\text{Anforderung der Klasse 2 in Bedienung} | n \text{ Anforderungen der Klasse 2 im Modell}\} \quad (4.1,12)$$

äquivalent.

ii) Wegen (4.1,11) gilt die Relation

$$0 \leq c_2(n) \leq 1, \quad (n \geq 1). \quad (4.1,13)$$

Die Koeffizienten bewirken demnach eine zustandsabhängige Reduktion der Bedienungsrate  $\mu_2$  und werden daher im folgenden auch als (zustandsabhängige) Reduktionskoeffizienten bezeichnet.

iii) Für den "Mittelwert" der Reduktionskoeffizienten gilt

$$\bar{c}_2 := \sum_{n=0}^{\infty} c_2(n) \cdot p_2(n) = \rho_2 \text{ für } c_2(0) \neq 0, \quad (4.1,14)$$

was man durch Ausrechnen der Summe leicht verifiziert. Im Falle der Vollast, ( $\rho_1 + \rho_2 = 1$ ), folgt insbesondere

$$\bar{c}_2 = 1 - \rho_1. \quad (4.1,15)$$

Die in diesem Abschnitt hergeleiteten Beziehungen für die Zustandsverteilungen der ersten und zweiten Klasse wurden auch numerisch bestätigt und motivieren zusammen mit den Ergebnissen für das Modell M/M/1/NONPRE zu den in Abschnitt 4.3 besprochenen einklassigen Ersatzmodellen. Für das Verständnis des folgenden Abschnitts ist schließlich noch anzumerken, daß die obigen Ergebnisse ihre Gültigkeit auch dann behalten, wenn endliche Speicherkapazitäten  $s_1$  und  $s_2$  vorliegen.

#### 4.1.2 Blockierungswahrscheinlichkeiten

In diesem Abschnitt wird für den Fall endlicher Speicherkapazitäten  $s_1$  und  $s_2$  auf ein effizientes Berechnungsverfahren der mit  $B_1$  und  $B_2$  bezeichneten Blockierungswahrscheinlichkeiten

der Klassen 1 und 2 eingegangen. Gemäß ihrer Definition aus Abschnitt 2.4 gilt für diese Größen

$$B_1 = P\{X_1=s_1+1\}, \quad (4.1,16a)$$

$$B_2 = P\{X_2=s_2\}. \quad (4.1,16b)$$

Hierin bezeichnet  $X_r$ , ( $r=1,2$ ), die Anzahl von Anforderungen der Klasse  $r$  im Modell  $M/M/1-s_1, s_2/PRE$ . In der Beziehung für  $B_2$  ist vorausgesetzt, daß auch im Falle der Bedienung einer Anforderung zweiter Klasse, deren Warteplatz für die Dauer der Bedienung nicht neu belegt werden darf. Dadurch ist bei einer Unterbrechung die Möglichkeit des Zurückspeicherns gesichert.

Zum Verständnis der weiter unten angegebenen Formeln zur Bestimmung der Blockierungswahrscheinlichkeit  $B_2$  sei ein leicht zu überprüfender Zusammenhang zwischen den Zustandswahrscheinlichkeiten  $\tilde{p}(x)$  im Modell  $M/M/1-s$  und den Zustandswahrscheinlichkeiten  $p(x)$  im Modell  $M/M/1$  angegeben:

$$\tilde{p}(x) = \frac{p(x)}{1-\sigma}, \quad \sigma = \sum_{x=s+2}^{\infty} p(x). \quad (4.1,17)$$

Diese Beziehung behält ihre Gültigkeit auch bei zustandsabhängigen Bedienungsraten und begründet zusammen mit den Ergebnissen des vorhergehenden Abschnitts sowie der bekannten Formel für die Blockierungswahrscheinlichkeit im Modell  $M/M/1-s_1$  mit konstanten Raten  $\mu_1, \lambda_1$  die folgenden Gleichungen:

$$B_1 = \frac{1-\rho_1}{s_1+2} \cdot \rho_1^{s_1+1}, \quad (4.1,18a)$$

$$B_2 = \frac{p(s_2)}{1-\sigma_2}, \quad (4.1,18b)$$

$$1-\sigma_2 = \sum_{x_2=0}^{s_2+1} p_2(x_2). \quad (4.1,18c)$$

Gemeinsam mit dem vorher erwähnten Algorithmus zur Bestimmung der Zustandswahrscheinlichkeiten  $p_2(x_2)$  steht damit ein neues

Berechnungsverfahren für die Größe  $B_2$  zur Verfügung, das der herkömmlichen Methode, zuerst die Verbundverteilung  $p(x_1, x_2)$  auszurechnen und über diese geeignet zu summieren, überlegen ist.

#### 4.1.3 Ausgangsprozesse

Die Untersuchung und Beschreibung von Ausgangsprozessen mittels analytischer Methoden hat sich als sehr schwierig erwiesen, und es überrascht daher nicht, daß sich die diesbezüglichen Kenntnisse für das Modell  $M/M/1-PRE$  im wesentlichen auf die beiden folgenden Ergebnisse beschränken:

- i) Der Ausgangsprozeß der ersten Klasse ist ein Poisson-Prozeß. Da für die erste Klasse Verhältnisse wie im Modell  $M/M/1$  vorliegen, folgt dies sofort mit dem bekannten Theorem von Burke [Burke 56].
- ii) Aus [Kelly 79] kann man entnehmen, daß im Falle  $\mu_1 = \mu_2$  der von den Anforderungen beider Klassen erzeugte allgemeine Ausgangsprozeß ein Poisson-Prozeß ist und daß der Ausgangsprozeß der zweiten Klasse kein Poisson-Prozeß ist.

Im folgenden werden nun zwei Möglichkeiten zur approximativen Bestimmung des Variationskoeffizienten  $c_{D2}$  der Ausgangsstände  $T_{D2}$  diskutiert. Beide Methoden beruhen auf der Anwendung der weiter oben angegebenen Beziehung (3.1,5), welche einen Zusammenhang zwischen den Variationskoeffizienten  $c_A, c_H$  und  $c_D$  im Modell  $GI/G/1$  herstellt. Sie unterscheiden sich in der Wahl der Größe  $c_{D2}^*$ , womit der Variationskoeffizient einer "virtuellen Bedienungszeit"  $T_{H2}^*$  bezeichnet sei. Diese neu eingeführte ZV  $T_{H2}^*$  ist so zu charakterisieren, daß die Unterbrechungszeiten der ersten Klasse eine möglichst gute Berücksichtigung finden.

Aufgrund der letztgenannten Forderung erscheint es sinnvoll,

$$T_{H2}^* = T_{C2} \quad (4.1,19a)$$



zu setzen, woraus wegen (3.1,13a-b) sofort die Beziehungen

$$h_2^* := E[T_{H2}^*] = \frac{h_2}{1-\rho_1}, \quad (4.1,19b)$$

$$\rho_2^* := h_2^* \lambda_2 = \frac{\rho_2}{1-\rho_1}, \quad (4.1,19c)$$

$$c_{H2}^{*2} = 1 + 2 \cdot \frac{h_1}{h_2} \cdot \frac{\rho_1}{1-\rho_1} \quad (4.1,19d)$$

folgen. Die Komplettierungszeit  $T_{C2}$  für eine Anforderung zweiter Klasse ist in Abschnitt 3.1.2 erläutert worden. Durch sinngemäßes Einsetzen von (4.1,19c-d) in (3.1,5) errechnet man nun leicht

$$c_{D2}^2 = 1 + 2 \cdot \frac{h_1}{h_2} \cdot \frac{\rho_1 \rho_2^2}{(1-\rho_1)^3}. \quad (4.1,20)$$

Dabei wurde berücksichtigt, daß der Variationskoeffizient  $c_{A2}$  gleich eins zu setzen ist, da der Strom ankommender Anforderungen als Poisson-Prozeß vorausgesetzt ist.

Ein interessanter Aspekt ergibt sich, wenn man die Beziehungen für die Größen  $h_2^*$ ,  $\rho_2^*$  und  $c_{H2}^*$  in die Gleichung (3.1,2a) einsetzt. Letztere geht im Falle  $c_A = 1$  in die exakt geltende Pollaczek-Khinchin-Formel für die mittlere Wartezeit im Warteschlangenmodell M/G/1 über. Es folgt dann nach kurzer Rechnung

$$w_2^* = \frac{h_1 \rho_1}{(1-\rho_1) \cdot (1-\rho_1-\rho_2)} \cdot \frac{\rho_2}{1-\rho_1} + \frac{h_2 \rho_2}{(1-\rho_1) \cdot (1-\rho_1-\rho_2)}. \quad (4.1,21)$$

Für  $\rho_1 + \rho_2 = 1$  ist  $\rho_2 / (1-\rho_1) = 1$ , und durch Vergleich von (4.1,21) mit (3.1,9) erkennt man:

$$w_2 = w_2^* \quad \text{für} \quad \rho_1 + \rho_2 = 1. \quad (4.1,22)$$

Diese durch eine sehr einfache Rechnung gewonnene Erkenntnis stimmt mit dem Ergebnis von Kaufman überein (siehe Abschnitt 3.2.3), welcher in [Kaufman 82] festgestellt hat, daß die

mittlere Komplettierungszeit nur für den Fall der Vollast mit der Zeit übereinstimmt, die eine Anforderung zweiter Klasse als ihre mittlere Bedienungszeit empfindet, wenn sie die Unterbrechungszeiten durch die erste Klasse "nicht bemerkt". Auch die Beziehung (4.1,15) aus Abschnitt 4.1.1 liefert diese Erkenntnis.

Der zweite Ansatz zur Approximation des Variationskoeffizienten  $c_{D2}$  lautet

$$h_2^* = \frac{h_2}{1-\rho_1}, \quad (4.1,23a)$$

$$\rho_2^* = \frac{\rho_2}{1-\rho_1}, \quad (4.1,23b)$$

$$c_{H2}^{*2} = 1 + 2 \cdot \frac{h_1 \rho_1}{h_2 \rho_2}. \quad (4.1,23c)$$

Er ist so gewählt, daß sich das erste Moment für die Wartezeit und die Verweilzeit beim Einsetzen der Gleichungen in die Pollaczek-Khinchin-Formel jeweils richtig reproduziert. Es gilt also  $w_2^* = w_2$  und  $f_2^* = f_2$ . Einsetzen von (4.1,23b-c) in (3.1,5) ergibt für  $c_{A2} = 1$ :

$$c_{D2}^2 = 1 + 2 \cdot \frac{h_1}{h_2} \cdot \frac{\rho_1 \rho_2}{(1-\rho_1)^2}. \quad (4.1,24)$$

Bei einer Diskussion der beiden Approximationen ist zunächst einmal festzustellen, daß sowohl (4.1,20) als auch (4.1,24) die Ungleichung

$$c_{D2}^2 \geq 1 \quad (4.1,25)$$

liefern. Dies ist eine richtige Beschreibung des Sachverhalts, daß wegen der größeren "Streuung" der virtuellen Bedienungszeiten die VFn der Ausgangsabstände  $T_{D2}$  hyperexponentiellen Charakter haben. Um eine fundiertere Beurteilung der Approximationsformeln vornehmen zu können, wurde deren Tauglichkeit mit Hilfe von Simulationsstudien überprüft. Einige Ergebnisse

sind in der Tabelle 4.1 wiedergegeben. Zur Unterscheidung der Verfahren ist dort von der "c t a" (completion time approximation) die Rede, falls (4.1,20) angewendet wurde und von der "f t a" (flow time approximation), falls (4.1,24) angewendet wurde. Diese Abkürzungen werden auch im weiteren Verlauf der Arbeit beibehalten.

Bei Betrachtung der Ergebnisse sieht man, daß die f t a der c t a fast immer überlegen ist. Das letztgenannte Approximationsverfahren versagt offenbar in den Fällen wo die mittleren Bedienungszeiten der zweiten Klasse wesentlich kleiner sind als diejenigen der ersten Klasse. Hieran erkennt man auch, daß nicht nur die Auslastungen  $\rho_1$  und  $\rho_2$  alleine einen wesentlichen Einfluß auf die Güte der Approximationen haben, sondern auch das Verhältnis  $h_1/h_2$  (vgl. auch [Kaufman 82]). Aus diesem Grunde sind bei den in dieser Arbeit zu findenden Ergebnissen die Parameter  $\lambda_1$ ,  $\lambda_2$ ,  $h_1$  und  $h_2$  stets getrennt angegeben.

Schließlich ist festzustellen, daß für den in der Praxis oft vorliegenden Fall  $h_1 \leq h_2$  der Variationskoeffizient  $c_{D2}$  nur relativ wenig von eins abweicht und der Ausgangsprozess dann näherungsweise als Poisson-Prozess angesehen werden kann.

4.2 Warteschlangenmodell M/M/1/NONPRE

4.2.1 Zustandsverteilungen

Dem Abschnitt 4.1.1 für das Modell mit unterbrechenden Prioritäten entsprechend dienen auch die nachfolgenden Untersuchungen der Randverteilungen  $p_1(m)$  und  $p_2(n)$  dem Ziel, Hinweise zur Konstruktion einklassiger Ersatzmodelle zu finden. Dabei ist zu bemerken, daß die zur Erreichung des Ziels angewendete Vorgehensweise im Falle endlicher Speicherkapazitäten  $s_1$  und  $s_2$  Ergebnisse liefert, die zu den hier vorgestellten äquivalent sind. Von dieser Tatsache wird wieder im nächsten Abschnitt Gebrauch gemacht, wo es um die Berechnung von Blockierungswahrscheinlichkeiten geht. Die Wahrscheinlichkeiten  $p_1(m)$  und  $p_2(n)$  sind durch die Gleichungen

		$c_{D2}$								
$\lambda_1$	$\lambda_2$	$h_1$	$h_2$	$\rho$	Simulation 95% Vertrauensinterv.	cta	rel. Fehler	fta	rel. Fehler	
0.1	0.6	1.0	1.0	0.7	1.0611 ± 0.0067	1.0482	-0.0122	1.0715	0.0098	
0.2	0.6	1.0	1.0	0.8	1.1638 ± 0.0060	1.1319	-0.0274	1.1726	0.0076	
0.3	0.6	1.0	1.0	0.9	1.3016 ± 0.0195	1.2766	-0.0192	1.3171	0.0119	
0.1	0.6	0.5	1.0	0.65	1.0126 ± 0.0068	1.0104	-0.0022	1.0165	0.0039	
0.3	0.6	0.5	1.0	0.75	1.0476 ± 0.0088	1.0430	-0.0044	1.0605	0.0123	
0.5	0.6	0.5	1.0	0.85	1.1223 ± 0.0080	1.1015	-0.0185	1.1255	0.0029	
0.7	0.6	0.5	1.0	0.95	1.2207 ± 0.0120	1.2078	-0.0106	1.2235	0.0023	
0.1	0.6	1.0	0.5	0.4	1.0544 ± 0.0054	1.0244	-0.0285	1.0715	0.0162	
0.3	0.6	1.0	0.5	0.6	1.2774 ± 0.0086	1.1467	-0.1023	1.3171	0.0311	
0.5	0.6	1.0	0.5	0.8	1.8246 ± 0.0425	1.5620	-0.1439	1.8439	0.0106	
0.6	0.6	1.0	0.5	0.9	2.2970 ± 0.0461	2.0917	-0.0894	2.3452	0.0210	

Tabelle 4.1: Approximation des Variationskoeffizienten  $c_{D2}$  der Ausgangsabstände  $T_{D2}$  für die zweite Klasse im Warteschlangenmodell M/M/1/PRE

$$p_1(m) := \sum_{n=0}^{\infty} p(m,n,1) + \sum_{n=1}^{\infty} p(m,n,2) + \delta_{0m} p(0,0,0), \quad (m \geq 0), \quad (4.2,1a)$$

$$p_2(n) := \sum_{m=1}^{\infty} p(m,n,1) + \sum_{m=0}^{\infty} p(m,n,2) + \delta_{0n} p(0,0,0), \quad (n \geq 0), \quad (4.2,1b)$$

definiert, wobei vereinbart sei, daß  $p(m,n,1)=0$  für  $m=0$  und  $p(m,n,2)=0$  für  $n=0$  gelte. Darüber hinaus erweist es sich als nützlich, die Abkürzungen

$$p(m, \cdot, s) := \sum_{n=0}^{\infty} p(m,n,s), \quad (m \geq 0; s=1,2), \quad (4.2,2a)$$

$$p(\cdot, n, s) := \sum_{m=0}^{\infty} p(m,n,s), \quad (n \geq 0; s=1,2), \quad (4.2,2b)$$

einzuführen.

Wird in den Gleichgewichtsgleichungen (3.1,18a-i) über  $n$  summiert, so liefert eine umfangreichere Rechnung:

$$\lambda_1 p_1(m-1) - \lambda_1 p(m) = \mu_1 \sum_{n=0}^{\infty} p(m,n,1) - \mu_1 \sum_{n=0}^{\infty} p(m+1,n,1), \quad (m \geq 1), \quad (4.2,3a)$$

$$\lambda_1 p_1(0) = \mu_1 \sum_{n=0}^{\infty} p(1,n,1). \quad (4.2,3b)$$

Hieraus folgt der Zusammenhang

$$p_1(m) = \rho_1^{-1} \cdot p(m+1, \cdot, 1), \quad (m \geq 0), \quad (4.2,4)$$

dessen Gültigkeit man durch Einsetzen in (4.2,3a-b) leicht bestätigt. Im Gegensatz zum Modell M/M/1/PRE, wo die entsprechende Gleichung (4.1,7) einen wertvollen Ansatz zur effizienten Bestimmung der Randverteilung  $p_2(n)$  lieferte, ist die Berechnung der Wahrscheinlichkeiten  $p(m, \cdot, 1)$  hier nicht so leicht möglich, so daß man die Randverteilung  $p_1(m)$  besser mittels der Formel (3.1,19) bestimmt. Die Beziehung (4.2,4) spielt jedoch eine wesentliche Rolle bei der im nächsten Abschnitt besprochenen Methode zur Approximation von Blockie-

rungrwahrscheinlichkeiten.

In diesem Abschnitt interessieren die folgenden, durch einfache Erweiterungen und Umformungen gewonnenen Gleichungen

$$0 = \lambda_1 p_1(m-1) - \left[ \lambda_1 + \mu_1 \frac{p(m, \cdot, 1)}{p_1(m)} \right] p_1(m) + \mu_1 \frac{p(m+1, \cdot, 1)}{p_1(m+1)} p_1(m+1), \quad (m \geq 1),$$

$$0 = -\lambda_1 p_1(0) + \mu_1 \frac{p(1, \cdot, 1)}{p_1(1)} p_1(1).$$

Mit Hilfe der Definitionen

$$\mu_1^*(m) := c_1(m) \cdot \mu_1, \quad (m \geq 1), \quad (4.2,5a)$$

$$c_1(m) := \frac{p(m, \cdot, 1)}{p_1(m)}, \quad (m \geq 1), \quad (4.2,5b)$$

wird daraus:

$$0 = \lambda_1 p_1(m-1) - (\lambda_1 + \mu_1^*(m)) p_1(m) + \mu_1^*(m+1) p_1(m+1), \quad (m \geq 1), \quad (4.2,6a)$$

$$0 = -\lambda_1 p_1(0) + \mu_1^*(1) p_1(1). \quad (4.2,6b)$$

Diese Gleichungen sind wieder sofort als die Gleichgewichtsgleichungen eines Warteschlangenmodells M/M/1 mit Ankunftsrate  $\lambda_1$  und zustandsabhängiger Bedienungsrate  $\mu_1^*(m)$  zu identifizieren, und es gilt

$$p_1(m) = \frac{1}{c_1(m) \mu_1} \cdot p_1(m-1), \quad (m \geq 1). \quad (4.2,7)$$

Vor der Zusammenstellung einiger Eigenschaften der Größen  $c_1(m)$  werden nun ohne weiteren Kommentar die entsprechenden Beziehungen für die zweite Klasse angegeben:

$$\lambda_2 p_2(n-1) - \lambda_2 p_2(n) = \mu_2 \sum_{m=0}^{\infty} p(m,n,2) - \mu_2 \sum_{m=0}^{\infty} p(m,n+1,2), \quad (n \geq 1), \quad (4.2,8a)$$

$$\lambda_2 p_2(0) = \mu_2 \sum_{m=0}^{\infty} p(m,1,2) \quad (4.2,8b)$$

$$p_2(n) = \rho_2^{-1} p(\cdot, n+1, 2), \quad (n \geq 0), \quad (4.2,9)$$

$$\mu_2^*(n) := c_2(n) \cdot \mu_2, \quad (n \geq 1), \quad (4.2,10a)$$

$$c_2(n) := \frac{p(\cdot, n, 2)}{p_2(n)}, \quad (n \geq 1), \quad (4.2,10b)$$

$$0 = \lambda_2 p_2(n-1) - (\lambda_2 + \mu_2^*(n)) p_2(n) + \mu_2^*(n+1) p_2(n+1), \quad (n \geq 1), \quad (4.2,11a)$$

$$0 = -\lambda_2 p_2(0) + \mu_2^*(1) p_2(1). \quad (4.2,11b)$$

Eigenschaften der Koeffizienten  $c_r(x_r)$ ,  $(r=1,2)$

i) Es gelten die alternativen Schreibweisen

1. Klasse:

$$c_1(m) = 1 - \frac{P\{X_1=m, X_2>0, S=2\}}{P\{X_1=m\}}, \quad (m \geq 1), \quad (4.2,12a)$$

$$c_1(m) = 1 - P\{X_2>0, S=2 | X_1=m\}, \quad (m \geq 1), \quad (4.2,12b)$$

$$c_1(m) = P\{S=1 | X_1=m\}, \quad (m \geq 1). \quad (4.2,12c)$$

2. Klasse:

$$c_2(n) = 1 - \frac{P\{X_1>0, X_2=n, S=1\}}{P\{X_2=n\}}, \quad (n \geq 1) \quad (4.2,13a)$$

$$c_2(n) = 1 - P\{X_1>0, S=1 | X_2=n\}, \quad (n \geq 1) \quad (4.2,13b)$$

$$c_2(n) = P\{S=2 | X_2=n\}, \quad (n \geq 1). \quad (4.2,13c)$$

ii) Aus i) folgen die Schranken

$$0 \leq c_1(m) \leq 1, \quad (4.2,14)$$

$$0 \leq c_2(n) \leq 1. \quad (4.2,15)$$

Demnach bewirken die Koeffizienten  $c_r(x_r)$  auch in diesem Falle eine zustandsabhängige Reduktion der Bedienungsraten  $\mu_r$ .

iii) Für die "Mittelwerte" der Reduktionskoeffizienten rechnet man aus:

$$\bar{c}_1 := \sum_{m=1}^{\infty} c_1(m) \cdot p_1(m) = \rho_1 \text{ für } c_1(0) \stackrel{!}{=} 0, \quad (4.2,16)$$

$$\bar{c}_2 := \sum_{n=1}^{\infty} c_2(n) \cdot p_2(n) = \rho_2 \text{ für } c_2(0) \stackrel{!}{=} 0. \quad (4.2,17)$$

Abschließend seien noch die im Anhang 1 hergeleiteten Gleichungen

$$\rho_2(\lambda_2 + \mu_2) p_2(n) - \rho_2 \lambda_2 p_2(n-1) = (\lambda + \mu_2) p(0, n+1, 2) - \lambda_2 p(0, n, 2), \quad (n \geq 1), \quad (4.2,18a)$$

$$\rho_2(\lambda_2 + \mu_2) p_2(0) = (\lambda + \mu_2) p(0, 1, 2), \quad (4.2,18b)$$

notiert, welche eine effiziente Berechnung der Randverteilung  $p_2(n)$  über die Wahrscheinlichkeiten  $p(0, n, 2)$  erlauben. Ein Verfahren zur Bestimmung der letzteren ist ebenfalls im Anhang 1 angegeben.

4.2.2 Blockierungswahrscheinlichkeiten

Für das Warteschlangenmodell M/M/1- $s_1, s_2$ /NONPRE mit endlichen Speicherkapazitäten  $s_1$  und  $s_2$  sind die Blockierungswahrscheinlichkeiten  $B_r$  der Klasse  $r$ ,  $(r=1,2)$ , durch die Gleichungen

$$B_1 = P\{(\tilde{X}_1 = s_1 + 1, S=1) \cup (\tilde{X}_1 = s_1, S=2)\}, \quad (4.2,19a)$$

$$B_2 = P\{(\tilde{X}_2 = s_2 + 1, S=2) \cup (\tilde{X}_2 = s_2, S=1)\}, \quad (4.2,19b)$$

bestimmt, in denen  $\tilde{X}_r$  die Anzahl von Anforderungen der Klasse  $r$  in diesem Modell angibt. Wird berücksichtigt, daß die Be-

ziehungen (4.2,4) und (4.2,9) auch für das Modell mit endlichen Speichern (Kennzeichnung der Zustandswahrscheinlichkeiten durch  $\tilde{\phantom{x}}$ ) gültig sind, so folgt:

$$B_1 = (1+\rho_1)\tilde{p}_1(s_1) - \rho_1\tilde{p}_1(s_1-1), \quad (4.2,20a)$$

$$B_2 = (1+\rho_2)\tilde{p}_2(s_2) - \rho_2\tilde{p}_2(s_2-1). \quad (4.2,20b)$$

Mit den Überlegungen aus Abschnitt 4.1.2 wird schließlich:

$$B_1 = \frac{1}{1-\sigma_1} [(1+\rho_1)p_1 + \rho_1p_1(s_1-1)], \quad (4.2,21a)$$

$$B_2 = \frac{1}{1-\sigma_2} [(1+\rho_2)p_2 + \rho_2p_2(s_2-1)], \quad (4.2,21b)$$

$$1-\sigma_r = \sum_{x_r=0}^{s_r+1} p_r(x_r), \quad (r=1,2). \quad (4.2,21c)$$

Damit steht auch im Falle unterbrechender Prioritäten ein effizientes Verfahren zur Berechnung der Blockierungswahrscheinlichkeiten  $B_1$  und  $B_2$  zur Verfügung.

#### 4.2.3 Ausgangsprozesse

Auch beim Warteschlangenmodell M/M/1/NONPRE erweist sich die analytische Untersuchung und Beschreibung von Ausgangsprozessen als außerordentlich schwierig. Als vermutlich einziges exaktes Ergebnis ist zu nennen, daß der allgemeine Ausgangsprozeß im Falle gleicher Bedienungsraten für beide Klassen ein Poisson-Prozeß ist. Wegen dieses Mangels an Informationen wurden in [Schmitt, Carl 83] u.a. eine Reihe von Simulationsstudien zur Ermittlung der VFn bzw. Variationskoeffizienten für die Ausgangsabstände  $T_{D1}$ ,  $T_{D2}$  und  $T_D$  durchgeführt. Wenngleich weder VFn noch die daraus ableitbaren Momente Abhängigkeiten erfassen, welche zwischen den Bedienungsenden von Anforderungen verschiedener Klassen bestehen, so sind sie doch ein wesentliches Hilfsmittel zur Charakterisierung der verschiedenen Ausgangsprozesse. In Tabelle 4.2 sind einige Ergebnisse für die Variationskoeffizienten  $c_{D1}$ ,  $c_{D2}$  und  $c_D$  zusammengestellt.

$\lambda_1$	$\lambda_2$	$h_1$	$h_2$	$\rho$	$c_{D1}$	$c_{D2}$	$c_D$
0.1	0.6	1.0	1.0	0.7	1.0029 ± 0.0146	1.0072 ± 0.0042	0.9990 ± 0.0042
0.2	0.6	1.0	1.0	0.8	1.0196 ± 0.0095	1.0539 ± 0.0081	0.9986 ± 0.0050
0.3	0.6	1.0	1.0	0.9	1.0370 ± 0.0072	1.1521 ± 0.0110	1.0000 ± 0.0052
0.1	0.6	0.5	1.0	0.65	1.0046 ± 0.0141	0.9838 ± 0.0038	1.0127 ± 0.0034
0.3	0.6	0.5	1.0	0.75	1.0417 ± 0.0067	0.9685 ± 0.0056	1.0446 ± 0.0064
0.5	0.6	0.5	1.0	0.85	1.0933 ± 0.0092	0.9734 ± 0.0062	1.0724 ± 0.0044
0.6	0.6	0.5	1.0	0.90	1.1217 ± 0.0063	0.9914 ± 0.0079	1.0838 ± 0.0051
0.1	0.6	1.0	0.5	0.4	1.0017 ± 0.0176	1.0314 ± 0.0070	1.0150 ± 0.0055
0.3	0.6	1.0	0.5	0.6	1.0020 ± 0.0054	1.2048 ± 0.0153	1.0456 ± 0.0044
0.5	0.6	1.0	0.5	0.8	1.0028 ± 0.0086	1.7452 ± 0.0242	1.0740 ± 0.0061
0.6	0.6	1.0	0.5	0.9	1.0113 ± 0.0087	2.2114 ± 0.0489	1.0842 ± 0.0043

Tabelle 4.2: Simulationsergebnisse für den Variationskoeffizienten der Ausgangsabstände der verschiedenen Klassen im Warteschlangenmodell M/M/1/NONPRE. (95% Vertrauensintervalle)

Bei einer Diskussion der Ergebnisse ist zunächst festzustellen, daß die Zeitspannen  $T_D$  in allen Fällen als  $(\lambda_1 + \lambda_2)$ -exponentialverteilt angesehen werden können ( $c_D \approx 1$ ). Entsprechendes gilt in etwas abgeschwächter Form für die Ausgangsabstände  $T_{D1}$  der ersten Klasse. Hier macht sich für  $h_1 = 0.5 \cdot h_2$  eine leichte Tendenz zur hyperexponentiellen VF bemerkbar ( $c_{D1} > 1$ ). Von den Zwischenzeiten  $T_{D2}$  kann man nur in den Fällen  $h_2 = 2 \cdot h_1$  und (mit gewissen Einschränkungen)  $h_2 = h_1$  annehmen, daß sie in guter Näherung  $\lambda_2$ -exponentialverteilt sind. Für  $h_2 = 0.5 \cdot h_1$  ist die Beschreibung durch eine  $H_2$ -Verteilung besser, wie eingehendere Untersuchungen gezeigt haben.

Es sei zum Abschluß bemerkt, daß verschiedene Versuche, die Variationskoeffizienten  $c_{D1}$  und  $c_{D2}$  in einer ähnlichen Weise zu approximieren, wie dies im Abschnitt 4.1.3 für unterbrechende Prioritäten erfolgte, zu keinen befriedigenden Ergebnissen geführt haben.

#### 4.3 Einklassige Ersatzmodelle mit zustandsabhängiger Bedienungsrate

Das für die Belange dieser Arbeit wichtigste Ergebnis der Untersuchungen über Zustandsverteilungen in Markoffschen Warteschlangenmodellen mit Prioritätsabfertigung ist, daß man für jede Klasse jeweils ein Modell des Typs M/M/1 angeben kann, dessen Zustandswahrscheinlichkeiten mit den im Prioritätsmodell gültigen Wahrscheinlichkeiten  $p_r(x_r)$  für die Zustände  $x_r$  der Klasse  $r$ , ( $r=1,2$ ), identisch sind. Die im allgemeinen zustandsabhängigen Bedienungsrate  $\mu_r^*(x_r)$  können sowohl für unterbrechende als auch für nichtunterbrechende Prioritäten in der einheitlichen Form

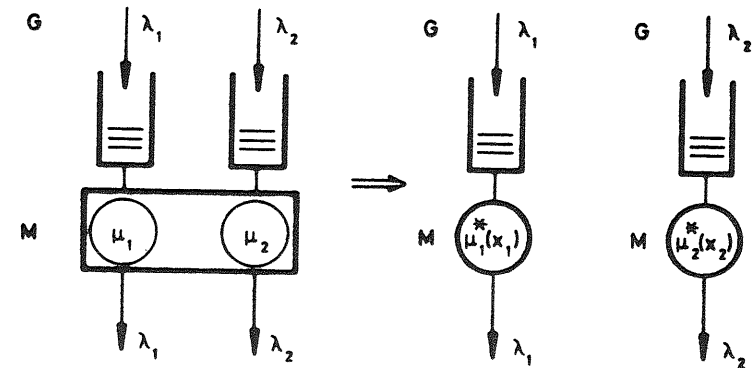
$$\mu_r^*(x_r) = c_r(x_r) \cdot \mu_r, \quad (x_r \in \mathbb{N}; r = 1, 2), \quad (4.3, 1a)$$

$$c_r(x_r) = P \{ \text{Anforderung der Klasse } r \text{ in Bedienung} \mid x_r \text{ Anforderungen der Klasse } r \text{ im Modell} \}, \quad (x_r \in \mathbb{N}; r = 1, 2), \quad (4.3, 1b)$$

geschrieben werden, was durch Vergleich mit den vorher gege-

benen Darstellungsweisen leicht zu bestätigen ist. Insbesondere liefert (4.3, 1b) bei unterbrechenden Prioritäten für die erste Klasse das Ergebnis  $c_1(x_1) \equiv 1$ .

Obwohl die Beziehungen (4.3, 1a-b) unter der Voraussetzung von Poisson-Ankunftsprozessen hergeleitet wurden, motivieren sie auch bei nichtmarkoffschen Ankunftsströmen zu der Zerlegung eines zweiklassigen Warteschlangenmodells mit Prioritätsabfertigung in zwei einklassige Warteschlangenmodelle ohne Prioritätsabfertigung entsprechend Bild 4.1. Diese Vorgehensweise wird in Abschnitt 5.2.1 des nächsten Kapitels noch deutlich untermauert werden.



$$\mu_r^*(x_r) := c_r(x_r) \cdot \mu_r,$$

$$c_r(x_r) := P \{ \text{Anforderung der Klasse } r \text{ in Bedienung} \mid x_r \text{ Anforderungen der Klasse } r \text{ im Modell} \},$$

$$(x_r \in \mathbb{N}; r = 1, 2).$$

Bild 4.1: Ersatzmodelle mit zustandsabhängig reduzierten Bedienungsrate für ein zweiklassiges Warteschlangenmodell mit Prioritätsabfertigung (PRE oder NONPRE)

In logischer Weiterführung der weiter oben angegebenen Bezeichnungen  $r o a$  und  $m-r o a$  wird das hier vorgestellte neue Zerlegungsverfahren als  $s d r o a$  (state dependent reduced occupancy approximation) bezeichnet.

Bemerkungen

- i) Für Poisson-Eingangsprozesse liefern nach [Muntz 72] die Ersatzmodelle auch stets Poisson-Ausgangsprozesse. Nach den vorher angegebenen Ergebnissen für Ausgangsprozesse in Prioritätsmodellen wird hieran deutlich, daß die Äquivalenz der Ersatzmodelle im allgemeinen nur in Bezug auf die Zustandsverteilungen gesichert ist.
- ii) Obwohl den Verfahren  $r o a$ ,  $m-r o a$  und  $s d r o a$  sehr unterschiedliche Beobachtungen zugrunde liegen, bestehen zwischen diesen Verfahren Zusammenhänge. Letztere werden durch die nachfolgende Aufstellung der im Falle unterbrechender Prioritäten gültigen Formeln für die Reduktionskoeffizienten verdeutlicht.

$r o a : c_2 = 1 - \rho_1 = P\{X_1=0\}$

$m-r o a : c_2 = \frac{P\{X_1=0, X_2>0\}}{P\{X_2>0\}}$

$s d r o a : c_2(x_2) = \frac{P\{X_1=0, X_2=x_2\}}{P\{X_2=x_2\}}, (x_2 \geq 1)$

Ergänzend ist festzustellen, daß auch bei unterbrechenden Prioritäten das Verfahren  $m-r o a$  in das Verfahren  $s d r o a$  übergeht, wenn man die in den Gleichungen (3.2,12a-b) jeweils in Zähler und Nenner auftauchenden Ausdrücke  $X_1>0$  bzw.  $X_2>0$  durch  $X_1=x_1$  bzw.  $X_2=x_2$  ersetzt. Die dort im Zusammenhang mit WSNen vorkommenden Indices  $i$  wurden hier aus Gründen der Übersichtlichkeit weggelassen.

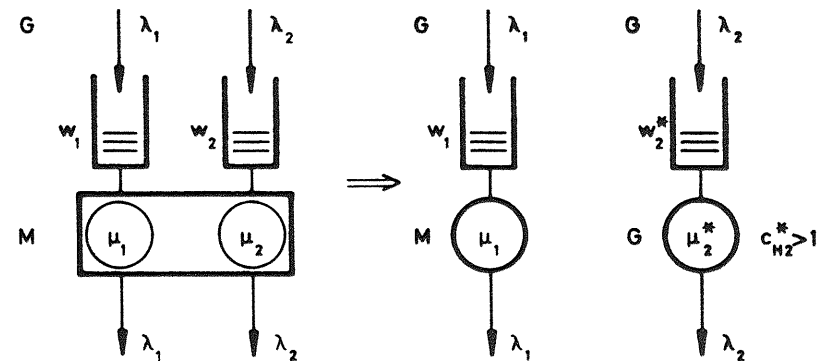
- iii) Die Koeffizienten  $c_r(x_r)$ , ( $r=1,2$ ), sind à priori unbekannt. D.h., man muß für ihre exakte Bestimmung erst das

Prioritätsmodell selbst analysieren (vgl. auch  $m-r o a$ ). Bei Anwendung der einklassigen Ersatzmodelle zur Analyse von WSNen, werden daher die Reduktionskoeffizienten durch geeignete Verfahren gegebenenfalls approximiert.

- iv) Die Formeln aus den Abschnitten 3.1.2 und 3.1.3 für die mittleren Verweilzeiten  $f_1$  und  $f_2$  sind auch für die Ersatzmodelle gültig. Dieses folgt aus der Gleichheit der Zustandsverteilungen und mit Little's Gesetz.

4.4 Einklassige Ersatzmodelle mit konstanter Bedienungsrate

Die teilweise sehr guten Ergebnisse, welche die Approximationsverfahren  $c t a$  und  $f t a$  aus Abschnitt 4.1.3 für den Variationskoeffizienten  $c_{D2}$  bei unterbrechenden Prio-



cta:  $\mu_2^* = (1 - \rho_1) \cdot \mu_2 ; c_{H2}^* = \sqrt{1 + 2 \frac{\mu_2}{\mu_1} \cdot \frac{\rho_1}{1 - \rho_1}}$

fta:  $\mu_2^* = (1 - \rho_1) \cdot \mu_2 ; c_{H2}^* = \sqrt{1 + 2 \frac{\mu_2}{\mu_1} \frac{\rho_1}{\rho_2}}$

Bild 4.2: Ersatzmodelle mit konstanten Bedienungsraten für ein zweiklassiges Warteschlangenmodell mit unterbrechenden Prioritäten

ritäten liefern, waren Anlaß für die Konstruktion der Ersatzmodelle nach Bild 4.2 . Dabei wurde davon ausgegangen, daß die Gründe für die Wahl der Variationskoeffizienten  $c_{H2}^*$  auch für nichtmarkoffsche Ankunftsprozesse ihre Gültigkeit behalten. Diese Annahme ist für die Methode c t a , wo  $c_{H2}^* = c_{C2}^*$  gesetzt ist, sogar richtig, da der Variationskoeffizient  $c_{C2}$  vom Typ des Ankunftsprozesses der zweiten Klasse unabhängig ist, wie ein Blick in den Beweis der Gleichung (3.1,13b) zeigt (siehe z.B. [Schassberger 73]).

Die Ersatzmodelle dieses Abschnitts sind darauf zugeschnitten, mit Hilfe des weiter oben skizzierten Momentenverfahrens von Kühn eine Analyse von WSNe mit unterbrechenden Prioritäten vornehmen zu können.

## 5. WARTESCHLANGENNETZE MIT PRIORITÄTEN

Hauptgegenstand dieses Kapitels ist die Beschreibung neu entwickelter Analyseverfahren für WSNe mit Prioritäten. Von diesen WSNe sei gefordert, daß sie außer BSen der Typen  $\cdot/M/1/PRE$  und/oder  $\cdot/M/1/NONPRE$  höchstens noch solche der im Kapitel 3 definierten Typen 1-4 enthalten dürfen. Bei offenen WSNe sollen außerdem die externen Ankunftsströme Poisson-Prozesse mit konstanten Ankunftsrate sein. Obwohl den folgenden Betrachtungen wieder das Vorhandensein zweier Prioritätsklassen zugrunde liegt, bieten sich in dieser Hinsicht Verallgemeinerungsmöglichkeiten an, auf die an gegebener Stelle hingewiesen werden wird.

### 5.1 Offene Warteschlangennetze

#### 5.1.1 Zustandsanalyse eines seriellen Warteschlangenmodells mit nichtunterbrechenden Prioritäten

Dieser Abschnitt enthält in knapper Form die Methodik für das Aufstellen und Lösen der Gleichgewichtsgleichungen, die den stationären Zustandsprozeß des in Bild 5.1 dargestellten seriellen Warteschlangenmodells  $M/M/1/NONPRE \rightarrow \cdot/M/1/NONPRE$  beschreiben. Schon bei einer solch einfachen Netzstruktur treten die wesentlichen Probleme auf, welche bei einer exakten Modellanalyse in vielen Fällen zu bewältigen sind.

Die ersten Schwierigkeiten bereitet das Aufstellen der Gleichgewichtsgleichungen und rühren daher, daß zur Beschreibung der Modellzustände sechs Variablen benötigt werden. Diese sind in dem 6-Tupel

$$(x_{11}, x_{12}, s_1, x_{21}, x_{22}, s_2)$$

zusammengefaßt. Darin geben die Komponenten  $x_{ir}$ , ( $x_{ir} \in \mathbb{N}_0$ ;  $i=1,2$ ;  $r=1,2$ ), an, wieviele Anforderungen der Klasse  $r$  sich in der BS  $i$  befinden. Die Größen  $s_i$ , ( $s_i \in \{0,1,2\}$ ), kennzeichnen wie im Abschnitt 3.1.3 ob bzw. durch welche Klasse die BE



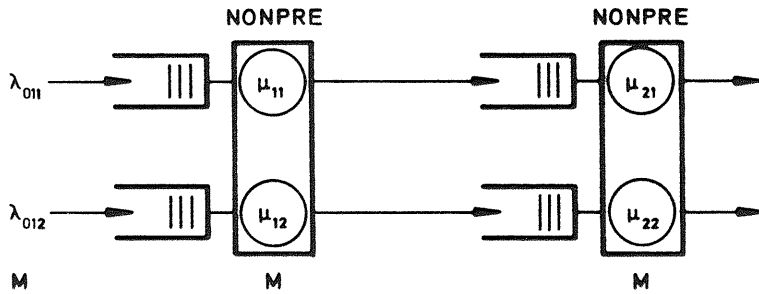
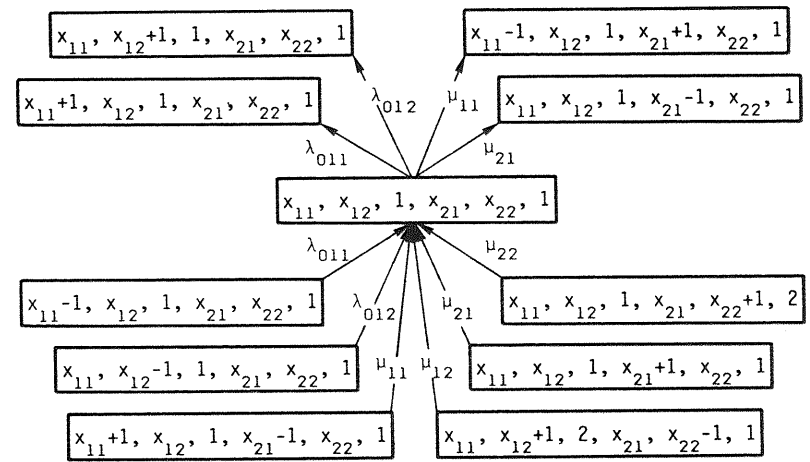


Bild 5.1: Serielles Warteschlangenmodell  $M/M/1/NONPRE \rightarrow *M/1/NONPRE$  mit zwei Prioritätsklassen

in der BS  $i$  belegt ist. Die Menge aller möglichen 6-Tupel bildet den (abzählbaren) Zustandsraum  $E$  des allgemeinen Zustandsprozesses. Eine geschlossene Darstellung des zugehörigen Übergangsgraphen, wie etwa bei den einstufigen Prioritätsmodellen, ist hier nicht möglich. Um dennoch eine gewisse Anschauungshilfe zu haben, hat es sich als sinnvoll erwiesen, eine Klasseneinteilung des Zustandsraumes vorzunehmen und für jede Klasse repräsentativ einen Ausschnitt aus dem Übergangsgraphen aufzuzeichnen, wie dies in Bild 5.2 exemplarisch gemacht wurde. Solch ein Ausschnitt enthält jeweils einen Repräsentanten der betrachteten Klasse sowie alle Zustände, die für diesen erreichbar sind (maximal vier) bzw. von denen aus er erreichbar ist (maximal sechs). Für weitere Einzelheiten zur Klasseneinteilung des Zustandsraumes sei auf den Anhang 2 dieser Arbeit und auf [Schmitt, Schneider 82] verwiesen.

Wegen der Komplexität der Gleichgewichtsgleichungen im Falle des vorliegenden Warteschlangenmodells erscheint es als ein aussichtsloses Unterfangen, mit einer der bekannten Methoden eine geschlossene Form der Lösung für dieses System gekoppelter, partieller Differenzgleichungen zu finden. Deshalb ist man darauf angewiesen, das dazu äquivalente homogene, lineare Gleichungssystem numerisch zu lösen. Letzteres umfaßt aufgrund der unendlich vielen Zustände auch unendlich viele Gleichungen zur Bestimmung der Zustandswahrscheinlichkeiten. Man setzt da-



$$x_{11} \geq 2, x_{12} \geq 1, x_{21} \geq 2, x_{22} \geq 1$$

Bild 5.2: Ausschnitt aus dem Übergangsgraphen für den allgemeinen Zustandsprozeß im Warteschlangenmodell  $M/M/1/NONPRE \rightarrow *M/1/NONPRE$

her üblicherweise

$$P\{X_{11} > \hat{x}_{11}, X_{12} > \hat{x}_{12}, S_1 = s_1, X_{21} > \hat{x}_{21}, X_{22} > \hat{x}_{22}, S_2 = s_2\} \stackrel{!}{=} 0,$$

$$(s_i \in \{0, 1, 2\}, i=1, 2).$$

Die Begrenzungswerte  $\hat{x}_{ir}$  hängen von den Auslastungen  $\rho_{ir}$  in den einzelnen Stufen ab und werden z.B. so gewählt, daß die aus den Zustandswahrscheinlichkeiten berechneten mittleren Wartezeiten  $\tilde{w}_{1r}$  in der BS 1 höchstens um ein vorgegebenes  $\epsilon > 0$  vom richtigen Wert  $w_{1r}$  abweichen. Dabei ist  $\hat{x}_{1r} = \hat{x}_{2r}$  anzunehmen. Definiert man nun

$$L := \max_{i,r} \{\hat{x}_{ir}\}, \quad (i=1, 2; r=1, 2),$$

so folgt damit für die Anzahl  $C$  der im Gleichungssystem zu berücksichtigenden Zustandswahrscheinlichkeiten

$$C = 4L^4 + 8L^3 + 8L^2 + 4L + 1.$$

Für L=20 erhält man z.B. C=707281. D.h., es wäre ein lineares Gleichungssystem mit dieser Anzahl von Unbekannten zu lösen. Dieses scheitert u.a. an der üblicherweise zur Verfügung stehenden Speicherkapazität der Rechenanlagen. Im Rahmen der Arbeit [Schmitt, Schneider 82] wurden die Zustandsgleichungen daher nur für L=12 mittels des Gauß-Seidel-Verfahrens (siehe z.B. [Törnig 79]) gelöst. Diesem Wert für L entsprechen C=97969 zu berechnende Zustandswahrscheinlichkeiten.

Einige Resultate für die mittleren Verweilzeiten  $f_r = f_{1r} + f_{2r}$  (r=1,2), im gesamten Modell, welche durch die hier skizzierte Vorgehensweise erzielt wurden und im numerischen Sinne als exakt gelten, sind zusammen mit approximativ gültigen Werten in der Tabelle 5.1 angegeben. Das den Näherungsergebnissen zugrunde liegende Verfahren ist in Abschnitt 5.1.3 beschrieben. Zu den Ergebnissen selbst sei an dieser Stelle lediglich bemerkt, daß die numerische Methode nur bis zu einer jeweiligen Auslastung der BEen von maximal 60 % (Fall 7) Ergebnisse liefert, bei denen die mit ihr berechneten mittleren Wartezeiten  $\tilde{w}_{1r}$  um weniger als 1 % von den exakten Werten  $w_{1r}$  abweichen. Für höhere Auslastungen wäre die Größe L entsprechend zu erhöhen.

Ein Resümee der aus diesem Beispiel gewonnenen Erkenntnisse zeigt, daß die besprochene Methode zur numerisch exakten Analyse von offenen WSnen mit Prioritäten zwei grundlegende Probleme enthält:

- i) Das Aufstellen der Gleichgewichtsgleichungen ist schon bei einfachen Strukturen sehr mühsam und nur schwer auf Fehler zu überprüfen.
- ii) Sind die Gleichgewichtsgleichungen bekannt, so ist man zur Bestimmung der Lösung auf numerische Methoden angewiesen. Deren Einsatz erfordert die Bereitstellung einer immens hohen Anzahl von Speicherplätzen im Rechner.

Fall	$\lambda_{011}$	$\lambda_{012}$	$h_{11}$	$h_{12}$	$h_{21}$	$h_{22}$	$f_1$		rel. Fehler	$f_2$		rel. Fehler
							exakt	approx.		exakt	approx.	
1	0.1	0.4	1.0	1.0	1.0	1.0	3.084	3.110	+0.84%	4.212	4.206	-0.14%
2	0.1	0.4	1.0	0.5	1.0	0.5	2.408	2.440	+1.33%	1.680	1.635	-2.68%
3	0.2	0.4	1.0	0.5	1.0	0.5	2.709	2.750	+1.51%	2.349	2.246	-4.40%
4	0.1	0.4	0.5	1.0	0.5	1.0	1.970	1.895	-3.81%	3.618	3.624	+0.17%
5	0.2	0.4	0.5	1.0	0.5	1.0	2.073	1.998	-3.61%	3.980	3.992	+0.30%
6	0.3	0.4	0.5	1.0	0.5	1.0	2.187	2.118	-3.16%	4.442	4.464	+0.50%
7	0.4	0.4	0.5	1.0	0.5	1.0	2.315	2.248	-2.89%	5.041	5.076	+0.69%

Tabelle 5.1: Numerische Ergebnisse für die mittleren Verweilzeiten  $f_1$  und  $f_2$  im seriellen Warteschlangensmodell M/M/1/NONPRE + \* / M/1/NONPRE. Die Näherungswerte wurden mit dem Verfahren m s d r o a berechnet.

zu i) Das erste Problem kann durch automatische Generierung der Gleichgewichtsgleichungen gelöst werden (siehe [Herfort, Müller-Clostermann 83]).

zu ii) Zur Lösung des zweiten Problems, das vor allem bei offenen WSNe eine große Rolle spielt, bietet es sich an, die im linearen Gleichungssystem berücksichtigten Zustandswahrscheinlichkeiten bei genügend hoher Ausnutzung nicht nur für große  $x_{ir}$ , sondern auch für kleine  $x_{ir}$  zu vernachlässigen. Damit würde nur die "Hauptmasse" der Wahrscheinlichkeiten eingerechnet. Für eine Einrechnung des durch die Vernachlässigung entstandenen Fehlers kann man ein Ergebnis aus [Takahashi 81] heranziehen. Dort wird u.a. gezeigt, daß die Zustandsverteilung im Warteschlangenmodell G/G/1 (G jeweils vom Phasentyp) für genügend große Zustandszahl  $x$  gegen eine geometrische Verteilung strebt.

Beide hier angedeuteten Verbesserungsmöglichkeiten erfordern jedoch bei entsprechender Vielfältigkeit der zugelassenen Netztypen einen ungemein hohen Programmieraufwand, der ihre Brauchbarkeit in Frage stellt. Wegen des Mangels anderer exakter Verfahren ist dieses wohl der Hauptgrund für die Entwicklung von Approximationsmethoden für die mathematische Analyse von WSNe mit Prioritäten.

5.1.2 Ein approximatives Momentenverfahren

Das nachfolgend erläuterte Approximationsverfahren ist nur im Falle unterbrechender Prioritäten anwendbar und macht von den einklassigen Ersatzmodellen des Abschnitts 4.4 Gebrauch. Es liefert in erster Linie Mittelwerte für die Verweilzeiten im Netz und enthält zwei Hauptschritte:

Schritt 1: Ersetzung aller BSen mit unterbrechenden Prioritäten durch Ersatzmodelle nach Abschnitt 4.4 (f t a oder c t a).

Schritt 2: Analyse des so entstandenen WSNe ohne Prioritätsklassen mit Hilfe des Momentenverfahrens von Kühn [Kühn 79].

Als einfaches Beispiel werde das serielle Warteschlangenmodell aus dem vorhergehenden Abschnitt jetzt für unterbrechende Prioritäten betrachtet. Es gelte für dieses Modell die Kurzbezeichnung  $M/M/1/PRE \rightarrow \cdot/M/1/PRE$ .

Die Durchführung des ersten Schrittes liefert das WSN nach Bild 5.3, welches zwei offene Ketten  $E_1$  und  $E_2$  enthält. Das zu einer Kette gehörende Teilnetz ist jeweils für sich analysierbar (Schritt 2). Hier interessiert vor allem die Analyse des Teilnetzes für die zweite Klasse. Mit den Ergebnissen der Abschnitte 3.1.1 (Gleichungen (3.1,2a-b), (3.1,3), (3.1,5)) und 4.4 (Bild 4.2) lauten die Gleichungen zur Bestimmung der mittleren Verweilzeit  $f_2$ :

$$f_2 = h_{12}^* + w_{12}^* + h_{22}^* + w_{22}^*$$

$$h_{i2}^* = \mu_{i2}^{*-1} = (1 - \rho_{i1})^{-1} \cdot h_{i2}, \quad (i=1,2),$$

$$\rho_{i1} = \lambda_{i1} \cdot h_{i1}, \quad \rho_{i2}^* = \lambda_{i2} \cdot h_{i2}^*, \quad (i=1,2),$$

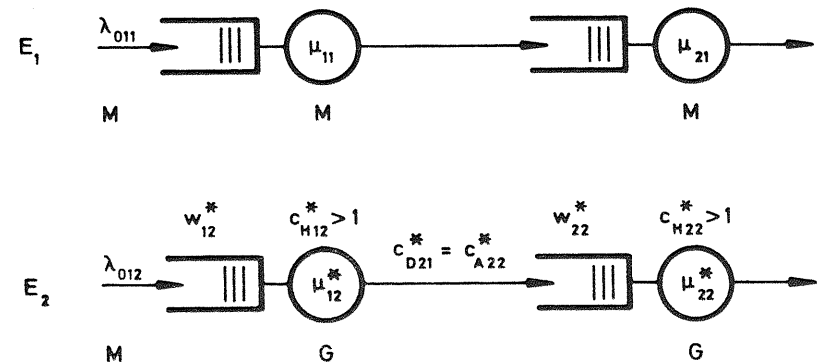


Bild 5.3: Ersatz-Warteschlangennetz für das zweiklassige, serielle Warteschlangenmodell  $M/M/1/PRE \rightarrow \cdot/M/1/PRE$

$$w_{12}^* = h_{12}^* \cdot \frac{\rho_{12}^*}{2(1-\rho_{12}^*)} \cdot (1+c_{H12}^{*2}),$$

$$w_{22}^* = h_{22}^* \cdot \frac{\rho_{22}^*}{2(1-\rho_{22}^*)} \cdot (c_{A22}^{*2} + c_{H22}^{*2}) \cdot g_{22},$$

$$g_{22} = \exp \left[ -(1-\rho_{22}^*) \cdot \frac{c_{A22}^{*2} - 1}{c_{A22}^{*2} + 4c_{H22}^{*2}} \right],$$

$$c_{A22}^{*2} = c_{D12}^{*2} = 1 + \rho_{12}^{*2} (c_{H12}^{*2} - 1),$$

$$c_{Hi2}^{*2} = 1 + 2 \cdot \frac{\mu_{i2}}{\mu_{i1}} \cdot \frac{\rho_{i1}}{1-\rho_{i1}}, \text{ falls c t a, } (i=1,2),$$

$$c_{Hi2}^{*2} = 1 + 2 \cdot \frac{\mu_{i2} \cdot \rho_{i1}}{\mu_{i1} \cdot \rho_{i2}}, \text{ falls f t a, } (i=1,2).$$

Es ist zu bemerken, daß bei WSNeN mit allgemeiner Struktur der Schritt 2 noch weitere Rechnungen erfordert, die im Zusammenhang mit Verzweigungen im Netz vor allem das Problem der Zerlegung und Überlagerung von sogenannten Erneuerungsprozessen betreffen. Hierzu sei jedoch auf [Kühn 79] verwiesen.

Die Bilder 5.4a-c enthalten verschiedene Kurven, welche mit dem Momentenverfahren berechnet wurden. Zum Vergleich sind zusätzlich einige Simulationsergebnisse eingetragen. Bei den gewählten Parametern zeigt sich, daß die f t a -Methode der c t a -Methode ungefähr gleichwertig ist, wenn  $h_{i1} = h_{i2}$  bzw.  $h_{i1} = 0.5 \cdot h_{i2}$ , ( $i=1,2$ ), gilt. Für  $h_{i1} = 2h_{i2}$  liefert die f t a -Methode jedoch bis auf den Bereich, in dem eine sehr hohe Gesamtauslastung der BEen vorliegt, die wesentlich besseren Resultate. Insgesamt gesehen überraschen die teilweise relativ großen Abweichungen gegenüber der Simulation. Diese waren aufgrund der überwiegend guten Ergebnisse bei der Approximation des Variationskoeffizienten  $c_{D2}$  (siehe Tabelle 4.1) nicht unbedingt zu erwarten.

Die Nachteile des Momentenverfahrens gegenüber der im näch-

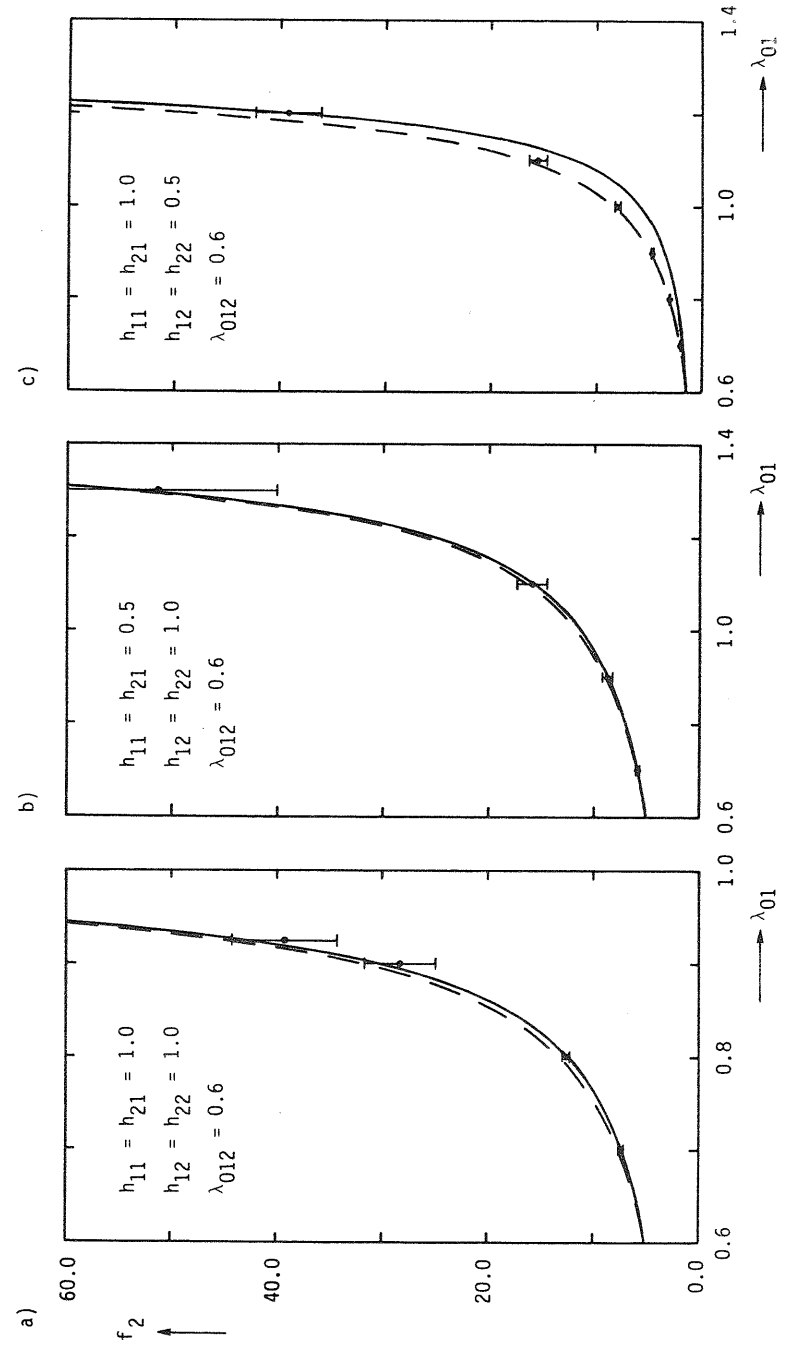


Bild 5.4: Mittlere Verweilzeit  $f_2 = f_{12} + f_{22}$  in Abhängigkeit von der Gesamtkunftsrate  $\lambda_{01} = \lambda_{011} + \lambda_{012}$

Modell: M/M/1/PRE + \*M/1/PRE. Analyseverfahren: — cta; - - - fta; I Simulation (95% Vertrauensinterv.).

sten Abschnitt vorgestellten Analyseverfahren liegen vor allem in der oft etwas schlechteren Genauigkeit der Ergebnisse, dem größeren Rechenaufwand und in der Tatsache, daß keine nichtunterbrechende Prioritäten berücksichtigt werden können. Dennoch ist anzunehmen, daß bei zukünftig zu entwickelnden Analyseverfahren für Prioritäten mit beliebigen VFN für die Bedienungszeiten Überlegungen eine Rolle spielen, die eine Charakterisierung der stochastischen Prozesse durch die beiden ersten Momente der betreffenden ZVen zum Ziel haben.

5.1.3 Approximation durch separable Warteschlangennetze

Das letzte in der vorliegenden Arbeit vorgestellte Analyseverfahren für offene WSNe mit Prioritäten basiert auf den im Abschnitt 4.3 beschriebenen Ersatzmodellen mit zustandsabhängigen Bedienungsraten. Deren Anwendung in einem WSN, das den zu Beginn dieses Kapitels genannten Voraussetzungen genügt, führt zu einem separablen WSN mit  $N^* > N$  BSen.

Als ein einfaches Beispiel betrachte man das serielle Warteschlangenmodell aus Abschnitt 5.1.1, welches durch die vorgeschlagene Zerlegung der Prioritätsmodelle in das Ersatz-

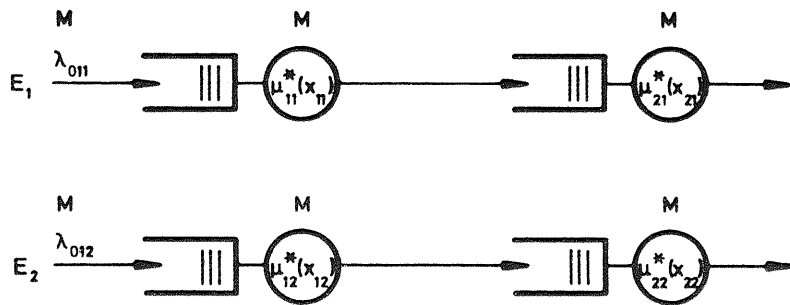


Bild 5.5: Ersatz-Warteschlangennetz für das zweiklassige, serielle Warteschlangenmodell  $M/M/1/NONPRE \rightarrow *M/1/NONPRE$

WSN nach Bild 5.5 übergeht. Wären darin die durch die Gleichungen (4.3,1a-b) definierten zustandsabhängigen Bedienungsraten  $\mu_{ir}^*(x_{ir}) = c_{ir}(x_{ir}) \cdot \mu_{ir}$ , ( $i=1,2; r=1,2$ ) bekannt, so könnten die unbekannt Zustandswahrscheinlichkeiten mit Hilfe des BCMP-Theorems leicht berechnet werden. Das verbleibende Problem besteht demnach in der Bestimmung der Koeffizienten  $c_{ir}(x_{ir})$ , die jeweils selbst durch eine bedingte Wahrscheinlichkeit definiert sind. Da jedoch eine exakte Berechnung der Reduktionskoeffizienten nicht möglich erscheint, ohne vorher die Gleichgewichtsgleichungen zu lösen, sind sowohl hier bei den offenen WSNen als auch im Abschnitt 5.2.3 bei den geschlossenen WSNen Näherungsannahmen zu treffen, die eine approximative Lösung des Problems ermöglichen.

Bei den offenen WSNen besteht der Näherungsansatz darin, in jeder BS  $i$  den Ankunftsprozeß einer jeden Klasse  $r$  als Poisson-Prozeß mit Parameter  $\lambda_{ir}$  anzunehmen. Die Ankunftsrate  $\lambda_{ir}$  sind dabei aus den Flußgleichungen (3.2,7a) bzw. (3.2,7b) exakt bestimmbar. Damit könnte jede BS isoliert analysiert und die gesuchten Koeffizienten  $c_{ir}(x_{ir})$  für eine genügend große Anzahl von Zuständen  $x_{ir} \in N_0$  approximativ bestimmt werden. Der nächste Schritt wäre, wie schon oben gesagt, die Analyse des Ersatz-WSNes durch Anwendung des BCMP-Theorems. Im allgemeinen wird man jedoch nicht in dieser Weise vorgehen müssen. Dies erklärt sich wie folgt: Die Berechnung der Zustandsverteilungen für die einzelnen BSen durch direkte Anwendung von Satz 3.1 führt zum gleichen Ergebnis wie die Analyse der Zustandsprozesse in den BSen in Isolation unter der jeweiligen Annahme eines Poisson-Ankunftsprozesses mit einer Ankunftsrate, die sich aus den Flußgleichungen ergibt (vgl. Bemerkung ii) zu Satz 3.1). Da die einklassigen Ersatzmodelle jedoch die Zustandsverteilungen der zweiklassigen Prioritätsmodelle exakt reproduzieren, ist es für eine Zustandsanalyse einfacher, die Prioritätsmodelle stufenweise mit den in Kapitel 3 und 4 vorgestellten Ergebnissen zu analysieren.

Insgesamt ergibt sich dann das folgende

Verfahren für die approximative Bestimmung von Zustandswahrscheinlichkeiten in offenen WSNe mit Prioritäten

Schritt 1: Ersetzung aller BSen des Typs  $\cdot/M/1/PRE$  und/oder  $\cdot/M/1/NONPRE$  durch Warteschlangenmodelle vom Typ  $\cdot/M/1$  mit à priori unbekanntem zustandsabhängigen reduzierten Bedienungsraten (siehe Abschnitt 4.3).

Schritt 2: Analyse des so entstandenen WSNe durch

i) Lösen der Flußgleichungen

$$\lambda_{js} = \lambda_{0js} + \sum_{(i,r) \in E_k} \lambda_{ir} \cdot p_{ir,js}, \quad ((j,s) \in E_k; k=1, \dots, K)$$

für jede Kette  $E_k$  des Ersatz-WSNe.

ii) Berechnung der gesuchten Zustandsverteilungen  $p_{ir}(x_{ir})$ , ( $i \in \{1, \dots, N^*\}$ ;  $r \in \{1, 2\}$ ), durch Analyse der betreffenden BSen. Die Ankunftsprozesse in diesen BSen sind näherungsweise als Poisson-Prozesse mit Parametern  $\lambda_{ir}$  anzunehmen.

Da das Verfahren auf der Zerlegungsmethode *s d r o a* beruht, wird es ebenso wie die hierzu analogen und an späterer Stelle besprochenen Analysemethoden für geschlossene WSNe als *m s d r o a* (modified state dependent reduced occupancy approximation) bezeichnet.

Bemerkungen

- i) Die Zustandswahrscheinlichkeiten für die Ersatzmodelle werden zweckmäßigerweise mit den Formeln (3.1,19) und (4.1,4a-b) bzw. mit den im Anhang 1 beschriebenen Algorithmen berechnet. Liegen endliche Speicherkapazitäten vor, so sind die Ergebnisse aus den Abschnitten 4.1.2 und 4.2.2 anwendbar.
- ii) Ist man nur an Mittelwerten interessiert, so können im

Schritt 2, ii) für die mittleren Verweilzeiten einer Anforderung in den Ersatzmodellen die Formeln (3.1,9), (3.1,13a) und (3.1,14) bzw. (3.1,15) und (3.1,16) angewendet werden. In diesem Fall ist auch eine Verallgemeinerung auf mehr als zwei Prioritätsklassen leicht möglich.

- iii) Das in Abschnitt 2.3.3 erwähnte Konzept der Einführung von Auftragsklassen zur möglichst wirklichkeitsgetreuen Modellierung der Verkehrslenkung ist auf das Ersatz-WSN aus Schritt 1 sofort anwendbar.
- iv) Der Schritt 2, i) ist auch für nichtmarkoffsche WSNe exakt. Insbesondere können mit den aus ihm resultierenden Ergebnissen die Auslastungen  $\rho_{ir} = \lambda_{ir} \cdot h_{ir}$ , ( $i=1, \dots, N$ ;  $r=1, \dots, R$ ) in jeder BS für jede Prioritätsklasse berechnet werden. Für Stabilität in einer BS muß die Bedingung (3.1,11) erfüllt sein.

Bei einer Diskussion des Verfahrens kann man einwenden, daß es die gleichen Ergebnisse liefert wie die in Abschnitt 3.2.3 als "knotenweise Zerlegung" bezeichnete Analysemethode. Dennoch bietet es in verschiedener Hinsicht Verbesserungen.

Diesbezüglich seien zuerst die neu entwickelten Algorithmen zur Berechnung von Zustands- und Blockierungswahrscheinlichkeiten genannt, welche vor allem eine Ersparnis des Bedarfs an Rechenzeit und Speicherplatz mit sich bringen. Beides sind Vorteile, die gerade bei der Analyse mehrstufiger Warteschlangenmodelle eine große Rolle spielen. Zweitens kann bei näherungsweise Berechnung der Reduktionskoeffizienten ein separables WSN angegeben werden, das, wie die späteren Ergebnisse zeigen werden, die Verhältnisse im ursprünglich gegebenen WSN zumindest hinsichtlich der Zustandsverteilungen gut approximiert. Von dieser Tatsache könnte man z.B. dann Gebrauch machen, wenn es um die Berechnung von VFn für die Verweilzeiten in WSNe mit Prioritäten geht und entsprechende Verfahren für separable WSNe (siehe z.B. [Kühn 83]) weiter entwickelt sind. Schließlich gestattet das hier vorgestellte Konzept die

Behandlung von offenen und geschlossenen WSNen nach einem gleichen, übergeordneten Prinzip, nämlich der Ersetzung von Prioritätsmodellen im WSN durch BSen mit zustandsabhängig reduzierten Bedienungsraten. Die letzteren werden allerdings nach unterschiedlichen Methoden approximiert.

Vor der Diskussion numerischer Ergebnisse sei noch eine Anmerkung zu der oben getroffenen Näherungsannahme gemacht. Die Approximation der Ankunftsprozesse in den BSen durch Poisson-Prozesse ist umso besser, je größer der Grad der Vermaschung des zu analysierenden WSNes ist und umso mehr Poisson-Quellen vorhanden sind. Diese Umstände bewirken, daß die Forderung nach unabhängigen Zuwächsen in den betreffenden Ankunftsprozessen (siehe Abschnitt 2.1.3, Forderung i)) weitgehend erfüllt ist. Rein serielle Warteschlangenmodelle der in Bild 5.1 gezeigten Art nehmen unter diesem Aspekt die Stellung eines "worst case" ein und dienen daher in dieser Arbeit oft als Testmodelle.

Abschließend werden in den Bildern 5.6 ... 5.9 einige numerische Ergebnisse für die seriellen Warteschlangenmodelle  $M/M/1/PRE \rightarrow \cdot/M/1/PRE$  und  $M/M/1/NONPRE \rightarrow \cdot/M/1/NONPRE$  vorgestellt. Dort sind jeweils die mittleren Verweilzeiten  $f_r = f_{1r} + f_{2r}$ , ( $r=1,2$ ), in Abhängigkeit von der gesamten Quellankunftsrate  $\lambda_{01} = \lambda_{011} + \lambda_{012}$  aufgetragen. Im Falle unterbrechender Prioritäten fehlen die Kurven für die erste Klasse, welche exakt berechnet werden können. Aus dem gleichen Grunde wurden bei dieser Unterbrechungsstrategie auch das Diagramm für den Fall gleicher mittlerer Bedienungszeiten für beide Klassen in beiden BSen (homogenes WSN) weggelassen. Für die Beurteilung des Näherungsverfahrens stehen Simulationswerte mit 95 % Vertrauensintervallen und die exakten Ergebnisse aus Tabelle 5.1 zur Verfügung, in der zusätzlich auch die zugehörigen Näherungswerte eingetragen sind.

Die vorgestellten Resultate zeigen, daß die relativen Fehler meistens weit unter 10 % liegen. Aus der Tabelle 5.1 wird insbesondere deutlich, daß im Falle  $h_{i1} = 0.5 \cdot h_{i2}$ , ( $i=1,2$ ), die Ergebnisse für die zweite Klasse erwartungsgemäß etwas genauer

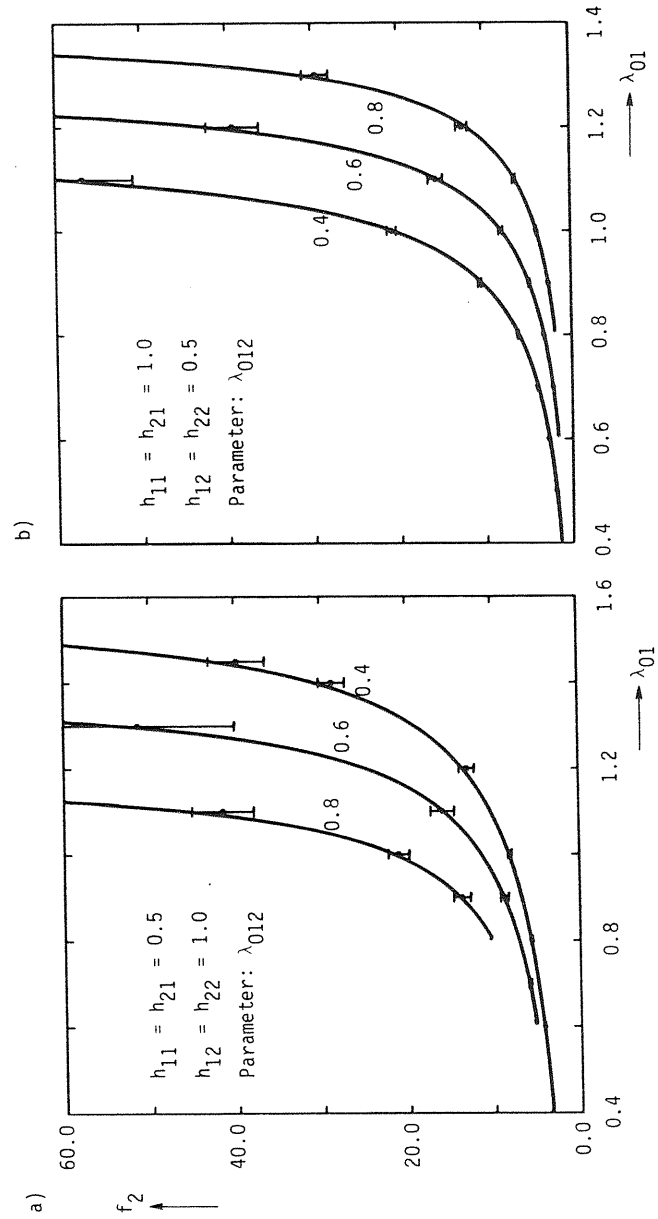
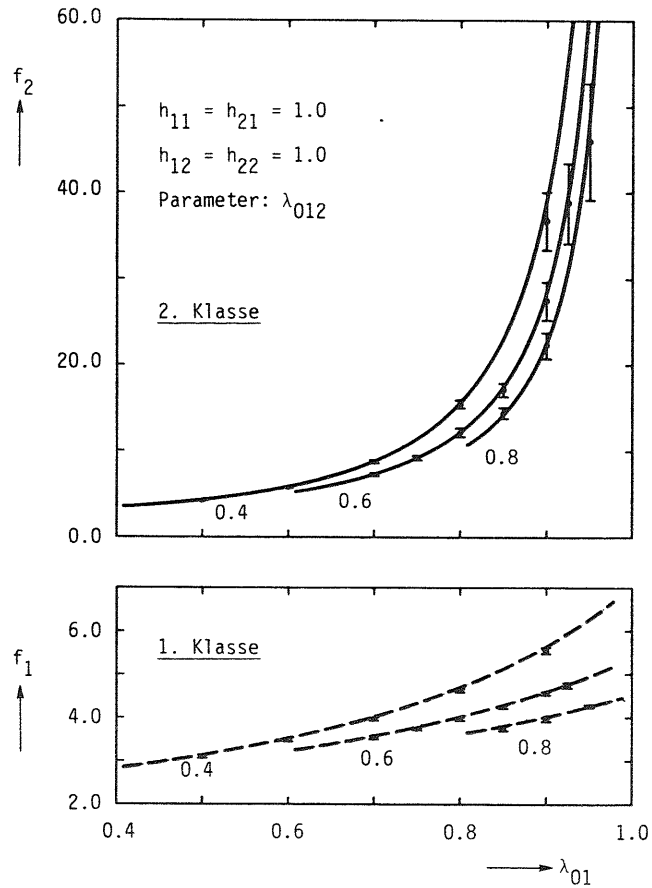


Bild 5.6: Mittlere Verweilzeit  $f_2 = f_{12} + f_{22}$  in Abhängigkeit von der Gesamtankunftsrate  $\lambda_{01} = \lambda_{011} + \lambda_{012}$ .  
 Modell:  $M/M/1/PRE \rightarrow \cdot/M/1/PRE$ . Analyseverfahren: ——— m s d r o a; [ Simulation (95% Vertrauensinterv.) ]

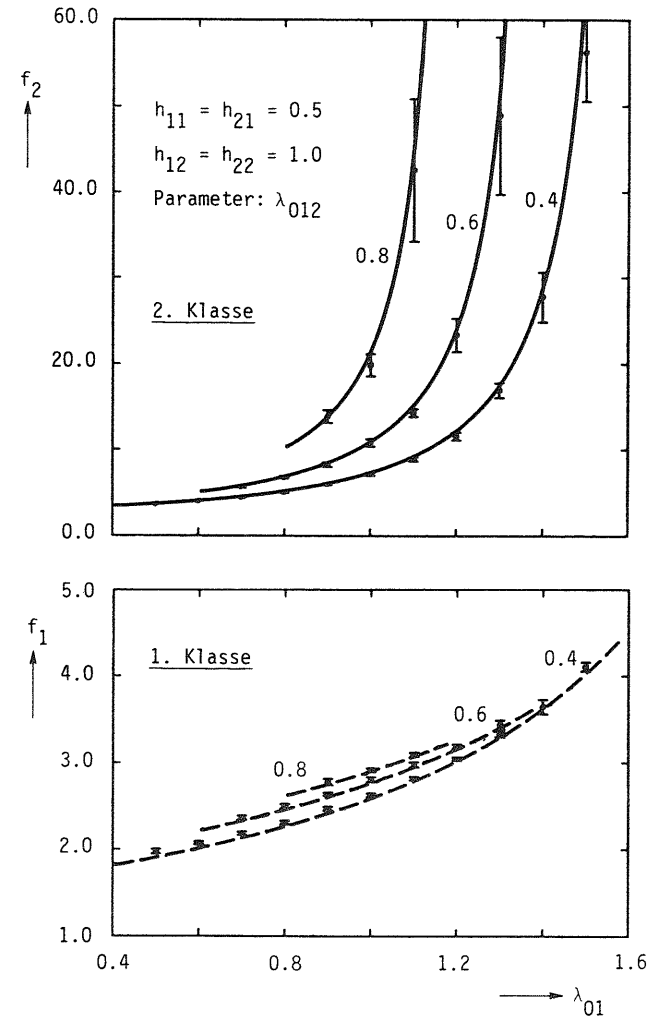


**Bild 5.7:** Mittlere Verweilzeiten  $f_r = f_{1r} + f_{2r}$ , ( $r = 1, 2$ ) in Abhängigkeit von der Gesamtankunftsrate  $\lambda_{01} = \lambda_{011} + \lambda_{012}$ .

**Modell:** M/M/1/NONPRE  $\rightarrow$  \*/M/1/NONPRE.

**Analyseverfahren:** — —, — m s d r o a;

⊥ Simulation (95% Vertrauensintervalle).



**Bild 5.8:** Mittlere Verweilzeiten  $f_r = f_{1r} + f_{2r}$ , ( $r = 1, 2$ ) in Abhängigkeit von der Gesamtankunftsrate  $\lambda_{01} = \lambda_{011} + \lambda_{012}$ .

**Modell:** M/M/1/NONPRE  $\rightarrow$  \*/M/1/NONPRE.

**Analyseverfahren:** — —, — m s d r o a;

⊥ Simulation (95% Vertrauensintervalle).



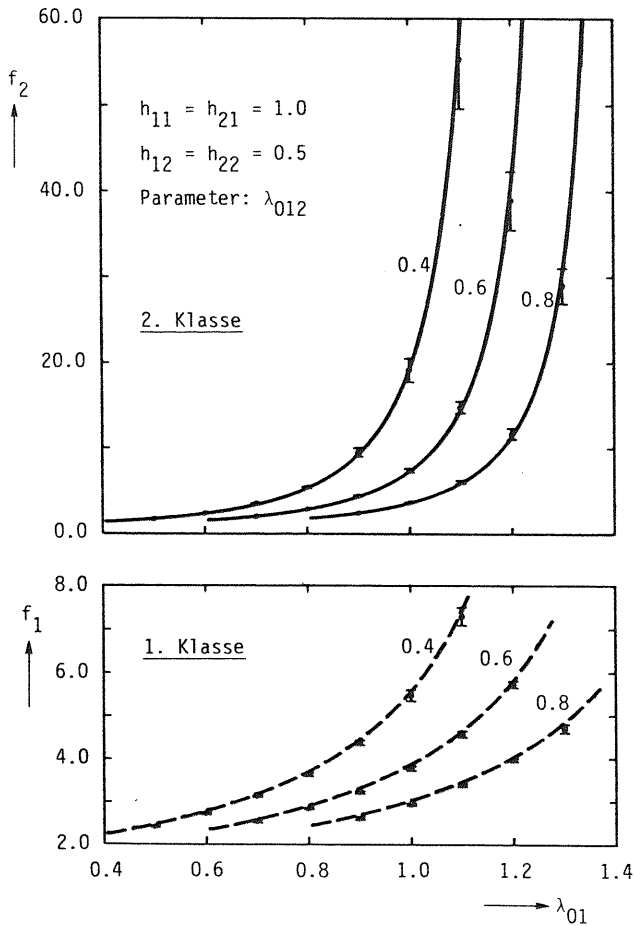


Bild 5.9: Mittlere Verweilzeiten  $f_r = f_{1r} + f_{2r}$ , ( $r = 1, 2$ ) in Abhängigkeit von der Gesamtankunftsrate  $\lambda_{01} = \lambda_{011} + \lambda_{012}$ .  
 Modell:  $M/M/1/NONPRE \rightarrow \cdot/M/1/NONPRE$ .  
 Analyseverfahren: —, —, — m s d r o a;  
 † Simulation (95% Vertrauensintervalle).

sind als im Fall  $h_{i1} = 2 \cdot h_{i2}$ . Häufigere Unterbrechungen durch kurze Bedienungsphasen von Anforderungen der ersten Klasse wirken sich demnach hier weniger ungünstig auf die Genauigkeit der Ergebnisse aus als seltenere, lange Unterbrechungsphasen. Das entsprechend Umgekehrte gilt bei unterbrechenden Prioritäten für die erste Klasse.

Weitere Ergebnisse für stärker vermaschte WSNe findet man in [Schmitt 83] und im letzten Kapitel dieser Arbeit. Dort wird das Warteschlangenmodell aus Bild 2.8 für den eingangs besprochenen RDC-Ring analysiert werden.

## 5.2 Geschlossene Warteschlangennetze

### 5.2.1 Zustandsanalyse eines zyklischen Warteschlangenmodells mit unterbrechenden Prioritäten

Im ersten Abschnitt dieses Unterkapitels über die Analyse geschlossener WSNe mit Prioritäten soll anhand des in Bild 5.10 dargestellten zyklischen Warteschlangenmodells  $\cdot/M/1/PRE \rightarrow \cdot/M/1/PRE$  gezeigt werden, daß die Voraussetzung von Poisson-Ankunftsprozessen nicht zwingend erfüllt zu sein braucht, um die Zerlegung von Prioritätsmodellen in einklassige Warte-

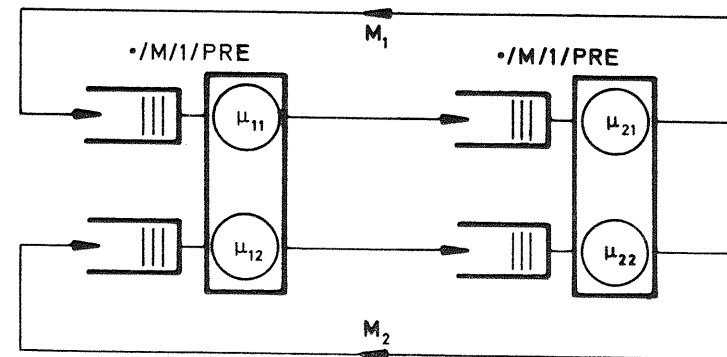


Bild 5.10: Zyklisches Warteschlangenmodell  $\cdot/M/1/PRE \rightarrow \cdot/M/1/PRE$  mit zwei Prioritätsklassen.

schlangenmodelle mit äquivalenten Zustandsverteilungen zu ermöglichen.

Da das Modell nur aus zwei BSen besteht, können die Gleichgewichtsgleichungen zur Bestimmung der Wahrscheinlichkeiten  $p(m,n)$  für das Auftreten der Zustände  $(m,n)$ , ( $0 \leq m \leq M_1$ ;  $0 \leq n \leq M_2$ ), in der BS 1 sofort angegeben werden:

$$0 = -(\mu_{11} + \mu_{21})p(m,n) + \mu_{11}p(m+1,n) + \mu_{21}p(m-1,n) \quad (0 < m < M_1, 0 \leq n \leq M_2), \quad (5.2,1a)$$

$$0 = -(\mu_{21} + \mu_{12})p(0,n) + \mu_{12}p(0,n+1) + \mu_{11}p(1,n) \quad (0 < n < M_2), \quad (5.2,1b)$$

$$0 = -(\mu_{11} + \mu_{22})p(M_1,n) + \mu_{22}p(M_1,n-1) + \mu_{21}p(M_1-1,n) \quad (0 < n < M_2), \quad (5.2,1c)$$

$$0 = -\mu_{21}p(0,0) + \mu_{11}p(1,0) + \mu_{12}p(0,1), \quad (5.2,1d)$$

$$0 = -(\mu_{12} + \mu_{21})p(0,M_2) + \mu_{11}p(1,M_2), \quad (5.2,1e)$$

$$0 = -(\mu_{11} + \mu_{22})p(M_1,0) + \mu_{21}p(M_1-1,0), \quad (5.2,1f)$$

$$0 = -\mu_{11}p(M_1,M_2) + \mu_{22}p(M_1,M_2-1) + \mu_{21}p(M_1-1,M_2). \quad (5.2,1g)$$

In [Morris 81] ist eine rekursive Lösung für die Differenzengleichung (5.2,1a) mit den Randbedingungen (5.2,1b-g) angegeben. Hier interessiert jedoch die Randverteilung  $p_2(n) := P\{X_{12}=n\}$ , für deren Berechnung zunächst in den obigen Gleichgewichtsgleichungen über  $n$  summiert wird. Dies führt zu den Beziehungen

$$0 = \mu_{22}(p(M_1,n-1) - p(M_1,n)) - \mu_{12}(p(0,n) - p(0,n+1)) \quad (0 < n < M_2), \quad (5.2,2a)$$

$$0 = \mu_{22}p(M_1,0) - \mu_{12}p(0,1), \quad (5.2,2b)$$

$$0 = \mu_{22}p(M_1,M_2-1) - \mu_{12}p(0,M_2). \quad (5.2,2c)$$

Durch eine einfache Rechnung folgt aus (5.2,2a):

$$\mu_{22} \cdot \left[ \frac{p(M_1,n-1)}{p_2(n-1)} \cdot p_2(n-1) - \frac{p(M_1,n)}{p_2(n)} \cdot p_2(n) \right] = \mu_{12} \cdot \left[ \frac{p(0,n)}{p_2(n)} \cdot p_2(n) - \frac{p(0,n+1)}{p_2(n+1)} \cdot p_2(n+1) \right], \quad (0 < n < M_2).$$

Mit den Vereinbarungen

$$\tilde{p}(m,n) := P\{X_{21}=m, X_{22}=n\}, \quad \tilde{p}_2(n) := P\{X_{22}=n\}$$

folgt hieraus unter Beachtung von

$$p(M_1,n) = \tilde{p}(0,M_2-n), \quad p_2(n) = \tilde{p}_2(M_2-n)$$

die Beziehung

$$\mu_{22} \cdot \frac{\tilde{p}(0,M_2-n+1)}{\tilde{p}_2(M_2-n+1)} \cdot p_2(n-1) - \mu_{22} \cdot \frac{\tilde{p}(0,M_2-n)}{\tilde{p}_2(M_2-n)} \cdot p_2(n) = \mu_{12} \cdot \frac{p(0,n)}{p_2(n)} \cdot p_2(n) - \mu_{12} \cdot \frac{p(0,n+1)}{p_2(n+1)} \cdot p_2(n+1), \quad (0 < n < M_2).$$

Definiert man jetzt

$$\mu_{12}^*(n) := c_2(n) \cdot \mu_{12}, \quad c_2(n) := \frac{p(0,n)}{p_2(n)}, \quad (0 < n \leq M_2), \quad (5.2,3a)$$

$$\mu_{22}^*(\tilde{n}) := \tilde{c}_2(\tilde{n}) \cdot \mu_{22}, \quad \tilde{c}_2(\tilde{n}) := \frac{\tilde{p}(0,\tilde{n})}{\tilde{p}_2(\tilde{n})}, \quad (0 < \tilde{n} \leq M_2), \quad (5.2,3b)$$

so erhält man schließlich die Gleichung

$$(\mu_{22}^*(M_2-n) + \mu_{12}^*(n)) \cdot p_2(n) = \mu_{22}^*(M_2-n+1) \cdot p_2(n-1) + \mu_{12}^*(n+1) \cdot p_2(n+1), \quad (0 < n < M_2). \quad (5.2,4a)$$

In analoger Weise zeigt man die Gültigkeit von

$$\mu_{22}^*(M_2) \cdot p_2(0) = \mu_{12}^*(1) \cdot p_2(1), \quad (5.2,4b)$$

$$\mu_{22}^*(1) \cdot p_2(M_2-1) = \mu_{12}^*(M_2) \cdot p_2(M_2). \quad (5.2,4c)$$

Insgesamt bilden (5.2,4a-c) die Gleichgewichtsgleichungen des zyklischen Warteschlangenmodells  $\cdot/M/1 \rightarrow \cdot/M/1$  mit zustandsabhängigen Bedienungsraten  $\mu_{12}^*(n)$  und  $\mu_{22}^*(n)$ . Letztere sind durch die gleiche Vorschrift definiert wie die Bedienungsraten der in Abschnitt 4.3 vorgeschlagenen Ersatzmodelle. Damit ist deren Gültigkeit im Sinne der Äquivalenz der Zustandswahrscheinlichkeiten für das hier behandelte zyklische Warteschlangenmodell gezeigt, in welchem die Ankunftsprozesse in den BSen aufgrund der vorhandenen Abhängigkeiten keine Poisson-Prozesse sind.

Das Ergebnis dieses Abschnitts, von dem man vermuten kann, daß es auch noch für andere geschlossene Netzstrukturen seine Gültigkeit behält, motiviert zusätzlich die Anwendung der neu entwickelten Ersatzmodelle zur klassenweisen Zerlegung von Prioritätsmodellen in geschlossenen WSNen.

### 5.2.2 Erweiterung des Gültigkeitsbereiches für das "Norton Theorem"

Die folgenden Ausführungen dienen der Vorbereitung eines im nächsten Abschnitt besprochenen Verfahrens zur Analyse geschlossener WSNe mit Prioritäten.

Im Abschnitt 3.2.1 war der als "Norton Theorem für WSNe" bekannte Satz 3.2 vorgestellt worden, welcher eine Möglichkeit aufzeigt, aus einem geschlossenen, mehrklassigen WSN ohne Klassenwechsel einen Teilkomplex herauszugreifen und den Einfluß des verbleibenden Restnetzes auf diesen Teilkomplex durch eine komplementäre BS vom Typ  $\cdot/M/1$  mit zustandsabhängigen Bedienungsraten zu berücksichtigen. Das Theorem enthält als wesentliche Voraussetzung für seine Gültigkeit die Forderung, daß das in Frage kommende WSN nur aus BSen der Typen 1...4 bestehen darf, wobei in [Kritzinger, Wyk, Krzesinski 82] für den Typ 1 jedoch keine zustandsabhängigen Bedienungsraten mit einbezogen sind. Wie aber die nachstehende Beweisskizze zeigt,

können auch zustandsabhängige Bedienungsraten zugelassen werden, ohne daß die Behauptung von Satz 3.2 eine Einschränkung erfährt.

Bei der nachfolgenden Argumentation wird auf den in [Kritzinger, Wyk, Krzesinski 82] zu findenden Beweis Bezug genommen, dessen vollständige Wiedergabe an dieser Stelle zuviel Raum in Anspruch nehmen würde und die für das weitere Verständnis auch nicht unbedingt erforderlich ist.

Dieser Beweis stützt sich u.a. auf die Produktlösung (3.2,5a) für BCMP-Netze, wobei die dort auftretenden Terme  $g_i(x_i)$ , ( $i=1, \dots, N$ ), allerdings nicht in Form der Gleichungen (3.2,5e-g), sondern durch

$$g_i(\underline{x}_i) = \beta_i(\underline{x}_i)^{-1} \prod_{r=1}^R e_{ir}^{x_{ir}}, \quad (i=1, \dots, N), \quad (5.2,5)$$

bestimmt sind. Die Größen  $e_{ir}$  genügen einer der Gleichungen (3.2,5h) oder (3.2,5i) und die Terme  $\beta_i(\underline{x}_i)$  sind rekursiv durch

$$\begin{aligned} \beta_i(x_{i1}, \dots, x_{iR}) &= \lambda_{ir}(x_{i1}, \dots, x_{iR}) \cdot \beta_i(x_{i1}, \dots, x_{iR-1}, \dots, x_{iR}), \\ \beta_i(0, \dots, 0) &= 1, \quad (i=1, \dots, N), \end{aligned} \quad (5.2,6)$$

erklärt. Darin wiederum ist mit  $\lambda_{ir}(x_{i1}, \dots, x_{iR})$  der zustandsabhängige Durchsatz der Klasse  $r$ , ( $r=1, \dots, R$ ), in der BS  $i$  bezeichnet, welcher vom Typ der jeweiligen BS abhängt. Für den Beweis der hier aufgestellten Behauptung ist zunächst die Richtigkeit des folgenden Hilfssatzes zu zeigen.

#### Hilfssatz 5.1

Die Größen  $\beta_i(\underline{x}_i)$  mögen der Rekursionsgleichung (5.2,6) genügen und es sei

$$\lambda_{ir}(\underline{x}_i) = \frac{x_{ir} \cdot \mu(x_i)}{x_i}, \quad (x_{ir} > 0; x_{is} \geq 0 \text{ für } s \neq r; r, s=1, \dots, R). \quad (5.2,7)$$

Dann gilt:

$$\beta_i(\underline{x}_i) = \frac{x_{i1}! \cdot x_{i2}! \cdot \dots \cdot x_{iR}!}{x_i!} \prod_{m=1}^{x_i} \mu_i^{(m)}, \quad (x_{ir} \geq 0; r=1, \dots, R). \quad (5.2,8)$$

Der Beweis für diesen Hilfssatz kann durch vollständige Induktion geführt werden und ist in Anhang 3 enthalten.

Durch Einsetzen von (5.2,8) in (5.2,5) folgt die Gleichung

$$g_i(\underline{x}_i) = \left[ x_i! \prod_{r=1}^R (e_{ir}/x_{ir}!) \right] \prod_{m=1}^{x_i} (\mu_i^{(m)})^{-1}, \quad (5.2,9)$$

welche mit (3.2,5e) übereinstimmt.

Insgesamt gesehen umfaßt demnach die im Beweis von Satz 3.2 verwendete Darstellung der Produktterme  $g_i(\underline{x}_i)$  nach (5.2,5) auch den Fall des Vorhandenseins von BSen des Typs 1 mit zustandsabhängigen Bedienungsrate. Die zustandsabhängigen Durchsätze  $\lambda_{ir}(\underline{x}_i)$  sind dann durch (5.2,7) definiert. Damit ist die Richtigkeit der oben aufgestellten Behauptung gezeigt.

### 5.2.3 Approximation durch separable Warteschlangennetze

Im letzten Abschnitt dieses Kapitels werden zwei Verfahren für die Analyse geschlossener WSNe mit Prioritäten vorgestellt. Beide Verfahren lassen sich in das nachfolgend angegebene Analysekonzept einordnen, welches schon bei den offenen WSNe angesprochen wurde.

#### Konzept zur approximativen Zustandsanalyse von geschlossenen WSNe mit Prioritäten

Schritt 1: Ersetzung aller BSen des Typs  $\cdot/M/1/PRE$  und/oder  $\cdot/M/1/NONPRE$  durch solche vom Typ  $\cdot/M/1$  mit à priori unbekanntem, zustandsabhängig reduzierten Bedienungsrate (siehe Abschnitt 4.3).

Schritt 2: Approximative Berechnung dieser Bedienungsrate.

Schritt 3: Analyse des separablen Ersatz-WSNe aus Schritt 1 durch Anwendung des BCMP-Theorems (Satz 3.1).

Während zu den Schritten 1 und 3 keine weiteren Erklärungen notwendig sind, beinhaltet der Schritt 2 ein Problem, dessen Lösung den wesentlichen Teil dieses Abschnitts in Anspruch nimmt. Wie bei den offenen WSNe schon erwähnt, werden die auf dem vorgestellten Konzept beruhenden Verfahren weiterhin mit *m s d r o a* abgekürzt.

Allgemein ist zu bemerken, daß die Approximation der gesuchten Bedienungsrate nach unterschiedlichen Methoden erfolgen kann, die unter Umständen von der Struktur des zu analysierenden WSNe abhängig sind. Im folgenden sind zwei Verfahren beschrieben, welche recht gute Ergebnisse liefern.

#### Erstes Verfahren zur Approximation der zustandsabhängig reduzierten Bedienungsrate

Das erste der vorzustellenden Verfahren ist in der vorliegenden Form nur für solche Netze geeignet, die ausschließlich BSen des Typs  $\cdot/M/1/PRE$  enthalten und in denen keine Klassenwechsel vorkommen. WSNe dieses Typs enthalten demnach zwei geschlossene, einklassige Ketten  $E_1$  (1. Klasse) und  $E_2$  (2. Klasse), von denen das zur ersten Kette gehörende Ersatz-WSN in bekannter Weise analysiert werden kann. Der Grundgedanke zur Approximation der gesuchten Bedienungsrate ist, den Zustandsprozeß in jeder BS isoliert unter der Annahme zu untersuchen, daß alle Zustandsänderungen nach exponentiell verteilten Zeitphasen stattfinden und der Zustandsprozeß somit eine M-K ist. Inwieweit diese Annahme vernünftig sein kann, sei anhand des Bildes 5.11a diskutiert.

In diesem Bild ist eine beliebig herausgegriffene BS  $i$ , ( $i \in \{1, \dots, N\}$ ), zusammen mit ihrer für Zustandsänderungen mitverantwortlichen Netzumgebung dargestellt. Unter den gegebenen Voraussetzungen ergibt sich diese Netzumgebung für die erste Klasse durch Anwendung des Norton Theorems als eine komplementäre BS vom Typ  $\cdot/M/1$  mit zustandsabhängigen Bedienungs-

raten  $\mu_{ic}(x_{ic})$ , ( $x_{ic}=1, \dots, M_1$ ). In ihr ereignen sich Bedienungsenden und somit Zustandsänderungen in BS  $i$  nach  $\mu_{ic}(x_{ic})$ -exponentialverteilten Zeitabständen, sofern mindestens eine Anforderung zur Bedienung ansteht. Für die zweite Klasse besteht die Netzumgebung aus einem fiktiven "Rest"-WSN, von dem angenommen sei, daß in ihm die Bedienungsenden in  $v_i$ -exponentialverteilten Zeitabständen stattfinden. Letzteres trifft nur näherungsweise zu und beschreibt die wahren Verhältnisse umso besser, je größer der Grad der Vermaschung und das Verhältnis  $M_2/M_1$  im gegebenen WSN ist. Bei Berücksichtigung der Prioritätsabfertigung folgt dann für den approximierten Zustandsprozeß in BS  $i$  ein Übergangsgraph mit der in Bild 5.11b dargestellten Struktur. Darin sind die unbekanntenen Übergangsraten  $v_i$  zunächst als freie Parameter zu betrachten. Die Lösung der Gleichgewichtsgleichungen kann entweder rekursiv (siehe z.B. [Chow 80]) oder durch Auflösen des äquivalenten linearen Gleichungssystems erfolgen, das wegen des endlichen Zustandsraumes ebenfalls endlich ist. In dieser Arbeit findet die letztgenannte Methode den Vorzug, da Rekursionsgleichungen aufgrund von Fehlerakkumulation unter Umständen ungünstige numerische Eigenschaften aufweisen [Schneemann 75]. Die folgenden Ausführungen betreffen nun die Berechnung der gesuchten Bedienungsraten  $\mu_{i2}^*(x_{i2})$  bzw. der Reduktionskoeffizienten  $c_{i2}(x_{i2})$ .

Zuerst wird die folgende Bedingung zur Festlegung der unbekanntenen Übergangsraten  $v_i$ , ( $i=1, \dots, N$ ), formuliert:

Es sei  $j$  die Nummer der BS mit der höchsten Auslastung  $\rho_{j1}$  durch Anforderungen der Klasse 1. Dann ist der Vektor  $\underline{v} := (v_1, \dots, v_i, \dots, v_N)^T$  so zu wählen, daß die Reduktionskoeffizienten  $c_{j2}(x_{j2})$ , ( $x_{j2}=0, \dots, M_2$ ), die Gleichung

$$\sum_{x_{j2}=0}^{M_2} c_{j2}(x_{j2}) \cdot p(x_{j2}) = 1 - \rho_{j1} \quad (5.2,10a)$$

mit

$$c_{j2}(0) := \frac{p(x_{j1}=0, x_{j2}=0)}{p_2(x_{j2}=0)} \quad (5.2,10b)$$

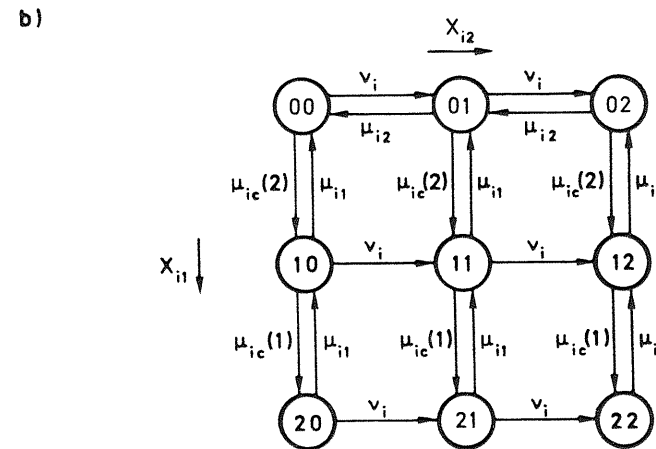
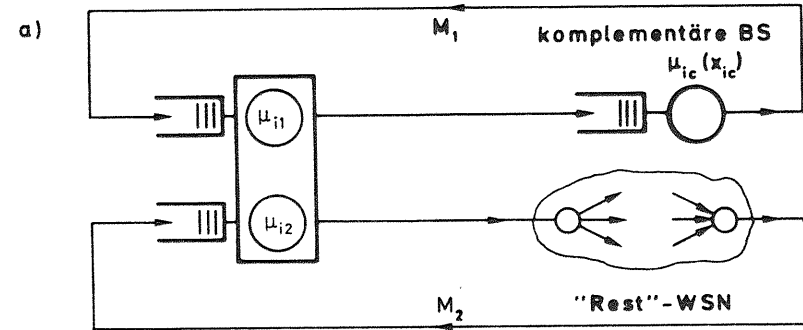


Bild 5.11: a) Bedienstation  $i$  mit komplementärer Bedienstation für Klasse 1 und "Rest"-Warteschlangennetz für Klasse 2. b) Übergangsgraph des approximierenden Zustandsprozesses in Bedienstation  $i$  für den Fall  $M_1 = M_2 = 2$ .

erfüllen.

Die Gleichung (5.10,a) entspricht der Beziehung (4.1,14). Es wurde hier lediglich eine andere Wahl für  $c_{j2}(0)$  getroffen. Der Grund, diese Gleichung für die BS mit höchster Auslastung durch die erste Klasse zu erfüllen, liegt darin, daß es diese BS ist, welche die größten Auswirkungen auf das Verkehrsgeschehen der zweiten Klasse hat ("Flaschenhalseffekt"). Die Bestimmung der Übergangsraten  $v_i$  und damit der Koeffizienten  $c_{i2}(x_{i2})$ , ( $i=1, \dots, N$ ), könnte nun prinzipiell mit einem auf Intervallschachtelung beruhenden Suchalgorithmus erfolgen, wie er etwa in [Shum 76] im Zusammenhang mit dem weiter oben besprochenen Verfahren "Erweiterte Produktformlösung" vorgeschlagen worden ist. Für das hier gestellte Problem erweist sich jedoch eine andere Methode als effektvoller, die aus dem nachstehenden Algorithmus hervorgeht.

Algorithmus zur approximativen Bestimmung der zustandsabhängig reduzierten Bedienungsraten  $\mu_{i2}^*(x_{i2})$ , ( $i=1, \dots, N$ ).

Schritt 1: Berechnung der zustandsabhängigen Bedienungsraten  $\mu_{ic}(x_{ic})$  (Norton Theorem) und der Auslastung  $\rho_{j1}$ .

Schritt 2: Bestimmung einer Anfangsnäherung  $\underline{v}^{(1)} := (v_1^{(1)}, \dots, v_N^{(1)})^T$  gemäß

$$v_i^{(1)} = \lambda_i \text{ roa}, \quad (i=1, \dots, N).$$

Mit  $\lambda_i \text{ roa}$  ist der mittels der  $\text{roa}$ -Technik (siehe Abschnitt 3.2.3) berechnete Durchsatz der zweiten Klasse in BS  $i$  bezeichnet.

Schritt 3: Lösung der Gleichgewichtsgleichungen des approximierenden Zustandsprozesses (Bild 5.11b) für jede BS  $i$  und anschließende Berechnung einer  $k$ -ten Näherung  $c_{i2}^{(k)}(x_{i2})$ , ( $k=1, 2, \dots$ ), für die gesuchten Reduktionskoeffizienten entsprechend (4.1,8b).

Schritt 4: Überprüfung, ob die Beziehung (5.2,10a) (bis auf

einen vorgebbaren Fehler  $\epsilon > 0$ ) erfüllt ist.

Wenn ja:  $\mu_{i2}^*(x_{i2}) = c_{i2}^{(k)}(x_{i2}) \cdot \mu_{i2}$ , ( $i=1, \dots, N$ ).

Wenn nein:

i) Falls  $\sum_{x_{j2}=0}^{M_2} c_{j2}(x_{j2}) \cdot p(x_{j2}) > 1 - \rho_{j1}$ :

Berechnung einer neuen Näherung  $\lambda_{i2}^{(k+1)}$  für die Durchsätze im Ersatz-WSN mit den Bedienungsraten  $\mu_{i2}^*(x_{i2})$ . Anschließend wird

$$v_i^{(k+1)} = \lambda_{i2}^{(k+1)}, \quad (i=1, \dots, N),$$

gesetzt. Fortsetzung mit Schritt 3 für  $k \rightarrow k+1$ .

ii) Sonst:

$$\underline{v}^{(k+1)} = \alpha \cdot \underline{v}^{(k)}, \quad (\alpha > 0).$$

Fortsetzung mit Schritt 3 für  $k \rightarrow k+1$ .

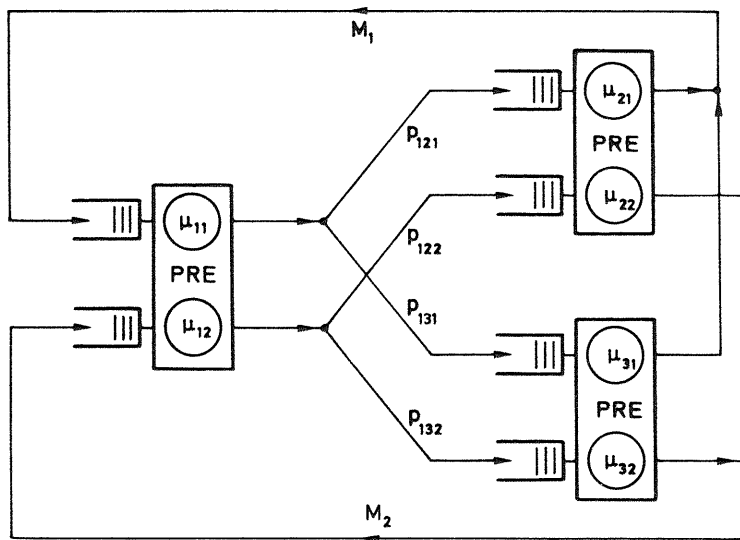
Dem Schritt 4 in diesem Algorithmus seien noch die folgenden Erläuterungen hinzugefügt. Würde man ohne Beachtung der eingangs formulierten Bedingung (5.2,10a) stets  $v_i^{(k+1)} = \lambda_{i2}^{(k+1)}$  setzen, so würde der Algorithmus ein Iterationsverfahren zur Lösung der Fixpunktgleichung

$$\underline{v} = T\underline{v}$$

beschreiben, in welcher  $T$  einen Operator bezeichnet, der die Operationen von Schritt 3 und die Berechnung der Durchsätze  $\lambda_{i2}^{(k+1)}$  in Schritt 4 erfaßt. Zum formalen Nachweis der Konvergenz müßte die Lipschitzeigenschaft (siehe z.B. [Törnig 79]) von  $T$  gezeigt werden. Da jedoch beim Umgang mit dem Verfahren keinerlei Konvergenzprobleme aufgetreten sind, wird auf die

Führung dieses Beweises verzichtet. Die Einarbeitung der Bedingung (5.2,10a), welche in einigen Fällen eine Verbesserung der Ergebnisse liefert, erfordert die im Schritt 4 eingeführten Fallunterscheidungen. Dabei bewirkt das Eintreten des Falles ii) eine Ablösung des Iterationsverfahrens durch ein Suchverfahren. Der Parameter  $\alpha > 0$  beeinflusst die Suchschrittweite. Für ihn hat sich eine Wahl von  $\alpha = 0.1$  als günstig erwiesen.

Nach dieser ausführlichen Beschreibung des ersten Verfahrens zur Approximation der zustandsabhängig reduzierten Bedienungsrate soll dessen Tauglichkeit anhand von Beispielen demonstriert werden. Als Testobjekte dienen dazu das zyklische Warteschlangenmodell aus Bild 5.10 und das dreistufige Warteschlangennetz nach Bild 5.12. Zyklische Warteschlangenmodelle



$$p_{ijr} := P\{\text{Anford. d. Kl. } r \text{ geht von BS } i \text{ nach BS } j\}, \quad (i, j \in \{1, 2, 3\}; r = 1, 2)$$

Bild 5.12: Testmodell mit drei Bedienstationen vom Typ  $\ast/M/1/PRE$

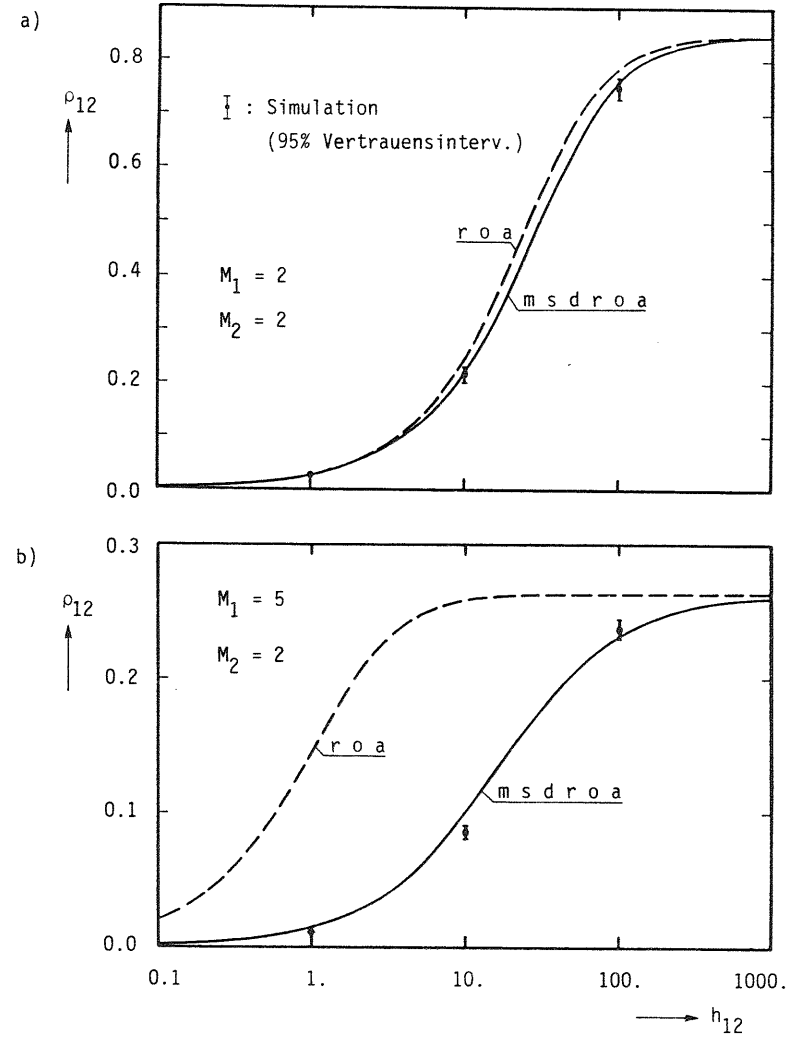
nehmen bei den geschlossenen WSNen aufgrund des oben Gesagten die Stellung eines "worst case" ein und sind daher als Testobjekte besonders gut geeignet.

In der Tabelle 5.2 sind für zehn verschiedene Parameterkonstellationen Ergebnisse für den Durchsatz  $\lambda_2$  der zweiten Klasse im Modell  $\ast/M/1/PRE \rightarrow \ast/M/1/PRE$  angegeben, welche mit der besprochenen Variante der Methode *m s d r o a* berechnet wurden. Dort findet man zum Vergleich auch exakte und mit den Verfahren *r o a* und *m-r o a* erzielte Resultate. Sie sind ebenso wie die Auswahl der Parameter aus [Morris 81] und [Kaufman 82] übernommen worden. Es sei nochmals daran erinnert, daß ein sinnvoller Vergleich nur zwischen den Verfahren *m-r o a* und *s d r o a* (beide beruhen auf à priori unbekanntem Reduktionskoeffizienten) sowie *r o a* und *m s d r o a* stattfinden kann.

Die aufgrund des Ergebnisses aus Abschnitt 5.2.1 zu erwartende Überlegenheit der Methode *s d r o a* wird durch die vorliegenden Zahlenwerte auch numerisch bestätigt und bedarf keiner weiteren Diskussion. Ein Vergleich der mittels *r o a*- und *m s d r o a*-Technik erzielten Resultate zeigt eine weitgehende Gleichwertigkeit dieser Verfahren, mit Ausnahme der Fälle 4 und 5, wo die letztgenannte Methode deutlich bessere Ergebnisse liefert. Eine eingehendere Analyse hat gezeigt, daß in den beiden Fällen die Gesamtauslastung  $\rho_i = \rho_{i1} + \rho_{i2}$ , ( $i=1,2$ ), von jeder BS 89 % (Fall 4) bzw. 75 % (Fall 5) beträgt. In allen anderen Fällen gilt  $\rho_i > 0.95$  für mindestens eine BS. Es liegt dann jeweils ein Systemengpaß vor, der das Verkehrsgeschehen im WSN bestimmt. Das gute Abschneiden der *r o a*-Technik überrascht dann nicht, da sie der zweiten Klasse die gesamte, durch die erste Klasse nicht genutzte Restkapazität der BE zugesteht, welche nur bei Systemengpässen auch tatsächlich voll beansprucht wird (ständig arbeitende BE). Man vergleiche hierzu auch die Beziehung (4.1,15) in Abschnitt 4.1.1 und das dort dazu Gesagte. Aus dem genannten Grunde folgt auch auf plausible Weise, daß das Verfahren *r o a* für Netztypen der hier besprochenen Art eine obere Schranke für die Verkehrsgrößen  $\rho_{i2}$  und  $\lambda_{i2}$  liefert.

Fall	M <sub>1</sub>	M <sub>2</sub>	h <sub>11</sub>	h <sub>12</sub>	h <sub>21</sub>	h <sub>22</sub>	Durchsatz λ <sub>2</sub> der Klasse 2			
							exakt s d r o a	r o a	m-r o a	m s d r o a
1	1	10	1.	10.	1.	10.	0.450 (-1)	0.455 (-1)	0.453 (-1)	0.452 (-1)
2	3	5	1.	33.33	1.	33.33	0.616 (-2)	0.625 (-2)	0.619 (-2)	0.618 (-2)
3	3	5	1.	33.33	1.	1.	0.750 (-2)	0.750 (-2)	0.750 (-2)	0.743 (-2)
4	3	5	1.	1.	1.	1.	0.139	0.208	0.160	0.153
5	3	5	1.	0.01	1.	0.01	0.408	20.8	0.672	0.272
6	1	10	5.	50.	1.	50.	0.333 (-2)	0.333 (-2)	0.333 (-2)	0.330 (-2)
7	3	5	5.	2000.	1.	100.	0.321 (-5)	0.321 (-5)	0.321 (-5)	0.320 (-5)
8	3	5	5.	10.	1.	100.	0.641 (-3)	0.641 (-3)	0.641 (-3)	0.637 (-3)
9	3	5	5.	2000.	1.	0.1	0.321 (-5)	0.321 (-5)	0.321 (-5)	0.320 (-5)
10	3	5	5.	1.	1.	0.1	0.608 (-2)	0.641 (-2)	0.608 (-2)	0.636 (-2)

Tabelle 5.2: Mit verschiedenen Verfahren berechneter Durchsatz λ<sub>2</sub> der zweiten Klasse im zyklischen Warteschlangenmodell •/M/1/PRE → •/M/1/PRE. (-x) bedeutet 10<sup>-x</sup>.



Parameter:

- a)  $\underline{h}_1 = (0.5, 1.0, 10.0)^T$ ;  $\underline{h}_2 = (h_{12}, 1.0, 10.0)^T$ ;  $p_{12r} = 0.7$ ;  $p_{13r} = 0.3$   
 b)  $\underline{h}_1 = (10.0, 10.0, 100.0)^T$ ;  $\underline{h}_2 = (h_{12}, 1.0, 10.0)^T$ ;  $p_{12r} = 0.9$ ;  $p_{13r} = 0.1$

Bild 5.13: Auslastung  $\rho_{12}$  in Abhängigkeit von der mittleren Bedienungszeit  $h_{12}$  (Warteschlangennetz nach Bild 5.12).



Die anhand des zyklischen Warteschlangenmodells gewonnenen Erkenntnisse werden durch die in Bild 5.13 dargestellten Diagramme weiter verdeutlicht. Dort sind für das WSN nach Bild 5.12 Näherungskurven für die Auslastung  $\rho_{12}$  in Abhängigkeit von der mittleren Bedienungszeit  $h_{12}$  (logarithmischer Maßstab) aufgetragen. Die übrigen Parameter sind so gewählt, daß im Fall a) für die zweite Klasse stets ein Engpaß in BS 3 besteht ( $\rho_{31}=0.92$ ), welcher im Fall b) nicht vorhanden ist ( $\rho_{i1}=0.7$ ,  $i=1, \dots, 3$ ). Der letzte Fall zeigt noch einmal die gravierende Überlegenheit der m s d r o a-Technik in solchen Situationen.

Als wesentlicher Nachteil des ersten Verfahrens zur Approximation der Reduktionskoeffizienten ist zu nennen, daß nur BSen vom Typ  $\cdot/M/1/PRE$  zugelassen sind. Diese Forderung schränkt seine Anwendbarkeit zur Lösung praktischer Probleme drastisch ein und lieferte daher die Motivation zur Entwicklung eines zweiten Verfahrens, das nachfolgend erläutert wird.

Zweites Verfahren zur Approximation der zustandsabhängig reduzierten Bedienungsraten

Für die Anwendbarkeit dieses Verfahrens ist ergänzend zu den am Beginn von Kapitel 5 formulierten Voraussetzungen lediglich gefordert, daß eine Anforderung ihre Prioritätsklasse nicht wechseln darf. Ferner wird aus Gründen der Übersichtlichkeit zunächst davon ausgegangen, daß das zu untersuchende WSN nur eine BS mit Prioritätsabfertigung (PRE oder NONPRE) enthält, welche durch die Nummer eins gekennzeichnet sei. Alle anderen BSen sind, wie in Bild 5.14a dargestellt, zu einem "Rest"-WSN zusammengefaßt. Letzteres wird dem Norton-Theorem (Satz 3.2) entsprechend durch eine komplementäre BS mit zustandsabhängigen Bedienungsraten  $\mu_{cr}(x_c)$ , ( $r=1,2$ ), ersetzt (Bild 5.14b). Da die Aussage des Norton-Theorems nicht mehr gilt, wenn in dem betrachteten WSN BSen mit Prioritätsabfertigung vorkommen, ist dieser Schritt eine Näherung. Die gesuchten Bedienungsraten  $\mu_{1r}^*(x_{1r})$  können nun durch Analyse des reduzierten WSNes nach Bild 5.14b in gewohnter Weise berechnet werden. Man siehe hierzu [Chow 80] sowie [Schmitt, Gotthardt 83].

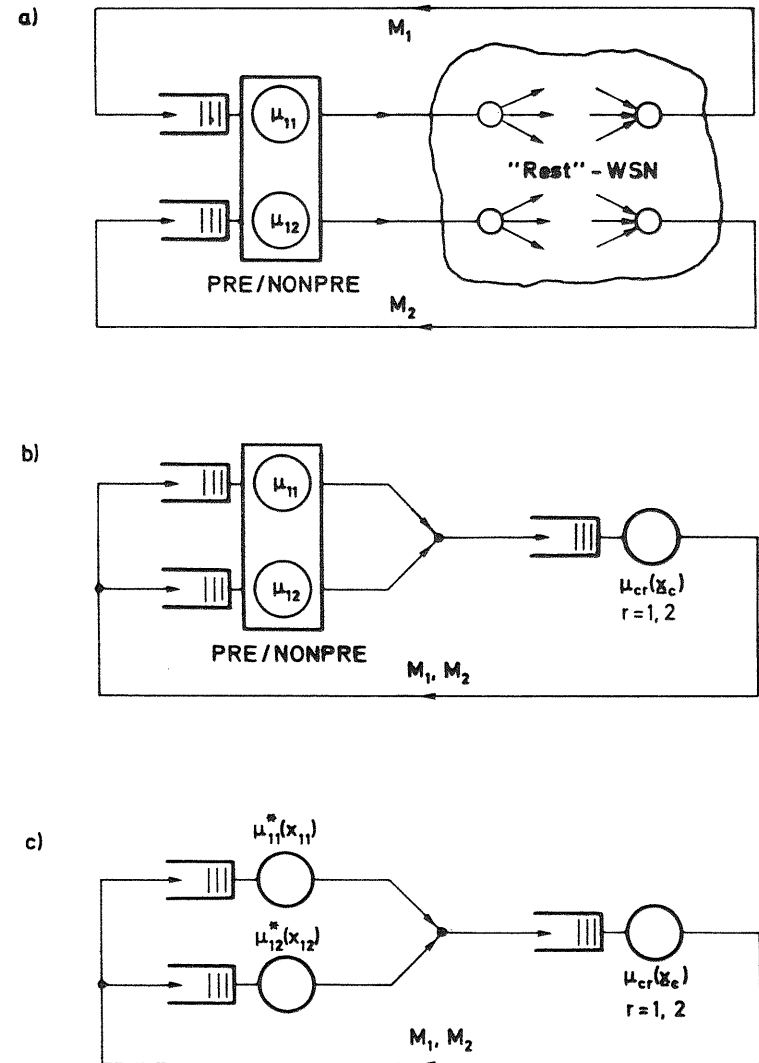


Bild 5.14: Zur Illustration des zweiten Approximationsverfahrens  
 a) Prioritätsstufe und "Rest"-Warteschlangennetz  
 b) Reduziertes Warteschlangennetz  
 c) Zerlegung der Prioritätsstufe im reduzierten Warteschlangennetz

Ein Vergleich des soeben beschriebenen Ansatzes zur approximativen Bestimmung der zustandsabhängig reduzierten Bedienungsrate mit dem in Abschnitt 3.2.3 skizzierten Aggregierungsverfahren zeigt einen Unterschied der Methoden nur in ihrer Zielsetzung. Beim Aggregierungsverfahren dient die Analyse des reduzierten WSNe der direkten Bestimmung von gesuchten Leistungsgrößen, während dieser Vorgang hier lediglich die Berechnung von Hilfsgrößen beinhaltet und damit einen Zwischenschritt im zuvor vorgestellten Gesamtanalysekonzept darstellt. Letzteres weist jedoch gegenüber dem Aggregierungsverfahren, welches fortan auch durch die Abkürzung  $c c a$  (composite centre approximation) bezeichnet wird, zwei entscheidende Vorteile auf:

- i) Es können Leistungsgrößen in allen BSen des WSNe berechnet werden (nicht nur in der BS mit Prioritätsabfertigung).
- ii) Die approximative Bestimmung der gesuchten Bedienungsrate kann auch in WSNe mit mehr als einer BS mit Prioritätsabfertigung durch wiederholte Anwendung des oben beschriebenen Verfahrens erfolgen.

Der zweite Sachverhalt soll durch den nachfolgenden Algorithmus näher erläutert werden. Dabei bezeichne  $\sigma_p := \{1, \dots, j\}$  die Menge der Stationsnummern, welche den BSen mit Prioritätsabfertigung zugeordnet sind, und es sei  $\bar{\sigma}_p := \{j+1, \dots, N\}$ .

Algorithmus zur approximativen Bestimmung der zustandsabhängig reduzierten Bedienungsrate  $\mu_{ir}^*(x_{ir})$ , ( $i \in \sigma_p$ ;  $r=1,2$ )

Sei  $i = 1$

Schritt 1: Aggregierung des  $\bar{\sigma}_p$  zugeordneten Teilnetzes entsprechend Satz 3.2.

Schritt 2: Hinzunahme der BS mit der Nummer  $i$  ergibt ein WSN entsprechend Bild 5.14 b. Berechnung der Bedienungsrate  $\mu_{ir}^*(x_{ir})$ , ( $r=1,2$ ), durch Analyse dieses Netzes.

Schritt 3: Falls  $i < j$ : Zerlegung der BS  $i$  in bekannter Weise (Bild 5.14c) und Aggregierung des neuen Teilnetzes entsprechend Satz 3.2 (man beachte an dieser Stelle das Ergebnis aus Abschnitt 5.2.2). Fortsetzung des Verfahrens mit Schritt 2 für  $i \rightarrow i+1$ .

Falls  $i=j$ : Ende des Algorithmus.

Zu diesem Algorithmus sei noch angemerkt, daß die Zwischenergebnisse für die komplementären BSen nutzbringend verwendet werden können, wenn man sich nur für die Verkehrsgrößen in einem Teil des zu untersuchenden WSNe interessiert. Hinsichtlich numerischer Ergebnisse zu dem zweiten Verfahren sei auf den Abschnitt 6.2 des nächsten Kapitels verwiesen. Das allgemeine Analysekonzept betreffend ist schließlich noch anzufügen, daß in dem Ersatz-WSN noch zusätzlich Auftragsklassen eingeführt werden können, um eine möglichst wirklichkeitsgetreue Modellierung der Verkehrslenkung zu erzielen.

Zum Ende dieses Kapitels sei noch kurz auf die Problematik der Berücksichtigung von mehr als zwei Prioritätsklassen eingegangen.

Prinzipiell ist es denkbar, durch Anwendung der Gleichungen (4.3,1a-b) für  $r > 2$  die Zerlegung einer Prioritätsstufe in ein-klassige Ersatzmodelle auf mehr als zwei Klassen auszudehnen. Der damit verbundene Rechenaufwand ist im allgemeinen jedoch so groß, daß eine solche Vorgehensweise wenig sinnvoll erscheint. Eine praktikablere Lösung ist das Zusammenfassen von mehreren Prioritätsklassen nach einer der in [Sauer, Chandy 75] oder [Chow, Yu 83] vorgeschlagenen Methoden, welche unter Umständen jedoch auch einen recht großen, zusätzlichen Rechenaufwand erfordern. In vielen Fällen liegt daher die beste Lösung darin, eine Zusammenfassung (bei sehr geringem Einfluß auch Vernachlässigung) von Prioritätsklassen schon im Zuge der Modellbildung vorzunehmen und somit von vornherein mit  $R=2$  zu rechnen. Beispielsweise können in einem Rechnersystem die Prioritätsklassen für zeitkritische und zeitunkritische Anforderungen nochmals mehrere Prioritätsebenen aufweisen.

Diese finden jedoch bei der Modellbildung keine Berücksichtigung, da sie für die Beurteilung der hauptsächlich interessierenden Verkehrsgrößen der beiden Hauptklassen keine Rolle spielen.

6. AUSGEWÄHLTE SYSTEMSTUDIEN

Die Approximationsverfahren aus den Abschnitten 5.1.3 und 5.2.3 wurden zur Analyse des Warteschlangenmodells für den RDC-Ring [Heger 83] und des "Central Server Models" [Buzen 71] angewendet. In diesem letzten Kapitel sind einige numerische Ergebnisse zu diesen Modellen angegeben, denen jeweils ein reales System zugrunde liegt. Zum Vergleich stehen wieder Simulationsergebnisse zur Verfügung. Beim CSM sind außerdem Resultate angegeben, die mit den Näherungsverfahren  $r o a$  und  $c c a$  berechnet wurden. Alle Modellparameter seien durch Bezug auf ihre Einheiten normiert und sind daher wie bisher als reine Zahlenwerte angegeben.

6.1 Kommunikationssystem mit Ringstruktur

Gegenstand der ersten Systemstudie ist das als RDC-Ring bekannte, ringförmige Kommunikationssystem aus Abschnitt 2.5.3. Das für dieses System in Bild 2.8b angegebene Warteschlangenmodell unterscheidet sich von den bisher untersuchten Modellen dieser Art dadurch, daß in den BSen jeweils ein Klassenwechsel stattfindet.

Die zur Analyse des Modells notwendige Zerlegung der Prioritätsstufen ist nochmals in Bild 6.1 verdeutlicht, wo das aus  $N^* = 2N$  BSen bestehende Ersatz-WSN gezeigt ist. Es enthält nur noch eine einklassige Kette. Die Zuordnung der Verzweigungswahrscheinlichkeiten  $p_{ir,js}$  bzw.  $p_{ir,0}$  des Originalmodells zu denen des Ersatzmodells wird am folgenden Beispiel klar:

$$p_{13} = p_{11,21} \quad , \quad p_{23} = p_{12,21}$$

$$p_{10} = p_{11,0} \quad , \quad p_{20} = p_{12,0} \quad .$$

Entsprechend gilt für die externen Ankunftsraten:

$$\lambda_{02} = \lambda_{012} \quad , \quad \lambda_{04} = \lambda_{022} \quad \text{usw.}$$

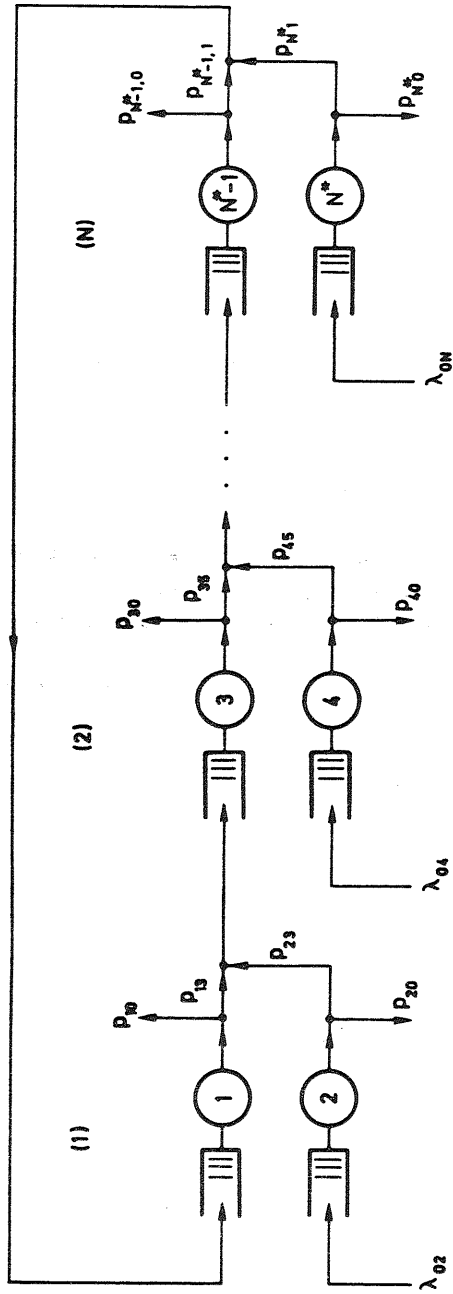


Bild 6.1.1: Ersatzmodell für das Warteschlangennetz aus Bild 2.8 b. (i) kennzeichnet die dort vorgesehene Numerierung der Bedienungsstationen.

Damit kann der zweite Schritt des in Abschnitt 5.1.3 beschriebenen Analyseverfahrens ausgeführt werden.

In den Bildern 6.2 ... 6.5 sind einige numerische Ergebnisse für den Fall von  $N = 10$  BSen angegeben. Dort ist jeweils die mittlere Verweilzeit  $f$  einer Anforderung im WSN über der externen Ankunftsrate  $\lambda_{012}$  in der BS 1 aufgetragen. Die Größe  $f$  berechnet sich gemäß

$$f = \lambda^{-1} \sum_{i=1}^{20} \bar{x}_i^*$$

wobei

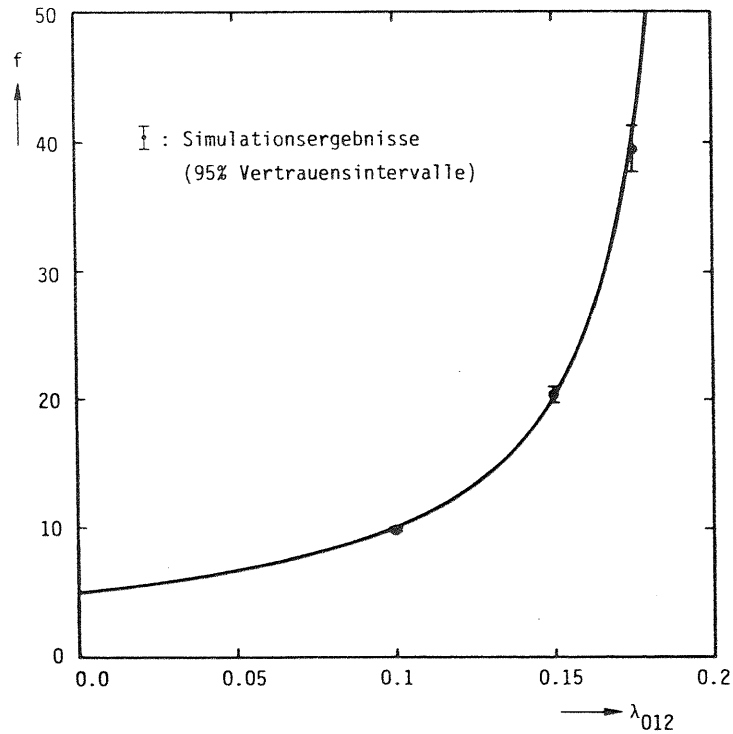
$$\lambda := \sum_{i=1}^{10} \lambda_{0i2}$$

die gesamte externe Ankunftsrate und  $\bar{x}_i^*$  die mittlere Anzahl von Anforderungen in der BS  $i$  des Ersatz-WSNes bezeichnet. Für die in allen BSen gleichen Bedienungsrate  $\mu_{ir}$  wurde immer der Wert eins gewählt. Hinsichtlich der Werte für die Größen  $\lambda_{0i2}$  und  $P_{ir,js}$  bzw.  $P_{ir,0}$  sind vier Fälle zu unterscheiden:

- Fall 1: symmetrische Last, symmetrische Verkehrslenkung
- Fall 2: unsymmetrische Last, symmetrische Verkehrslenkung
- Fall 3: symmetrische Last, unsymmetrische Verkehrslenkung
- Fall 4: unsymmetrische Last, unsymmetrische Verkehrslenkung.

"symmetrische Last" bedeutet die Gleichheit von allen externen Ankunftsrate. Mit "symmetrischer Verkehrslenkung" ist gemeint, daß im realen System die Wahrscheinlichkeit dafür, daß ein beliebiger Teilnehmerrechner  $i$  eine Nachricht an einen anderen Teilnehmerrechner  $j \neq i$  sendet, gleich ist für alle  $i$ . Die Bezeichnung "unsymmetrisch" kennzeichnet dementsprechend das Nichterfülltsein der genannten Bedingungen. In den Fällen unsymmetrischer Parameterwahl liegt jeweils ein Betriebsmittelengpaß in der BS 1 vor.

Ein Blick auf die Ergebnisse zeigt in allen vier Fällen eine



Parameter

Anzahl der Bedienstungen :  $N = 10$   
 Bedienungsrate :  $\mu_{ir} = 1.0$  ,  $(i=1, \dots, 10; r=1,2)$   
 Externe Ankunftsrate :  $\lambda_{0i2} = \lambda_{012}$  ,  $(i=1, \dots, 10)$

Verzweigungswahrscheinlichkeiten:

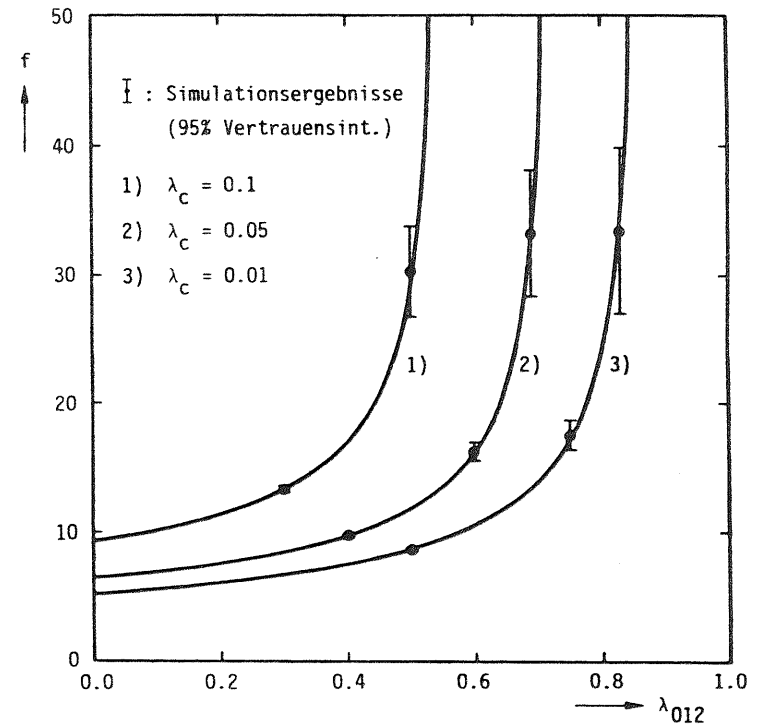
$$P_{i1,0} = 2/9$$

$$P_{i1,j1} = 7/9$$

$$P_{i2,0} = 1/9$$

$$P_{i2,j1} = 8/9$$

Bild 6.2: Mittlere Verweilzeit  $f$  in Abhängigkeit von der externen Ankunftsrate  $\lambda_{012}$ . Fall 1: symmetrische Last, symmetrische Verkehrslenkung. Modell: RDC-Ring. Analyseverfahren: m s d r o a.



Parameter

Anzahl der Bedienstungen :  $N = 10$   
 Bedienungsrate :  $\mu_{ir} = 1.0$  ,  $(i=1, \dots, 10; r=1,2)$   
 Externe Ankunftsrate :  $\lambda_{0i2} = (\lambda_{012}, \lambda_c, \dots, \lambda_c)^T$   
 Verzweigungswahrscheinlichkeiten:

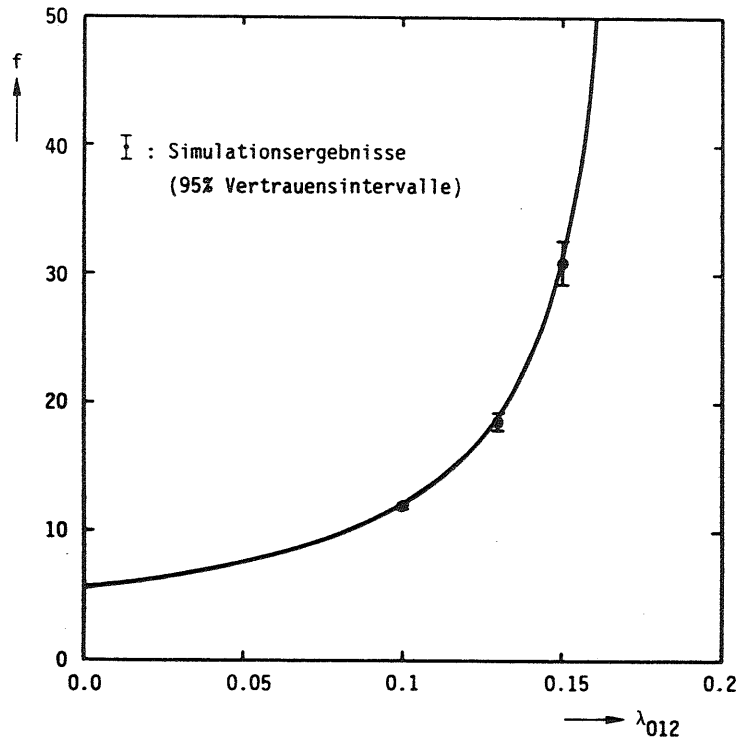
$$P_{i1,0} = 2/9$$

$$P_{i1,j1} = 7/9$$

$$P_{i2,0} = 1/9$$

$$P_{i2,j1} = 8/9$$

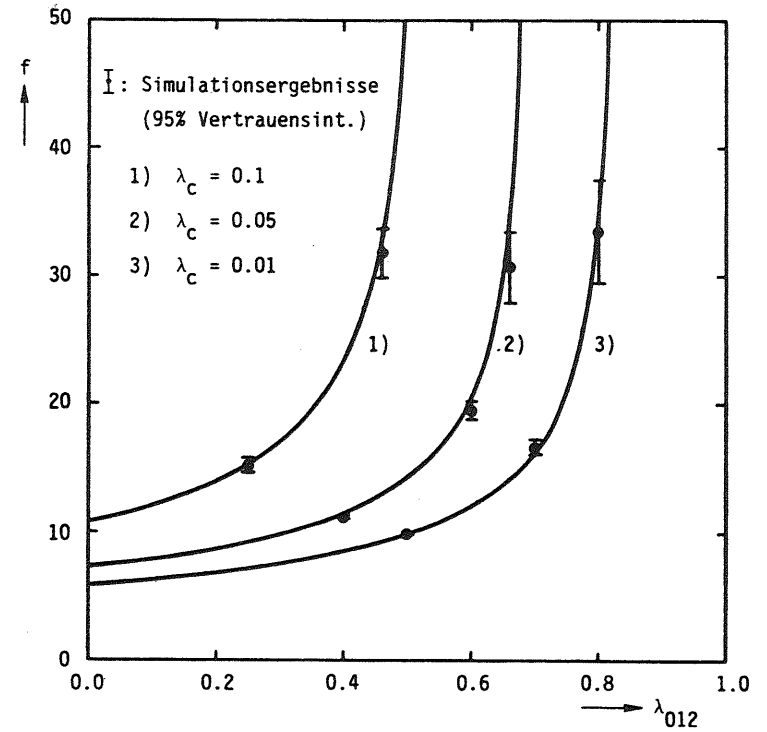
Bild 6.3: Mittlere Verweilzeit  $f$  in Abhängigkeit von der externen Ankunftsrate  $\lambda_{012}$ . Fall 2: unsymmetrische Last, symmetrische Verkehrslenkung. Modell: RDC-Ring. Analyseverfahren: m s d r o a.



Parameter

Anzahl der Bedienstungen :  $N = 10$   
 Bedienungsraten :  $\mu_{ir} = 1.0$  ,  $(i=1, \dots, 10; r=1,2)$   
 Externe Ankunftsrate :  $\lambda_{0i2} = \lambda_{012}$  ,  $(i=1, \dots, 10)$   
 Verzweigungswahrscheinlichkeiten:  
 $P_{i1,0} = 0.1$  ,  $(i=1,2)$   
 $P_{i1,0} = 0.2$  ,  $(i=3, \dots, 10)$   
 $P_{i1,j1} = 0.9$  ,  $(i=1,2; j=i+1)$   
 $P_{i1,j1} = 0.8$  ,  $(i=3, \dots, 10; j=i+1 \text{ für } i < 10, j=1 \text{ für } i=10)$   
 $P_{i2,0} = 0.2$  ,  $(i=1, \dots, 10)$   
 $P_{i2,0} = 0.8$  ,  $(i=1, \dots, 10; j=i+1 \text{ für } i < 10, j=1 \text{ für } i=10)$

Bild 6.4: Mittlere Verweilzeit  $f$  in Abhängigkeit von der externen Ankunftsrate  $\lambda_{012}$ . Fall 3: symmetrische Last, unsymmetrische Verkehrslenkung. Modell: RDC-Ring. Analyseverfahren: m s d r o a.



Parameter

Anzahl der Bedienstungen :  $N = 10$   
 Bedienungsraten :  $\mu_{ir} = 1.0$  ,  $(i=1, \dots, 10; r=1,2)$   
 Externe Ankunftsrate :  $\lambda_{0i2} = (\lambda_{012}, \lambda_c, \dots, \lambda_c)^T$   
 Verzweigungswahrscheinlichkeiten:  
 $P_{i1,0} = 0.1$  ,  $(i=1,2)$   
 $P_{i1,0} = 0.2$  ,  $(i=3, \dots, 10)$   
 $P_{i1,j1} = 0.9$  ,  $(i=1,2; j=i+1)$   
 $P_{i1,j1} = 0.8$  ,  $(i=3, \dots, 10; j=i+1 \text{ für } i < 10, j=1 \text{ für } i=10)$   
 $P_{i2,0} = 0.2$  ,  $(i=1, \dots, 10)$   
 $P_{i2,0} = 0.8$  ,  $(i=1, \dots, 10; j=i+1 \text{ für } i < 10, j=1 \text{ für } i=10)$

Bild 6.5: Mittlere Verweilzeit  $f$  in Abhängigkeit von der externen Ankunftsrate  $\lambda_{012}$ . Fall 4: unsymmetrische Last, unsymmetrische Verkehrslenkung. Modell: RDC-Ring. Analyseverfahren: m s d r o a.

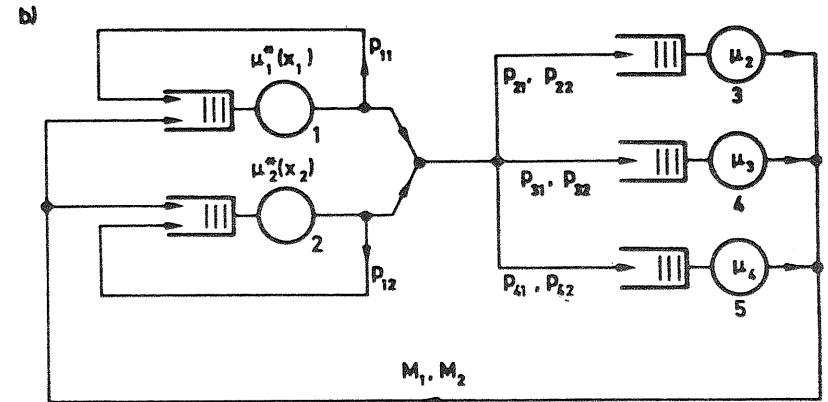
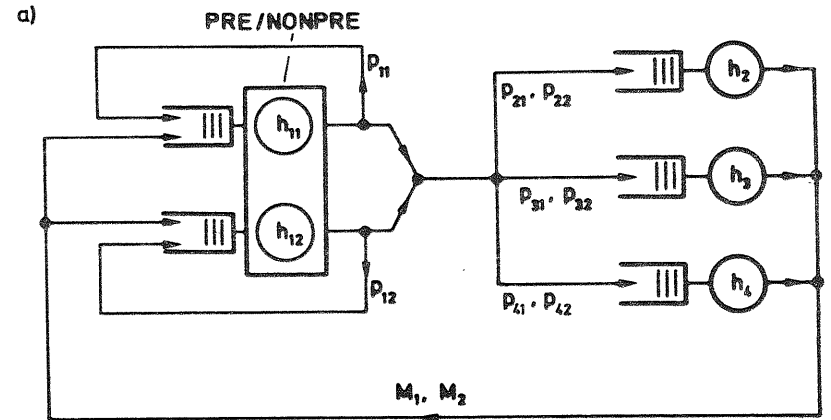
sehr gute Übereinstimmung der Näherungswerte mit den Simulationsresultaten. Trotz der im Prinzip seriellen Struktur des Modells kommt diese Übereinstimmung nicht unerwartet zustande, da für die zweite Klasse immer Poisson-Ankunftsprozesse vorliegen. Der Hauptvorteil des Näherungsverfahrens gegenüber der Simulation zeigt sich beim Vergleich der Rechenzeiten. So wurden zur Erstellung der drei Diagramme in Bild 6.5 (dreihundertsiebenundneunzig Wertepaare) nur 1.34 Sekunden CPU-Zeit auf einer Rechananlage vom Typ Cyber 76 benötigt, gegenüber einer CPU-Zeit von 150 Sekunden für nur einen Simulationswert.

Zum Abschluß sei bemerkt, daß das Ersatz-WSN nach Bild 6.1 im Rahmen der für die Gültigkeit des BCMP-Theorems vorauszusetzenden Bedingungen ohne Probleme erweitert werden kann, um weitere Details (z.B. Verzögerungszeiten in einer Teilnehmerstation) zu berücksichtigen.

### 6.2 "Multiprogramming" in Rechnern mit Speicherhierarchie

Gegenstand der zweiten Systemstudie ist das in Abschnitt 2.5.4 beschriebene CSM, welches eine Analyse des Verkehrsgeschehens in Rechnern mit Multiprogramming gestattet. Das Modell ist nochmals in Bild 6.6a wiedergegeben. Die dort eingetragenen Werte für die Verzweigungswahrscheinlichkeiten und die mittleren Bedienungszeiten in den Ein-/Ausgabe-Kanälen sind aus [Sevcik 77] übernommen worden. In dieser Arbeit sind numerische Ergebnisse für den Mittelwert  $f_1$  der Zeitspanne angegeben, welche zwischen dem Zeitpunkt des erstmaligen Eintritts eines Programmes der Klasse 1 in die CPU-Warteschlange und dem Zeitpunkt des endgültigen Verlassens der CPU (Programmende) vergeht. Die Ergebnisse wurden mit Hilfe der Methode *r o a* berechnet und dienen hier zu Vergleichszwecken.

Das CSM nach Bild 6.6a berücksichtigt unterbrechende oder nichtunterbrechende Prioritäten in der CPU und hat drei Ein-/Ausgabe-Kanäle. Durch Zerlegung der Prioritätsstation entsteht das WSN nach Bild 6.6b. Dieses enthält zwei einklassige Ketten  $E_1$  (BSen 1,3,4,5) und  $E_2$  (BSen 2,3,4,5), welche die erste bzw.



Parameter (nach [Sevcik 77]):

$$h_2 = 0.01, \quad h_3 = 0.02, \quad h_4 = 0.03; \quad \mu_i = h_i^{-1}, \quad (i = 2, 3, 4)$$

$$p_{11} = p_{12} = 0.01, \quad p_{21} = p_{22} = 0.49, \quad p_{31} = p_{32} = 0.3, \quad p_{41} = p_{42} = 0.2$$

Bild 6.6: a) "Central Server Model" mit Prioritätsabfertigung (PRE oder NONPRE) in der CPU  
b) Ersatzmodell zu a)

zweite Prioritätsklasse repräsentieren. Die Zuordnung der Parameter des gegebenen WSnes zum Ersatz-WSN ist aus den genannten Bildern ersichtlich. Für die Approximation der unbekannt Bedienungsraten  $\mu_1^*(x_1)$  und  $\mu_2^*(x_2)$  kommt nur das zweite Verfahren aus dem Abschnitt 5.2.3 des vorhergehenden Kapitels in Frage. Zur Analyse des Ersatz-WSnes ist das am Lehrstuhl für Informatik IV der Universität Dortmund entwickelte Programmsystem COPE verwendet worden, welches u.a. die exakte Analyse von BCMP-Netzen gestattet.

Die folgenden Bilder 6.7a ... 6.18a zeigen Ergebnisse für die Auslastung  $\rho_{1r}$  der CPU durch Anforderungen der Klasse r, (r=1,2), in Abhängigkeit vom Multiprogramminggrad  $M_1$  der Klasse 1. Zum Zwecke der Validierung sind außerdem Simulationsergebnisse mit 95 % Vertrauensintervallen eingetragen. Ferner ist in den Bildern 6.7b ... 6.18b für die Verfahren c c a, r o a und m s d r o a jeweils der Betrag der relativen Abweichung a (in %) der Näherungswerte von den Simulationsergebnissen angegeben. Den für das Verfahren r o a eingetragenen Werten liegen die Ergebnisse aus [Sevcik 77] für die oben definierte Größe  $f_1$  zugrunde. Zwischen dieser mittleren Verweilzeit und der Auslastung  $\rho_{11}$  besteht der einfache Zusammenhang:

$$\rho_{11} = \frac{M_1 \cdot h_{11}}{P_{11} \cdot f_1} .$$

Er ergibt sich im wesentlichen aus der Überlegung, daß die Anzahl der Zyklen, die ein Programm bis zum endgültigen Verlassen der CPU durchläuft, geometrisch verteilt ist. Die Unvollständigkeit der Eintragungen für das Verfahren r o a erklärt sich aus der Tatsache, daß in der Arbeit von Sevcik nur Ergebnisse für die erste Klasse im Falle unterbrechender Prioritäten angegeben sind. Wie schon in Abschnitt 5.2.3 bemerkt, ergeben sich die Resultate für das Verfahren c c a aus einem Zwischenschritt im Verfahren m s d r o a .

Bei der Diskussion der Ergebnisse hinsichtlich ihrer Genauigkeit sind die Fälle  $h_{11} > h_{12}$  (Bilder 6.7 ... 6.10),  $h_{11} = h_{12}$

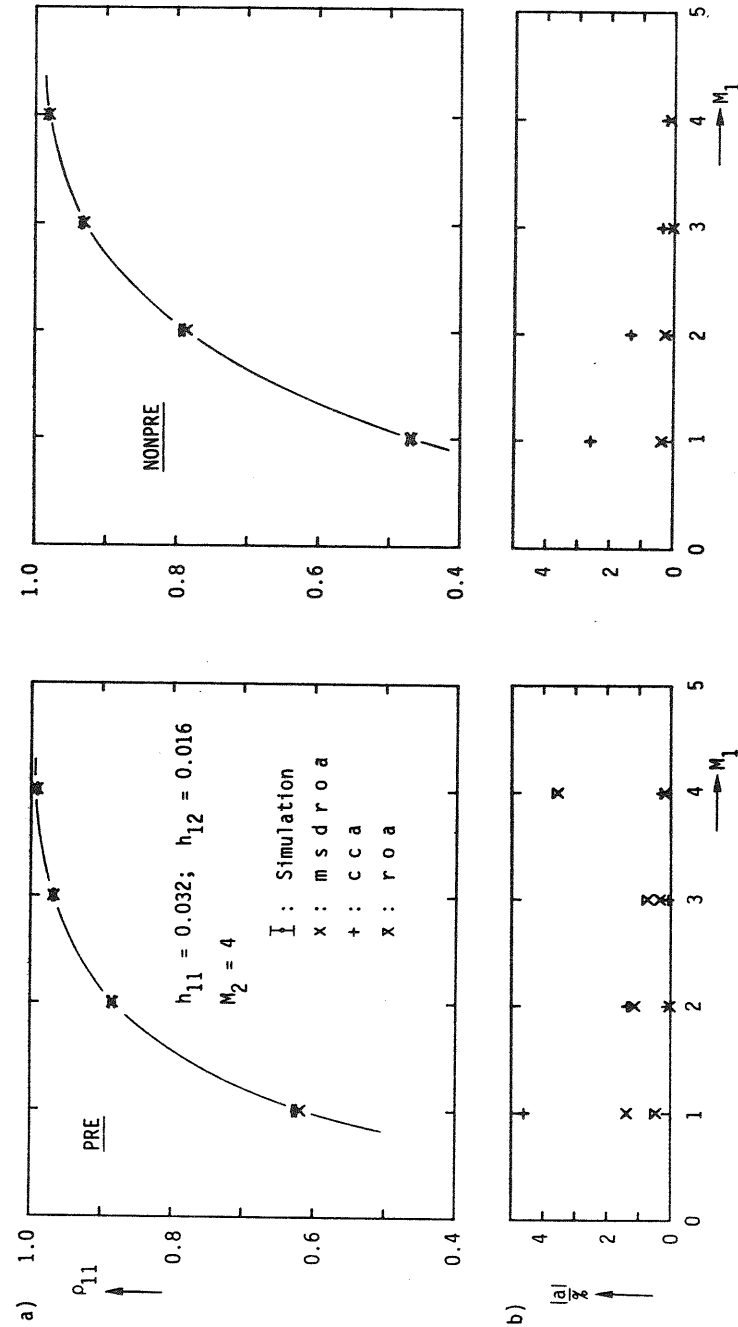


Bild 6.7: a) Auslastung  $\rho_{11}$  der CPU durch die Klasse 1 in Abhängigkeit vom Multiprogramminggrad  $M_1$   
 b) Betrag der relativen Abweichung a der Näherungswerte von den Simulationsergebnissen

Modell: CSM nach Bild 6.6a



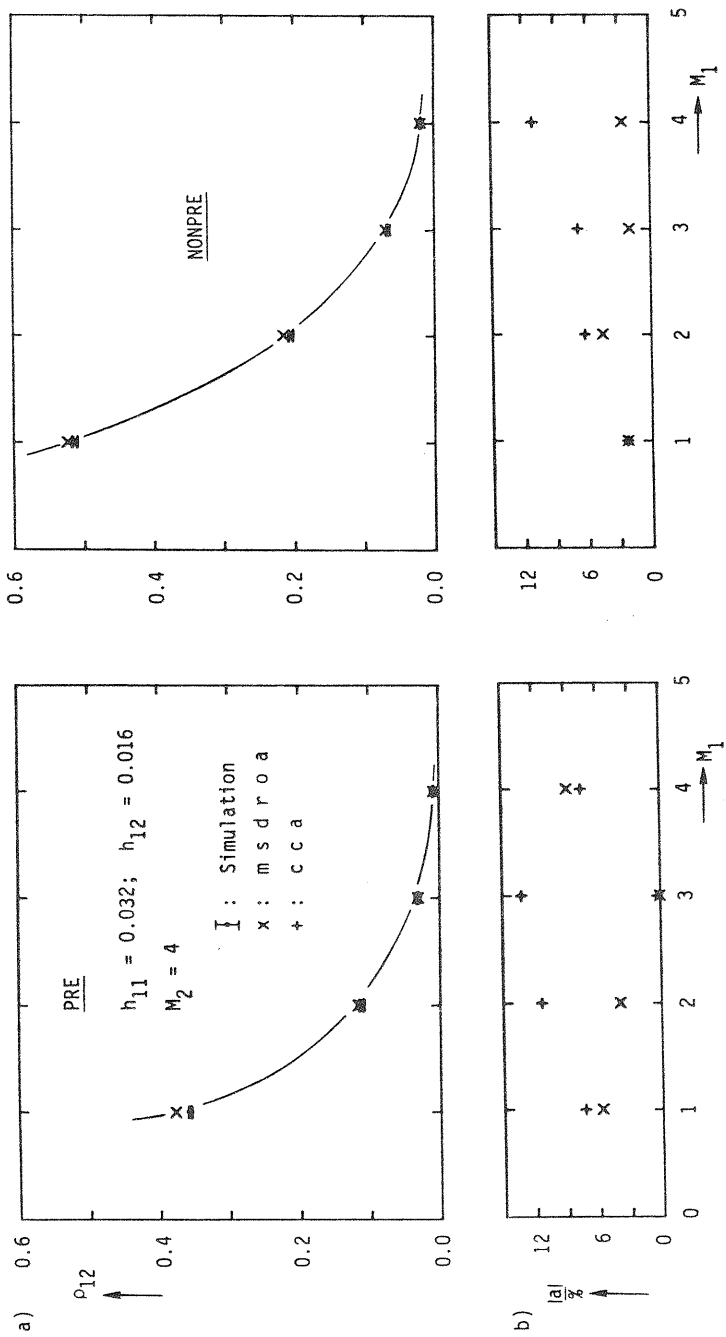


Bild 6.8: a) Auslastung  $\rho_{12}$  der CPU durch die Klasse 2 in Abhängigkeit vom Multiprogrammgrad  $M_1$   
 b) Betrag der relativen Abweichung  $a$  der Näherungswerte von den Simulationsergebnissen  
Modell: CSM nach Bild 6.6a

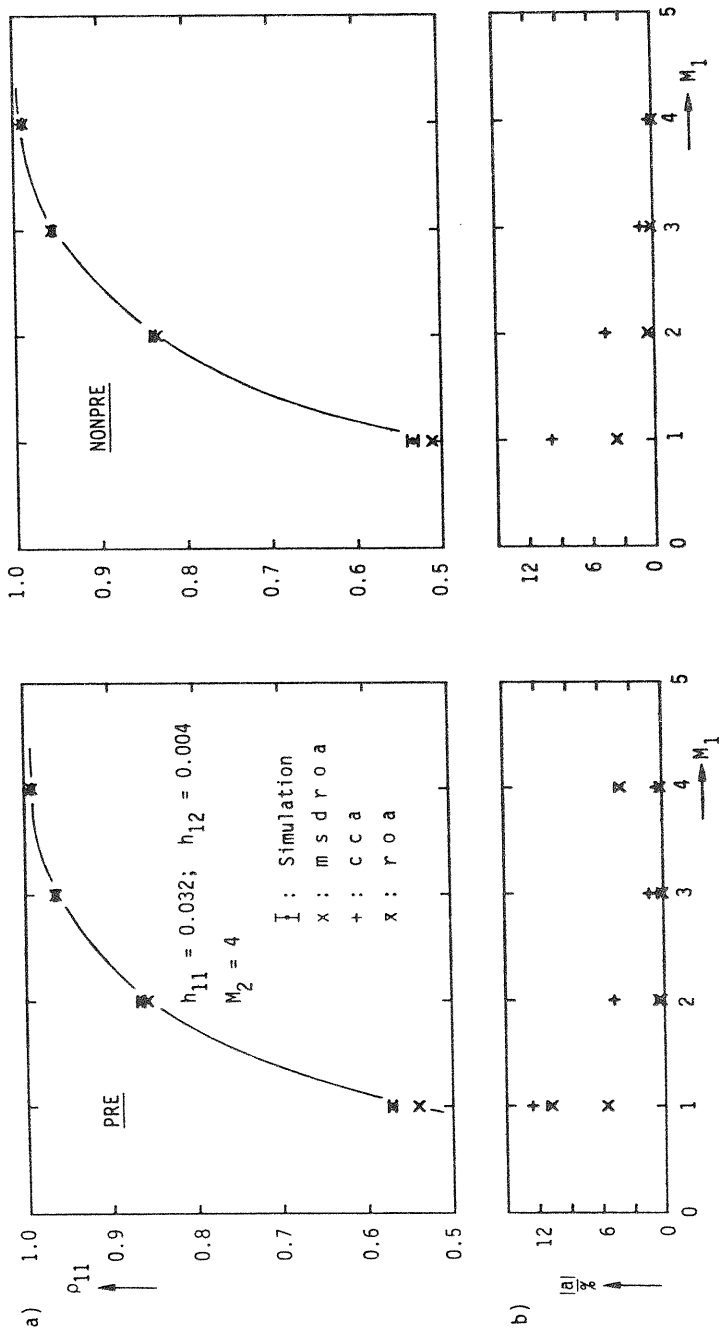


Bild 6.9: a) Auslastung  $\rho_{11}$  der CPU durch die Klasse 1 in Abhängigkeit vom Multiprogrammgrad  $M_1$   
 b) Betrag der relativen Abweichung  $a$  der Näherungswerte von den Simulationsergebnissen  
Modell: CSM nach Bild 6.6a

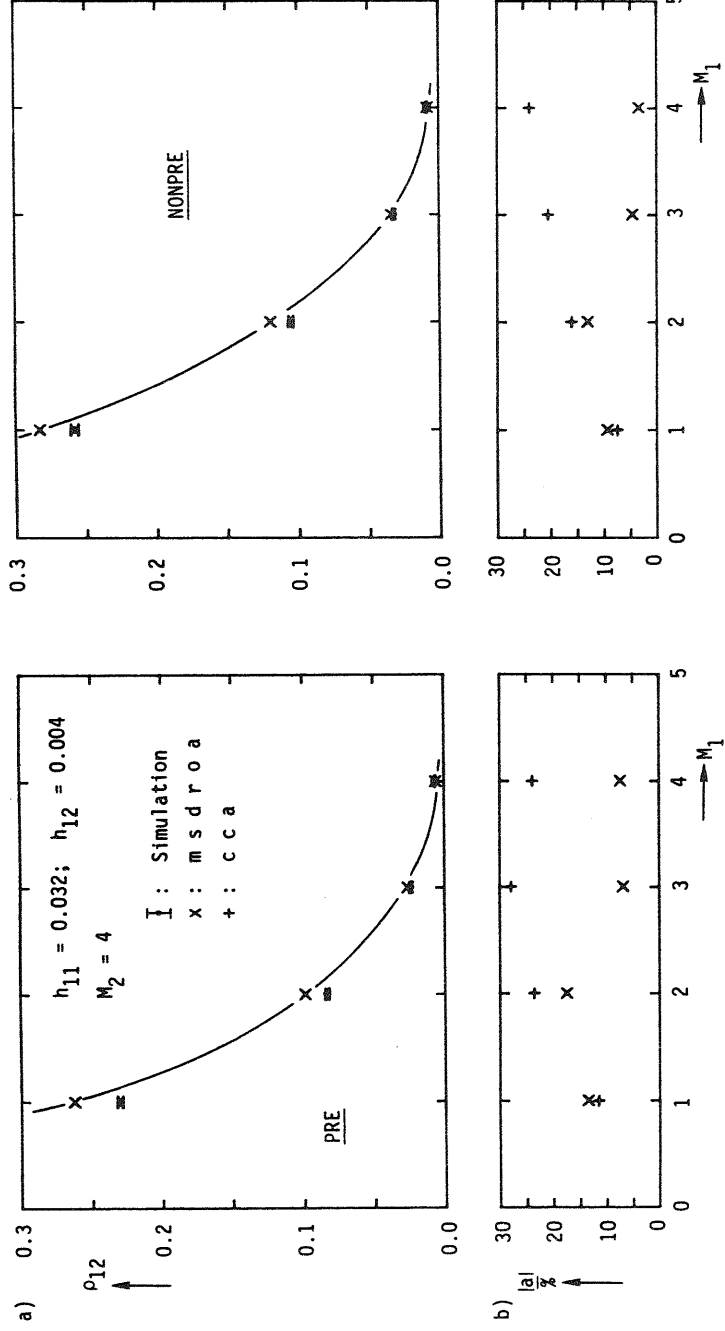


Bild 6.10: a) Auslastung  $\rho_{12}$  der CPU durch die Klasse 2 in Abhängigkeit vom Multiprogrammgrad  $M_1$ .  
 b) Betrag der relativen Abweichung  $a$  der Näherungswerte von den Simulationsergebnissen  
 Modell: CSM nach Bild 6.6a

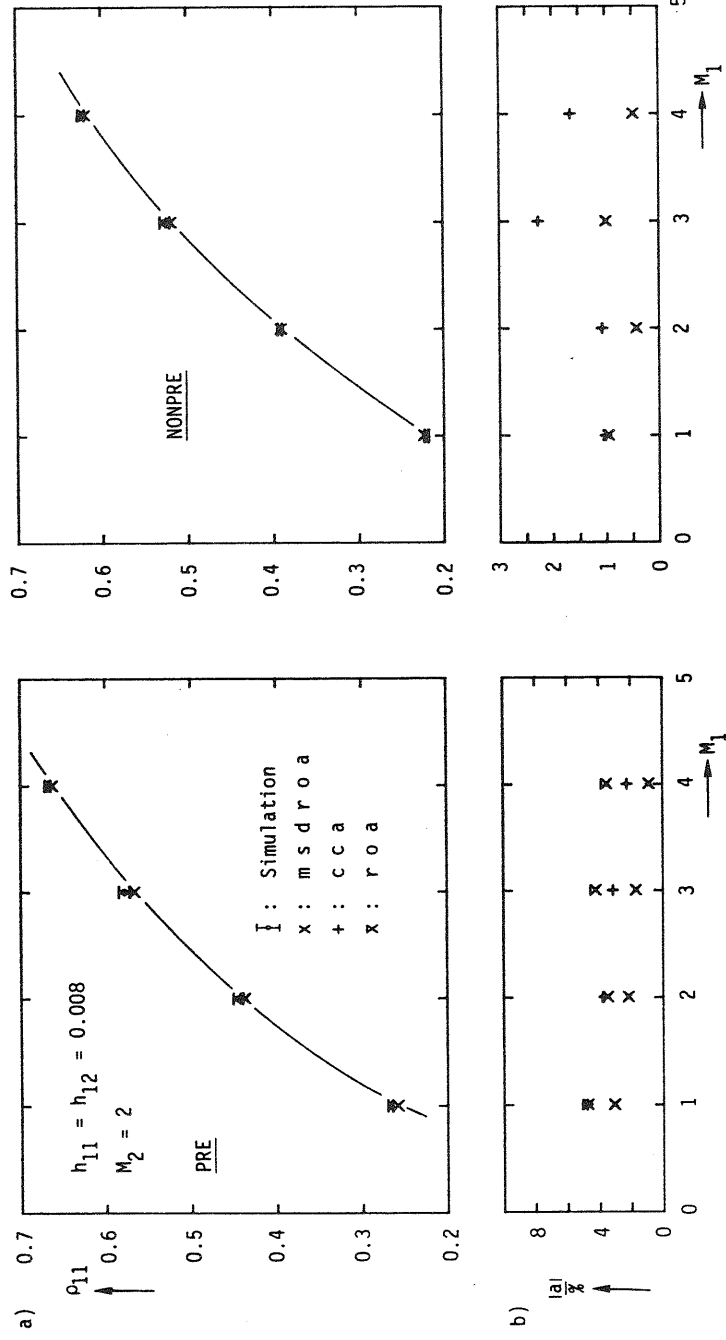


Bild 6.11: a) Auslastung  $\rho_{11}$  der CPU durch die Klasse 1 in Abhängigkeit vom Multiprogrammgrad  $M_1$ .  
 b) Betrag der relativen Abweichung  $a$  der Näherungswerte von den Simulationsergebnissen  
 Modell: CSM nach Bild 6.6a

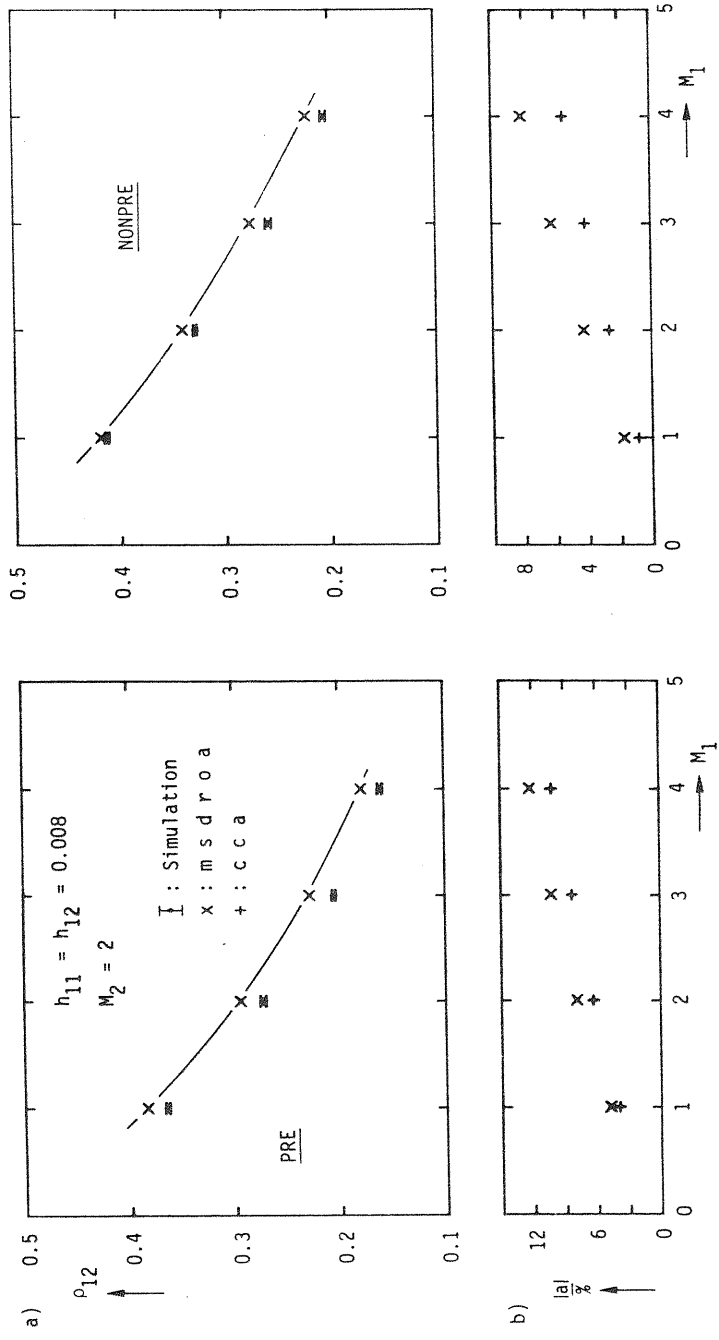


Bild 6.12: a) Auslastung  $\rho_{12}$  der CPU durch die Klasse 2 in Abhängigkeit vom Multiprogrammgrad  $M_1$   
 b) Betrag der relativen Abweichung a der Näherungswerte von den Simulationsergebnissen

Modell: CSM nach Bild 6.6a

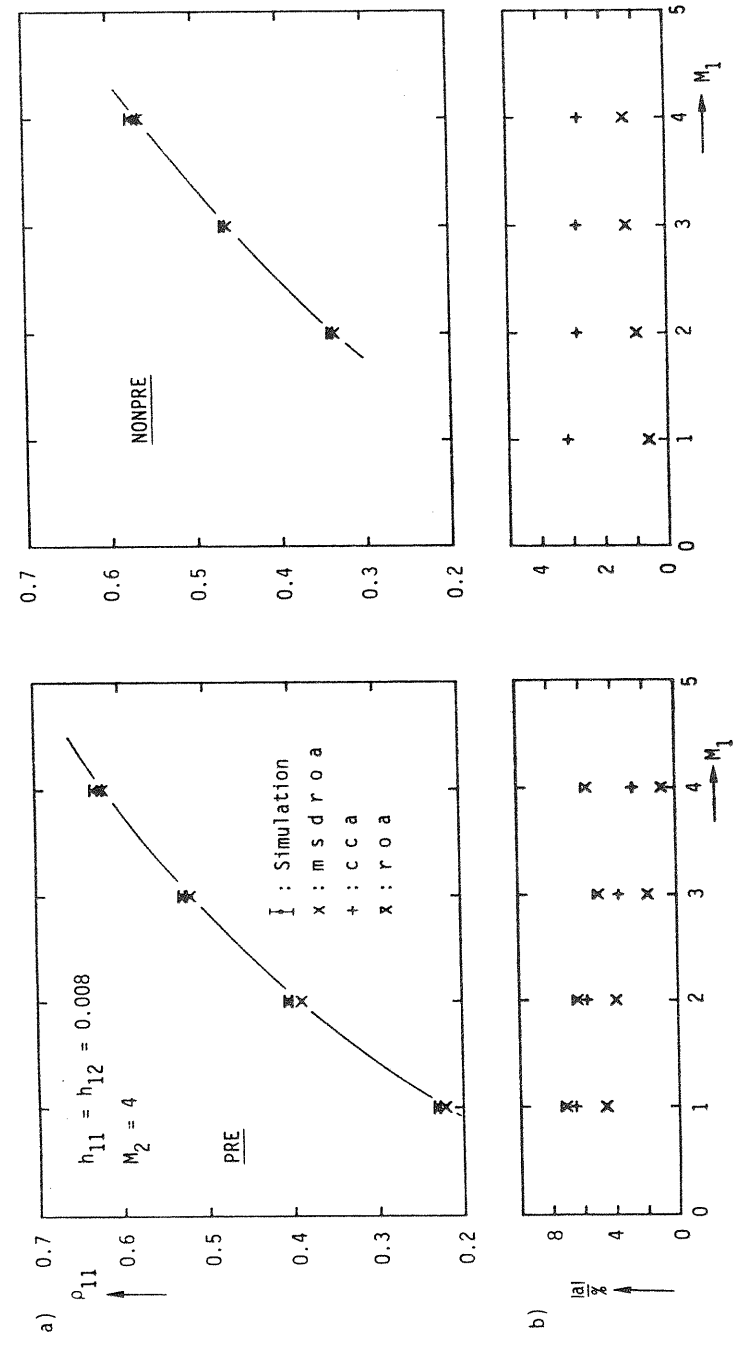


Bild 6.13: a) Auslastung  $\rho_{11}$  der CPU durch die Klasse 1 in Abhängigkeit vom Multiprogrammgrad  $M_1$   
 b) Betrag der relativen Abweichung a der Näherungswerte von den Simulationsergebnissen

Modell: CSM nach Bild 6.6a

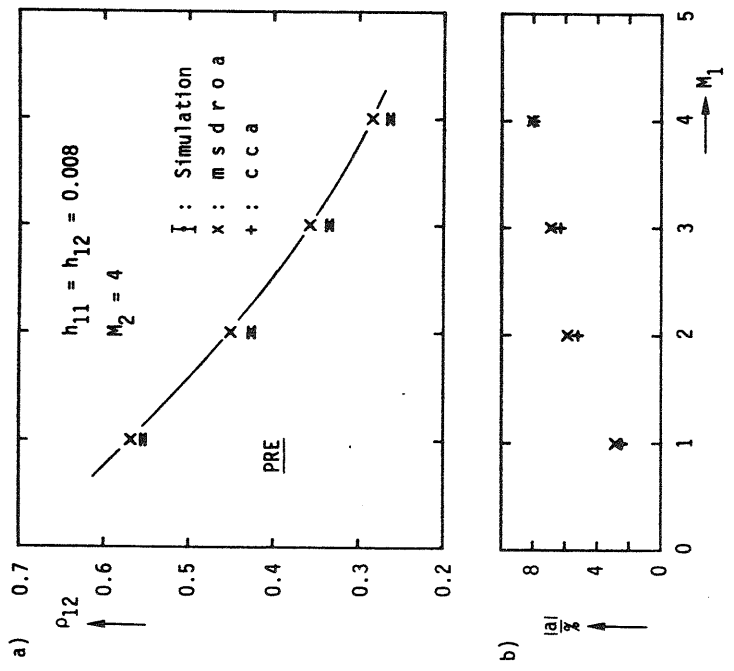
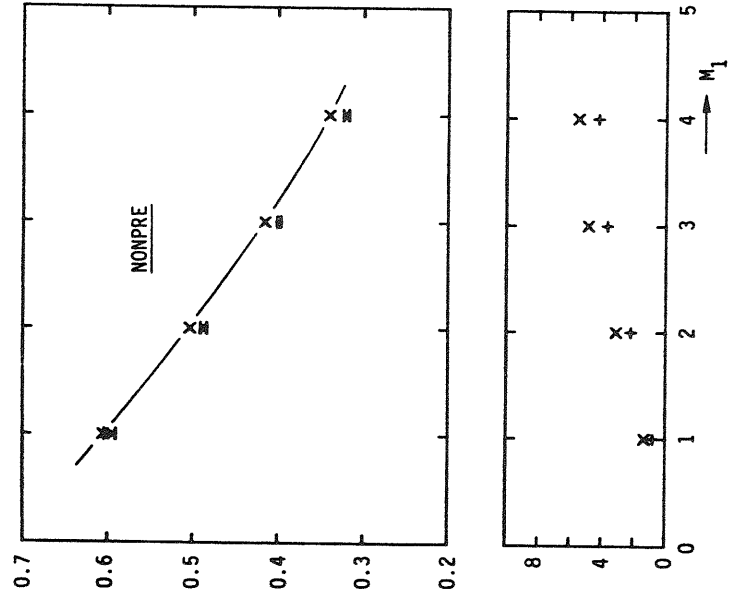


Bild 6.14: a) Auslastung  $\rho_{12}$  der CPU durch die Klasse 2 in Abhängigkeit vom Multiprogrammgrad  $M_1$   
 b) Betrag der relativen Abweichung a der Näherungswerte von den Simulationsergebnissen  
 Modell: CSM nach Bild 6.6a

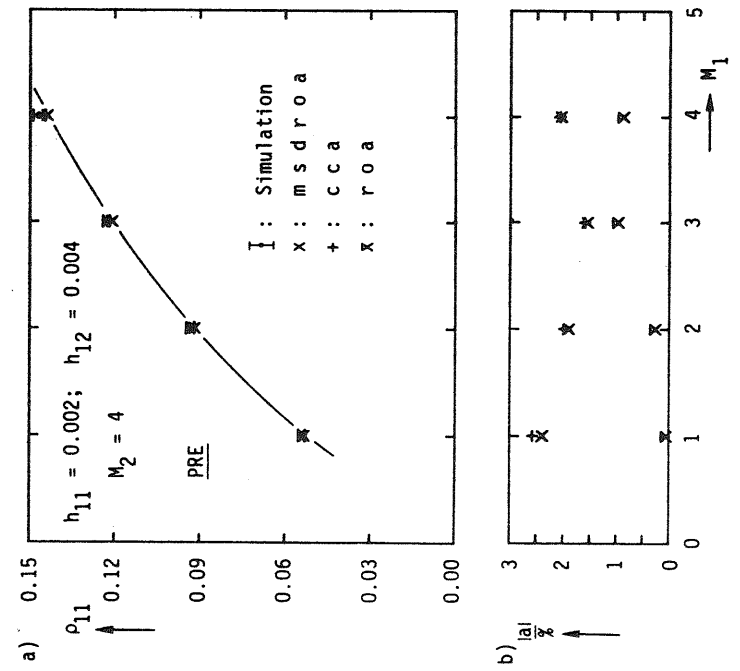
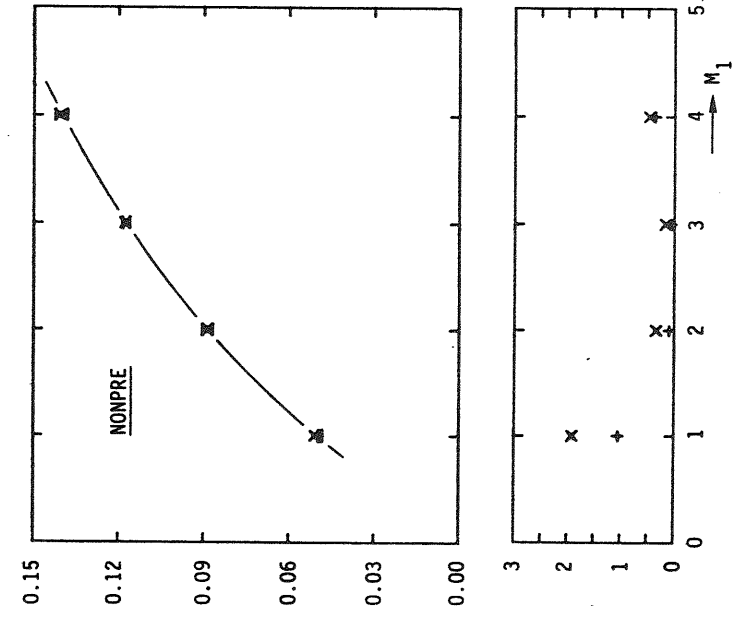
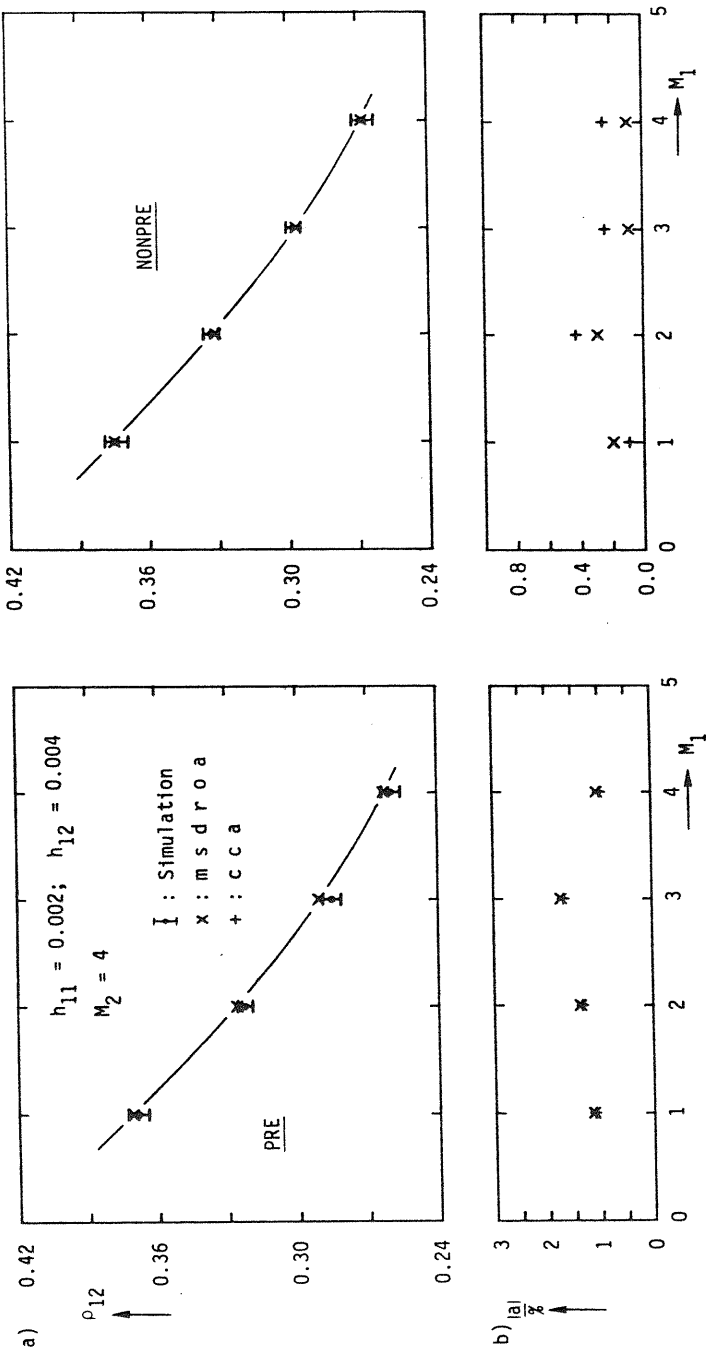
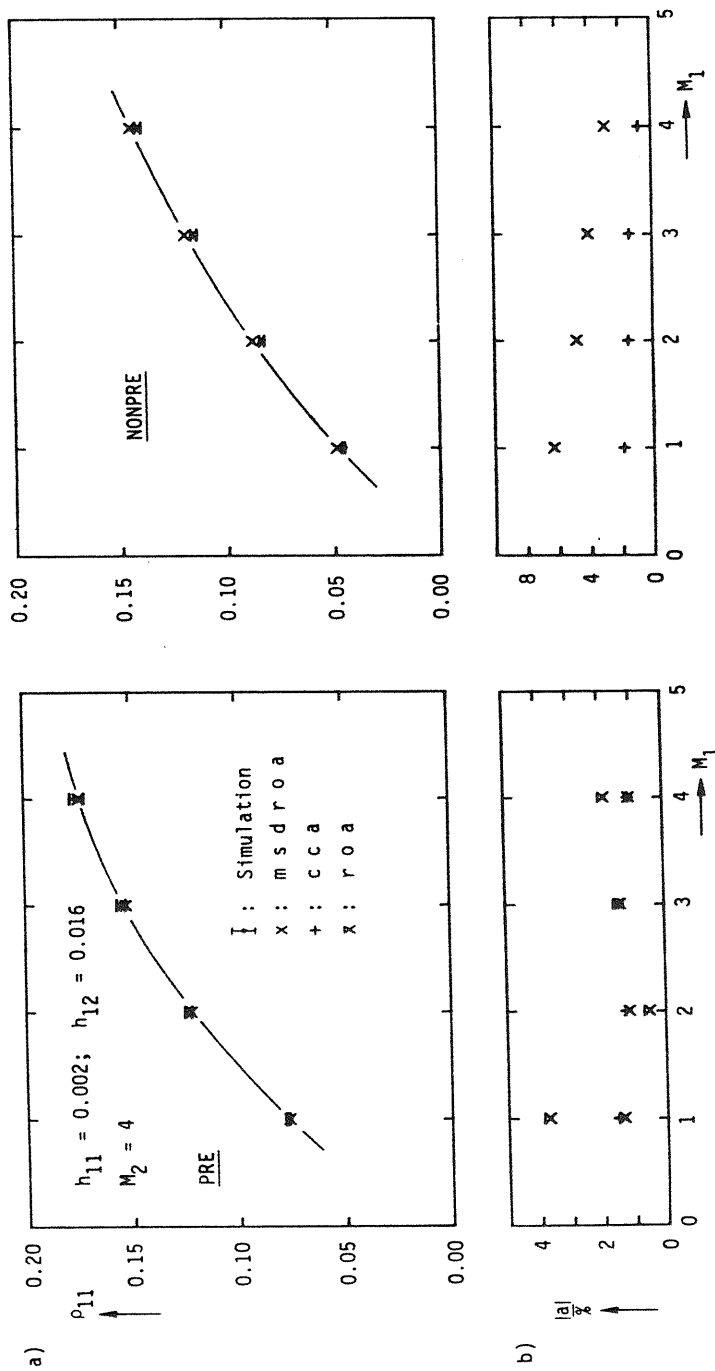


Bild 6.15: a) Auslastung  $\rho_{11}$  der CPU durch die Klasse 1 in Abhängigkeit vom Multiprogrammgrad  $M_1$   
 b) Betrag der relativen Abweichung a der Näherungswerte von den Simulationsergebnissen  
 Modell: CSM nach Bild 6.6a



**Bild 6.16:** a) Auslastung  $\rho_{12}$  der CPU durch die Klasse 2 in Abhängigkeit vom Multiprogrammgrad  $M_1$   
 b) Betrag der relativen Abweichung a der Näherungswerte von den Simulationsergebnissen  
 Modell: CSM nach Bild 6.6a



**Bild 6.17:** a) Auslastung  $\rho_{11}$  der CPU durch die Klasse 1 in Abhängigkeit vom Multiprogrammgrad  $M_1$   
 b) Betrag der relativen Abweichung a der Näherungswerte von den Simulationsergebnissen  
 Modell: CSM nach Bild 6.6a

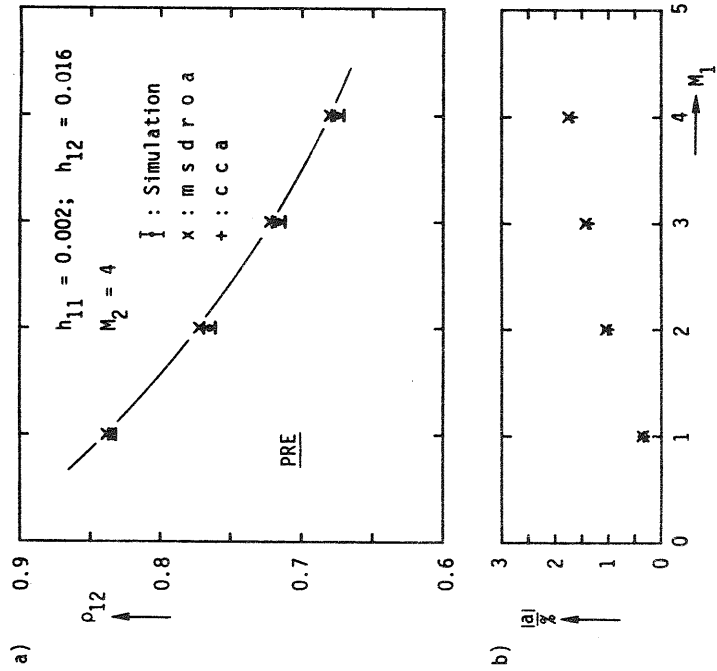
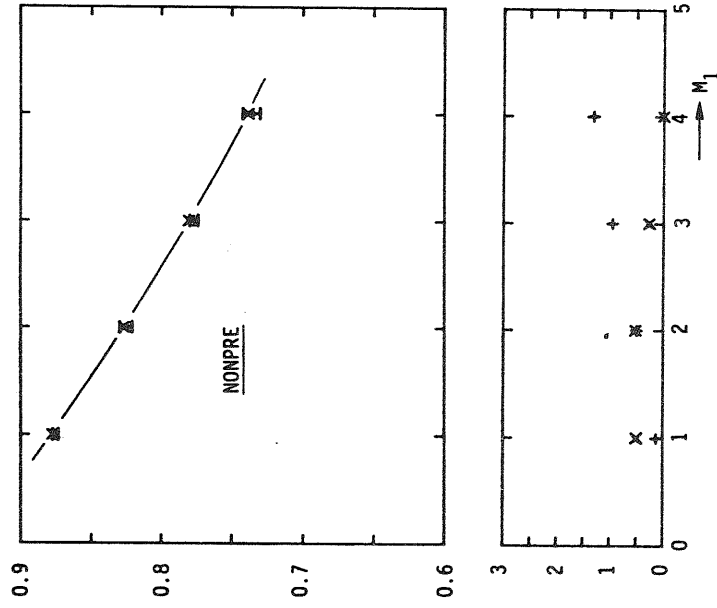


Bild 6.18: a) Auslastung  $\rho_{12}$  der CPU durch die Klasse 2 in Abhängigkeit vom Multiprogrammgrad  $M_1$   
 b) Betrag der relativen Abweichung  $\delta$  der Näherungswerte von den Simulationsergebnissen  
 Modell: CSM nach Bild 6.6a

(Bilder 6.11 ... 6.14) und  $h_{11} < h_{12}$  (Bilder 6.15 ... 6.18) zu unterscheiden.

i)  $h_{11} > h_{12}$

In diesem Fall liefern erwartungsgemäß alle Näherungsverfahren genauere Ergebnisse für die erste als für die zweite Klasse. Die mit den Verfahren *m s d r o a* erzielten Resultate weisen nahezu ohne Ausnahme die größte Genauigkeit auf (in manchen Fällen Verbesserung  $>10\%$  gegenüber *c c a*). Zu den für  $M_1=4$  angegebenen Werten des relativen Fehlers sei angemerkt, daß sie von geringerer Zuverlässigkeit sind, da in diesem Fall die Simulation nur relativ wenige Ereignisse liefert ( $\rho_{12} \approx 0$ ). Ein Vergleich der *r o a*-Ergebnisse mit den *m s d r o a*-Ergebnissen fällt zugunsten des letzteren Verfahrens aus.

ii)  $h_{11} = h_{12}$

Auch hier weisen die Näherungsergebnisse für die erste Klasse durchweg eine größere Genauigkeit auf als für die zweite Klasse. Während sich das Verfahren *m s d r o a* bei den Resultaten für die erste Klasse gegenüber seinen Konkurrenten als überlegen erweist, liefert das Verfahren *c c a* für die zweite Klasse die etwas genaueren Werte. Insgesamt können in diesem Fall *c c a* und *m s d r o a* als ungefähr gleichwertig bezeichnet werden. Die Methode *r o a* liefert hier die schlechteste Übereinstimmung mit der Simulation.

iii)  $h_{11} < h_{12}$

Bei dieser Parameterkonstellation liegen die Verhältnisse im Prinzip umgekehrt wie im Fall i). Die relativen Fehler fallen jedoch durchweg geringer aus ( $<8\%$ ) als dort, und die Genauigkeit der Ergebnisse ist weniger klassenabhängig. Außerdem erweisen sich die Verfahren *c c a* und *m s d r o a* als weitgehend gleichwertig. Eine Ausnahme liegt diesbezüglich nur im Falle der Para-

meterwahl nach Bild 6.17a vor, bei der für unterbrechende Prioritäten der relative Fehler für die Auslastung  $\rho_{11}$  bei der Methode *m s d r o a* bis zu 4.3 % höher liegt als bei der Methode *c c a*. Das Verfahren *r o a* liefert im vorliegenden Fall für die erste Klasse meistens die genauesten Resultate.

Insgesamt gesehen erweist sich damit das neu entwickelte Verfahren *m s d r o a* über die schon im Abschnitt 5.2.3 festgestellten Vorteile hinaus auch hinsichtlich der Genauigkeit der mit ihm erzielbaren Ergebnisse als mindestens gleichwertig gegenüber den Methoden *r o a* und *c c a*.

ANHANG 1: Zur Berechnung der Randverteilungen  $p_2(n)$  für die Warteschlangenmodelle M/M/1/PRE und M/M/1/NONPRE

a) Warteschlangenmodell M/M/1/PRE

Aufgrund des Zusammenhangs  $p_2(n) = \rho_2^{-1} \cdot p(0, n+1)$ , ( $n \geq 0$ ), sind die Gleichgewichtsgleichungen (3.1,8a-d) für die Verbundwahrscheinlichkeiten  $p_2(0, n+1)$  zu lösen. Dies kann mit Hilfe von Standardmethoden aus der Theorie der Differenzgleichungen (siehe z.B. [Jordan 75]) erfolgen und führt nach [Marks 73] zu dem Ergebnis:

$$p(0, n) = B_{0n}, \quad (n \geq 0).$$

Die Koeffizienten  $B_{0n}$  sind rekursiv zu berechnen gemäß:

$$B_{00} = 1 - \rho_1 - \rho_2,$$

$$B_{01} = \frac{\beta-1}{\gamma} B_{00},$$

$$B_{0n} = \frac{\omega}{\gamma} B_{0, n-1} - \rho_2 B_{0, n-2} - \frac{\alpha}{\gamma} \sum_{k=0}^{n-1} B_{k, n-1}, \quad (n \geq 2),$$

$$B_{in} = \frac{1}{i(\beta-\alpha)} \left[ \gamma \rho_2 B_{i-1, n-1} + \sum_{k=i+1}^n \binom{k}{i-1} (\alpha + \beta(-1)^{k-i+1}) B_{kn} \right],$$

( $i = n, n-1, \dots, 1; n \geq 1$ ).

Darin sind die Koeffizienten  $\gamma$ ,  $\alpha$ ,  $\beta$  und  $\omega$  durch die Beziehungen

$$\gamma := \frac{\mu_2}{\mu_1},$$

$$\alpha = \frac{1}{2} (a-b), \quad \beta = \frac{1}{2} (a+b), \quad \text{mit}$$

$$a := \rho_1 + \gamma \rho_2 + 1, \quad b := \sqrt{(\rho_1 + \gamma \rho_2 + 1)^2 - 4\rho_1},$$

$$\omega := \rho_1 + \gamma \rho_2 + \gamma,$$

bestimmt. Die Größen  $\alpha$  und  $\beta$  sind die Wurzeln der zu (3.1,8c)

gehörenden charakteristischen Gleichung.

b) Warteschlangenmodell M/M/1/NONPRE

Zuerst seien die Beziehungen (4.2,18a-b) hergeleitet.

Summiert man in Gleichung (3.1,18b) über m und addiert beiderseits den Term  $-(\lambda+\mu_2) p(0,n,2) + \lambda_2 p(0,n-1,2)$ , so folgt

$$-(\lambda_2+\mu_2) \sum_{m=0}^{\infty} p(m,n,2) + \lambda_2 \sum_{m=0}^{\infty} p(m,n-1,2) = -(\lambda+\mu_2) p(0,n,2) + \lambda_2 p(0,n-1,2), \quad (n \geq 2).$$

Summierung in Gleichung (3.1,18i) über m und beiderseitige Addition von  $-(\lambda+\mu_2) p(0,1,2)$  ergibt:

$$-(\lambda_2+\mu_2) \sum_{m=0}^{\infty} p(m,1,2) = -(\lambda+\mu_2) p(0,1,2).$$

Unter Beachtung der Definition (4.2,2b) und der Beziehung (4.2,9) folgt aus den beiden Gleichungen sofort die Behauptung.

Mit Hilfe der Gleichungen (4.2,18a-b) kann die Randverteilung  $p_2(n)$  berechnet werden, wenn die Verbundverteilung  $p(0,n,2)$  bekannt ist. Letztere lässt sich ähnlich wie im Falle unterbrechender Prioritäten aus den Gleichgewichtsgleichungen (3.1,18a-i) bestimmen. In [Marks 73] ist für den Fall  $\mu_1 = \mu_2$  ein diesbezügliches Berechnungsverfahren angegeben, welches in [Schmitt, Brockert 83] für ungleiche Bedienungsraten verallgemeinert wurde. Danach gilt:

$$p(0,n,2) = \sum_{k=0}^{n-1} A_{kn}, \quad (n \geq 1).$$

Die Berechnung der Koeffizienten  $A_{kn}$  erfolgt rekursiv nach dem folgendem

Algorithmus:

Es seien die Größen  $\gamma, \alpha, \beta$  und  $\omega$  wie unter a) definiert und

es sei ferner vereinbart:

$$\beta := \frac{\rho_1}{\rho_1 + \gamma \rho_2 + \gamma} = \frac{\rho_1}{\omega},$$

$$p_0 := p(0,0,0) = 1 - \rho_1 - \rho_2.$$

Mit  $n_{\max}$  sei der größte Wert von n bezeichnet.

Schritt 1: Berechnung der Koeffizienten  $A_{01}, A_{02}, A_{12}, B_{00}, B_{01}, B_{11}, C_{00}, C_{01}$  und  $C_{11}$ .

$$A_{01} = \frac{\beta-1}{\gamma} \frac{n+(\gamma-1)}{\alpha+(\gamma-1)} p_0,$$

$$B_{00} = \frac{\rho_1 + \gamma \rho_2 + (\gamma-1)}{\alpha+(\gamma-1)} p_0,$$

$$C_{00} = -\frac{\beta-1}{\alpha+(\gamma-1)} p_0,$$

$$A_{12} = \frac{\rho_2}{\rho_1} \gamma n A_{01},$$

$$B_{11} = \frac{\gamma \rho_2}{\beta-\alpha} B_{00},$$

$$C_{11} = -\frac{\gamma}{n+(\gamma-1)} A_{12},$$

$$\begin{bmatrix} A_{02} \\ B_{01} \\ C_{01} \end{bmatrix} = c \begin{bmatrix} \alpha \beta n(n+(\gamma-1)) & -\alpha n(n+(\gamma-1)) & \alpha n(\alpha-n) \\ \gamma n^2(\alpha+\beta+(\gamma-1)) & \gamma n(\gamma-1) & -\gamma n(\alpha+\beta) \\ -\gamma \alpha \beta n & \gamma \alpha n & \gamma \alpha(\beta+n) \end{bmatrix} \cdot \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

mit  $c := [\gamma \alpha n(\alpha+\beta(n+\gamma-1)+n(\gamma-1))]^{-1}$

und



$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} := \begin{bmatrix} (\alpha\beta)/\eta & 0 & 0 & -\alpha & 0 & -n & -\rho_2\gamma \\ 0 & \gamma\eta & \rho_2\gamma\alpha & \alpha(\alpha-\beta) & \rho_2\gamma\eta & 2n^2-\eta(\alpha+\beta) & 0 \\ 0 & 0 & 0 & 0 & -\rho_2\gamma\eta & -(n^2-\alpha\beta) & 0 \end{bmatrix} \cdot \begin{bmatrix} A_{01} \\ A_{12} \\ B_{00} \\ B_{11} \\ C_{00} \\ C_{11} \\ p_0 \end{bmatrix}.$$

Schritt 2: Berechnung der Koeffizienten  $A_{n,n+1}$ ,  $B_{nn}$  und  $C_{nn}$  für  $n = 2, \dots, n_{\max}$ .

$$A_{n,n+1} = \frac{\rho_2}{n\rho_1} \gamma \eta A_{n-1,n},$$

$$B_{nn} = \frac{\rho_2\gamma}{n(\beta-\alpha)} B_{n-1,n-1},$$

$$C_{nn} = -\frac{\gamma}{\eta+(\gamma-1)} A_{n,n+1}.$$

Schritt 3:  $n = 1$  setzen.

Schritt 4:  $n$  um 1 erhöhen.

Schritt 5: Berechnung der Koeffizienten  $A_{i,n+1}$ ,  $B_{in}$  und  $C_{in}$  für  $i = n-1, n-2, \dots, 1$ .

$$A_{i,n+1} = \frac{1}{i} \left[ \sum_{k=1}^{n-i} (-1)^{k+1} \binom{k+i}{i-1} A_{k+i,n+1} + \frac{\rho_2}{\rho_1} \gamma \eta A_{i-1,n} \right],$$

$$B_{in} = \frac{1}{i(\beta-\alpha)} \left[ \sum_{k=1}^{n-i} \binom{k+i}{i-1} B_{k+i,n} [(-1)^{k+1} \beta + \alpha] + \rho_2\gamma B_{i-1,n-1} \right],$$

$$C_{in} = -\frac{1}{\eta(\eta+(\gamma-1))} \sum_{k=1}^{n-i} \binom{k+i}{i} [(-1)^k \alpha + \eta^2] C_{k+i,n} - \left[ \frac{\gamma}{\eta+(\gamma-1)} (A_{i,n+1} + \rho_2 C_{i,n-1}) \right].$$

Schritt 6: Berechnung der Hilfsgrößen  $F_{1n}$ ,  $F_{2n}$  und  $F_{3n}$ .

$$F_{1n} = -\rho_2\gamma A_{0,n-1} + \frac{\alpha\beta}{\eta} A_{0n} - \alpha \sum_{k=1}^n B_{kn} - \eta \sum_{k=1}^n C_{kn},$$

$$F_{2n} = -\gamma\eta \sum_{k=1}^n A_{k,n+1} - \sum_{k=1}^n \alpha [2^k \alpha - (\alpha+\beta)] B_{kn} - \sum_{k=1}^n \eta [2^k \eta - (\alpha+\beta)] C_{kn} - \rho_2\gamma \left[ \alpha \sum_{k=0}^{n-1} B_{k,n-1} + \eta \sum_{k=0}^{n-1} C_{k,n-1} \right],$$

$$F_{3n} = -\rho_2\gamma\eta C_{0,n-1} - \sum_{k=1}^n [\eta^2 + (-1)^k \alpha\beta] C_{kn}.$$

Schritt 7: Berechnung der Koeffizienten  $A_{0,n+1}$ ,  $B_{0n}$  und  $C_{0n}$ .

$$A_{0,n+1} = \frac{\alpha\eta}{D} [\beta[\eta+(\gamma-1)] F_{1n} + [\eta+(\gamma-1)] F_{2n} - (\eta-\alpha) F_{3n}],$$

$$B_{0n} = \frac{\eta}{D} [\gamma\eta[\alpha+\beta+(\gamma-1)] F_{1n} - \gamma(\gamma-1) F_{2n} - \gamma(\alpha+\beta) F_{3n}],$$

$$C_{0n} = \frac{\alpha}{D} [-\gamma\eta\beta F_{1n} - \gamma\eta F_{2n} + \gamma(\beta+\eta) F_{3n}]$$

$$\text{mit } D = \alpha\gamma\eta^2[\alpha+\beta+(\gamma-1)] + \alpha\beta\gamma\eta(\gamma-1).$$

Schritt 8: Falls  $n < n_{\max}$ , Fortsetzung mit Schritt 4.

Falls  $n = n_{\max}$ , Ende des Algorithmus.

**ANHANG 2:** Zur Klasseneinteilung des Zustandsraumes für das serielle Warteschlangenmodell M/M/1/NONPRE →  
•/M/1/NONPRE

Das der Klasseneinteilung zugrunde liegende Prinzip ist am leichtesten anhand der Tabelle A1 zu erklären. Man erkennt dort zunächst eine Grobzerlegung des Zustandsraumes in neun verschiedenen Gruppen A...I, welche sich durch die Belegung der Größen  $s_1$  und  $s_2$  unterscheiden. Die Gruppe I enthält lediglich den Zustand (0,0,0,0,0,0). Alle anderen Gruppen sind nochmals in sechzehn (Gruppen A...D) bzw. vier (Gruppen E...H) disjunkte Klassen unterteilt. Für jede dieser Klassen kann man genau einen repräsentativen Ausschnitt aus dem Übergangsgraphen (entsprechend Bild 5.2) mit der zugehörigen Gleichgewichtsgleichung angeben. Die Klassen unterscheiden sich demnach durch den Gültigkeitsbereich der jeweils mit ihnen korrespondierenden Gleichgewichtsgleichung. Mit Hilfe der Tabelle A1 läßt sich nachprüfen, daß in den Klassen alle möglichen Zustände erfaßt worden sind und daß sie daher das vollständige System der Gleichgewichtsgleichungen erzeugen. Auf eine explizite Angabe dieses gekoppelten Systems partieller Differenzgleichungen mit den zugehörigen Randbedingungen wird in dieser Arbeit verzichtet. Es sei hierzu auf die Arbeit [Schmitt, Schneider 82] verwiesen. Dort findet man die zu den einundachtzig Gleichungen gehörenden Ausschnitte aus dem Übergangsgraphen, welche die nötige Information liefern.

Gruppe A				Gruppe B				Gruppe C				Gruppe D								
Kl.	$x_{11}$	$x_{12}$	$s_1$	$x_{21}$	$x_{22}$	$s_2$	Kl.	$x_{11}$	$x_{12}$	$s_1$	$x_{21}$	$x_{22}$	$s_2$	Kl.	$x_{11}$	$x_{12}$	$s_1$	$x_{21}$	$x_{22}$	$s_2$
A <sub>0</sub>	z2	z1	1	z2	z1	1	B <sub>0</sub>	z2	z1	1	z1	z2	2	D <sub>0</sub>	z1	z2	2	z1	z2	2
A <sub>1</sub>	=1	z1	1	z1	z2	2	B <sub>1</sub>	=1	z1	1	z1	z2	2	D <sub>1</sub>	=0	z2	2	z1	z2	2
A <sub>2</sub>	z2	z1	1	=1	z1	1	B <sub>2</sub>	z2	z1	1	=0	z2	2	D <sub>2</sub>	z1	z2	2	=0	z2	2
A <sub>3</sub>	=1	z1	1	=1	z1	1	B <sub>3</sub>	z2	z1	1	=0	=1	z1	D <sub>3</sub>	=0	z2	2	=0	z2	2
A <sub>4</sub>	z2	z1	1	z2	=0	1	B <sub>4</sub>	z2	z1	1	z1	=1	z1	D <sub>4</sub>	z1	z2	2	z1	=1	z1
A <sub>5</sub>	z2	z1	1	=1	=0	1	B <sub>5</sub>	=1	z1	1	z1	=1	z1	D <sub>5</sub>	=0	z2	2	z1	=1	z1
A <sub>6</sub>	z2	=0	1	z2	z1	1	B <sub>6</sub>	z2	=0	1	z1	z2	2	D <sub>6</sub>	z1	z2	2	=0	=1	z1
A <sub>7</sub>	=1	=0	1	z2	z1	1	B <sub>7</sub>	=1	=0	1	z1	z2	2	D <sub>7</sub>	=0	z2	2	=0	=1	z1
A <sub>8</sub>	z2	=0	1	z2	=0	1	B <sub>8</sub>	z2	=0	1	=0	z2	2	D <sub>8</sub>	z1	=1	z1	z2	2	z1
A <sub>9</sub>	=1	=0	1	z2	=0	1	B <sub>9</sub>	=1	=0	1	=0	z2	2	D <sub>9</sub>	=0	=1	z1	z2	2	z1
A <sub>10</sub>	z2	=0	1	=1	z1	1	B <sub>10</sub>	z2	=0	1	z1	=1	z1	D <sub>10</sub>	z1	=1	z1	z2	2	z1
A <sub>11</sub>	=1	=0	1	=1	z1	1	B <sub>11</sub>	=1	=0	1	z1	=1	z1	D <sub>11</sub>	=0	=1	z1	z2	2	z1
A <sub>12</sub>	z2	=0	1	=1	=0	1	B <sub>12</sub>	z2	=0	1	=0	=1	z1	D <sub>12</sub>	z1	=1	z1	z2	=0	z1
A <sub>13</sub>	=1	=0	1	=1	=0	1	B <sub>13</sub>	=1	=0	1	=0	=1	z1	D <sub>13</sub>	=0	=1	z1	z2	=0	z1
A <sub>14</sub>	=1	z1	1	z2	=0	1	B <sub>14</sub>	=1	z1	1	=0	z2	2	D <sub>14</sub>	z1	=1	z1	z2	=0	z1
A <sub>15</sub>	=1	z1	1	=1	=0	1	B <sub>15</sub>	=1	z1	1	=0	=1	z1	D <sub>15</sub>	=0	=1	z1	z2	=0	z1

Gruppe E				Gruppe F				Gruppe G				Gruppe H								
Kl.	$x_{11}$	$x_{12}$	$s_1$	$x_{21}$	$x_{22}$	$s_2$	Kl.	$x_{11}$	$x_{12}$	$s_1$	$x_{21}$	$x_{22}$	$s_2$	Kl.	$x_{11}$	$x_{12}$	$s_1$	$x_{21}$	$x_{22}$	$s_2$
E <sub>0</sub>	z2	z1	1	0	0	0	F <sub>0</sub>	0	0	0	z2	z1	1	H <sub>0</sub>	0	0	0	z1	z2	2
E <sub>1</sub>	=1	z1	1	0	0	0	F <sub>1</sub>	0	0	0	=1	z1	1	H <sub>1</sub>	0	0	0	=0	z2	2
E <sub>2</sub>	z2	=0	1	0	0	0	F <sub>2</sub>	0	0	0	z2	=0	1	H <sub>2</sub>	0	0	0	z1	=1	z1
E <sub>3</sub>	=1	=0	1	0	0	0	F <sub>3</sub>	0	0	0	=1	=0	1	H <sub>3</sub>	0	0	0	=0	=1	z1

Gruppe I:	
I <sub>0</sub>	=0 =0 0 =0 =0 =0

ANHANG 3: Beweis von Hilfssatz 5.1

Der Beweis wird durch vollständige Induktion über die Zustände  $x_{ir}$ , ( $x_{ir}=0,1,\dots,M_r$ ;  $i=1,\dots,N$ ;  $r=1,\dots,R$ ), geführt.

Induktionsanfang:

Aus den Gleichungen (5.2,6) und (5.2,7) folgt offenbar

$$\beta_i(0, \dots, x_{ir}, \dots, 0) = \prod_{m=1}^{x_i} \mu_i(m), \quad (r=1, \dots, R).$$

Induktionsannahme:

Es gelte

$$\beta_i(x_{i1}, \dots, x_{ir}, \dots, x_{iR}) = \frac{x_{i1}! \cdot x_{i2}! \cdot \dots \cdot x_{iR}!}{x_i!} \prod_{m=1}^{x_i} \mu_i(m).$$

Induktionsschluß:

$$\begin{aligned} \beta_i(x_{i1}, \dots, x_{ir}+1, \dots, x_{iR}) &\stackrel{(5.2,6)}{=} \lambda_{ir}(x_{i1}, \dots, x_{ir}+1, \dots, x_{iR}) \cdot \\ &\quad \beta_i(x_{i1}, \dots, x_{ir}, \dots, x_{iR}) \\ \text{---} &= \frac{x_{ir}+1}{x_i+1} \mu_i(x_i+1) \frac{x_{i1}! \cdot \dots \cdot x_{ir}! \cdot \dots \cdot x_{iR}!}{x_i!} \prod_{m=1}^{x_i} \mu_i(m) \\ \text{---} &= \frac{x_{i1}! \cdot \dots \cdot (x_{ir}+1)! \cdot \dots \cdot x_{iR}!}{(x_i+1)!} \prod_{m=1}^{x_i+1} \mu_i(m), \end{aligned}$$

q. e. d.

