

Institut für Nachrichtenvermittlung und Datenverarbeitung
Universität Stuttgart
Prof. Dr.-Ing. A. Lotze

26. Bericht über verkehrstheoretische Arbeiten

Über die verkehrsgerechte Dimensionierung
der Vermittlungsknoten
in Daten- und Rechnernetzen

von
Werner Bux

Institute of Switching and Data Technics
University of Stuttgart
Prof. Dr.-Ing. A. Lotze

26 th Report on Studies in Congestion Theory

Traffic-Based Dimensioning
of Switching Nodes in Data and
Computer Networks

by
Werner Bux

1980

© 1980 Institut für Nachrichtenvermittlung und Datenverarbeitung Universität Stuttgart
Druck: W. Schräpler
ISBN 3 - 922403 - 36 - 0

ABSTRACT

This report deals with traffic-theoretical problems of store-and-forward networks; it focuses on the performance evaluation of the switching nodes. Models which describe the operation of the nodes to a sufficient level of detail are analyzed in order to provide dimensioning tools.

The nodes are modeled by appropriate queueing networks describing the packet buffers, the switching processor(s), and the data links including link control procedures. Two different node architectures are studied: (1) centralized systems consisting of one switching module only (i.e. one packet buffer memory and one switching processor) and (2) decentralized systems with several switching modules. Decentralized systems can be organized in numerous ways; two different strategies are investigated in this report, named origin-oriented and destination-oriented traffic distribution, respectively.

Exact solutions for the relevant performance measures are derived under the assumption of Poisson arrival processes of the packets and exponentially distributed packet transmission and processing times. Furthermore, approximate solutions are developed for generally distributed transmission and processing times. The results provide a fundamental insight into how the architecture of the nodes, together with the system parameters (buffer memory size, processor speed, transmission rates, etc.) affect the performance measures: buffer overflow probability, throughput, and delay. Subsequently, a brief outline of the report is given.

SECTION 1: Introduction - Overview of report

This section describes the objectives of the report and provides a brief outline of its contents.

SECTION 2: Structure and organization principles of data and computer networks

In this section the basic concepts of data and computer networks are briefly described. Topics discussed include switching principles: circuit switching and store-and-forward techniques, protocols and packet switching networks: network structure, virtual circuits and datagrams, routing and flow control.

SECTION 3: Traffic flow in switching nodes with one switching module

Based on the description of the structure and organization of a switching node, an appropriate traffic model is developed. The following major components of the node and its environment are modeled by a suitable queueing network: packet buffer memory, consisting of a finite number of buffers in each of which one packet of maximum length can be temporarily stored, switching processor and data links connected to the node including link control.

The finite number of buffers in the node is reflected by an upper-bound imposed on the number of packets in the queueing network. Essential for the accurate modeling of a node is the correct description of the link control procedures employed. Therefore this problem is discussed in some detail based on the experience gained from a former study of the standard link control procedure HDLC, c.f. Refs. /11, 12/.

The analysis of the model under the assumption of a Poisson arrival process and exponentially distributed packet transmission and processing times is described in Section 3.3. According to a general theorem for queueing networks /2/, the steady-state probabilities of the network are given by a product form solution from which all interesting performance measures can be derived: buffer overflow probability, mean time between overflows, mean duration of blocking states, buffer load, mean time of packets in buffer, throughput, mean queueing times, and mean queue lengths at the processor and the transmission channels. For the numerical computation of these quantities the convolution algorithm /13,80/, initially developed for closed queueing networks is applied.

Since actual distributions of processing and transmission times may significantly deviate from the exponential distribution, an approximate analysis of the model under the assumption of generally distributed transmission and processing times is provided in Section 3.4. The approximation is based on two assumptions: (1) It is conjectured that also in case of non-exponentially distributed service times, the steady-state probabilities of the considered kind of queueing networks can be described with sufficient accuracy by a product-form solution. (2) The factors of this solution are calculated by analyzing the queues isolated from the network under the

assumption of Poisson arrival processes. In computing the steady-state probabilities of the isolated queues, special algorithms are used to overcome numerical stability problems. Comparisons with simulation results demonstrate that the described approximation approach yields accurate results (see Section 3.5). An interesting outcome of this part of the study is that certain performance measures, for example the buffer overflow probability, are very sensitive to the type of the transmission time distribution.

Aside from this particular result, Section 3.5 contains numerous graphs which characterize the behavior of the switching nodes with respect to the various performance measures mentioned above.

SECTION 4: Traffic flow in switching nodes with several switching modules

Following a brief discussion of the considerations which lead to decentralized node architectures, two different possibilities for distributing the traffic among the modules of the node are described in Section 4.1. The first strategy - named "origin-oriented traffic distribution" - has been suggested as part of the node architecture described in /46/. The concept is to allocate a packet to a certain module depending on the particular line over which it is received. An inherent problem of this scheme is the need for contention resolution when different modules attempt to transmit packets over the same outgoing channel. This conflict can be avoided if - at the expense of some extra buffering and processing capability at the communication interfaces - the second strategy "destination-oriented traffic distribution" is used. In this scheme the communication interfaces which provide the interface function between the communication lines and the system bus are equipped with sufficient buffering and processing capabilities such that they are able to route an incoming packet to a certain module according to its destination address. Each switching module stores and processes only those packets which are forwarded over links, which it is responsible for. Hence, the access conflict described above is avoided.

In the balance of Section 4 both methods are analyzed. The analysis of the destination-oriented strategy in Section 4.2 is straightforward: The model of the entire node can be subdivided into independent submodels of the switching modules for which the solutions derived in Section 3 can be directly applied. Some numerical examples in Section 4.2.3 demonstrate the usefulness of this kind of decentralization.

In contrast to destination orientation, the model of a node employing origin-oriented traffic distribution is relatively complex, as shown in Section 4.3. The node is modeled by a queueing network with as many different classes of packets as switching modules. The number of packets of the same class within the queueing network is upper-bounded by the size of the packet buffer memory of the corresponding module. Despite the complexity of the model, an exact solution for the steady-state probabilities can again be derived with the aid of the queueing network theorem /2/ if the arrival processes are Poisson and transmission and processing times, are exponentially distributed.

However, because of computing time and memory problems, the numerical evaluation of the exact solution is restricted to relatively small systems with respect to the number of modules and packet buffers. Therefore an approximative method is also developed which allows to evaluate systems of almost arbitrary size. The major idea of this solution is to consider one switching module with its finite buffer memory while assuming that all other modules possess infinitely many packet buffers. Applied to each module of the node, this approach yields performance results of excellent accuracy as shown in Section 4.3.5 by comparison to the exact solution and to simulation results.

The approximative solution obtained under the above assumption is of the same type as the solution for nodes with one module. This similarity suggests to apply the approximation developed in Section 3 for the analysis of node models with generally distributed transmission and processing times also to this kind of system. By comparison to simulation results (see Section 4.3.5) it turns out that the combination of both approaches serves indeed as a very reliable method for analyzing nodes with several modules, origin-oriented traffic distribution, and generally distributed transmission and processing times.

SECTION 5: Summary

In this section the major results of the report are briefly summarized.

APPENDIX

The appendix contains a concise representation of known results from the theory of queueing networks which can be applied to the kind of models studied in this investigation. Furthermore, it includes several mathematical derivations and a brief description of the convolution algorithm.

INHALTSVERZEICHNIS

	Seite
Schrifttumsverzeichnis	4
Abkürzungen	11
1. EINLEITUNG - ÜBERBLICK ÜBER DIE ARBEIT	24
2. AUFBAU- UND BETRIEBSPRINZIPIEN VON DATEN- UND RECHNERNETZEN	27
2.1 Vermittlungsverfahren	27
2.2 Protokolle	29
2.3 Paketvermittlungstechnik	31
2.3.1 Struktur von Paketvermittlungsnetzen	31
2.3.2 Virtuelle Verbindungen und Datagrams	32
2.3.3 Wegelenkung und Datenfluß-Steuerung	34
3. VERKEHR IN VERMITTLUNGSKNOTEN MIT EINEM VERMITTLUNGSMODUL	36
3.1 Aufbau und Funktion der Vermittlungsknoten	36
3.2 Verkehrsmodell	38
3.2.1 Modellkomponenten	38
3.2.2 Verkehrsablauf	39
3.2.3 Verkehrsparameter	43
3.3 Analyse bei exponentialverteilten Übertragungs- und Prozessorbelegungszeiten	47
3.3.1 Zustandswahrscheinlichkeiten	47
3.3.2 Charakteristische Verkehrsgrößen	50
3.3.3 Numerische Auswertung	55
3.3.3.1 Zusammenfassung der Verzögerungsstationen	56
3.3.3.2 Numerische Auswertung der Produktlösung	57
3.4 Analyse bei allgemein verteilten Übertragungs- und Prozessorbelegungszeiten	62
3.4.1 Motivation	62
3.4.2 Analyse von Warteschlangennetzen mit allgemein verteilten Bedienungszeiten	63
3.4.3 Näherungsverfahren	63
3.4.3.1 Grundlagen des Verfahrens	63
3.4.3.2 Analyse der isolierten Wartestationen	66
3.4.3.3 Numerische Auswertung der Produktlösung	72
3.4.3.4 Charakteristische Verkehrsgrößen	72
3.4.3.5 Zusammenfassung des Näherungsverfahrens	73
3.5 Ergebnisse	73

	Seite
4. VERKEHR IN VERMITTLUNGSKNOTEN MIT MEHREREN VERMITTLUNGSMODULN	93
4.1 Aufbau und Funktion der Vermittlungsknoten	93
4.1.1 Gründe für die Modularisierung	93
4.1.2 Möglichkeiten der Verkehrsaufteilung	94
4.2 Zielorientierte Verkehrsaufteilung	97
4.2.1 Verkehrsmodell	97
4.2.2 Analyse	99
4.2.3 Ergebnisse	99
4.3 Herkunftsorientierte Verkehrsaufteilung	106
4.3.1 Verkehrsmodell	106
4.3.2 Exakte Analyse bei exponentialverteilten Übertragungs- und Prozessorbelegungszeiten	107
4.3.2.1 Zustandswahrscheinlichkeiten	108
4.3.2.2 Charakteristische Verkehrsgrößen	111
4.3.2.3 Numerische Auswertung	115
4.3.3 Näherungslösung bei exponentialverteilten Übertragungs- und Prozessorbelegungszeiten	118
4.3.3.1 Motivation	118
4.3.3.2 Beschreibung des Approximationsverfahrens	119
4.3.4 Näherungslösung bei allgemein verteilten Übertragungs- und Prozessorbelegungszeiten	122
4.3.5 Ergebnisse	123
5. ZUSAMMENFASSUNG	132
ANHANG	134
Anhang 1 : Theorem über Warteschlangennetze	134
Anhang 2 : Zusammenfassung der Verzögerungsstationen	138
Anhang 3 : Umformung der Rekursion für die stationären Zustandswahrscheinlichkeiten des Wartesystems $M/G/1,s$	139
Anhang 4 : Marginalverteilung für einen endlich großen Speichermodul bei unendlich großen Speichern der übrigen Moduln	141
Anhang 5 : Faltungsalgorithmus	144

SCHRIFTTUMSVERZEICHNIS

[1] ANSI: American National Standard for Advanced Data Communication Control Procedures (ADCCP). Document ANSI/B5R 3.66.

[2] Baskett, F., Chandy, K.M., Muntz, R.R., Palacios, F.G.: Open, Closed, and Mixed Networks of Queues with Different Classes of Customers. J.ACM 22(1975)2, S. 248-260.

[3] Bocker, P.: Datenübertragung I und II. Springer-Verlag, Berlin, 1977.

[4] Bocker, P.: Möglichkeiten und Grenzen von Paketvermittlungsnetzen. telcom report 1(1978), S. 110-117.

[5] Boell, H.-P.: Datenfernübertragungsnetze - Eine Komponentendarstellung. NTZ 30(1977)6, S. 483-490.

[6] Brandt, G.J., Chretien, G.J.: Methods to Control and Operate a Message Switching Network. Symposium on Computer-Communications Networks and Teletraffic, Brooklyn, 4.-6. April, 1972.

[7] Brizzi, M., Grillo, D.: Implementation of Algorithms for Performance Analysis of a Class of Multiprogrammed Computers. Proc. of the 3rd International Symposium on Modelling and Performance Evaluation of Computer Systems, Bonn-Bad Godesberg, 3.-5. Okt. 1977, North-Holland Publ. Comp., Amsterdam/New York/Oxford, 1977, S. 131-150.

[8] Brockmeyer, E., Halstrøm, H.L., Jensen, A.: The Life and Work of A.K. Erlang. Acta Polytechnica Scandinavia, 1960.

[9] Bux, W., Herzog, U.: The Phase-Concept: Approximation of Measured Data and Performance Analysis. Proc. of the International Symposium on Computer Performance Modeling, Measurement and Evaluation, Yorktown Heights, N.Y., 16.-18. August, 1977, Hrsg.: K.M. Chandy und M. Reiser, North-Holland Publ. Comp., Amsterdam/New York/Oxford, 1977, S. 23-38.

[10] Bux, W., Kühn, P., Kümmerle, K.: Throughput Considerations in a Multiprocessor Packet-Switching Node. IEEE-COM 27(1979)4, S. 745-750.

[11] Bux, W., Kümmerle, K., Truong, H.L.: Results on the Performance of Balanced HDLC Procedures. Proc. National Telecommunications Conference, Birmingham, Alabama, 4.-6. Dezember 1978, S. 28.3.1-28.3.7.

[12] Bux, W., Truong, H.L.: A Queueing Model for HDLC-Controlled Data Links. Proc. of the International Symposium on Flow Control in Computer Networks, Versailles, 12.-14. Februar 1979, Hrsg.: J.-L. Grange und M. Gien, North-Holland Publ. Comp. Amsterdam/New York/Oxford, 1979, S. 287-306.

[13] Buzen, J.P.: Computational Algorithms for Closed Queueing Networks with Exponential Servers. C.ACM 16(1973)9, S.527-531.

[14] CCITT: Recommendation X.3: Packet assembly/disassembly facility (PAD) in a public data network. Circular letter no. 35, Genf, Juli 1977.

[15] CCITT: Recommendation X.21: General purpose interface between data terminal equipment and data circuit-terminating equipment for synchronous operation on public data networks. Orange Book, Vol. VIII.2, ITU, Genf, 1977.

[16] CCITT: Recommendation X.25: Interface between data terminal equipment and data circuit-terminating equipment for terminals operating in the packet mode on public data networks. Orange Book, Vol. VIII.2, ITU, Genf, 1977.

[17] Chandy, K.M., Herzog, U., Woo, L.: Approximate Analysis of General Queueing Networks. IBM Journ. of Res. and Dev. 19(1975)1, S. 43-49.

[18] Chu, W.W.: Optimal Message Block Size for Computer Communications with Error Detection and Retransmission Strategies. IEEE-COM 22(1974)10, S. 1516-1525.

[19] Clipsham, B., Narraway, M.: Datapac: A Public, Shared Data Network for Canada. Tesis (1976), S. 130-136.

[20] Cooper, R.B.: Introduction to Queueing Theory. The Macmillan Company, New York, 1972.

[21] Cox, D.R.: A Use of Complex Probabilities in the Theory of Stochastic Processes. Proc. Camb. Phil. Soc. 51(1955), S. 313-319.

[22] Crowther, W.R., Heart, F.E., McKenzie, A.A., McQuillan, J.M., Walden, D.C.: Issues in Packet Switching Network Design. Proc. National Computer Conference, 1975, S. 116-175.

[23] Danet, A., Despres, R., Le Rest, A., Pichon, G., Ritzenthaler, S.: The French Public Packet Switching Service: the Transpac Network. Proc. ICC 76 Toronto, 3.-6. August 1976, S. 251-260.

|24| Davies, D.W.,
Barber, D.L.A.:
Communications Networks for Computers.
J. Wiley & Sons, London/New York, 1973.

|25| Emmons, W.F.:
Data Network Protocol Standards. Proc.
of the Computer Network Symposium,
Gaithersburg, Maryland, 15. Dezember
1977, S. 9-17.

|26| European Computer and Communications
Markets 1973-1985. PA International
Management Consultants and Quantum
Science Corp., London/New York, 1973.

|27| Fayolle, G.,
Gelenbe, E.,
Pujolle, G.:
An Analytic Evaluation of the Perform-
ance of the "Send and Wait" Protocol.
IEEE-COM 26(1978), S. 313-319.

|28| Feldmann, J.,
Jilek, P.,
Nowak, R.:
Das Datenfernverarbeitungssystem TRANS-
DATA von Siemens. telcom report 1(1978),
S. 70-77.

|29| Frank, H.,
Chou, W.:
Routing in Computer Networks. Networks
1(1971), S. 99-112.

|30| Frank, H.,
Kahn, R.E.,
Kleinrock, L.:
Computer Communications Network Design:
Experience with Theory and Practice.
Networks 2(1972), S. 135-166.

|31| Fultz, G.L.:
Adaptive Routing Techniques for Message
Switching Computer-Communication Net-
works. Ph.D. Dissertation, University
of California, School of Engineering
and Applied Science, Los Angeles, 1972.

|32| Giffin, W.C.:
Transform Techniques for Probability
Modeling. Academic Press, New York, 1975.

|33| Gnedenko, B.W.,
Kowalenko, I.N.:
Einführung in die Bedienungstheorie.
R. Oldenbourg Verlag, München/Wien, 1971.

|34| Gray, J.P.:
Line Control Procedures. IEEE-COM 60
(1972), S. 1301-1312.

|35| Guenther, K.:
Vermeidung von Deadlocks im allgemeinen
und von Pufferdeadlocks bei Store-and-
Forward-Communication im besonderen.
Gesellschaft für Mathematik und Daten-
verarbeitung, GMD-Dokument No. 20, 1975.

|36| Heart, F.E.,
Kahn, R.F.,
Ornstein, S.M.,
Crowther, W.R.,
Walden, D.C.:
The Interface Message Processor for the
ARPA Computer Network. Proc. AFIPS SJCC
36(1970), S. 551-567.

|37| Heart, F.E.,
Ornstein, S.M.,
Crowther, W.R.,
Barker, W.B.:
A New Minicomputer/Multiprocessor for
the ARPA Network. AFIPS Conf. Proc.,
1973 Nat. Comp. Conf., Vol. 42, Mont-
vale, NJ, AFIPS Press, 1973, S. 529-
537.

|38| Herzog, U.:
Verkehrsfluß in Datennetzen. Habilita-
tionsschrift, Universität Stuttgart, 1973.

|39| IBM (Hrsg.)
IBM Synchronous Data Link Control -
General Information. GA 270930, 1974.

|40| Irland, M.I.:
Buffer Management in a Packet Switch.
IEEE-COM 26(1978), S. 328-337.

|41| ISO:
Data communication - High level data
link control procedures - Frame struc-
ture. International Standard ISO 3309.

|42| ISO:
Data communication - High level data
link control procedures - Elements of
procedures (independent numbering).
International Standard ISO 4335.

|43| ISO:
HDLC - Proposed unbalanced classes of
procedures. Document ISO/TC 97/SC 6 -
No. 1339.

|44| ISO:
HDLC - Proposed balanced class of proce-
dures. Document ISO/TC 97/SC 6 - No.
1444.

|45| Jackson, J.R.:
Jobshop-Like Queueing Systems. Manage-
ment Science 10(1963), S. 131-142.

|46| Jenny, C.J.,
Kümmerle, K.:
Distributed Processing within an Inte-
grated Circuit/Packet-Switching Node.
IEEE-COM 24(1976), S. 1089-1100.

|47| Johnson, T.:
Packet Switching Services and the Data
Communications User. Ovum Ltd., London,
1976.

|48| Kamoun, F.,
Kleinrock, L.:
Analysis of Shared Storage in a Computer
Network Environment. Proc. of the 9th
HICSS, Honolulu (Hawaii), 1976, S. 89-
92.

|49| Karp, H.R.:
Basics of Data Communications. McGraw-
Hill, 1976.

|50| Kelly, P.T.F.,
Lee, E.J.B.:
The Telecommunications Network for EURO-
NET. POEEJ 70(1977), S. 208-216.

|51| Kleinrock, L.,
Naylor, W.E.:
On Measured Behavior of the ARPA Network.
Nat. Comp. Conf. 1974, AFIPS Conf. Proc.
Vol. 43, AFIPS Press, Montvale, NJ, 1974,
S. 767-780.

|52| Kleinrock, L.:
Queueing Systems. Vol. I: Theory.
J. Wiley & Sons, New York, 1975.

|53| Kleinrock, L.:
Queueing Systems. Vol. II: Computer
Applications. J. Wiley & Sons, New York,
1976.

|54| Kobayashi, H.:
Application of the Diffusion Approxima-
tion to Queueing Networks. IBM Journ. of
Res. and Dev. 19(1974)2, S. 316-328.

- [55] Kohl, H.: Systems Network Architecture: Rechner-netze (1), (2). IBM Nachrichten 26(1976), S. 396-407 und IBM Nachrichten 27(1977), S. 63-70.
- [56] KtK: Kommission für den Ausbau des technischen Kommunikationssystems - Telekommunikationsbericht. Hrsg.: Bundesministerium für das Post- und Fernmeldewesen, Bonn, 1976.
- [57] Kühn, P.: Analysis of Complex Queueing Networks by Decomposition. IEEE-COM 27(1979)1, S. 113-126.
- [58] Kümmerle, K., Rudin, H.: Packet and Circuit Switching: A Comparison of Cost and Performance. Nat. Telecomm. Conf., Dallas, Texas, 29. Nov. - 1. Dez. 1976.
- [59] Lam, S.S.: Store-and-Forward Buffer Requirements in a Packet Switching Network. IEEE-COM 24(1976)4, S. 394-403.
- [60] Lam, S.S., Reiser, M.: Congestion Control of Store-and-Forward Networks by Input Buffer Limits. Proc. Nat. Telecomm. Conf. 77, 5.-7. Dez. 1977, Los Angeles, S. 12:1-1 - 12:1-6.
- [61] Lotze, A.: Loss Formula, Artificial Traffic Checks, and Quality Standards for Characterizing One Stage Gradings. 3rd International Teletraffic Congress, Paris 1961, Doc. 28.
- [62] Lotze, A.: Verlust und Güteigenschaften einstufiger Mischungen. NTZ 9(1961), S. 449-453.
- [63] Lotze, A.: Nachrichtenvermittlung II (Verkehrstheorie). Vorlesung an der Universität Stuttgart.
- [64] Mair, E., Schaffer, B., Schubert, R.: Struktur des Siemens-Systems EDS. Siemens-Zeitschrift 51(1977), S. 82-87.
- [65] Manning, E.G.: On Datagram Service in Public Packet-Switched Networks. Computer Networks 2(1978), S. 79-83.
- [66] Martin, J.: Telecommunications and the Computer. 2nd Ed., Prentice-Hall, 1976.
- [67] McFayden, J.H.: Systems Network Architecture: An Overview. IBM Systems Journal 15(1976), S. 4-23.
- [68] McQuillan, J.M.: Adaptive Routing Algorithms for Distributed Computer Networks. BBN Report No. 2831, Mai 1974.
- [69] Mills, D.L.: Communication Software. IEEE-COM 60(1972), S. 1333-1341.

- [70] Neuts, M.: Algorithms for the Waiting Time Distribution under Various Queue Disciplines in the M/G/1 Queue with Service Time Distribution of Phase Type. Management Science (Special Issue on Algorithmic Methods in Probability) 1977.
- [71] Opderbeck, H., Hovey, R.B.: TELENET - Network Features and Interface Protocols. NTG-Fachtagung Daten-netze, Baden-Baden, 18.-20. Febr. 1976, VDE-Verlag, Berlin, 1976, S. 145-156.
- [72] Ornstein, S.M.: The Terminal IMP for the ARPA Computer Network. AFIPS SJCC 40(1972), S. 243-254.
- [73] Ornstein, S.M.: Pluribus - A Reliable Multiprocessor. AFIPS Conf. Proc., 1975 Nat. Comp. Conf., Vol. 45, Montvale, NJ, AFIPS Press, 1975, S. 551-559.
- [74] Palm, C.: Intensitätsschwankungen im Fernsprech-verkehr. Ericsson Technics, 6(1943), S. 1-189.
- [75] Petersen, J.: Remarks on the Implementation of the Packet Level Protocols of Public Packet Switching Networks. Proc. of the Computer Network Protocols Symposium, Liège, 13.-15. Febr. 1978, S. A2-1 - A2-7.
- [76] Pollaczek, F.: Über eine Aufgabe der Wahrscheinlich-keitstheorie I, II. Math. Zeitschrift, 32(1930), S. 64-100 und 729-750.
- [77] Port, E., Kümmerle, K., Rudin, H., Jenny, C., Zafiropulo, P.: A Network Architecture for the Integra-tion of Circuit and Packet Switching. Proc. ICCO 76, Toronto, 3.-6. August 1976, S. 505-522.
- [78] Pouzin, L.: Virtual Circuits vs. Datagrams: Techni-cal and Political Problems. Proc. Nat. Comp. Conf., AFIPS Press, Montvale, NJ, 1976, S. 483-493.
- [79] Raubold, E., Haenle, J.: A Method of Deadlock-Free Resource Al-location and Flow Control in Packet Networks. Proc. ICCO 76, Toronto, 3.-6. August 1976, S. 483-487.
- [80] Reiser, M., Kobayashi, H.: Horner's Rule for the Evaluation of General Closed Queueing Networks. C.ACM 18(1975)10, S. 592-593.
- [81] Reiser, M.: Interactive Modeling of Computer Systems. IBM Systems Journal 15(1976)4, S. 309-327.

|82| Rudin, H.: On Routing and "Delta Routing": A Taxonomy and Performance Comparison of Techniques for Packet-Switched Networks. IEEE-COM 24(1976), S. 43-59.

|83| Schassberger, R.: Warteschlangen. Springer-Verlag, Wien/New York, 1973.

|84| Shum, A.W., Buzen, J.P.: The EPF Technique: A Method for Obtaining Approximate Solutions to Closed Queueing Networks with General Service Times. Proc. of the 3rd International Symposium on Modelling and Performance Evaluation of Computer Systems, Bonn-Bad Godesberg, 3.-5. Okt. 1977, North-Holland Publ. Comp., Amsterdam/New York/Oxford, 1977, S. 201-220.

|85| Schwager, A.O.: The Hewlett-Packard Distributed System Network. Hewlett-Packard Journal, März 1978, S. 2-5.

|86| Schweitzer, P.J., Lam, S.S.: Buffer Overflow in a Store-and-Forward Network Node. IBM Journ. of Res. and Dev. 20(1976)6, S. 542-550.

|87| Teicholtz, N.A.: Digital Network Architecture, Proc. of the European Conference on Communications Networks, London, 1975, S. 13-24.

|88| Thomas, K.: DCP und TELCON-Bausteine zur Realisierung großer Rechnernetze. Datascope 26 (1978), S. 42-50.

|89| Zimmermann, H., Naffah, N.: On Open Systems Architecture. Proc. ICCS, Kyoto, September 1978, S. 669-674.

ABKÜRZUNGEN

a) Allgemein verwendete Abkürzungen

$P(A)$	Wahrscheinlichkeit für das Auftreten des Ereignisses A
$E[X]$	Erwartungswert der Zufallsvariablen X
\wedge	logisches "und"
$\{x E(x)\}$	Menge der x mit der Eigenschaft E(x)
$A \setminus B$	Differenz der Mengen A und B
\mathbb{N}_0	Menge der ganzen nichtnegativen Zahlen 0,1,2,...
$\binom{x}{x_1, x_2, \dots, x_n}$	Multinomialkoeffizient = $\frac{x!}{x_1! x_2! \dots x_n!}$ ($x = x_1 + x_2 + \dots + x_n$)
A/B/n,s	Kurznotation für ein Wartesystem mit Verteilungsfunktion A der Ankunftsabstände, Verteilungsfunktion B der Bedienungszeiten, n Bedienungseinheiten und s Warteplätzen. Gebräuchliche Abkürzungen für die Verteilungsfunktionen A bzw. B:
	M: Exponentialverteilung
	D: Konstante Verteilung
	E_k : Erlang-Verteilung der Ordnung k
	H_k : Hyperexponentialverteilung der Ordnung k
	G: Allgemeine Verteilung
c	Variationkoeffizient
t	allgemeine Zeitvariable

b) Abkürzungen in Kapitel 2

VK	Vermittlungsnoten
TP	Terminalprozessor
T	Datenstation (Rechner oder Terminal)
VL	Verbindungsleitung
AL	Anschlußleitung
TIP	Terminal Interface Processor
PAD	Packet Assembly/Disassembly facility

c) Abkürzungen in Kapitel 3

n	Anzahl der Paketspeicherplätze im Vermittlungsknoten
ACK	positive Quittung (<u>acknowledgment</u>)
NAK	negative Quittung (<u>negative acknowledgment</u>)
$T_{res}^{(i)}$	resultierende Übertragungszeit des Pakets Nr. i
λ	Paket-Ankunftsrate
T_a	zufälliger Ankunftsabstand der Pakete
L	zufällige Paketlänge
l	realisierte Paketlänge
R	Anzahl der abgehenden Übertragungskanäle
b_1	mittlere Prozessorbelegungszeit
b_{2r}	mittlere Belegungsdauer des Übertragungskanals Nr. r
b_{3r}	mittlere Verzögerungszeit der negativen Quittungen für Übertragungskanal Nr. r
b_{4r}	mittlere Verzögerungszeit der positiven Quittungen für Übertragungskanal Nr. r
p_r	Wahrscheinlichkeit, daß ein Paket über den Übertragungskanal Nr. r übertragen wird
q_r	Wiederholungswahrscheinlichkeit bezüglich Übertragungskanal Nr. r
X_1, x_1	Zufallsvariable bzw. deren Realisierung für die Anzahl gleichzeitiger Anforderungen in Prozessor und Prozessorwarteschlange
X_{2r}, x_{2r}	Zufallsvariable bzw. deren Realisierung für die Anzahl gleichzeitiger Pakete auf Übertragungskanal Nr. r und in dessen Warteschlange
X_{3r}, x_{3r}	Zufallsvariable bzw. deren Realisierung für die Anzahl der gleichzeitigen Anforderungen in Verzögerungsstufe Nr. $(3, r)$
X_{4r}, x_{4r}	Zufallsvariable bzw. deren Realisierung für die Anzahl der gleichzeitigen Anforderungen in Verzögerungsstufe Nr. $(4, r)$
λ^*	zustandsabhängige Paket-Ankunftsrate
\hat{X}, \hat{x}	Zufallsvariable bzw. deren Realisierung für die Gesamtzahl der Pakete im Speicher

\vec{x}	Zustandsvektor
$p_e(\vec{x})$	Zustandswahrscheinlichkeit des Warteschlangennetzes bei exponentialverteilten Übertragungs- und Prozessorbelegungszeiten
$d(\vec{x})$	Ankunftsprozeßspezifischer Faktor in der Produktlösung bei exponentialverteilten Übertragungs- und Prozessorbelegungszeiten
$\left. \begin{matrix} \varepsilon_1 \\ \vdots \\ \varepsilon_{4R} \end{matrix} \right\}$	Stationsspezifische Faktoren in der Produktlösung bei exponentialverteilten Übertragungs- und Prozessorbelegungszeiten
C_e	Normierungskonstante bei exponentialverteilten Übertragungs- und Prozessorbelegungszeiten
$\left. \begin{matrix} e_1 \\ \vdots \\ e_{4R} \end{matrix} \right\}$	Erwartungswerte für die Anzahl von Durchläufen der Anforderungen durch die Stationen Nr. $1, \dots, (4, R)$ des Warteschlangennetzes
A_1	Angebot an Prozessor
A_{2r}	Angebot an Übertragungskanal Nr. r
A_{3r}	Angebot an Verzögerungsstufe Nr. $(3, r)$
A_{4r}	Angebot an Verzögerungsstufe Nr. $(4, r)$
Z	Zustandsraum (Menge aller möglichen Zustände)
B	Speicherüberlauf-Wahrscheinlichkeit
$\Psi(\hat{x})$	Menge aller Zustände, für welche die Gesamtzahl der Pakete im Speicher gleich \hat{x} ist
t_A	mittlerer Abstand zwischen aufeinanderfolgenden Blockierungszuständen
t_B	mittlere Dauer eines Blockierungszustandes
T_A	zufälliger Abstand zwischen aufeinanderfolgenden Blockierungszuständen
T_B	zufällige Dauer eines Blockierungszustandes
T	Beobachtungszeit-Intervall
z_1	mittlere Anzahl der Übergänge der Zufallsvariablen \hat{X} vom Zustand $\hat{X}=n-1$ in den Zustand $\hat{X}=n$ während der Beobachtungszeit T

z_2	mittlere Anzahl der Übergänge der Zufallsvariablen \hat{X} vom Zustand $\hat{X}=n$ in den Zustand $\hat{X}=n-1$ während der Beobachtungszeit T
Y_{SP}	Speicherbelastung
t_{SP}	mittlere Speicherverweilzeit
D	Paket-Durchsatz
Ω_1	mittlere Warteschlangenlänge vor dem Prozessor
$\Phi_1(\eta)$	Menge aller Zustände \vec{x} mit η gleichzeitigen Anforderungen in Prozessor und Prozessorwarteschlange
w_1	mittlere Wartezeit vor dem Prozessor
Y_1	Prozessor-Belastung
Ω_{2r}	mittlere Warteschlangenlänge vor dem Übertragungskanal Nr. r
$\Phi_{2r}(\eta)$	Menge aller Zustände \vec{x} mit η gleichzeitigen Paketen auf Übertragungskanal Nr. r und in dessen Warteschlange
w_{2r}	mittlere Wartezeit vor Übertragungskanal Nr. r
Y_{2r}	Belastung des Übertragungskanals Nr. r
A_o	resultierendes Angebot der Ersatzstation für alle Verzögerungsstationen im Warteschlangennetz
X_o, x_o	Zufallsvariable bzw. deren Realisierung für die gleichzeitige Anzahl der Anforderungen in der Ersatzstation
\vec{x}_o	reduzierter Zustandsvektor des Warteschlangennetzes
Z_o	reduzierter Zustandsraum (Menge aller möglichen reduzierten Zustandsvektoren \vec{x}_o)
u	allgemeine Stationsnummer
U	Gesamtzahl der Stationen im Warteschlangennetz
$\left. \begin{matrix} y_1 \\ \vdots \\ y_U \end{matrix} \right\}$	Ersatz-Zustandsvariable für Faltungsalgorithmus
$f_u(y_u)$	stationsspezifischer Faktor in der Produktlösung
\vec{y}	Vektor der Ersatz-Zustandsvariablen y_1, \dots, y_U
$Z(U, n)$	Menge der Vektoren \vec{y} mit Komponentensumme höchstens gleich n
$D(U, n)$	Menge der Vektoren \vec{y} mit Komponentensumme gleich n

$G(U, i)$	Summe der Produkte aller Faktoren $f_u(y_u)$ mit Summe der y_u gleich i
A	allgemeines Verkehrsangebot
K	Normierungskonstante der Zustandswahrscheinlichkeiten bei M/M/1 bzw. M/M/1,s Wartesystemen
$p_a(\vec{x})$	Zustandswahrscheinlichkeit des Warteschlangennetzes in der Näherungslösung für allgemein verteilte Übertragungs- und Prozessorbelegungszeiten
C_a	Normierungskonstante in der Näherungslösung für allgemein verteilte Übertragungs- und Prozessorbelegungszeiten
$p_1(x_1)$	allgemeine Zustandsverteilung für den Prozessor in der Näherungslösung
$p_{2r}(x_{2r})$	allgemeine Zustandsverteilung für den Übertragungskanal Nr. (2,r) und dessen Warteschlange in der Näherungslösung
$p_{3r}(x_{3r})$	allgemeine Zustandsverteilung für die Verzögerungsstation Nr. (3,r) in der Näherungslösung
$p_{4r}(x_{4r})$	allgemeine Zustandsverteilung für die Verzögerungsstation Nr. (4,r) in der Näherungslösung
t_n	Zeitpunkt, unmittelbar nachdem eine Anforderung ein Wartesystem M/G/1,s verlassen hat
$X(t_n)$	zufällige Anzahl der Anforderungen in einem Wartesystem M/G/1,s, unmittelbar nachdem eine Anforderung die Bedienungseinheit verlassen hat
$p^*(x)$	Wahrscheinlichkeit, daß sich in einem Wartesystem M/G/1,s unmittelbar nach Verlassen einer Anforderung noch x Anforderungen befinden
a_i	Wahrscheinlichkeit, daß während einer zufälligen Bedienungszeit i Anforderungen aus einem Poisson-Prozess eintreffen
$\bar{\lambda}$	Ankunftsrate an das betrachtete, isolierte M/G/1,s Wartesystem
$b(t)$	Dichte der Bedienungszeit-Verteilung des betrachteten, isolierten M/G/1,s Wartesystems

$p(x)$	Wahrscheinlichkeit, daß sich beim Eintreffen einer neuen Anforderung x Anforderungen in einem Wartesystem $M/G/1,s$ befinden
\bar{b}_k	Hilfsgröße für die rekursive Berechnung der Zustandswahrscheinlichkeiten des Wartesystems $M/G/1,s$
b_{pi}	mittlere Verweilzeit in der i -ten Phase bei einer allgemeinen Phasentyp-Verteilung
ϵ_i	Enderate der i -ten Phase bei einer allgemeinen Phasentyp-Verteilung
v	Anzahl der Phasen bei einer allgemeinen Phasentyp-Verteilung
α_i, β_i	Aufteilungswahrscheinlichkeiten einer allgemeinen Phasentyp-Verteilung
$a^*(z)$	Erzeugende Funktion der Wahrscheinlichkeiten a_i
$B^*(s)$	Laplace-Transformierte der Wahrscheinlichkeitsdichte $b(t)$
δ_j	Wahrscheinlichkeit, daß eine Bedienungszeit aus j Phasen besteht
$Q(j,i)$	Hilfsgröße zur Bestimmung der Wahrscheinlichkeit a_i
$p_g(\vec{x})$	Zustandswahrscheinlichkeiten des geschlossenen Warteschlangennetzes für die Bestimmung des Maximaldurchsatzes
Z_g	Zustandsraum (Menge aller möglichen Zustände) des geschlossenen Warteschlangennetzes für die Bestimmung des Maximaldurchsatzes
C_g	Normierungskonstante für die Bestimmung des Maximaldurchsatzes
D_{max}	maximaler Paketdurchsatz
S	Schiefe des Verkehrsangebots für die abgehenden Übertragungskanäle
Δp	Inkrement der Verzweigungswahrscheinlichkeiten bei schiefem Verkehr

d) Abkürzungen in Kapitel 4

d1) gemeinsam für beide Arten der Verkehrsaufteilung

M	Anzahl der Vermittlungsmoduln im Vermittlungsknoten
m	allgemeine Nummer eines Vermittlungsmoduls
n	Gesamtzahl der Paketspeicherplätze im Vermittlungsknoten
R	Gesamtzahl der abgehenden Übertragungskanäle
n_m	Anzahl der Paketspeicherplätze im Modul Nr. m
λ	gesamte Paketankunftsrate des Vermittlungsknotens
λ_m	Paketankunftsrate für Vermittlungsmodul Nr. m
$b_{m;1}$	mittlere Belegungszeit des Prozessors im Vermittlungsmodul Nr. m

d2) zielorientierte Verkehrsaufteilung

$b_{m;2r}$	mittlere Belegungszeit des Übertragungskanals Nr. $(m;2,r)$
$b_{m;3r}$	mittlere Verzögerungszeit der negativen Quittungen für Übertragungskanal Nr. $(m;2,r)$
$b_{m;4r}$	mittlere Verzögerungszeit der positiven Quittungen für Übertragungskanal Nr. $(m;2,r)$
$P_{m;r}$	Wahrscheinlichkeit, daß ein Paket des Vermittlungsmoduls Nr. m über den Übertragungskanal Nr. $(m;2,r)$ übertragen wird
$q_{m;r}$	Wiederholungswahrscheinlichkeit bezüglich Übertragungskanal Nr. $(m;2,r)$
R_m	Anzahl der abgehenden Übertragungskanäle, welche dem Vermittlungsmodul Nr. m zugeordnet sind

d3) herkunftsorientierte Verkehrsaufteilung

b_{2r}	mittlere Belegungszeit des Übertragungskanals Nr. $(2,r)$
b_{3r}	mittlere Verzögerungszeit der negativen Quittungen für Übertragungskanal Nr. $(2,r)$

b_{4r}	mittlere Verzögerungszeit der positiven Quittungen für Übertragungskanal Nr. (2,r)
$p_{m;r}$	Wahrscheinlichkeit, daß ein Paket des Vermittlungsmoduls Nr. m über den Übertragungskanal Nr. (2,r) übertragen wird
q_r	Wiederholungswahrscheinlichkeit der Pakete bezüglich Übertragungskanal Nr. (2,r)
$X_{m;sr}$, $x_{m;sr}$	Zufallsvariable bzw. deren Realisierung für die Anzahl gleichzeitiger Anforderungen der Klasse m in der Station Nr. (s,r)
\hat{X}_m, \hat{x}_m	Zufallsvariable bzw. deren Realisierung für die Gesamtzahl der Pakete im Speicher des Vermittlungsmoduls Nr. m
$\lambda_m^*(\hat{x}_m)$	zustandsabhängige Paketankunftsrate für Vermittlungsmodul Nr. m
\vec{x}'	Zustandsvektor
$d'(\vec{x}')$	Ankunftsprozeßspezifischer Faktor in der Produktlösung bei exponentialverteilten Übertragungs- und Prozessorbelegungszeiten
$\left. \begin{matrix} \xi_{1m} \\ \xi_{21} \\ \vdots \\ \xi_{4R} \end{matrix} \right\}$	Stationsspezifische Faktoren in der Produktlösung bei exponentialverteilten Übertragungs- und Prozessorbelegungszeiten und mehreren Anforderungsklassen
$p'(\vec{x}')$	Zustandswahrscheinlichkeiten des Warteschlangennetzes bei exponentialverteilten Übertragungs- und Prozessorbelegungszeiten
C'	Normierungskonstante bei exponentialverteilten Übertragungs- und Prozessorbelegungszeiten
Z'	Zustandsraum (Menge aller möglichen Zustandsvektoren \vec{x}')
$A_{m;1m}$	Angebot an den Prozessor im Vermittlungsmodul Nr. m
$A_{m;2r}$	Angebot der Klasse m an den Übertragungskanal Nr. (2,r)
$A_{m;3r}$	Angebot der Klasse m an die Verzögerungsstation Nr. (3,r)
$A_{m;4r}$	Angebot der Klasse m an die Verzögerungsstation Nr. (4,r)

B_m	Speicherüberlauf-Wahrscheinlichkeit des Vermittlungsmoduls Nr. m
$\Psi_m(\hat{x}_m)$	Menge aller Zustände \vec{x}' , für welche die Anzahl der Pakete im Speichermodul Nr. m gleich \hat{x}_m ist
$t_{A;m}$	mittlerer Abstand zwischen aufeinanderfolgenden Blockierzuständen des Vermittlungsmoduls Nr. m
$t_{B;m}$	mittlere Dauer eines Blockierzustandes des Speichermoduls Nr. m
$Y_{SP;m}$	Speicherbelastung des Vermittlungsmoduls Nr. m
$t_{SP;m}$	mittlere Speicherverweilzeit der Pakete im Vermittlungsmodul Nr. m
D_m	Paket-Durchsatz des Vermittlungsmoduls Nr. m
$\Omega_{m;1m}$	mittlere Warteschlangenlänge vor dem Prozessor des Vermittlungsmoduls Nr. m
$\Phi_{m;1m}(\xi)$	Menge aller Zustände \vec{x}' mit ξ gleichzeitigen Anforderungen im Prozessor Nr. m und dessen Warteschlange
$w_{m;1m}$	mittlere Wartezeit vor dem Prozessor des Vermittlungsmoduls Nr. m
$Y_{m;1m}$	Belastung des Prozessors im Vermittlungsmodul Nr. m
Ω_{2r}	mittlere Warteschlangenlänge vor dem Übertragungskanal Nr. (2,r)
$\Phi_{2r}(\xi)$	Menge aller Zustände \vec{x}' mit ξ gleichzeitigen Paketen auf dem Übertragungskanal Nr. (2,r) und in dessen Warteschlange
w_{2r}	mittlere Wartezeit vor dem Übertragungskanal Nr. (2,r)
Y_{2r}	Belastung des Übertragungskanals Nr. (2,r)
u	allgemeine Stationsnummer
U	Gesamtzahl der Stationen im Warteschlangennetz
$\left. \begin{matrix} y_{1;u} \\ \vdots \\ y_{M;u} \end{matrix} \right\}$	Ersatz-Zustandsvariablen für Faltungsalgorithmus
y_u	Vektor der Ersatz-Zustandsvariablen $y_{1;u}, \dots, y_{M;u}$
$f_u(y_u)$	stationsspezifischer Faktor in der Produktlösung
\vec{y}	Ersatz-Zustandsvektor
$Z_U(n_1, \dots, \dots, n_M)$	Menge der Zustände \vec{y} , für welche die Vektoren y_u die Eigenschaft besitzen, daß die Summe der Komponenten $y_{m;1}$ bis $y_{m;U}$ höchstens gleich n_m ist

$D_U(i_1, \dots, \dots, i_M)$	Menge der Zustände \vec{y} , für welche die Vektoren y_u die Eigenschaft besitzen, daß die Summe der Komponenten $y_{m;1}$ bis $y_{m;U}$ gleich i_m ist
$G_U(i_1, \dots, \dots, i_M)$	Summe all jener Produkte der Faktoren $f_u(y_u)$, für welche der Zustandsvektor \vec{y} zu $D_U(i_1, \dots, i_M)$ gehört
μ	Nummer des für die Näherungslösung betrachteten Vermittlungsmoduls
B_μ^*	Speicherüberlauf-Wahrscheinlichkeit des Vermittlungsmoduls Nr. μ , falls alle übrigen Moduln unendlich große Paketspeicher besitzen
\vec{x}_μ	Vektor der Anzahlen von Anforderungen der Klasse μ in jeder Station des Warteschlangennetzes
$p_\mu(\vec{x}_\mu)$	Marginalverteilung der Anzahlen von Anforderungen der Klasse μ in jeder Station des Warteschlangennetzes unter der Voraussetzung, daß alle Vermittlungsmoduln $m \neq \mu$ unendlich große Paketspeicher besitzen
C_μ	Normierungskonstante der Marginalverteilung $p_\mu(\vec{x}_\mu)$
Z_μ	Zustandsraum des Warteschlangennetzes bezüglich der Klasse μ (Menge aller möglichen Zustandsvektoren \vec{x}_μ)
$A_{0;2r}$	Summe der Angebote aller Anforderungsklassen außer μ an den Übertragungskanal Nr. (2,r)
b_{2r}^*	fiktive mittlere Belegungszeit des Übertragungskanals Nr. (2,r) in der Näherungslösung
$p_\mu^!(\vec{x}_\mu)$	Näherung für die Marginalverteilung der Anzahlen von Anforderungen der Klasse μ in jeder Station des Warteschlangennetzes unter der Voraussetzung, daß alle übrigen Moduln unendlich große Paketspeicher besitzen (allgemein verteilte Übertragungs- und Prozessorbelegungszeiten)
$C_\mu^!$	Normierungskonstante der Näherung für die Marginalverteilung bei allgemein verteilten Übertragungs- und Prozessorbelegungszeiten
$P_{\mu;1\mu}$	allgemeine Zustandsverteilung für den Prozessor und dessen Warteschlange im Vermittlungsmodul Nr. μ in der Näherungslösung für allgemein verteilte Übertragungs- und Prozessorbelegungszeiten

$P_{\mu;2r}$ allgemeine Zustandsverteilung für den Übertragungskanal Nr. (2,r) und dessen Warteschlange in der Näherungslösung für allgemein verteilte Übertragungs- und Prozessorbelegungszeiten

e) Abkürzungen im Anhang 1

U	Anzahl der Bedienungsstationen im Warteschlangennetz
K	Anzahl der Anforderungsklassen
$q_{k;i,j}$	Wahrscheinlichkeit, daß eine Anforderung der Klasse k nach Verlassen der Station Nr. i zur Station Nr. j gelangt
λ_k	externe Ankunftsrate der Klasse k
ξ_k	Anzahl gleichzeitiger Anforderungen der Klasse k im Warteschlangennetz
$q_{k;i}$	Wahrscheinlichkeit, daß eine Anforderung der Klasse k über Station Nr. i in ein offenes Warteschlangennetz eintritt
\vec{y}	Zustandsvektor des Warteschlangennetzes
y_i	Zustandsvektor für die Station Nr. i
$y_{k;i}$	Anzahl der Anforderungen der Klasse k in der Station Nr. i
$\mu_{k;i}$	Bedienungsrate der Klasse k in der Station Nr. i
$p(\vec{y})$	stationäre Zustandswahrscheinlichkeit des Warteschlangennetzes
C	Normierungskonstante
$d(\vec{y})$	Ankunftsprozeßspezifischer Faktor in der Produktlösung
$g_i(y_i)$	stationsspezifischer Faktor in der Produktlösung
$e_{k;i}$	Erwartungswert für die Anzahl der Durchläufe einer Anforderung der Klasse k durch die Station Nr. i

f) Abkürzungen im Anhang 4

μ	Nummer des betrachteten Vermittlungsmoduls
$X_{0;1m}$	Zufallsvariable bzw. deren Realisierung für die Anzahl der Anforderungen im Prozessor des Vermittlungsmoduls Nr. m und in dessen Warteschlange
$x_{0;1m}$	

$X_{0;sr}$	Zufallsvariable bzw. deren Realisierung für die Gesamtzahl der Anforderungen, welche nicht zur Klasse μ gehören und sich gleichzeitig in der Station Nr. (s,r) befinden
\vec{x}	Zustandsvektor des Warteschlangennetzes mit zwei Anforderungsklassen
$p(\vec{x})$	Zustandswahrscheinlichkeiten des Warteschlangennetzes mit zwei Anforderungsklassen
C	Normierungskonstante für die Zustandswahrscheinlichkeiten des Warteschlangennetzes mit zwei Anforderungsklassen
Z	Zustandsraum für das Warteschlangennetz mit zwei Anforderungsklassen (Menge aller möglichen Zustandsvektoren \vec{x})
$A_{\mu;1\mu}$	Angebot an den Prozessor im Vermittlungsmodul Nr. μ
$A_{0;1m}$	Angebot an den Prozessor im Vermittlungsmodul Nr. m
$A_{\mu;2r}$	Angebot der Klasse μ an den Übertragungskanal Nr. (2,r)
$A_{0;2r}$	resultierendes Angebot aller Klassen $m \neq \mu$ an den Übertragungskanal Nr. (2,r)
$A_{\mu;3r}$	Angebot der Klasse μ an die Verzögerungsstation Nr. (3,r)
$A_{0;3r}$	resultierendes Angebot aller Klassen $m \neq \mu$ an die Verzögerungsstation Nr. (3,r)
$A_{\mu;4r}$	Angebot der Klasse μ an die Verzögerungsstation Nr. (4,r)
$A_{0;4r}$	resultierendes Angebot aller Klassen $m \neq \mu$ an die Verzögerungsstation Nr. (4,r)
\vec{x}_{μ}	Vektor der Anzahlen von Anforderungen der Klasse μ in jeder Station des Warteschlangennetzes
$p_{\mu}(\vec{x}_{\mu})$	Marginalverteilung der Anzahlen von Anforderungen der Klasse μ in jeder Station des Warteschlangennetzes, unter der Voraussetzung, daß alle Vermittlungsmoduln $m \neq \mu$ unendlich große Paketspeicher besitzen
C_{μ}	Normierungskonstante der Verteilung $p_{\mu}(\vec{x}_{\mu})$
Z_{μ}	Zustandsraum des Warteschlangennetzes bezüglich der Klasse μ (Menge aller möglichen Zustandsvektoren \vec{x}_{μ})

g) Abkürzungen im Anhang 5

U	Anzahl der Faktoren
$\left. \begin{matrix} y_1 \\ \vdots \\ y_U \end{matrix} \right\}$	allgemeine ganzzahlige und nichtnegative Variablen
$\left. \begin{matrix} f_1(y_1) \\ \vdots \\ f_U(y_U) \end{matrix} \right\}$	allgemeine Faktoren
\vec{y}	Vektor der Variablen y_1, \dots, y_U
$D(U, i)$	Menge der Vektoren \vec{y} mit Komponentensumme gleich i
$G(U, i)$	Summe der Produkte der Faktoren $f_1(y_1), \dots, f_U(y_U)$, für welche die Summe der Variablen y_1, \dots, y_U gleich i ist
h	Hilfsgröße

1. EINLEITUNG - ÜBERBLICK ÜBER DIE ARBEIT

Während des letzten Jahrzehnts hat die Übertragung und Vermittlung von Daten wachsende Bedeutung erlangt. Da diese Entwicklung aller Voraussicht nach anhalten, bzw. sich sogar noch beschleunigen wird [26,56], unternehmen sowohl die Datenverarbeitungs- und Fernmeldeindustrie als auch die Fernmeldeverwaltungen und -betriebsgesellschaften verstärkte Anstrengungen, um dem Bedürfnis nach Datenkommunikation nachzukommen. Dabei können Aktivitäten in verschiedenen Bereichen beobachtet werden [47,4]:

- 1) Neuerdings bieten alle größeren Rechnerhersteller sogenannte Systemarchitekturen an, mit deren Hilfe sich die Kunden ihre privaten Rechnernetze aufbauen können [55, 67,28,85,87,88].
- 2) Schon seit einigen Jahren entwickeln und betreiben Benutzergemeinschaften für ihre spezifischen Zwecke der Datenkommunikation Gemeinschaftsnetze, wie zum Beispiel Fluggesellschaften [6] oder Forschungszentren [22,30,50].
- 3) Um praktisch jedermann Datenübertragungs- und Datenvermittlungsdienste anbieten zu können, beginnen Fernmeldeverwaltungen und -betriebsgesellschaften in vielen Ländern, öffentliche Datennetze in Betrieb zu nehmen [23,19, 71,64].

Für diejenigen Netze, bei denen der direkte Anschluß von Rechenanlagen eine wichtige Rolle spielt, hat sich eine bestimmte Form der Datenvermittlung als besonders günstig herausgestellt, die sogenannte Teilstreckenvermittlung (store-and-forward switching). Bei dieser Technik werden die Nachrichten in Blöcken stafettenweise durch das Netz transportiert, wobei sie in jedem durchlaufenden Vermittlungsknoten solange zwischengespeichert werden, bis der nächste Übertragungsabschnitt verfügbar geworden ist. Die Nachrichtenblöcke führen dabei ihre Zieladresse im sog. Nachrichtenkopf mit sich. Diese Technik des "Speicherns und Weiterleitens" wird heute sowohl in den privaten Rechnernetzen und Gemeinschaftsnetzen als auch in der speziellen Version der Paketvermittlungstechnik in öffentlichen Datennetzen angewandt.

Es ist offensichtlich, daß beim Entwurf solcher Datennetze eine

Fülle theoretischer und praktischer Probleme zu lösen ist. Diese Arbeit beschäftigt sich mit verkehrstheoretischen Fragestellungen in Teilstreckenvermittlungsnetzen und zwar speziell mit der Bestimmung der Leistungsfähigkeit der Vermittlungsknoten aus verkehrsmäßiger Sicht. Ihr Ziel ist es, wirklichkeitsnahe Verkehrsmodelle zu entwickeln und zu analysieren, um damit Hilfsmittel für die praktische Bemessung von Datennetzen bereitzustellen.

Nach dieser Einleitung wird im 2. Abschnitt der Arbeit ein kurzer Überblick über die Aufgaben der Datenvermittlungstechnik gegeben. Ferner wird auf einige wesentliche Eigenschaften von Paketvermittlungsnetzen eingegangen.

Die Kapitel 3 und 4 sind die Hauptabschnitte der Arbeit. Im 3. Abschnitt wird das Ablaufgeschehen in einem Vermittlungsknoten¹⁾ und seiner Umgebung im Detail beschrieben und daraus ein Verkehrsmodell abgeleitet. Dabei wird in diesem Abschnitt davon ausgegangen, daß der Knoten nur einen Vermittlungsmodul (Paket-speicher und Prozessor) besitzt. Es stellt sich heraus, daß das Ablaufgeschehen auf der Ebene von Paketen als spezielles Warteschlangennetz modelliert werden kann. Das Modell beschreibt in vereinfachter Form den endlich großen Paketspeicher, den Vermittlungsprozessor sowie die abgehenden Anschluß- und Verbindungsleitungen mit ihren Übertragungsprozeduren. Unter der Voraussetzung exponentialverteilter Übertragungs- und Prozessorbelegungszeiten gelingt es, die interessierenden Verkehrsgrößen (Speicherüberlaufwahrscheinlichkeit, Durchsatz, Warte- und Transferzeiten, Belastungen, usw.) exakt zu bestimmen. Für den Fall allgemeinverteilter Übertragungs- und Prozessorbelegungszeiten wird eine Näherungslösung entwickelt und deren gute Genauigkeit mit Hilfe von Simulationen gezeigt. Den Abschluß von Kapitel 3 bilden die Darstellung und Diskussion einiger wichtiger Ergebnisse zur Dimensionierung der modellierten Vermittlungsknoten.

Aus Gründen der besseren Verfügbarkeit, günstigeren Ausbaufähig-

1) Die Bezeichnung Vermittlungsknoten in der Literatur über Daten- und Rechnernetze entspricht der Bezeichnung Vermittlungsstelle in der Fernsprechvermittlungstechnik.

keit, niedrigerer Kosten und nicht zuletzt höherer Verkehrsleistung besteht in neuen Entwicklungen die Tendenz, überzugehen von zentralisierten Systemen mit einem Speicher und einem Steuerungsprozessor zu modular aufgebauten Systemen mit mehreren Prozessoren und Speichermoduln. Abschnitt 4 der Arbeit beschäftigt sich mit der Verkehrsleistung solcher Datenvermittlungssysteme. Hierzu werden zwei Strategien der Verkehrsaufteilung unterschieden: zielorientiert und herkunftsorientiert. Bei zielorientierter Verkehrsaufteilung werden die Pakete aufgrund ihres nächsten Ziels, d.h. gemäß der abgehenden Anschluß- oder Verbindungsleitung, über welche sie den Knoten verlassen werden, von einem bestimmten Vermittlungsmodul bearbeitet. Dagegen findet die Auswahl des Moduls bei herkunftsorientierter Speicherzuweisung aufgrund der Anschluß- oder Zwischenleitung statt, auf welcher die Pakete ankommen.

Für beide Systeme werden die Konsequenzen für ihre Organisation beschrieben, Verkehrsmodelle entwickelt und eine Analyse durchgeführt. Während sich der Fall der zielorientierten Zuweisung auf den Fall mit einem Vermittlungsmodul abbilden läßt, muß für die herkunftsorientierte Zuweisung eine neue Lösung entwickelt werden. Zwar gelingt es, für exponentialverteilte Übertragungs- und Prozessorbelegungszeiten die Zustandswahrscheinlichkeiten explizit zu bestimmen, jedoch gestaltet sich die numerische Auswertung für Systeme mit einer größeren Zahl von Moduln äußerst problematisch. Aus diesem Grund wird aus der exakten Lösung eine Approximation abgeleitet, mit deren Hilfe Systeme mit beliebiger Zahl von Moduln zuverlässig analysiert werden können. Auf der Basis dieser Näherung wird danach außerdem für allgemeinverteilte Übertragungs- und Prozessorbelegungszeiten eine näherungsweise Analyse durchgeführt und mit Hilfe der Simulation überprüft. Besonderer Wert wird auch im 4. Kapitel auf die Darstellung und Diskussion von numerischen Ergebnissen zur praktischen Dimensionierung gelegt.

Der abschließende 5. Abschnitt gibt eine Zusammenfassung der Arbeit.

Im Anhang befinden sich eine kurze Darstellung der bis heute bekannten Lösungen für Warteschlangennetze sowie mehrere mathematische Ableitungen und Algorithmen.

2. AUFBAU- UND BETRIEBSPRINZIPIEN VON DATEN- UND RECHNERNETZEN

In diesem Kapitel wird ein kurzer Abriß über diejenigen Aspekte des Aufbaus und Betriebs von Daten- und Rechnernetzen gegeben, welche für die Untersuchungen dieser Arbeit von Bedeutung sind.

2.1 Vermittlungsverfahren

Für die Vermittlung von Daten lassen sich verschiedene Verfahren anwenden [4,24,66]. Die heute realisierten Prinzipien sind im Bild 2.1 aufgezählt und werden im folgenden kurz charakterisiert.

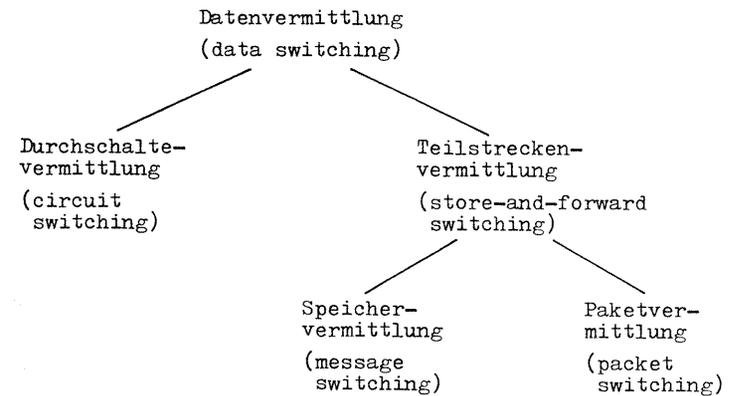


Bild 2.1: Vermittlungsverfahren für Datennetze

In Netzen mit Durchschaltevermittlung werden drei verschiedene Phasen einer Verbindung unterschieden: Verbindungsaufbau, Datenübertragung, Verbindungsabbau. In der Verbindungsaufbauphase wird die rufende mit der gerufenen Datenstation durch einen physikalischen Übertragungsweg verbunden. Dieser Weg steht während der nachfolgenden Übertragungsphase ausschließlich diesen Datenstationen zur Verfügung. Während der Übertragungsphase verhält sich

das Netz in der Regel vollkommen passiv, d.h. die zwischen den Stationen fließenden Daten werden in keiner Weise vom Netz beeinflusst (keine netzseitige Codesicherung, Codewandlung, Geschwindigkeitswandlung). Aufgrund eines Auslösesignals wird die Durchschaltung schließlich in der Verbindungsabbauphase aufgehoben.

In Netzen mit Teilstreckenvermittlung werden keine exklusiven Übertragungskanäle zwischen den Datenstationen geschaltet, sondern die Nachrichten werden in den Vermittlungsknoten zwischengespeichert und dann weitergeleitet, wenn der nächste Übertragungsabschnitt zur Verfügung steht. Die Auswahl des weiterführenden Weges, die sogenannte Wegelenkung (routing), erfolgt aufgrund der Zielangabe, die jede Nachricht in ihrem Nachrichtenkopf mit sich führt (vgl. Abschnitt 2.3.3.). Abhängig davon, ob die Nachrichten der Datenteilnehmer als ganzes durch das Netz transportiert werden oder in Datenpakete vorgeschriebener maximaler Länge aufgeteilt werden, bezeichnet man das Verfahren als Speichervermittlung (message switching) oder Paketvermittlung (packet switching).

Ein charakteristisches Merkmal von Speichervermittlungsnetzen (message switching) ist die Verwendung von relativ langsamen Massenspeichern (Magnetplatten- oder Magnettrommelspeicher) als Zwischenpuffer für die Nachrichten. Diese Technik ist deshalb für die Bedürfnisse eines universellen Datennetzes für Dialogverkehr weniger geeignet.

Dagegen werden Paketvermittlungsnetze so ausgelegt, daß mit ihnen von Teilnehmer zu Teilnehmer Durchlaufzeiten erzielt werden können, die deutlich unter einer Sekunde liegen und somit für einen Dialogbetrieb Terminal - Rechner in Frage kommen. Dies wird, neben weiteren Maßnahmen, durch den Einsatz schneller Speichermedien, wie Halbleiterspeicher, ermöglicht. Auf die Paketvermittlungstechnik wird im Abschnitt 2.3 ausführlicher eingegangen.

Neben den beschriebenen "reinen" Formen von Vermittlungsverfahren sind in jüngster Zeit auch Konzepte für eine Kombination von Durchschalte- und Paketvermittlung in einem Netz entwickelt worden, bei denen angestrebt wird, die Vorteile beider Verfahren zu vereinigen [46,77].

2.2 Protokolle

Die Hauptaufgabe eines Datennetzes ist der zuverlässige und möglichst schnelle Transport von Nachrichten zwischen den Datenend-einrichtungen. Da im Netz durch schwankende Verkehrsbelastungen zufällige Zeitverzögerungen entstehen und außerdem durch Übertragungsfehler die übermittelten Nutz- und Steuerdaten gestört werden können, sind in Daten- und Rechnernetzen eine Vielzahl von Funktionen notwendig, um einen zuverlässigen Informationsaustausch zu gewährleisten. Die Regeln, denen diese Funktionen gehorchen, werden als Protokolle bezeichnet.

Um die vielfältigen Aufgaben der Benutzer, die unterschiedlichen Eigenschaften der Übertragungswege, sowie die verschiedenen Realisierungsformen von Terminals und Rechnern beherrschen zu können, wird die Menge der Regeln dadurch strukturiert, daß sogenannte Protokollebenen (protocol layers) definiert werden. Eine mögliche Architektur von Rechnernetzprotokollen zeigt Bild 2.2,

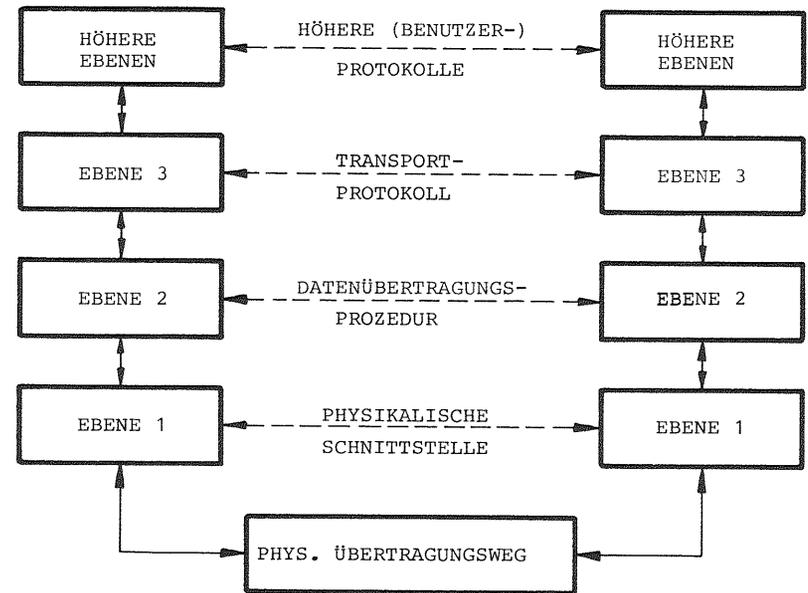


Bild 2.2: Hierarchie der Protokollebenen

in welchem die für diese Arbeit interessierenden Ebenen skizziert sind |4,28,25 |.

Die niedrigste, direkt auf den physikalischen Übertragungsweg bezogene Ebene enthält die notwendige Definition der physikalischen Schnittstelle, wie z.B. die Anzahl und Art der Leitungen, die elektrischen Eigenschaften (Pegelfestlegungen, Art der verwendeten Signale) usw. Eine solche Schnittstellendefinition ist beispielsweise enthalten in der CCITT-Empfehlung X.21 |15|.

Diese physikalische Schnittstelle wird von der zweiten Protokollebene, der Datenübertragungsprozedur benützt. Sie definiert die Regeln, mit deren Hilfe über einen möglicherweise gestörten Übertragungsweg ein geordneter, gegen unerkannte Übertragungsfehler gesicherter Austausch von Datenblöcken organisiert wird. Hierzu wird die zu übertragende Information blockweise mit einem Codesicherungsteil und einem Steuerungsteil versehen. Beginn und Ende eines Blockes werden in der Regel durch eine spezielle Bitkombination (flag) gekennzeichnet. Die Hauptaufgabe der Übertragungsprozedur besteht darin, mit Hilfe der Prüfbits eventuelle Übertragungsfehler beim Empfänger zu erkennen und eine wiederholte Übertragung der gestörten Blöcke zu veranlassen. Bekannte Beispiele für Übertragungsprozeduren sind HDLC (High Level Data Link Control) |41 - 44 |, SDLC (Synchronous Data Link Control) |39|, ADCCP (Advanced Data Communications Control Procedure) |1|, DDCCMP (Digital Data Communications Message Protocol) |87|. Auf spezielle Realisierungsmöglichkeiten von Übertragungsprozeduren wird im Abschnitt 3.2.2 eingegangen.

Auf der zweiten Protokollebene baut die nächsthöhere, die sogenannte Transportprotokoll-Ebene auf. Diese Ebene enthält alle Vereinbarungen und Funktionen, welche notwendig sind, um Informationen in einem Netz aus mehreren Knoten und Übertragungsstrecken zuverlässig austauschen zu können. Dazu gehören u.a. alle Vereinbarungen über das Format der Nachrichten und die Bedeutung der Signalisierungsinformationen, die Regeln der Datenflußsteuerung und der Wegelenkung, sowie gegebenenfalls die Prozeduren für den Auf- und Abbau virtueller Verbindungen (s. Abschnitt 2.3.2).

Die höheren Protokollebenen bedienen sich des Transportprotokolls zur Kommunikation zwischen Prozessen, welche innerhalb der Datenstationen ablaufen. Sie enthalten die Steuerungsfunktionen für die stationsbezogene Datenbehandlung, wie Unterstützung eines virtuellen Terminals, Textaufbereitung, Datenformatierung usw. |28,89|. Diese höheren Protokolle sind i.a. so komplex, daß sie sinnvollerweise wiederum in verschiedene Ebenen unterteilt werden. Da sie mit der Übertragungs- und Vermittlungsfunktion eines Netzes nur indirekt zusammenhängen, werden die höheren Protokolle im folgenden nicht betrachtet.

Dagegen werden die Funktionen der unteren Protokollebenen noch im einzelnen diskutiert und bezüglich ihrer verkehrsmäßigen Eigenschaften analysiert werden.

2.3 Paketvermittlungstechnik

In diesem Abschnitt wird die Technik der Paketvermittlung etwas detaillierter beschrieben. Die meisten der angeführten Begriffe besitzen auch außerhalb der eigentlichen Paketvermittlungsnetze ihre Bedeutung, beispielsweise in privaten Rechnernetzen.

2.3.1 Struktur von Paketvermittlungsnetzen

Paketvermittlungsnetze bestehen im wesentlichen aus den Datenendeinrichtungen (Terminals oder Anwendungsrechner), den Vermittlungsknoten und den Übertragungswegen zwischen diesen Einheiten (vgl. Bild 2.3) |4,24,38 |.

Die Vermittlungsknoten bilden zusammen mit den sie verbindenden Übertragungswegen (Verbindungsleitungen) das "Vermittlungsnetz". Der übrige Teil des Netzes wird als "Anschlußnetz" bezeichnet. Zu ihm zählen die Datenstationen, die Anschlußleitungen, sowie unter Umständen sogenannte Terminal-Prozessoren. Diese speziellen Einrichtungen, welche oft auch als TIP (Terminal Interface Processor) oder PAD (Packet Assembly/Disassembly facility) bezeichnet werden, ermöglichen es, auch solche Terminals an ein Paket-

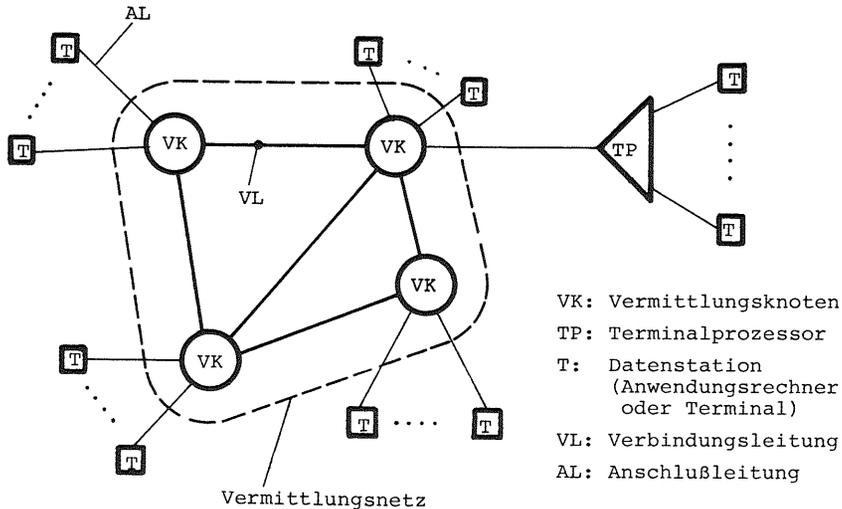


Bild 2.3: Struktur eines Paketvermittlungsnetzes

vermittlungnetz anzuschließen, die nicht selbständig Daten in Paketform senden und empfangen können [72,14].

2.3.2 Virtuelle Verbindungen und Datagrams

Kennzeichnend für die Paketvermittlungstechnik ist, daß die Übertragungswege im Netz nur während der reinen Übertragung der Pakete belegt werden. Nach Beendigung der Paketübertragung steht der betreffende Kanal für die Übertragung von Paketen anderer Teilnehmer zur Verfügung. Dieses Verfahren ist für alle Anwendungen günstig, bei denen stoßweise Daten übertragen werden, wie z.B. bei der interaktiven Benutzung einer Rechenanlage von einem entfernten Terminal aus. Der bei diesen Anwendungen entstehende Verkehr besteht typischerweise aus kürzeren Übertragungsphasen,

die sich mit relativ langen "Denkpausen" abwechseln [58]. An diesem Beispiel der interaktiven Rechnerbenutzung wird zugleich ein weiteres charakteristisches Merkmal von vielen Verkehrsbeziehungen in einem Datennetz deutlich, daß nämlich häufig zwei Datenteilnehmer nicht nur eine einzige Nachricht, sondern oft eine ganze Folge von Nachrichten (z.B. Anfragen und Antworten) austauschen. In einem solchen Fall kann es auch angebracht sein, innerhalb eines Paketvermittlungsnetzes logische oder "virtuelle" Verbindungen zwischen den beteiligten Dateneinrichtungen einzurichten. Unter einer virtuellen Verbindung versteht man ein Dienstmerkmal eines Paketvermittlungsnetzes, das den Datenteilnehmern bis zu einem gewissen Grad die Eigenschaften einer real durchgeschalteten Verbindung bietet. Dazu gehört insbesondere, daß dem Teilnehmer garantiert wird, daß alle von ihm gesendeten Datenpakete fehlergesichert und in der Sendereihenfolge zum Zielteilnehmer gelangen. Die letztgenannte Eigenschaft läßt sich am einfachsten dadurch erreichen, daß man alle Pakete einer Verbindung in strenger Reihenfolge auf einem festen Weg durch das Netz transportiert. Dies ist die Methode, für die ursprünglich die Bezeichnung Verfahren der virtuellen Verbindungen benützt wurde. Neuerdings wird der Begriff bevorzugt verwendet, um ein Dienstmerkmal eines Paketvermittlungsnetzes in dem oben beschriebenen Sinn zu charakterisieren.

In der CCITT-Empfehlung X.25 [16] ist die Schnittstelle zwischen Dateneinrichtung und Datenübertragungseinrichtung für öffentliche Paketvermittlungsnetze beschrieben, welche virtuelle Verbindungen anbieten. Diese Empfehlung ist in drei Ebenen gegliedert:

1. Physikalische Schnittstelle
2. Datenübertragungsprozedur
3. Paketübermittlung auf der Basis von virtuellen Verbindungen.

(Die Schnittstellendefinition reicht also bis in die dritte Protokollebene in Bild 2.2 hinein.)

Es werden drei Phasen einer Verbindung unterschieden: Verbindungsaufbau, Datenübertragung, Verbindungsabbau. Die Signalisierung für Verbindungsauf- und -abbau geschieht mit Hilfe speziell gekennzeichnete Pakete. Ein Vorteil der X.25 Schnittstelle

beim Anschluß eines Anwendungsrechners ist, daß über dessen eine Anschlußleitung gleichzeitig mehrere virtuelle Verbindungen unterhalten werden können; die Unterscheidung zwischen diesen Verbindungen erfolgt mit Hilfe "logischer Kanalnummern" im Steuerungsteil der Pakete.

Eine Alternative zum Verfahren der virtuellen Verbindungen stellt die Datagram-Betriebsweise dar. In einem reinen Datagram-Netz werden die von einer Datenendeinrichtung in das Netz abgegebenen Pakete aufgrund der in jedem Paketkopf enthaltenen Adresse vom Startknoten aus unabhängig voneinander, eventuell auf verschiedenen Wegen, durch das Netz zum Zielknoten transportiert. Ein Eintreffen der Pakete beim Empfänger in einer anderen als der Sendereihenfolge ist daher prinzipiell möglich; außerdem ist ein Verlust einzelner Pakete im Falle von Netzüberlastungen nicht ausgeschlossen.

Als Vorteil der Datagram-Betriebsweise gegenüber derjenigen mit virtuellen Verbindungen gelten eine einfachere Teilnehmer-Schnittstelle, einfachere Prozeduren innerhalb des Netzes, kürzere Transitzeiten für Einzelpakete und ein einfacherer Umstieg zwischen verschiedenen Netzen [78,65]. Eine Möglichkeit, beide Verfahren vorteilhaft zu kombinieren, besteht darin, zwischen den jeweiligen lokalen Vermittlungsknoten der Datenteilnehmer virtuelle Verbindungen zu realisieren, innerhalb des Netzes die Pakete aber nach dem Datagram-Verfahren zu transportieren [19].

2.3.3 Wegelenkung und Datenfluß-Steuerung

Der Weg, den die Pakete im Netz zurücklegen, wird in der folgenden Weise festgelegt: Jedesmal wenn ein Paket in einem Vermittlungsknoten zwischengespeichert ist, wird entweder aufgrund der Zielangabe im Paketkopf oder aufgrund der Zugehörigkeit des Pakets zu einer bestimmten virtuellen Verbindung der nächste Übertragungsabschnitt für das Weiterleiten des Pakets bestimmt.

Die Auswahl des weiterführenden Pfades kann nach verschiedenen Strategien erfolgen: Die Wegelenkung kann von einer zentralen

Stelle des Netzes aus erfolgen oder sie kann lokal in jedem einzelnen Vermittlungsknoten vorgenommen werden; es kann eine feste oder verkehrsabhängige - alternative oder adaptive - Wegelenkung realisiert werden. Die Wegelenkung kann gemeinsam für alle Pakete einer Verbindung erfolgen oder sie kann im Datagram-Betrieb auch individuell für jedes Paket vorgenommen werden [82,29,31,68].

Ziel der Wegelenkung ist es, möglichst kurze Transferzeiten der Pakete durch das Netz zu erzielen, sowie einen guten Verkehrsausgleich im Netz zu erreichen, um möglichst weitgehend Stausituationen zu verhindern.

Allerdings ist auch die beste Wegelenkungsstrategie nicht in der Lage, diejenigen Überlastsituationen zu verhindern, welche durch ein momentanes starkes Verkehrsüberangebot verursacht werden. Gefährlich können insbesondere Situationen werden, in denen eine Datenendeinrichtung Pakete in sehr viel schnellerer Folge absetzt, als sie die Ziel-Datenendeinrichtung abnehmen kann. Ohne geeignete Steuerungsmaßnahmen könnte in diesem Fall die sendende Datenendeinrichtung die Netzabschnitte im Verbindungsweg verstopfen. Es muß deshalb Mechanismen in einem Paketvermittlungnetz geben, die sowohl den Verkehrszustrom zum Netz als auch den Verkehrsfluß innerhalb des Netzes regulieren können. Diese Mechanismen werden als Datenfluß-Steuerung (flow-control) bezeichnet [24,53,60].

Ein häufig angewandtes Verfahren zur Datenfluß-Steuerung ist die "Window"-Technik, bei der eine Verkehrsquelle (Datenendeinrichtung oder Vermittlungsknoten) in ununterbrochener Folge nur eine gewisse Anzahl (die sog. "Fensterbreite") von Paketen absenden darf und danach auf eine ausdrückliche Genehmigung warten muß, um weitersenden zu dürfen. Das Netz hat ferner die Möglichkeit, durch Änderung der Fensterbreite den Verkehrsfluß zu regulieren. Diese Steuerung kann besonders gezielt durchgeführt werden, wenn innerhalb des Netzes einzelne Verbindungen, die für Stausituationen verantwortlich sind, unterschieden werden können [75].

3. VERKEHR IN VERMITTLUNGSKNOTEN MIT EINEM VERMITTLUNGSMODUL

In diesem Kapitel wird der Verkehrsfluß in einem Vermittlungsknoten und seiner Umgebung beschrieben. Daraus wird ein Verkehrsmodell entwickelt und analysiert. Dabei beschränken wir uns hier auf solche Vermittlungsknoten, die nur einen Prozessor und einen gemeinsamen Paketspeicher besitzen, die zusammen als Vermittlungsmodul bezeichnet werden. Systeme mit mehreren Vermittlungsmodulen werden im Kapitel 4 behandelt.

Der Einfachheit halber werden im folgenden die Begriffe aus der Paketvermittlungstechnik, wie "Paket", "Paketspeicher", usw. verwendet. Die Untersuchungen gelten jedoch auch für gleichartige Netze, bei denen hierfür andere Bezeichnungen üblich sind, wie z.B. private Rechnernetze.

3.1 Aufbau und Funktion der Vermittlungsknoten

Aufgrund seiner Aufgabe, die im einzelnen weiter unten beschrieben wird, besitzt ein Vermittlungsknoten einen Aufbau, der sehr ähnlich dem eines üblichen Rechners ist (vgl. Bild 3.1). Die Zentraleinheit aus Prozessor und Speicher (der Vermittlungsmodul) wird ergänzt durch Interface-Baugruppen (Leitungsinterfaces), welche zum Anschluß der Übertragungstechnischen Einrichtungen an das zentrale Rechnersystem dienen [24]. Die Funktion dieser Interfaces kann sehr unterschiedlich sein; sie kann von der reinen Anpassung und Pufferung der elektrischen Signale bis hin zur Übernahme solcher Verarbeitungsaufgaben reichen, die sich dezentral erledigen lassen. Zum Beispiel lassen sich diese Einrichtungen so ausbauen, daß sie die Aufgaben der Codesicherung und Übertragungssteuerung (d.h. Teile der Übertragungsprozedur) weitestgehend übernehmen und dadurch den zentralen Prozessor entlasten. Die Leitungsinterfaces unterscheiden sich ferner danach, ob am anderen Ende der Leitung ein Vermittlungsknoten, ein Paketterminal oder ein synchrones bzw. asynchrones Terminal ohne Paketfunktionen angeschlossen ist.

Im Bild 3.1 ist angenommen, daß die Baugruppen des Vermittlungsknotens, wie dies allgemein in heutigen Systemen der Fall ist, über einen Bus verbunden sind. Dem Bus kommt eine zentrale Rolle zu, da seine Geschwindigkeit mitentscheidend für die Leistungsfähigkeit des Vermittlungssystems ist. Bei den heutigen Übertragungsgeschwindigkeiten in Datennetzen stellen jedoch die Busse im allgemeinen keine Probleme bezüglich eventueller Verkehrsengpässe dar [46].

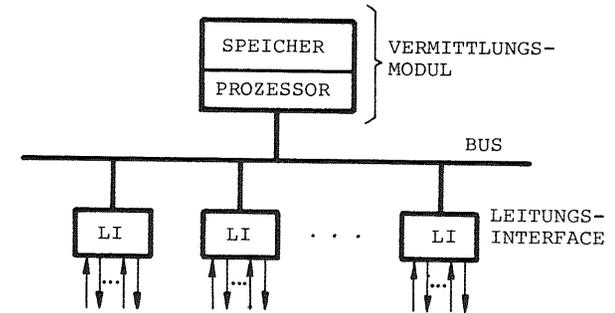


Bild 3.1: Prinzipieller Aufbau eines Vermittlungsknotens mit einem Vermittlungsmodul

Die Hauptaufgaben des Vermittlungsknotens sind:

- Einspeichern ankommender Pakete
- Realisierung der Übertragungsprozeduren, d.h. im einzelnen:
 - Erkennen von Übertragungsfehlern
 - Quittierung von Paketen
 - wiederholte Übertragung gestörter Pakete
 - Datenfluß-Steuerung
- Auswertung der Zielinformation im Paketkopf zur Wegelenkung
- Übertragung der Pakete
- Löschen quittierter Pakete im Speicher
- gegebenenfalls Auf- und Abbau von virtuellen Verbindungen
- gegebenenfalls Terminalunterstützung (Paketierungsfunktionen, Umcodierung, usw.)
- Gebührenerfassung, Verkehrsmessungen, laufende Überwachung und Prüfung aller Einrichtungen .

3.2 Verkehrsmodell

Ein Verkehrsmodell dient dazu, alle für die Untersuchung eines Systems wesentlichen Bedienungs- und Transportvorgänge in realitätsnaher Weise nachzubilden. Ob ein Vorgang als wesentlich anzusehen ist, folgt aus der jeweiligen Fragestellung der Analyse, genauer, aus dem angestrebten Detaillierungsgrad der Untersuchung.

Die folgende Verkehrsuntersuchung beschreibt das Ablaufgeschehen auf der Ebene von Paketen; der Detaillierungsgrad des Modells läßt sich mit demjenigen der meisten Fernsprechverkehrsuntersuchungen vergleichen, welche auf der Ebene von Rufen, nicht auf derjenigen von Steuerungsaufrufen, Wählziffern, usw. angesiedelt sind. Das Modell soll das Ablaufgeschehen in einem Vermittlungsknoten so beschreiben, daß damit Aussagen bezüglich der richtigen Bemessung des Paketspeichers, der Steuerung und der Übertragungskanäle in Abhängigkeit von den Eigenschaften des angebotenen Verkehrs gemacht werden können.

3.2.1 Modellkomponenten

Bild 3.2 zeigt das prinzipielle Modell des Vermittlungsknotens mit einem Vermittlungsmodul. Es besitzt vier verschiedenen Arten von Komponenten:

- 1) Paketspeicher: Er besteht aus n Plätzen, in denen jeweils genau ein Paket maximaler Länge zwischengespeichert werden kann. Speicherplatzverschnitt durch kürzere Pakete wird nicht anderweitig ausgenutzt.
- 2) Prozessor: Er wird mit einer Bedienungseinheit nachgebildet, die von jedem Paket einmal für seine Wegelenkung belegt wird.
- 3) Übertragungskanäle sind entweder Anschlußleitungen zu den lokalen Dateneneinrichtungen oder Verbindungsleitungen zu anderen Vermittlungsknoten.
- 4) Verzögerungsstufen für positive bzw. negative Quittungen: Sie bilden die Zeitspanne nach zwischen dem Absenden eines Pakets und dem Empfang einer positiven bzw. negativen Quittung.

Wie lange und in welcher Reihenfolge diese Einzelkomponenten des Modells belegt werden, wird im folgenden beschrieben.

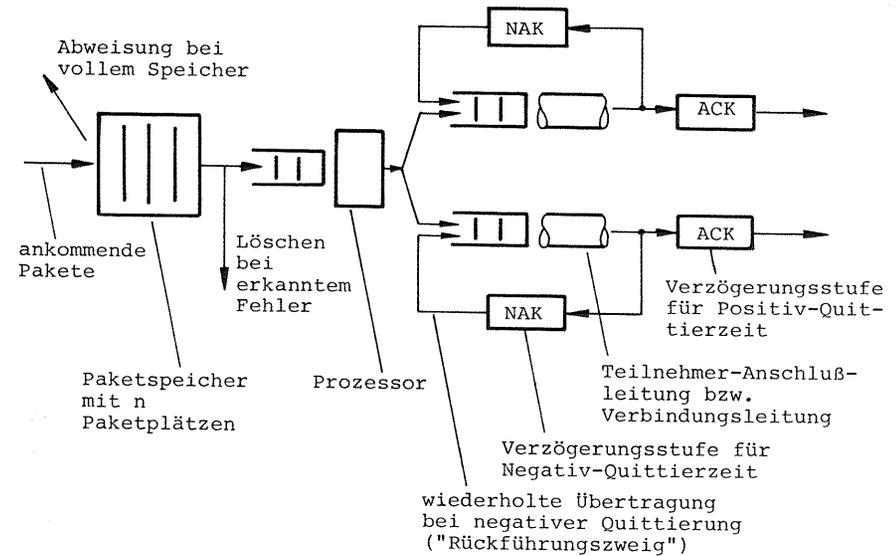


Bild 3.2: Prinzipielles Modell für den Verkehrsfluß in einem Vermittlungsknoten mit einem Vermittlungsmodul

3.2.2 Verkehrsablauf

Der durch das Modell wiedergegebene Verkehrsablauf läßt sich folgendermaßen beschreiben.

Einspeichern

Alle an dem betrachteten Vermittlungsknoten ankommenden Pakete werden zu einem Ankunftsstrom zusammengefaßt. Ein ankommendes Paket, das nicht alle Speicherplätze belegt vorfindet, belegt einen Speicherplatz zunächst für die Dauer seiner Einspeicherung.

Dieser Teil der Speicherbelegungszeit kann maximal gleich der gesamten Übertragungszeit des Pakets sein; er kann aber auch praktisch verschwinden, falls die Speicherplatzzuweisung erst dann erfolgt, wenn das letzte Bit eines Pakets eingetroffen ist. Das letztere gilt insbesondere, wenn die Pakete während ihres Eintreffens zwischengepuffert werden.

Wenn im Modell ein ankommendes Paket auf einen voll belegten Speicher trifft, wird es abgewiesen und verschwindet aus dem System. Dies sieht zunächst wie eine zu starke Vereinfachung des Modells aus, denn in Wirklichkeit werden solche Pakete nach einer gewissen Zeit von neuem übertragen. Genau wie in der Fernsprechverkehrstheorie kann man den Effekt der Wiederholungen im Besetztfall jedoch auch hier gut dadurch einbeziehen, daß man von den durchgesetzten Paketen als ursprünglich vorhandenem Verkehrsbedürfnis ausgeht. Beim Fernsprechen entspricht dies einem Ausgehen von der Belastung anstelle des Angebots [61,62].

Während des Einlaufens eines Pakets wird die Codeprüfung vorgenommen. Falls ein Fehler erkannt wird, wird das eingespeicherte Paket wieder gelöscht.

Prozessor-Bearbeitung

Im fehlerfreien Fall dagegen wird eine Anforderung in die Prozessorwarteschlange eingetragen. Der Prozessor bearbeitet die wartenden Pakete in der Reihenfolge ihres Eintreffens, indem er die Steuerungsinformation jedes Pakets interpretiert und daraus den abgehenden Kanal bestimmt, über den das Paket übertragen werden soll. Als Ergebnis dieses Wegelenkungsvorgangs wird ein Paket in die Warteschlange vor dem betreffenden Kanal eingetragen.

Übertragung

Wenn ein Kanal frei wird, meldet sich sein Leitungsinterface beim Prozessor, der daraufhin das am längsten in der Warteschlange stehende Paket zur Übertragung bestimmt. Nach erfolgter Übertragung kann das Paket noch nicht im Speicher gelöscht werden, da jetzt die Übertragungsprozedur wirksam wird, die dafür sorgt, daß ein eventuell bei der Übertragung entstehender Fehler durch Wiederholung beseitigt wird.

Übertragungsprozedur

Die heute angewandten Übertragungsprozeduren lassen sich in zwei Kategorien einteilen:

- Send-and-Wait Automatic Repeat Request (ARQ),
- Continuous Automatic Repeat Request [34,18,49].

Beim Send-and-Wait ARQ Verfahren, das heute bei einfacheren Terminals benutzt wird, wartet der Sender prinzipiell nach einem übertragenen Paket ¹⁾ so lange, bis er eine positive Quittung (ACK) des Empfängers erhalten hat. Im Bedarfsfall wird ein gestörtes Paket nach Ablauf einer Karenzzeit (time-out) wiederholt (vgl. Bild 3.3).

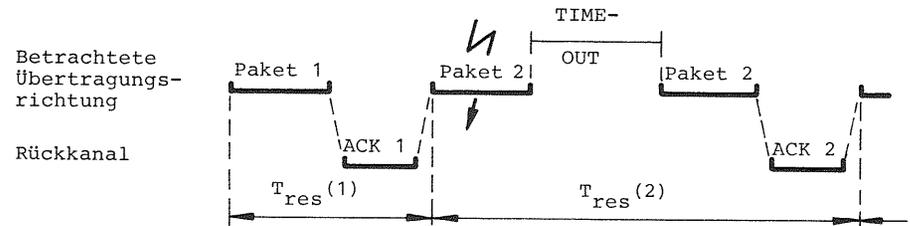


Bild 3.3: Send-and-Wait Übertragungsprozedur

In unserem Modell beschreiben wir die Wirkung einer Send-and-Wait Prozedur dadurch, daß wir den betreffenden Übertragungskanal für die gesamte Zeitdauer $T_{res}(i)$ vom Absenden des Pakets Nr. i bis zu dessen positiver Quittierung belegen. Diese resultierende Übermittlungszeit läßt sich bei bekannter Blockfehlerwahrscheinlichkeit des Kanals bestimmen [27]. Die Verzögerungszeiten im Bild 3.2 werden hier gleich null gesetzt.

Das Continuous ARQ Verfahren ist aufwendiger und wird heute bei neueren Terminals und insbesondere zwischen den Vermittlungs-

¹⁾ Der Einfachheit halber wird auch in diesem Zusammenhang die Bezeichnung "Paket" verwendet, obwohl für die Zwecke der Übertragungsprozedur das reine Paket noch mit einem weiteren Steuerungs- und einem Codesicherungsteil versehen wird.

knoten, sowie zwischen den Vermittlungsknoten und den Anwendungsrechnern eingesetzt. Bei dieser Methode werden die Pakete kontinuierlich hintereinander übertragen, ohne daß die Quittierung jedes Pakets abgewartet wird.

Im Fehlerfall werden zwei verschiedene Wiederholungsstrategien angewandt, die im Bild 3.4 dargestellt sind: Entweder wird nur das gestörte Paket selektiv wiederholt ("Selective Reject" bei HDLC) oder das gestörte Paket und alle seine in der Zwischenzeit gesendeten Nachfolger werden wiederholt übertragen ("Reject" bei HDLC). Die Wiederholungsaufforderung erfolgt mit Hilfe einer speziellen negativen Quittung (NAK).

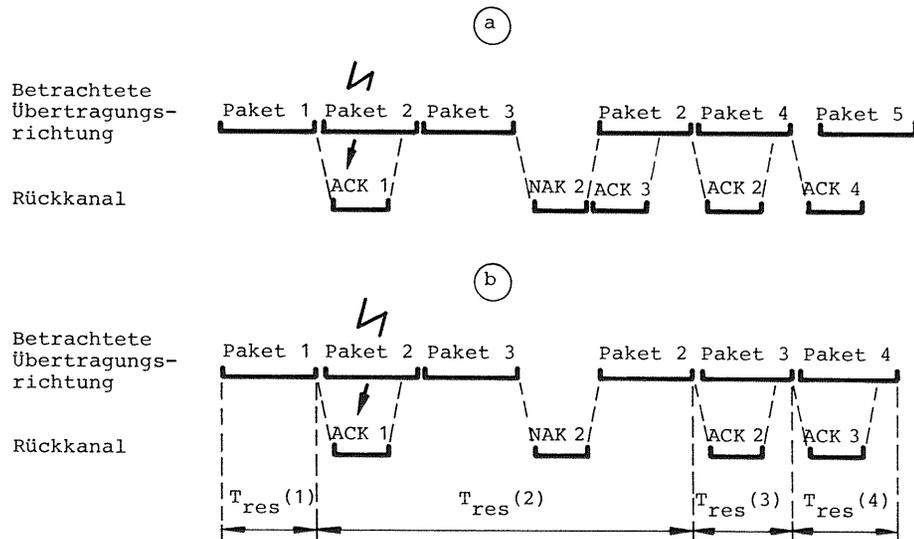


Bild 3.4: Continuous-Übertragungsprozedur
a) mit selektiver Wiederholung
b) mit nichtselektiver Wiederholung

Die Wirkungen dieser Übertragungsprozeduren werden in unserem Modell nach Bild 3.2 folgendermaßen beschrieben:

Bei selektiver Wiederholung wird im Fehlerfall nach einer gewissen "Negativ-Quittierzeit" eine Wiederholung des Pakets veranlaßt

(Rückführungszeit im Bild 3.2). Bei fehlerfreier Übertragung kommt nach einer gewissen Zeit die positive Quittung als Aufforderung zum Löschen des betreffenden Pakets im Speicher an. Die Länge dieser Quittierzeiten kann aus den Eigenschaften der Übertragungsstrecke, der Prozeduren und der Vermittlungsknoten bestimmt werden [12]. Im vorliegenden Modell werden diese Zeiten als veränderliche Parameter berücksichtigt. Die Belegungszeiten des Kanals im Modell entspricht hier der reinen Übertragungszeit.

Im Fall der nichtselektiven Wiederholung wird der Kanal im Modell je Paket für eine resultierende Übermittlungszeit $T_{res}(i)$ belegt, wie im Bild 3.4b angedeutet. Diese besteht aus der reinen Übertragungszeit des Pakets Nr. i , sowie eventuell derjenigen Zeitdauer, welche im Fehlerfall bis zur wiederholten und erfolgreichen Übertragung des Pakets Nr. i verstreicht. Im Modell wird außerdem die Zeit bis zur Quittierung eines fehlerfrei empfangenen Pakets durch eine Verzögerungszeit nachgebildet. Beide Zeiten können bei bekannter Blockfehlerwahrscheinlichkeit des Kanals berechnet oder durch eine gesonderte Simulation bestimmt werden [11,12]. Die Rückführungen und Negativ-Quittierzeiten im Bild 3.2 entfallen hier.

Wie aus dieser Beschreibung hervorgeht, besitzt das im Bild 3.2 dargestellte Modell eine Struktur, welche die Nachbildung aller beschriebenen drei Übertragungsprozeduren erlaubt, indem die Belegungszeiten der einzelnen Komponenten (Übertragungskanäle, Verzögerungsstufen) sowie die Rückführungswahrscheinlichkeiten geeignet gewählt werden. Im folgenden wird deshalb das alle Möglichkeiten umfassende Modell mit Rückführungen und Verzögerungsstufen analysiert.

3.2.3 Verkehrsparameter

Anhand des im Bild 3.5 jetzt ausschließlich mittels verkehrstheoretischer Symbole "Warteschlangen" und "Bedienungseinheiten" dargestellten Modells sollen nun die Verkehrsparameter des Systems

beschrieben werden, das heißt die zugrundegelegten Annahmen über die stochastischen Ankunfts- und Bedienungsprozesse.

Wie man sieht, fehlt im Bild 3.5 ein Symbol für den Paketspeicher, da - wie oben beschrieben - im folgenden angenommen wird, daß die Speicherplatzzuweisung erst beim Eintreffen des letzten Bits eines Pakets erfolgt. Die endliche Größe des Speichers wird im Modell dadurch berücksichtigt, daß externe Ankünfte nur erfolgreich sind, solange die Anzahl aller Anforderungen (Pakete) im gesamten Modell (d.h. in allen vorhandenen Warteschlangen und Bedienungseinheiten) kleiner ist als die Anzahl n von Paket-speicherplätzen.

Der Grund für diese Festlegung ist, daß für jede Anforderung, die sich in irgendeiner Komponente des Modells befindet, genau ein physikalischer Speicherplatz belegt ist. Voraussetzung für diese Beschreibung ist, daß für die Listenführung (Anlegen der Warteschlangen) im Vermittlungsrechner soviel Speicherplatz vorgesehen ist, daß hierdurch kein Engpaß entstehen kann. Da für diese Listen ungleich weniger Platz notwendig ist als für die physikalische Speicherung der Pakete, ist dies in aller Regel der Fall.

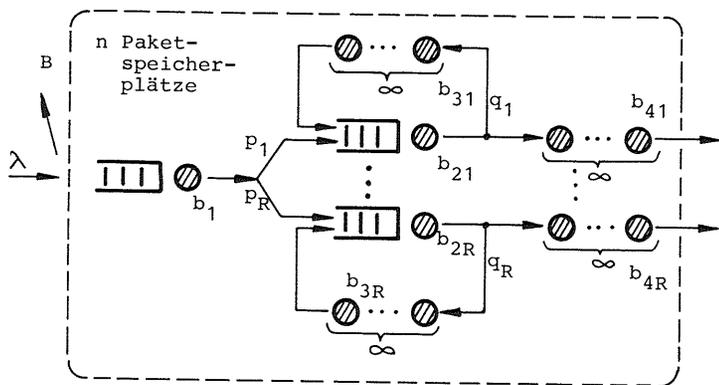


Bild 3.5: Untersuchtes Verkehrsmodell eines Vermittlungsknotens mit einem Vermittlungsmodul (Bedeutung der einzelnen Komponenten siehe Bild 3.2 und Text)

Als Ankunftsprozeß wird ein Poisson-Prozeß mit Ankunftsrate λ angenommen, d.h. die Ankunftsabstände der Pakete sind exponentialverteilt gemäß:

$$P(T_a \leq t) = 1 - e^{-\lambda t} \quad (t \geq 0) \quad (3.1)$$

Diese Annahme erscheint aufgrund der vorausgesetzten Zusammenfassung aller am betrachteten Vermittlungsknoten ankommenden Pakete in einem Ankunftsstrom gerechtfertigt ¹⁾.

Für den ersten - exakten - Teil der Analyse wird angenommen, daß die Belegungszeiten des Prozessors und der Übertragungskanäle exponentialverteilt sind; diese Einschränkung wird im zweiten - approximativen - Teil der Analyse fallengelassen.

Sieht man von Übertragungsfehlern ab, deren Einfluß in der Regel gering ist, so besteht die Belegungszeit der Übertragungskanäle - zumindest bei der wichtigeren Continuous ARQ Methode - aus der reinen Übertragungszeit der Pakete. Diese ergibt sich aus dem Quotienten von Paketlänge und Übertragungsgeschwindigkeit. Bild 3.6 zeigt die im ARPA-Netz gemessene Paketlängenverteilung [51], sowie zum Vergleich die Exponentialverteilung mit gleichem Mittelwert. Wie man sieht, stimmen beide Verteilungen, mit Ausnahme von kleinen Paketlängen relativ gut überein. Daher besitzt die Annahme exponentialverteilter Übertragungszeiten der Pakete, wie sie mit wenigen Ausnahmen (z.B. [38]) gemacht wurde [59,60, 86], durchaus ihre Berechtigung. Dennoch erscheint es aus verschiedenen Gründen notwendig, die Analyse auch für nicht exponentialverteilte Übertragungszeiten durchzuführen (näheres siehe Abschnitt 3.4.1).

Die reinen Verzögerungszeiten für positive bzw. negative Quittungen werden mit Bedienstationen nachgebildet, die aus unendlich vielen Bedienungseinheiten bestehen. Jede dieser Bedienungseinheiten besitzt eine Bedienstzeit, die der vorgegebenen Verzögerung entspricht. Da bei unendlich vielen Bedienungseinheiten kein Warten auftreten kann, wird somit jede Anforderung in einer solchen "Infinite Server"-Station um genau diese vorgeschriebene Zeit verzögert.

¹⁾ Der resultierende Summenprozeß aus unendlich vielen unabhängigen, aber beliebigen Einzelprozessen ist ein Poisson-Prozeß [33].

Die Belegungszeiten der Verzögerungsstufen können im Modell durch beliebige Phasentyp-Verteilungen vorgeschrieben werden, d.h. die Verteilungen müssen eine rationale Laplace-Transformierte besitzen. Mit Hilfe dieser Verteilungen läßt sich auch jede andere Verteilung beliebig genau approximieren [21,9].

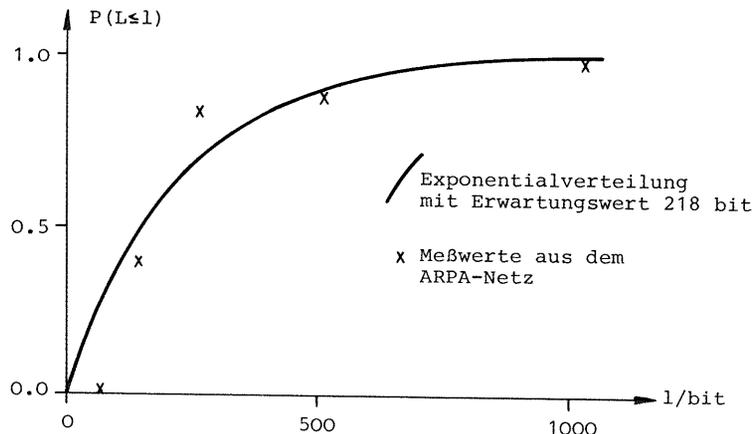


Bild 3.6: Beispiel für Paketlängen-Verteilung [51]

Die einzelnen Bedienungseinheiten besitzen die im Bild 3.5 angegebenen mittleren Belegungszeiten $b_1, b_{21}, \dots, b_{4R}$.

Die Bedienungseinheiten mit ihren Warteschlangen im Bild 3.5 werden im folgenden auch als "Bedienungsstationen" (oder kurz "Stationen") des Warteschlangennetzes bezeichnet. Unter der Nummer einer Station wird der im Bild 3.5 angegebene jeweilige Index (oder Doppelindex) der mittleren Bedienungszeit verstanden.

Es wird angenommen, daß sich die Pakete nach Verlassen des Prozessors gemäß fest vorgegebenen Verzweigungswahrscheinlichkeiten p_1, \dots, p_R auf die R abgehenden Übertragungskanäle aufteilen. Mit der Wiederholungswahrscheinlichkeit q_r wird ein Paket nicht erfolgreich über den Kanal Nr. r übertragen und muß deshalb wiederholt werden.

3.3 Analyse bei exponentialverteilten Übertragungs- und Prozessorbelegungszeiten

Das im Bild 3.5 dargestellte und im vorigen Abschnitt beschriebene Modell eines Vermittlungsknotens stellt einen Spezialfall jener Bedienungssysteme dar, die man als Warteschlangennetze bezeichnet. Die Theorie und Anwendung dieser Netze hat in jüngster Zeit im Rahmen der Rechnerverkehrstheorie einen großen Aufschwung genommen.

Im Anhang 1 sind die für unsere Anwendung wichtigsten exakten Ergebnisse aus der Theorie der Warteschlangennetze zusammengestellt. Das dort beschriebene Theorem [45,2] wird im folgenden auf unser Problem angewandt. In diesem Abschnitt gehen wir davon aus, daß die Belegungszeiten des Prozessors und die Übertragungszeiten exponentialverteilt sind und daß ferner die ankommenden Pakete exponentialverteilte Ankunftsabstände besitzen.

Ziel der Analyse ist die Bestimmung der charakteristischen Verkehrsgrößen des Vermittlungsknotens, wie z.B. sein Durchsatz, Warte- und Verweilzeiten der Pakete, Speicherüberlauf-Wahrscheinlichkeit, usw. Wie in vielen Beispielen der Verkehrstheorie führt der Weg zur Bestimmung dieser Größen auch hier über die Berechnung der stationären Zustandswahrscheinlichkeiten des Systems. Ihre Definition und Bestimmung wird im folgenden Abschnitt beschrieben.

3.3.1 Zustandswahrscheinlichkeiten

Zur Berechnung der Zustandswahrscheinlichkeiten des Warteschlangennetzes im Bild 3.5 wird das im Anhang 1 beschriebene Theorem herangezogen. Da es an dieser Stelle der Arbeit zum ersten Mal benutzt wird, soll seine Anwendung auf unser Problem hier etwas ausführlicher dargestellt werden.

Es erweist sich als zweckmäßig, zur Beschreibung des Momentan-

zustandes des zu untersuchenden Warteschlangennetzes, die Anzahlen von Anforderungen zu benutzen, welche sich gleichzeitig in den verschiedenen Stationen des Warteschlangennetzes befinden. Das heißt, wir bezeichnen mit $X_1, X_{21}, \dots, X_{2R}, X_{31}, \dots, \dots, X_{3R}, X_{41}, \dots, X_{4R}$ die zufällige Anzahl vorhandener Anforderungen in den jeweiligen Bedienstationsstationen Nr. 1, (2,1),, (4,R) gemäß Bild 3.5 und mit den zugehörigen kleinen Buchstaben $x_1, x_{21}, \dots, x_{4R}$ die entsprechenden Realisierungen dieser Zufallsvariablen. Ein bestimmter Zustand des Systems ist durch den Vektor aller so definierten Variablen bestimmt.

Unser Modell ist ein sogenanntes offenes Warteschlangennetz, da bei ihm Anforderungen (Pakete) von externen Verkehrsquellen eintreffen. Bedingt durch den endlich großen Paketspeicher (n Plätze) werden jedoch eintreffende Anforderungen abgewiesen, falls sich bereits n Anforderungen im Warteschlangennetz befinden (vgl. Abschnitt 3.2.3). Dies können wir (im Einklang mit der Beschreibung im Anhang 1) als zustandsabhängige externe Ankunftsrate auffassen:

$$\lambda^* = \lambda^*(\vec{x}) = \begin{cases} \lambda & \text{für } \hat{x} = x_1 + \sum_{s=2}^4 \sum_{r=1}^R x_{sr} < n \\ 0 & \text{sonst} \end{cases} \quad (3.2)$$

Nach dem im Anhang 1 beschriebenen Theorem besitzt die Lösung für die stationären Zustandswahrscheinlichkeiten eine Produktform gemäß:

$$\begin{aligned} P(X_1=x_1, X_{21}=x_{21}, \dots, X_{2R}=x_{2R}, X_{31}=x_{31}, \dots, X_{3R}=x_{3R}, X_{41}=x_{41}, \dots, X_{4R}=x_{4R}) \\ = p_e(x_1, x_{21}, \dots, x_{2R}, x_{31}, \dots, x_{3R}, \dots, x_{41}, \dots, x_{4R}) = p_e(\vec{x}) \\ = d(\vec{x}) \cdot g_1(x_1) \cdot g_{21}(x_{21}) \cdot \dots \cdot g_{4R}(x_{4R}) \cdot C_e \end{aligned} \quad (3.3)$$

Im folgenden wird auf die Bedeutung der Faktoren dieser Lösung im einzelnen eingegangen.

Der erste Faktor $d(\vec{x})$ ist eine Funktion des Zustandsvektors \vec{x}

und der zustandsabhängigen Ankunftsrate λ^* nach Gl.(3.2):

$$\begin{aligned} d(\vec{x}) &= d(x_1, \dots, x_{4R}) \\ &= \prod_{\xi=0}^{x_1 + \dots + x_{4R} - 1} \lambda^*(\xi) = \lambda^{x_1 + \sum_{s=2}^4 \sum_{r=1}^R x_{sr}} \end{aligned} \quad (3.4)$$

Die Faktoren g_1, \dots, g_{4R} in Gl.(3.3) stehen jeweils für eine Station des Netzes und hängen ab von deren mittlerer Bedienungszeit, der Anzahl gleichzeitiger Anforderungen in der Station, vom Typ der Bedienstationsstation, sowie vom Erwartungswert der Anzahl von Durchläufen einer Anforderung durch diese Station. Für unser Beispiel gilt:

$$\left. \begin{aligned} g_1(x_1) &= (e_1 b_1)^{x_1} \\ g_{2r}(x_{2r}) &= (e_{2r} b_{2r})^{x_{2r}} \\ g_{3r}(x_{3r}) &= \frac{(e_{3r} b_{3r})^{x_{3r}}}{x_{3r}!} \\ g_{4r}(x_{4r}) &= \frac{(e_{4r} b_{4r})^{x_{4r}}}{x_{4r}!} \end{aligned} \right\} r \in \{1, \dots, R\} \quad (3.5a)$$

Die mittleren Anzahlen e_1, \dots, e_{4R} von Durchläufen der Anforderungen durch die Stationen bestimmt man wie folgt:

- 1) $e_1 = 1$ (Station Nr. 1 wird genau einmal von jeder Anforderung durchlaufen.)
- 2) $e_{2r} = e_1 p_r + e_{3r}$ (Station Nr. (2r) wird nach Station Nr. 1 mit Wahrscheinlichkeit p_r durchlaufen und ferner mit Sicherheit im Anschluß an Station Nr. (3r).)
- 3) $e_{3r} = q_r e_{2r}$ (Station Nr. (3r) wird nach Station Nr. (2r) mit Wahrscheinlichkeit q_r durchlaufen.)
- 4) $e_{4r} = (1 - q_r) e_{2r}$ (Station Nr. (4r) wird nach Station Nr. (2r) mit Wahrscheinlichkeit $(1 - q_r)$ durchlaufen.)

Aus diesen Gleichungen erhält man unmittelbar:

$$e_1 = 1 ; e_{2r} = \frac{p_r}{1-q_r} ; e_{3r} = \frac{p_r q_r}{1-q_r} ; e_{4r} = p_r \quad r \in \{1, \dots, R\} \quad (3.5b)$$

Der letzte Faktor C_e in Gl.(3.3) ist die Normierungskonstante, die so gewählt werden muß, daß die Summe aller Zustandswahrscheinlichkeiten gleich 1 ist.

Hiermit ergibt sich die Lösung für die stationären Zustandswahrscheinlichkeiten des betrachteten Modells nach Bild 3.5 zu:

$$p_e(\vec{x}) = C_e \cdot A_1^{x_1} \prod_{r=1}^R A_{2r}^{x_{2r}} \frac{A_{3r}^{x_{3r}}}{x_{3r}!} \frac{A_{4r}^{x_{4r}}}{x_{4r}!} \quad (3.6)$$

$$\text{mit: } A_1 = \lambda b_1 ; A_{2r} = \frac{p_r}{1-q_r} \lambda b_{2r} ; A_{3r} = \frac{p_r q_r}{1-q_r} \lambda b_{3r} ; A_{4r} = p_r \lambda b_{4r} \quad r \in \{1, \dots, R\}$$

$$C_e = \left(\sum_{\vec{x} \in Z} A_1^{x_1} \prod_{r=1}^R A_{2r}^{x_{2r}} \frac{A_{3r}^{x_{3r}}}{x_{3r}!} \frac{A_{4r}^{x_{4r}}}{x_{4r}!} \right)^{-1}$$

$$Z = \left\{ (x_1, x_{21}, \dots, x_{4R}) \mid x_1, x_{21}, \dots, x_{4R} \in \mathbb{N}_0 \wedge x_1 + \sum_{s=2}^4 \sum_{r=1}^R x_{sr} \leq n \right\}$$

Im folgenden werden die Größen $A_1, A_{2r}, A_{3r}, A_{4r}$ in Gl.(3.6) als Angebote an die einzelnen Stationen Nr. 1, (2r), (3r), (4r) bezeichnet.

3.3.2 Charakteristische Verkehrsgrößen

Mit Hilfe der Zustandswahrscheinlichkeiten lassen sich die interessierenden charakteristischen Verkehrsgrößen des Vermittlungsknotens bestimmen. (Numerische Ergebnisse hierzu siehe Abschnitt 3.5.)

Speicherüberlauf-Wahrscheinlichkeit

Die Speicherüberlauf-Wahrscheinlichkeit gibt an, mit welcher Wahrscheinlichkeit ein ankommendes Paket im betrachteten Vermittlungsknoten abgewiesen wird, weil es auf einen vollen Paket-

speicher trifft. Wegen des vorausgesetzten Poisson-Ankunftsprozesses ist diese Wahrscheinlichkeit gleich der Wahrscheinlichkeit, daß sich gleichzeitig n Pakete im Speicher befinden:

$$B = P \left(X_1 + \sum_{s=2}^4 \sum_{r=1}^R X_{sr} = n \right) = \sum_{\vec{x} \in \Psi(n)} p_e(\vec{x}) \quad (3.7)$$

Dabei soll hier wie im folgenden gelten:

$$\Psi(\hat{x}) = \left\{ \vec{x} \mid \vec{x} \in Z \wedge x_1 + \sum_{s=2}^4 \sum_{r=1}^R x_{sr} = \hat{x} \right\} \quad (3.8)$$

Mittlerer Abstand zwischen aufeinanderfolgenden Blockierzuständen und mittlere Dauer eines Blockierzustandes

Für die praktische Beurteilung des Verkehrsverhaltens eines Vermittlungsknotens sind zwei Größen wegen ihrer guten anschaulichen Deutbarkeit besonders geeignet (vgl. Bild 3.7):

- der mittlere Abstand zwischen aufeinanderfolgenden Blockierzuständen t_A , sowie
- die mittlere Dauer eines Blockierzustandes t_B .

Beide Zeiten lassen sich mit Hilfe der folgenden Überlegung bestimmen.

Es sei \hat{X} die zufällige Gesamtzahl der Anforderungen in unserem Warteschlangennetz, d.h. die Anzahl belegter Paketspeicherplätze:

$$\hat{X} = X_1 + \sum_{s=2}^4 \sum_{r=1}^R X_{sr} \quad (3.9)$$

Ferner sei z_1 die mittlere Anzahl der Übergänge der Zufallsvariablen \hat{X} vom Zustand $\hat{X} = n-1$ in den Zustand $\hat{X} = n$, sowie z_2 die mittlere Anzahl der Übergänge vom Zustand $\hat{X} = n$ in den Zustand $\hat{X} = n-1$, jeweils während einer Beobachtungszeit T. Dann gilt:

$$z_1 = (\text{gesamte Verweilzeit im Zustand } \hat{X} = n-1) \cdot (\text{Übergangsrate von } \hat{X} = n-1 \text{ nach } \hat{X} = n) \\ = P(\hat{X} = n-1) \cdot T \cdot \lambda$$

$$z_2 = \frac{\text{(gesamte Verweilzeit im Zustand } \hat{X} = n)}{t_B} = \frac{P(\hat{X} = n) \cdot T}{t_B}$$

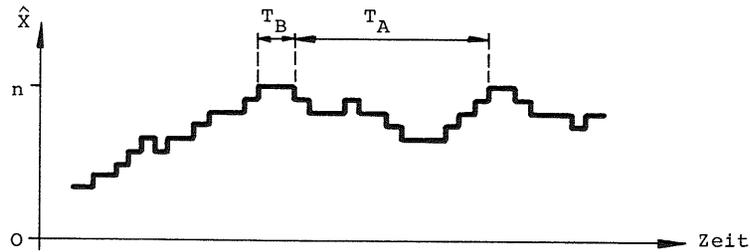


Bild 3.7: Zeitlicher Verlauf der Gesamtzahl von Paketen im Speicher

T_A : zufälliger Abstand zwischen aufeinanderfolgenden Blockierzuständen

T_B : zufällige Dauer eines Blockierzustandes

Bei Stationarität des Prozesses und für $T \rightarrow \infty$ muß $z_1 = z_2$ sein. Daraus folgt für die mittlere Dauer eines Blockierzustandes:

$$t_B = \frac{P(\hat{X} = n)}{P(\hat{X} = n-1)} \cdot \frac{1}{\lambda} = \frac{B}{\sum_{\vec{x} \in \Psi(n-1)} p_e(\vec{x})} \cdot \frac{1}{\lambda} \quad (3.10)$$

Außerdem muß natürlich gelten (vgl. Bild 3.7):

$$P(\hat{X} = n) = \frac{t_B}{t_B + t_A} \quad (3.11)$$

Wegen $P(\hat{X} = n) = B$ ergibt sich somit der mittlere Abstand zwischen aufeinanderfolgenden Blockierzuständen zu:

$$t_A = \frac{1-B}{B} t_B = \frac{1-B}{P(\hat{X} = n-1)} \cdot \frac{1}{\lambda} = \frac{1-B}{\sum_{\vec{x} \in \Psi(n-1)} p_e(\vec{x})} \cdot \frac{1}{\lambda} \quad (3.12)$$

Speicherbelastung

Die Speicherbelastung ist definiert als die mittlere Anzahl von Paketen, die sich gleichzeitig im Speicher befinden:

$$Y_{SP} = \sum_{x=0}^n \hat{x} \cdot P\left(\chi_1 + \sum_{s=2}^4 \sum_{r=1}^R \chi_{sr} = \hat{x}\right) = \sum_{x=0}^n \hat{x} \cdot \sum_{\vec{x} \in \Psi(\hat{x})} p_e(\vec{x}) \quad (3.13)$$

Mittlere Speicherverweilzeit

Als Speicherverweilzeit bezeichnen wir die Zeitdauer vom Einspeichern eines Pakets im betrachteten Vermittlungsknoten bis zu seiner Löschung infolge einer positiven Quittung vom Nachbarknoten. Ihr Erwartungswert ist gegeben durch:

$$t_{SP} = \frac{Y_{SP}}{\lambda(1-B)} \quad (3.14)$$

Durchsatz

Als Durchsatz wird die mit Erfolg pro Zeiteinheit vermittelte und übertragene Anzahl von Paketen bezeichnet. Er stellt eine wichtige Größe bei der Beurteilung der Verkehrsleistung eines

Vermittlungsknotens dar.

$$D = \lambda(1-B) \quad (3.15)$$

Mittlere Warteschlangenlänge und mittlere Wartezeit vor dem Prozessor

Die mittlere Warteschlangenlänge vor dem Prozessor ist der Erwartungswert der Anzahl von Paketen, welche auf die Bedienung durch den Prozessor warten:

$$\begin{aligned} \Omega_1 &= \sum_{\xi=1}^{n-1} \xi \cdot P(X_1 = \xi+1) \\ &= \sum_{\xi=1}^{n-1} \xi \cdot \sum_{\vec{x} \in \Phi_1(\xi+1)} p_e(\vec{x}) \end{aligned} \quad (3.16)$$

mit $\Phi_1(\eta) = \{\vec{x} \mid \vec{x} \in Z \wedge x_1 = \eta\}$

Die mittlere Wartezeit vor dem Prozessor ergibt sich damit zu:

$$w_1 = \frac{\Omega_1}{\lambda(1-B)} \quad (3.17)$$

Belastung des Prozessors

Die Prozessorbelastung ist gleich der Wahrscheinlichkeit, daß der Prozessor belegt ist. Sie kann aus den bereits abgeleiteten Verkehrsgrößen berechnet werden:

$$Y_1 = \lambda(1-B)b_1 = Db_1 \quad (3.18)$$

Mittlere Warteschlangenlängen und mittlere Wartezeiten vor den Übertragungskanälen

Völlig analoge Überlegungen wie für die entsprechenden Größen des Prozessors führen zu:

$$\Omega_{2r} = \sum_{\xi=1}^{n-1} \xi \cdot P(X_{2r} = \xi+1)$$

$$\Omega_{2r} = \sum_{\xi=1}^{n-1} \xi \cdot \sum_{\vec{x} \in \Phi_{2r}(\xi+1)} p_e(\vec{x}) \quad (3.19)$$

mit $\Phi_{2r}(\eta) = \{\vec{x} \mid \vec{x} \in Z \wedge x_{2r} = \eta\}$

$$w_{2r} = \frac{\Omega_{2r}}{\lambda(1-B)} \cdot \frac{1-q_r}{P_r} \quad r \in \{1, \dots, R\} \quad (3.20)$$

Belastung der Übertragungskanäle

Entsprechend zur Prozessorbelastung gilt für die Belastung der Übertragungskanäle:

$$\begin{aligned} Y_{2r} &= \lambda(1-B) \frac{P_r}{1-q_r} b_{2r} \\ &= D \frac{P_r}{1-q_r} b_{2r} \quad r \in \{1, \dots, R\} \end{aligned} \quad (3.21)$$

3.3.3 Numerische Auswertung

Die Berechnung der interessierenden Verkehrsgrößen erfolgt, wie bereits beschrieben, mit Hilfe der Zustandswahrscheinlichkeiten des Prozesses. Die numerische Auswertung der Lösung für die Zustandswahrscheinlichkeiten erfordert zunächst eine Berechnung der Summe aller Produkte in Gl. (3.6) für die Bestimmung der normierenden Konstanten C_e . Aufgrund der für reale Anwendungen riesigen Anzahl von Zuständen verbietet sich eine direkte Berechnung der Summe ¹⁾.

Glücklicherweise kann für unser Problem eine für geschlossene Warteschlangennetze entwickelte numerische Methode sinngemäß angewandt werden [13,80]. Dieser sog. "Faltungsalgorithmus" er-

¹⁾ In unserem Fall gibt es, wie sich leicht zeigen läßt, insgesamt $\binom{n+3R+1}{3R+1}$ Zustände. Dies sind z.B. ungefähr $6,7 \cdot 10^{60}$ Zustände bei $R=30$ abgehenden Leitungen und $n=110$ Speicherplätzen.

möglichst eine äußerst effektive Berechnung der normierenden Konstanten in Gl. (3.6) sowie der charakteristischen Verkehrsgrößen. Seine Anwendung auf unser Problem wird im Abschnitt 3.3.3.2 beschrieben.

Zuvor soll aber im nächsten Punkt auf eine Möglichkeit eingegangen werden, welche die numerische Rechnung zu vereinfachen hilft, nämlich die Zusammenfassung der Verzögerungsstationen.

3.3.3.1 Zusammenfassung der Verzögerungsstationen

In unserem Modell nach Bild 3.5 kommen bei insgesamt R abgehenden Übertragungskanälen 2R reine Verzögerungsstationen vor, die die Quittierzeiten nachbilden. Dies führt in der Lösung für die Zustandswahrscheinlichkeiten nach Gl. (3.6) zu 2R Faktoren des Typs

$$\frac{A_{3r}^{x_{3r}}}{x_{3r}!} \text{ bzw. } \frac{A_{4r}^{x_{4r}}}{x_{4r}!} \quad r \in \{1, \dots, R\}$$

Die einzelne Berücksichtigung aller dieser Terme erfordert selbst bei Benutzung des Faltungsalgorithmus relativ viel Rechenzeit und kann u.U. zu numerischen Genauigkeitsproblemen führen.

Einen Ausweg aus dieser Schwierigkeit liefert die folgende Überlegung. Wie im Anhang 2 gezeigt wird, ist es möglich, alle Verzögerungsstationen in unserem Modell zu einer fiktiven Verzögerungsstation zusammenzufassen. Dazu müssen in der Formel (3.6) für die Zustandswahrscheinlichkeiten alle Faktoren $A_{sr}^{x_{sr}}/x_{sr}!$ entfernt werden und stattdessen ein einziger Faktor $A_0^{x_0}/x_0!$ eingefügt werden. Das resultierende Angebot A_0 an diese Ersatzstation ist die Summe der Angebote an alle Verzögerungsstationen im Modell:

$$A_0 = \sum_{r=1}^R (A_{3r} + A_{4r}) \quad (3.22)$$

Damit stellt sich die Lösung für die Zustandswahrscheinlichkeiten in kürzerer Form dar:

$$\begin{aligned} p_0(x_0, x_1, x_{21}, \dots, x_{2R}) &= p_0(\vec{x}_0) \\ &= C_e \frac{A_0^{x_0}}{x_0!} A_1^{x_1} \prod_{r=1}^R A_{2r}^{x_{2r}} \end{aligned} \quad (3.23)$$

$$C_e = \left(\sum_{\vec{x}_0 \in Z_0} \frac{A_0^{x_0}}{x_0!} A_1^{x_1} \prod_{r=1}^R A_{2r}^{x_{2r}} \right)^{-1}$$

$$Z_0 = \left\{ \vec{x}_0 \mid x_0, x_1, x_{21}, \dots, x_{2R} \in \mathbb{N}_0 \wedge x_0 + x_1 + \sum_{r=1}^R x_{2r} \leq n \right\}$$

A_0 gemäß Gl. (3.22), $A_1, A_{21}, \dots, A_{2R}$ gemäß Gl. (3.6)

3.3.3.2 Numerische Auswertung der Produktlösung

Um die Zustandswahrscheinlichkeiten und damit die interessierenden Verkehrsgrößen numerisch berechnen zu können, muß zunächst die normierende Konstante C_e in Gl. (3.23) bestimmt werden. Hierzu wird ein Faltungsalgorithmus [13,80] herangezogen, der im folgenden beschrieben wird. Danach wird gezeigt, wie die numerische Bestimmung der Verkehrsgrößen erfolgt.

Um die Notation zu vereinfachen, wird für die folgende Beschreibung die Produktlösung gemäß Gl. (3.23) in einer etwas abgewandelten Form dargestellt, indem die Stationen des Warteschlangennetzes fortlaufend numeriert werden. Die Ersatzstation erhält die Nummer $u = 1$, der Prozessor erhält die Nummer $u = 2$, der Übertragungskanal Nr. (2,1) im Bild 3.5 erhält die Nummer $u = 3$, usw., und schließlich erhält der Übertragungskanal Nr. (2,R) die Nummer $u = U = R+2$.

Mit Hilfe dieser Nummern werden auch die Zustandsvariablen umbenannt:

$$x_0 = y_1 ; \quad x_1 = y_2 ; \quad x_{2r} = y_{2+r} \quad (r \in \{1, \dots, R\}) \quad (3.24)$$

Schließlich sollen noch die zu den einzelnen Stationen gehörenden Faktoren in Gl. (3.23) folgendermaßen abgekürzt werden:

$$\begin{aligned} f_1(y_1) &= \frac{A_0^{y_1}}{y_1!} \\ f_2(y_2) &= A_1^{y_2} \\ f_u(y_u) &= A_{2,u-2}^{y_u} \quad u \in \{3, 4, \dots, U\} \end{aligned} \quad (3.25)$$

Berechnung der Normierungskonstanten

Mit den oben eingeführten Abkürzungen läßt sich die Normierungskonstante C_e in Gl. (3.23) in der folgenden Form darstellen:

$$C_e = \left(\sum_{\vec{y} \in Z(U,n)} \prod_{u=1}^U f_u(y_u) \right)^{-1}$$

$$\vec{y} = (y_1, \dots, y_U) \quad (3.26a)$$

$$Z(U,n) = \{ \vec{y} \mid y_1, \dots, y_U \in \mathbb{N}_0 \wedge \sum_{u=1}^U y_u \leq n \}$$

Mit Hilfe der Definition

$$D(U,i) = \{ \vec{y} \mid y_1, \dots, y_U \in \mathbb{N}_0 \wedge \sum_{u=1}^U y_u = i \} \quad (3.26b)$$

läßt sich C_e zerlegen in:

$$C_e = \left(\sum_{i=0}^n G(U,i) \right)^{-1} \quad (3.26c)$$

$$G(U,i) = \sum_{\vec{y} \in D(U,i)} \prod_{u=1}^U f_u(y_u) \quad (3.26d)$$

Die Größen $G(U,i)$ lassen sich in sehr effektiver Weise mit Hilfe des im Anhang 5 allgemein beschriebenen Faltungsalgorithmus [13, 80] berechnen. Der Klarheit wegen soll die Anwendung dieses Verfahrens auf das hier vorliegende Problem im folgenden gezeigt werden. Die Größen $G(U,i)$ in Gl. (3.26d) lassen sich folgendermaßen berechnen:

(1) Setze für $u \in \{1, 2, \dots, U\}$

$$G(u,0) = 1 \quad (3.27a)$$

(2) Setze für $i \in \{1, 2, \dots, n\}$

$$G(1,i) = \frac{A_0^i}{i!} \quad (3.27b)$$

(3) Berechne für $i \in \{1, 2, \dots, n\}$ rekursiv

$$G(2,i) = G(1,i) + A_1 G(2,i-1) \quad (3.27c)$$

(4) Berechne für $u \in \{3, 4, \dots, U\}$
und $i \in \{1, 2, \dots, n\}$ rekursiv

$$G(u,i) = G(u-1,i) + A_{2,u-2} G(u,i-1) \quad (3.27d)$$

Somit sind alle Größen $G(U,i)$ und mit Hilfe von Gl. (3.26c) auch die Normierungskonstante C_e sehr effektiv berechenbar.

Berechnung der Verkehrsgrößen

Wenn die Normierungskonstante bestimmt ist, stellt sich noch die Frage, wie die Verkehrsgrößen gemäß Gl. (3.7) bis (3.21) numerisch bestimmt werden können. Eine unmittelbare Ausrechnung durch Summation aller relevanten Einzelwahrscheinlichkeiten verbietet sich in den meisten Fällen wiederum wegen der zu großen Zustandszahl.

Wie im folgenden gezeigt wird, lassen sich jedoch alle interessierenden Verkehrsgrößen dadurch numerisch bestimmen, daß sie auf Terme zurückgeführt werden, die bei der Berechnung der Normierungskonstanten mit Hilfe des Faltungsalgorithmus sowieso bestimmt werden.

Die Speicherüberlaufwahrscheinlichkeit nach Gl. (3.7) berechnet sich folgendermaßen:

$$B = P(X_0 + X_1 + \sum_{r=1}^R X_{2r} = n)$$

$$= P\left(\sum_{u=1}^U y_u = n\right) \quad (3.28)$$

$$= C_e \sum_{\vec{y} \in D(U,n)} \prod_{u=1}^U f_u(y_u)$$

$$= C_e G(U,n)$$

Zur Bestimmung des mittleren Abstandes zwischen aufeinanderfolgenden Blockierzuständen und der mittleren Dauer eines Blockierzustandes wird die Wahrscheinlichkeit benötigt, daß sich gleichzeitig $(n - 1)$ Pakete im Speicher befinden. Diese berechnet sich analog zu oben:

$$P(X_0 + X_1 + \sum_{r=1}^R X_{2r} = n-1) = C_e G(U, n-1) \quad (3.29)$$

Auch alle Terme, welche für die Berechnung der Speicherbelastung benötigt werden, fallen während der Bestimmung der Normierungskonstanten von selbst an:

$$\begin{aligned} Y_{sp} &= \sum_{x=0}^n \hat{x} \cdot P(X_0 + X_1 + \sum_{r=1}^R X_{2r} = \hat{x}) \\ &= C_0 \sum_{\hat{x}=0}^n \hat{x} G(U, \hat{x}) \end{aligned} \quad (3.30)$$

Die mittlere Warteschlangenlänge vor dem Prozessor läßt sich in der folgenden Weise auf die Terme des Faltungsalgorithmus zurückführen:

$$\begin{aligned} \Omega_1 &= \sum_{\xi=1}^{n-1} \xi \cdot P(X_1 = \xi + 1) \\ &= \sum_{\xi=1}^n \xi \cdot P(X_1 = \xi) - P(X_1 > 0) \\ &= E[X_1] - Y_1 \end{aligned} \quad (3.31)$$

Der Erwartungswert von X_1 kann folgendermaßen berechnet werden:

$$\begin{aligned} E[X_1] &= \sum_{\eta=1}^n \eta P(y_2 = \eta) \\ &= \sum_{\eta=1}^n P(y_2 \geq \eta) \end{aligned}$$

$$\begin{aligned} E[X_1] &= \sum_{\eta=1}^n C_e \sum_{\substack{\vec{y} \in Z(U, n) \\ \wedge y_2 \geq \eta}} \prod_{u=1}^U f_u(y_u) \\ &= \sum_{\eta=1}^n C_e \sum_{v=0}^n \sum_{\substack{\vec{y} \in D(U, v) \\ \wedge y_2 \geq \eta}} \prod_{u=1}^U f_u(y_u) \\ &= \sum_{\eta=1}^n C_e \sum_{v=\eta}^n A_1^v G(U, v-\eta) \end{aligned} \quad (3.32)$$

Hiermit ergibt sich schließlich die mittlere Warteschlangenlänge vor dem Prozessor zu:

$$\Omega_1 = C_e \sum_{\eta=1}^n A_1^{\eta} \sum_{v=\eta}^n G(U, v-\eta) - \lambda(1-B)b_1 \quad (3.33)$$

Durch eine völlig analoge Überlegung findet man die mittlere Warteschlangenlänge vor den Übertragungskanälen:

$$\Omega_{2r} = C_e \sum_{\eta=1}^n A_{2r}^{\eta} \sum_{v=\eta}^n G(U, v-\eta) - \lambda(1-B)b_{2r} \frac{p_r}{1-q_r} \quad (3.34)$$

$r \in \{1, \dots, R\}$

Alle übrigen Verkehrsgrößen lassen sich, wie im Abschnitt 3.3.2 beschrieben, aus den hier angegebenen Größen direkt berechnen.

3.4 Analyse bei allgemein verteilten Übertragungs- und Prozessorbelegungszeiten

3.4.1 Motivation

In diesem Abschnitt befassen wir uns mit einem Näherungsverfahren für die Berechnung der Verkehrsgrößen von Vermittlungsknoten, wenn die Übertragungszeiten und die Belegungszeiten des Prozessors nicht ausschließlich exponentiell sondern beliebig anders verteilt sind. Maßgebende Gründe für diese Untersuchung sind:

1. Wie wir im Abschnitt 3.2.2 gesehen haben, ist die Exponentialverteilung in recht guter Übereinstimmung mit den dort wiedergegebenen gemessenen Paketlängen für ein Paketvermittlungsknoten. Im Sinne einer allgemeinen Anwendbarkeit der hier durchgeführten Untersuchungen erscheint es dennoch wünschenswert, das Verfahren so zu erweitern, daß auch nicht exponentialverteilte Paketlängen behandelt werden können, denn es ist nicht auszuschließen, daß durch spezielle Anwendungen in einem Netz die Paketlängenverteilung erheblich beeinflußt wird. Dies kann insbesondere in speziellen Rechnernetzen der Fall sein, wie zum Beispiel in Auskunftssystemen.
- Dieselbe Überlegung trifft im Prinzip auch für die Prozessorbelegungszeiten zu.
2. Im Abschnitt 3.2.1.2 wurde gezeigt, wie sich der Einfluß der Übertragungsprozedur auf den Verkehrsablauf modellieren läßt. Sowohl beim "Send-and-Wait" als auch beim "Continuous"-Verfahren mit nichtselektiver Wiederholung geschieht dies dadurch, daß man als Belegungszeit des Kanals die sogenannte resultierende Übermittlungszeit eines Pakets betrachtet. Diese Zeit setzt sich zusammen aus der reinen Übertragungszeit plus eventuellen zusätzlichen Zeitanteilen für negative Quittungen, Time-outs und Übertragungswiederholungen im Falle von Störungen.

Wie die resultierenden Übermittlungszeiten bestimmt werden

können ist in [27] bzw. [12] beschrieben. Da diese Ersatzbelegungszeiten in der Regel nicht exponentialverteilt sind, ergibt sich auch aus dieser Sicht das Bedürfnis, unser Modell eines Vermittlungsknotens für allgemein verteilte Kanalbelegungszeiten zu analysieren.

3.4.2 Analyse von Warteschlangennetzen mit allgemein verteilten Bedienungszeiten

Geht man bei der Analyse von Warteschlangennetzen von der Voraussetzung exponentialverteilter Bedienungszeiten in den Stationen mit einer Bedienungseinheit und first-come-first-served Abfertigungsdisziplin (Typ 1 im Anhang 1) ab, so wird die mathematische Analyse äußerst schwierig, was sich u.a. darin ausdrückt, daß - abgesehen von Sonderfällen - bis heute keine exakten theoretischen Ergebnisse dafür bekannt sind. Im Rahmen gewisser Einschränkungen bezüglich der Art der Verteilungsfunktionen, der Topologie der Netze und der Anzahl der Anforderungen im Netz sind jedoch approximative Lösungen entwickelt worden, und zwar sowohl für offene als auch für geschlossene Netze [17,57,54].

Für diejenige Art von Warteschlangennetzen, wie sie unser Modell darstellt, nämlich offene Netze mit einer beschränkten Gesamtzahl von Anforderungen im Netz, wird in den folgenden Abschnitten ein approximatives Rechenverfahren beschrieben. Dieses Verfahren ist verwandt mit der Methode in [84] für geschlossene Warteschlangennetze.

3.4.3 Näherungsverfahren

3.4.3.1 Grundlagen des Verfahrens

Ausgangspunkt des Näherungsverfahrens ist eine Beobachtung, welche bei der exakten Lösung für Warteschlangennetze mit exponentialverteilten Bedienungszeiten zutrifft: Das Ergebnis für die Zustandswahrscheinlichkeiten in Gl. (3.6) läßt sich durch einfaches Erweitern von Zähler und Nenner in der folgenden Weise darstellen:

$$p_e(\vec{x}) = \frac{\prod_{r=1}^R K_r A_r^{x_r} e^{-A_r}}{\sum_{\vec{x} \in Z} \prod_{r=1}^R K_r A_r^{x_r} e^{-A_r}} \quad (3.35)$$

$$A_1 = \lambda b_1; A_{2r} = \frac{p_r}{1-q_r} \lambda b_{2r}; A_{3r} = \frac{p_r q_r}{1-q_r} \lambda b_{3r}; A_{4r} = p_r \lambda b_{4r}$$

Wie man sieht, setzt sich das Ergebnis aus einzelnen Faktoren zusammen, welche jeweils zu einer Station des Netzes gehören. Diese Faktoren entsprechen - nach der obigen Erweiterung - genau den Zustandswahrscheinlichkeiten, die man erhält, wenn man die Stationen aus dem Netz isoliert und sie mit einem Poisson-Prozess mit einer solchen Rate speist, daß sich die in Gl. (3.35) angegebenen Angebote $A_1, A_{2r}, A_{3r}, A_{4r}$ einstellen ¹⁾.

Diese Produktlösungseigenschaft wollen wir nun als heuristischen Ansatz für die Analyse unseres Modells auch für den Fall nicht-exponentialverteilter Bedienungszeiten übernehmen, d.h. wir postulieren folgende Gestalt der Lösung für die Zustandswahrscheinlichkeiten:

$$p_a(\vec{x}) = C_a p_1(x_1) \prod_{r=1}^R p_{2r}(x_{2r}) p_{3r}(x_{3r}) p_{4r}(x_{4r}) \quad (3.36)$$

¹⁾ Die Zustandswahrscheinlichkeiten (Wahrscheinlichkeit x Anforderungen gleichzeitig im System) eines $M/M/1$ Wartesystems mit endlich oder unendlich großem Wartespeicher besitzen die folgende Form [8]:

$$p(x) = K A^x \quad (3.37)$$

wobei A das Verkehrsangebot und K die Normierungskonstante ist. (Bei unendlich großem Wartespeicher gilt: $K = 1 - A$.)

Für die Zustandswahrscheinlichkeiten eines $M/G/\infty$ Systems mit Angebot A gilt [8]:

$$p(x) = e^{-A} \frac{A^x}{x!} \quad (3.38)$$

Die Faktoren für die einzelnen Stationen in Gl. (3.36) bestimmen wir durch die Berechnung der Zustandswahrscheinlichkeiten der isolierten Systeme mit angebotenen Poisson-Verkehr und allgemein verteilten Bedienungszeiten. Die Verkehrsangebote dieser isolierten Systeme sind dieselben wie die in Gl. (3.35) für die jeweiligen Stationen angegebenen Angebote $A_1, A_{2r}, A_{3r}, A_{4r}$. Im einzelnen bedeutet dies:

- a) Die Faktoren $p_1, p_{21}, \dots, p_{2R}$ für die Stationen Nr. 1, (2,1), ..., (2,R) mit einer Bedienungseinheit (vgl. Bild 3.5) berechnen wir aus den Zustandswahrscheinlichkeiten der entsprechenden isolierten Wartesysteme vom Typ $M/G/1, n-1$ mit Poisson-Ankunftsprozess und der für diese Station vorgeschriebenen Bedienungszeitverteilung. Da die Gesamtzahl der Anforderungen im System maximal gleich n ist, sieht man bei der Analyse der isolierten Stationen konsequenterweise einen endlich großen Wartespeicher mit $s = n-1$ Wartepätzen vor.
- b) Die Faktoren p_{31}, \dots, p_{3R} und p_{41}, \dots, p_{4R} bestimmen wir in analoger Weise durch die isolierte Betrachtung von Systemen des Typs $M/G/\infty$. Wie bereits erwähnt, hängen die Zustandswahrscheinlichkeiten dieses Bedienungssystems nur vom Erwartungswert der Bedienungszeit, nicht aber von den sonstigen Eigenschaften der Bedienungszeitverteilung ab. (Dasselbe gilt im übrigen auch für das Verlustsystem $M/G/n$ mit vollkommener Erreichbarkeit.) Für die Faktoren p_{31}, \dots, p_{4R} setzen wir also dieselbe Lösung wie im exponentiellen Fall an.

Zusammengefaßt besteht unsere Näherungslösung aus dem folgenden Ansatz:

$$p_a(\vec{x}) = C_a p_1(x_1) \prod_{r=1}^R p_{2r}(x_{2r}) \frac{A_{3r}^{x_{3r}}}{x_{3r}!} \frac{A_{4r}^{x_{4r}}}{x_{4r}!} \quad (3.39)$$

$$C_a = \left(\sum_{\vec{x} \in Z} p_1(x_1) \prod_{r=1}^R p_{2r}(x_{2r}) \frac{A_{3r}^{x_{3r}}}{x_{3r}!} \frac{A_{4r}^{x_{4r}}}{x_{4r}!} \right)^{-1}$$

$$Z = \left\{ \vec{x} \mid x_1, \dots, x_{4R} \in \mathbb{N}_0 \wedge x_1 + \sum_{s=2}^4 \sum_{r=1}^R x_{sr} \leq n \right\}$$

$p_1(x_1)$: Zustandswahrscheinlichkeiten des Systems $M/G_1/1, n-1$ mit Angebot $A_1 = \lambda b_1$ (G_1 : vorgeschriebene Bedienungszeitverteilung des Prozessors)

$p_{2r}(x_{2r})$: Zustandswahrscheinlichkeiten des Systems $M/G_{2r}/1, n-1$ mit Angebot $A_{2r} = \frac{p_r}{1-q_r} \lambda b_{2r}$ (G_{2r} : vorgeschriebene Bedienungszeitverteilung für Kanal Nr. r)

$$A_{3r} = \frac{p_r q_r}{1-q_r} \lambda b_{3r} \quad ; \quad A_{4r} = p_r \lambda b_{4r}$$

Die Bestimmung der Faktoren $p_1(x_1), p_{21}(x_{21}), \dots, p_{2R}(x_{2R})$ wird im folgenden Abschnitt beschrieben.

3.4.3.2 Analyse der isolierten Wartestationen

Verfahren der eingebetteten Markoffkette

Zur Berechnung der beschriebenen Näherungslösung bei allgemein verteilten Bedienungszeiten der Stationen mit einer Bedienungseinheit ist die Bestimmung der Zustandswahrscheinlichkeiten der entsprechenden $M/G/1, s$ Wartesysteme notwendig. Für die Analyse solcher Wartesysteme hat sich die Methode der "eingebetteten Markoffkette" als sehr wirksames theoretisches Hilfsmittel erwiesen [74,76]. Dieses Verfahren nützt aus, daß zu den Zeitpunkten t_n , bei denen eine Anforderung gerade die Bedienungseinheit verlassen hat, die zufälligen Anzahlen der Anforderungen im System $X(t_n)$ eine Markoffkette bilden. Das heißt, daß der Systemzustand in den nachfolgenden Kettenpunkten $X(t_{n+i})$ nur vom gegenwärtigen Zustand $X(t_n)$ abhängt, nicht aber von Zuständen zu früheren Zeitpunkten $X(t_{n-i})$.

Aus diesem Grund gehorchen die Wahrscheinlichkeiten $p^*(x)$, daß sich zu einem Kettenzeitpunkt bei Stationarität x Anforderungen im System befinden, dem folgenden Gleichungssystem:

$$\begin{aligned} p^*(0) &= p^*(0)a_0 + p^*(1)a_0 \\ p^*(x) &= p^*(x)a_1 + p^*(0)a_x + \sum_{j=1}^{x-1} p^*(j)a_{x-j+1} + p^*(x+1)a_0 \quad (3.40) \\ \sum_{x=0}^s p^*(x) &= 1 \quad x \in \{1, \dots, s-1\} \end{aligned}$$

Die Größen a_i in Gl. (3.40) stellen die Übergangswahrscheinlichkeiten der Markoffkette dar; a_i ist die Wahrscheinlichkeit, daß während einer zufälligen Bedienungszeit insgesamt i neue Anforderungen eintreffen. Bei einem Poisson-Ankunftsprozeß mit Rate $\bar{\lambda}$ berechnen sich diese nach der folgenden Beziehung, wenn wir mit $b(t)$ die Dichte der Bedienungszeit-Verteilung bezeichnen:

$$a_i = \int_0^{\infty} \frac{(\bar{\lambda}t)^i}{i!} e^{-\bar{\lambda}t} b(t) dt \quad (3.41)$$

Eine schnelle und zuverlässige Methode zur Auswertung dieses Integrals für eine wichtige Klasse von Verteilungen wird weiter unten angegeben.

Für unsere Lösung interessieren uns allerdings nicht die Zustandswahrscheinlichkeiten $p^*(x)$ (d.h. die Zustandswahrscheinlichkeiten nach einem Bedienungsende), sondern die sog. Antreffwahrscheinlichkeiten $p(x)$, das heißt die Wahrscheinlichkeit, daß eine ankommende Anforderung x Anforderungen im System vorfindet.

Der Zusammenhang zwischen $p^*(x)$ und $p(x)$ läßt sich durch eine Durchsatzbetrachtung finden [20]. Es gilt für die Antreffwahrscheinlichkeiten (A sei das Angebot an das $M/G/1, s$ System):

$$\begin{aligned} p(x) &= \frac{p^*(x)}{p^*(0) + A} \quad x \in \{0, 1, \dots, s\} \\ p(s+1) &= \frac{p^*(0) + A - 1}{p^*(0) + A} \end{aligned} \quad (3.42)$$

Somit lassen sich die gesuchten Antreffwahrscheinlichkeiten $p(x)$ leicht bestimmen, sofern die Zustandswahrscheinlichkeiten der Markoffkette $p^*(x)$ bekannt sind. Wie man aus Gl. (3.40) ersieht, können die Zustandswahrscheinlichkeiten $p^*(x)$, beginnend mit $x = 0$, rekursiv berechnet werden:

$$\begin{aligned} p^*(1) &= \frac{1}{a_0} (1 - a_0) p^*(0) \\ p^*(x+1) &= \frac{1}{a_0} \left[(1 - a_1) p^*(x) - a_x p^*(0) - \sum_{j=1}^{x-1} a_{x-j+1} p^*(j) \right] \quad (3.43) \\ & \quad x \in \{1, \dots, s-1\} \end{aligned}$$

Numerische Behandlung der Rekursion

Bei unserer Anwendung ist es unverzichtbar, daß die Zustandswahrscheinlichkeiten auch für große Werte von x und damit normalerweise kleinen Werten von $p^*(x)$ numerisch genau berechnet werden können. Wie die Erfahrung zeigt, kann hierbei die Berechnung der Differenz in Gl. (3.43) zu numerischen Schwierigkeiten führen.

Dieses Problem kann durch die folgende Substitution vollkommen vermieden werden [70]. Wir definieren als Hilfsgrößen:

$$\sigma_k = \sum_{i=k}^{\infty} a_i \tag{3.44}$$

Wie im Anhang 3 gezeigt wird, lassen sich die rekursiven Beziehungen (3.43) mit Hilfe der Hilfsgrößen σ_k folgendermaßen umformen:

$$p^*(x+1) = \frac{1}{a_0} \left[\sigma_{x+1} p^*(0) + \sum_{j=1}^x \sigma_{x-j+2} p^*(j) \right] \tag{3.45}$$

$x \in \{0, 1, \dots, s-1\}$

Da die Berechnung der Größen σ_k numerisch problemlos ist und die Rekursion für die Wahrscheinlichkeiten $p^*(x)$ jetzt keine Differenzbildung mehr erfordert, ist die numerisch stabile Bestimmung der Zustandswahrscheinlichkeiten der isolierten $M/G/1, s$ Wartesysteme hiermit gewährleistet.

Voraussetzung der geschilderten Berechnung ist allerdings die Kenntnis der Übergangswahrscheinlichkeiten a_i der Markoffkette. Im folgenden Punkt wird ein Verfahren beschrieben, wie diese Übergangswahrscheinlichkeiten für die wichtige Klasse der allgemeinen Phasentyp-Verteilungen schnell und numerisch zuverlässig berechnet werden können.

Bestimmung der Übergangswahrscheinlichkeiten

Für beliebig verteilte Bedienungszeiten können die Übergangswahrscheinlichkeiten prinzipiell mit Hilfe der Gl. (3.41) be-

stimmt werden. Allerdings zeigt die Erfahrung, daß die numerische Auswertung des Integrals in (3.41) für gewisse Verteilungsfunktionen erhebliche Schwierigkeiten mit sich bringt. Im folgenden wird deshalb ein neues Verfahren entwickelt, welches eine effektive und numerisch stabile Berechnung der Übergangswahrscheinlichkeiten erlaubt.

Voraussetzung des Verfahrens ist, daß die Bedienungszeit mit Hilfe einer Phasentyp-Verteilung exakt oder näherungsweise dargestellt ist. Eine Verteilung ist vom Phasentyp, wenn sie sich mittels exponentialverteilter Zufallsvariablen erzeugen läßt. Dabei geschieht der Übertritt von einer Phase zu einer anderen gemäß vorgegebenen Aufteilungswahrscheinlichkeiten. Wie Cox [21] zeigen konnte, lassen sich alle denkbaren Phasentyp-Verteilungen in der normierten Darstellung gemäß Bild 3.8 darstellen.

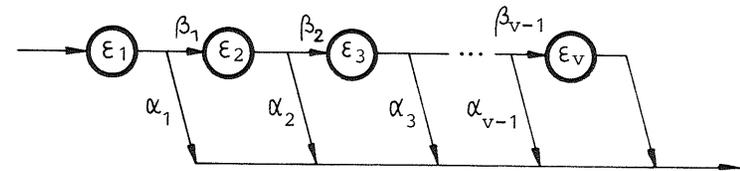


Bild 3.8: Allgemeine Anordnung exponentialverteilter Phasen nach Cox

Dabei denkt man sich eine zufällige Bedienungszeit in der folgenden Weise erzeugt: Die Bedienungszeit besteht mindestens aus einem exponentialverteilten Zeitabschnitt (Phase) mit Mittelwert $b_{p1} = \epsilon_1^{-1}$. Sie ist mit Wahrscheinlichkeit α_1 danach beendet, oder sie verlängert sich mit Wahrscheinlichkeit $\beta_1 = 1 - \alpha_1$ um eine weitere Phase mit Mittelwert $b_{p2} = \epsilon_2^{-1}$, usw.

Bekannte Vertreter der Phasentyp-Verteilungen sind die Familien der Erlangverteilungen und der Hyperexponentiellen Verteilungen. Für die praktische Anwendung der Phasentyp-Verteilungen ist die Tatsache wichtig, daß allgemeine Verteilungsfunktionen mit beliebiger Genauigkeit mit ihrer Hilfe approximiert werden können. Ein Approximationsverfahren auf der Basis von Phasentyp-Vertei-

lungen wird in |9| vorgeschlagen.

Für allgemeine Bedienungszeit-Verteilungen vom Phasentyp lassen sich die Übergangswahrscheinlichkeiten a_i in Gl. (3.41) in der folgenden Weise bestimmen.

Die erzeugende Funktion der Verteilung a_i ist definiert durch:

$$a^*(z) = \sum_{i=0}^{\infty} a_i z^i \quad (3.46)$$

Einsetzen von Gl. (3.41) in (3.46) ergibt:

$$a^*(z) = \sum_{i=0}^{\infty} \left\{ \int_0^{\infty} \frac{(\bar{\lambda}t)^i}{i!} e^{-\bar{\lambda}t} b(t) dt \right\} z^i \quad (3.47)$$

Durch Vertauschen von Summations- und Integrationsreihenfolge erhält man:

$$\begin{aligned} a^*(z) &= \int_0^{\infty} \sum_{i=0}^{\infty} \frac{(\bar{\lambda}tz)^i}{i!} e^{-\bar{\lambda}t} b(t) dt \\ &= \int_0^{\infty} e^{-(\bar{\lambda}-\bar{\lambda}z)t} b(t) dt \\ &= B^*(\bar{\lambda}-\bar{\lambda}z) \end{aligned} \quad (3.48)$$

mit
$$B^*(s) = \int_0^{\infty} e^{-st} b(t) dt \quad (3.49)$$

Wie man sieht, stellt $B^*(s)$ die Laplace-Transformierte der Bedienungszeit-Verteilung dar. Diese ist, wie sich aus Bild 3.8 ablesen läßt, gegeben durch

$$\begin{aligned} B^*(s) &= \mathcal{L}\{b(t)\} \\ &= \sum_{j=1}^v \gamma_j \prod_{k=1}^j \frac{\epsilon_k}{s + \epsilon_k} \end{aligned} \quad (3.50)$$

mit
$$\gamma_j = \prod_{k=1}^{j-1} \beta_k \alpha_j \quad (3.51)$$

Eingesetzt in Gl. (3.48) ergibt dies die folgende erzeugende Funktion der Verteilung a_j :

$$a^*(z) = \sum_{j=1}^v \gamma_j \prod_{k=1}^j \frac{\epsilon_k}{\bar{\lambda} - \bar{\lambda}z + \epsilon_k} \quad (3.52)$$

Mit Hilfe des Faltungssatzes bei erzeugenden Funktionen |32| erhält man aus (3.52) durch Rücktransformation:

$$\begin{aligned} a_i &= \sum_{j=1}^v \gamma_j \sum_{i_1+\dots+i_j=i} \prod_{k=1}^j \left(\frac{\bar{\lambda}}{\bar{\lambda} + \epsilon_k} \right)^{i_k} \frac{\epsilon_k}{\bar{\lambda} + \epsilon_k} \\ &= \sum_{j=1}^v \gamma_j \prod_{k=1}^j \frac{\epsilon_k}{\bar{\lambda} + \epsilon_k} Q(j, i) \end{aligned} \quad (3.53)$$

mit
$$Q(j, i) = \sum_{i_1+\dots+i_j=i} \prod_{k=1}^j \left(\frac{\bar{\lambda}}{\bar{\lambda} + \epsilon_k} \right)^{i_k} \quad (3.54)$$

Entscheidend für die praktische Berechnung ist nun die Tatsache, daß sich die Berechnung der Terme $Q(j, i)$ in Gl. (3.54) sehr effektiv und numerisch stabil mit Hilfe des im Anhang 5 allgemein beschriebenen Faltungsalgorithmus durchführen läßt. Angewandt auf unser Problem läßt sich der Algorithmus folgendermaßen in drei Schritten durchführen:

- (1) Für $i \in \{1, 2, \dots\}$ setze $Q(0, i) = 0$
- (2) Für $j \in \{1, 2, \dots, v\}$ setze $Q(j, 0) = 1$
- (3) Für $j \in \{1, 2, \dots, v\}$ und $i \in \{1, 2, \dots\}$ berechne rekursiv

$$Q(j, i) = Q(j-1, i) + \frac{\lambda}{\lambda + \epsilon_j} Q(j, i-1) \quad (3.55)$$

Durch diesen extrem einfachen und schnellen Algorithmus können somit die Übergangswahrscheinlichkeiten der eingebetteten Markoff-Kette mit minimalem Speicherplatz- und Rechenaufwand

berechnet werden. Die Kenntnis dieser Übergangswahrscheinlichkeiten ist die Voraussetzung zur Bestimmung der Zustandswahrscheinlichkeiten $p^*(x)$ gemäß Gl. (3.45) und der Antreffwahrscheinlichkeiten $p(x)$ gemäß Gl. (3.42).

3.4.3.3 Numerische Auswertung der Produktlösung

Wenn die einzelnen Faktoren der Gesamtlösung aus der Betrachtung der isolierten Stationen bestimmt sind, so stellt sich natürlich auch hier die Aufgabe, die gesamte Produktlösung auszuwerten, das heißt insbesondere die normierende Konstante und die Verkehrsgrößen numerisch zu bestimmen. Hierzu wird geschickterweise wiederum der im Anhang 5 beschriebene Faltungsalgorithmus benutzt. Allerdings muß hier wegen des allgemeinen Charakters der Faktoren $p_1(x_1)$ die allgemeine Version des Faltungsalgorithmus herangezogen werden, so daß die numerische Berechnung nicht mehr ganz so effizient abläuft wie im exponentiellen Fall. Wie die Erfahrung zeigt, lassen sich aber auch hier bei vertretbarem Speicherplatz- und Rechenzeitaufwand die für die Praxis interessanten Fälle beherrschen (vgl. Abschnitt 3.5).

3.4.3.4 Charakteristische Verkehrsgrößen

Die interessierenden Verkehrsgrößen sind hier selbstverständlich dieselben wie im Falle exponentialverteilter Übertragungs- und Prozessorbelegungszeiten, nämlich Speicherüberlauf-Wahrscheinlichkeit, Speicherbelastung, Durchsatz, Warte- und Verweilzeiten, usw. (vgl. Abschnitt 3.3.2). Wegen des prinzipiell gleichen Aufbaus beider Lösungen werden die Verkehrsgrößen hier völlig analog aus den Zustandswahrscheinlichkeiten bestimmt, das heißt, es gelten nach wie vor die Gleichungen (3.7) bis (3.21), dies natürlich im Rahmen der postulierten Näherungslösung.

3.4.3.5 Zusammenfassung des Näherungsverfahrens

Das entwickelte Näherungsverfahren beruht auf dem Postulat, daß sich die Zustandswahrscheinlichkeiten des gesamten Warteschlangennetzes aus den Einzellösungen für die isolierten Stationen multiplikativ zusammensetzen lassen. Die Zustandswahrscheinlichkeiten der isolierten Stationen werden dabei unter der Voraussetzung eines Poisson-Ankunftsprozesses bestimmt.

Mit der beschriebenen effektiven Bestimmung der Übergangswahrscheinlichkeiten der eingebetteten Markoffkette und der numerisch stabilen rekursiven Berechnung der Zustandswahrscheinlichkeiten der isolierten M/G/1,s Stationen können sämtliche Faktoren der postulierten Produktlösung sicher und schnell berechnet werden. Zur Berechnung der normierenden Konstanten und der Verkehrsgrößen wird auch hier der Faltungsalgorithmus verwendet, so daß insgesamt eine speicherplatz- und rechenzeiteffektive Berechnung der interessierenden Verkehrsgrößen möglich ist.

Um die Genauigkeit der entwickelten Näherungslösung im vorgesehenen Anwendungsbereich festzustellen, kann hier ausschließlich die Simulationstechnik herangezogen werden; die einzige bekannte exakte Lösung bei exponentialverteilten Bedienungszeiten ist als Sonderfall in unserer Lösung enthalten. Mit Hilfe eines zeittreuen Simulationsprogrammes wurde die Genauigkeit der Näherungslösung über eine große Bandbreite von Konfigurationen und Verteilungsfunktionen getestet. Typische Ergebnisse werden im folgenden Abschnitt 3.5 gezeigt und diskutiert.

3.5 Ergebnisse

In diesem Abschnitt werden numerische Ergebnisse für das Verkehrsmodell eines Vermittlungsknotens mit einem Vermittlungsmodul vorgestellt und diskutiert.

Die zuerst gezeigten Ergebnisse wurden für exponentialverteilte Übertragungs- und Prozessorbelegungszeiten durch exakte Rechnung gewonnen (Bilder 3.11 bis 3.18). Anschließend werden Ergebnisse der beschriebenen Näherungsrechnung für allgemein verteilte Übertragungs- und Prozessorbelegungszeiten mit Simulationsergebnissen verglichen (Bilder 3.19 bis 3.21).

Allen Ergebnissen ist ein Modell gemäß Bild 3.5 zugrundegelegt. Um die Vielfalt der möglichen Parameterkombinationen etwas einzuschränken, wurden für alle gezeigten Ergebnisse gemeinsam die folgenden Annahmen gemacht:

- 1) Die mittlere Belegungszeit des Prozessors pro durchgesetztem Paket betrage $b_1 = 5$ msec .
- 2) Die mittlere Paketlänge betrage 1000 bit. Bei angenommenen Übertragungsgeschwindigkeiten von 2.4 kbit/sec der Teilnehmeranschlußleitungen und 48 kbit/sec der Verbindungsleitungen führt dies zu mittleren Übertragungszeiten pro Paket von $b_{2r} = 417$ msec bzw. $b_{2r} = 21$ msec .
- 3) Die Verzögerungszeiten b_{4r} für die positiven Quittungen seien im Mittel doppelt so lang wie die Übertragungszeiten b_{2r} . Die Verzögerungszeiten b_{3r} für negative Quittungen seien im Mittel dreimal so lang wie die Übertragungszeiten b_{2r} . Diese Annahmen beruhen auf entsprechenden, etwas vereinfachten Ergebnissen für die Übertragungsprozedur HDLC [11,12] .
- 4) Die Wiederholungswahrscheinlichkeiten q_r werden für alle Kanäle einheitlich zu $q_1 = \dots = q_R = 0.01$ angenommen.

Die übrigen Struktur- und Verkehrsparameter werden jeweils einzeln für jedes Diagramm angegeben.

Für die Bilder 3.11 und 3.13 bis 3.17 wurde ferner angenommen, daß an dem untersuchten Vermittlungsknoten insgesamt 40 Anschlußleitungen mit 2.4 kbit/sec und 2 Verbindungsleitungen mit 48 kbit/sec Übertragungsgeschwindigkeit angeschlossen sind. Die Verzweigungswahrscheinlichkeiten p_r wurden so gewählt, daß alle abgehenden Leitungen gleich stark belastet sind. Dies heißt, daß über die um einen Faktor 20 schnelleren Verbindungsleitungen 20 mal so viel Pakete übertragen werden wie über eine Anschluß-

leitung. Selbstverständlich kann jede andere Verkehrsaufteilung ebenfalls untersucht werden (siehe z.B. das Ergebnis für schiefes Angebot im Bild 3.18).

Bild 3.11 zeigt den Verlauf des Paketdurchsatzes D über der Ankunftsrate λ der Pakete für verschiedene Speichergrößen n . Wie man sieht, folgt der Durchsatz für kleine Ankunftsraten λ der idealen 45°-Charakteristik, biegt dann ab und erreicht bei endlich großem Speicher abhängig von der Speicherplatz-Anzahl n asymptotisch für $\lambda \rightarrow \infty$ einen Höchstwert. Dieser Höchstwert des Durchsatzes der Pakete bei unendlich großer Ankunftsrate und endlich großem Paketspeicher kann exakt durch die folgende Überlegung bestimmt werden.

Das ursprünglich offene Warteschlangennetz im Bild 3.5 wird durch einen Rückkopplungszweig geschlossen, siehe Bild 3.10 . Im jetzt geschlossenen Warteschlangennetz befinden sich ständig die Maximalzahl von n Anforderungen. Die durch den künstlichen Rückkopplungszweig fließende Rate ist der maximale Paketdurchsatz des Vermittlungsknotens. Die Begründung für diese Tatsache läßt sich auf anschauliche Weise geben:

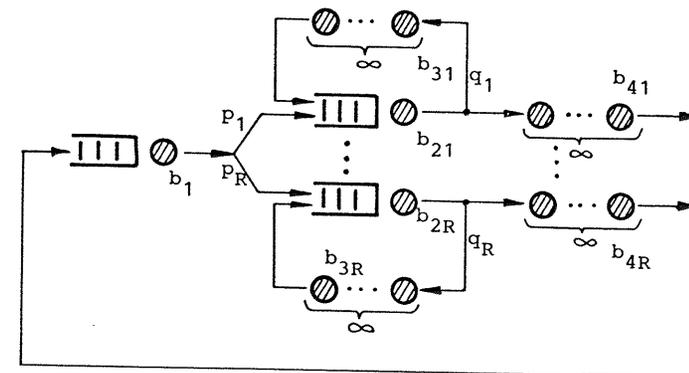


Bild 3.10: Geschlossenes Warteschlangennetz zur Bestimmung des maximalen Paketdurchsatzes

Die Rückkopplung kann auch so interpretiert werden, als ob durch eine unendlich große Ankunftsrate λ sofort ein neues Paket ankommt, sobald auch nur einer der n Speicherplätze frei wird.

Die Lösung für das geschlossene Warteschlangennetz nach Bild 3.11 kann mit dem Theorem im Anhang 1 angegeben werden. Die Zustandswahrscheinlichkeiten ergeben sich zu:

$$\begin{aligned}
 & P(X_1 = x_1, X_{21} = x_{21}, \dots, X_{2R} = x_{2R}, X_{31} = x_{31}, \dots, X_{3R} = x_{3R}, X_{41} = x_{41}, \dots, X_{4R} = x_{4R}) \\
 &= p_g(x_1, x_{21}, \dots, x_{4R}) \\
 &= p_g(\vec{x}) \\
 &= C_g b_1^{x_1} \prod_{r=1}^R \left\{ \left(\frac{p_r}{1-q_r} b_{2r} \right)^{x_{2r}} \left(\frac{p_r q_r}{1-q_r} b_{3r} \right)^{x_{3r}} \frac{1}{x_{3r}!} (p_r b_{4r})^{x_{4r}} \frac{1}{x_{4r}!} \right\} \quad (3.56)
 \end{aligned}$$

$$\sum_{\vec{x} \in Z_g} p_g(\vec{x}) = 1$$

$$Z_g = \left\{ \vec{x} \mid x_1, \dots, x_{4R} \in \mathbb{N}_0 \wedge x_1 + \sum_{s=2}^4 \sum_{r=1}^R x_{sr} = n \right\}$$

Der maximale Durchsatz D_{\max} berechnet sich mit Hilfe dieser Lösung folgendermaßen:

$$D_{\max} = \frac{Y_1}{b_1} \quad (3.57)$$

$$Y_1 = P(x_1 > 0) = \sum_{\xi_1=1}^n \sum_{\vec{x} \in \Psi(\xi_1)} p_g(\vec{x})$$

$$\Psi(\xi_1) = \left\{ \vec{x} \mid \vec{x} \in Z_g \wedge x_1 = \xi_1 \right\}$$

Die gesamte Übertragungsrate aller abgehenden Leitungen im Beispiel des Bildes 3.10 beträgt $(40 \cdot 2,4 + 2 \cdot 48)$ kbit/sec = 192 kbit/sec. Wie zu erkennen ist, ist es nicht möglich, bei endlich großem Speicher die Übertragungskapazität aller Leitungen voll auszunutzen, das heißt, daß trotz voll belegtem Speicher ein nicht kleiner Anteil der Kanäle im Mittel leer steht. Um dies deutlich zu machen, wurde im Bild 3.11 die Durchsatzcharakteristik für den fiktiven Fall des unendlich großen Speichers ebenfalls eingezeichnet. Der Grund für dieses Verhalten liegt in dem für alle Richtungen gemeinsamen Speicher: Pakete, welche vor einem gerade belegten Übertragungskanal warten, belegen Speicherplätze, ohne daß sie momentan zum Durchsatz beitragen, während unter Umständen gleichzeitig eine gewisse Anzahl anderer Kanäle wegen Speicherplatzmangel leer steht.

Bild 3.12 zeigt den Verlauf des maximalen Durchsatzes D_{\max} für einen Vermittlungsknoten mit rein lokalem Verkehr als Funktion der Speichergröße n für verschiedene Anzahlen angeschlossener Leitungen R . Mit wachsender Leitungszahl und Speichergröße nimmt der maximale Durchsatz zu, kann jedoch das Maximum nicht überschreiten, welches durch die maximale Prozessorkapazität von 200 Paketen pro Sekunde in diesem Beispiel vorgegeben ist.

Im Bild 3.13 ist die Überlaufwahrscheinlichkeit B als Funktion der Paket-Ankunftsrate λ für verschiedene Werte der Speichergröße n aufgetragen. Eine etwas abgewandelte Darstellung desselben Zusammenhangs zeigt Bild 3.14. Darin ist für vier feste Werte der Speicherüberlauf-Wahrscheinlichkeit B die erforderliche Speichergröße n als Funktion der Paket-Ankunftsrate λ aufgetragen. Ein Diagramm dieser Art kann direkt zur Speicherdimensionierung verwendet werden.

Bild 3.15 zeigt die auf die Speicherplatz-Anzahl normierte Speicherbelastung Y_{SP} als Funktion des Paketdurchsatzes D für verschiedene Speichergrößen n . Ein Vergleich mit Bild 3.13 zeigt,

daß für die angestrebten kleinen Werte der Überlaufwahrscheinlichkeit zwangsläufig eine relativ niedrige Speicherplatzausnutzung in Kauf genommen werden muß. Ferner sollte beachtet werden, daß die dargestellte Speicherbelastung den sogenannten Speicherverschnitt durch unvollständig ausgenützte Paketspeicherplätze nicht enthält.

Im Bild 3.16 ist die mittlere Speicherverweilzeit t_{SP} als Funktion der Paketankunftsrate λ für verschiedene Werte der Speichergröße n aufgetragen. Das zunächst paradox anmutende Verhalten, daß für sehr hohe Ankunftsraten ein kleinerer Speicher zu einer kürzeren Speicherverweilzeit führt, ist dadurch zu erklären, daß diese Zeit auf die durchgesetzten Pakete bezogen ist. Diejenige Zeit, welche durch wiederholte Übertragungen infolge der in diesem Bereich hohen Speicherüberlauf-Wahrscheinlichkeit verbraucht wird, ist in dieser Darstellung nicht enthalten.

Bild 3.17 gibt die Abhängigkeit des mittleren Abstandes zwischen zwei aufeinanderfolgenden Blockierzuständen t_A von der Paketankunftsrate λ und der Speichergröße n wieder. Wie zu erkennen ist, kann dieser Abstand durch geeignete Dimensionierung in die Größenordnung von Stunden gelegt werden. Da man die Bemessung selbstverständlich für die Hauptverkehrsstunde vornimmt, wird aus diesem Zusammenhang deutlich, daß durch geeignete Dimensionierung Speicherüberläufe weitgehend vermieden werden können. Dies deutet darauf hin, daß bei großzügiger Bemessung der Speicher und Übertragungskanäle Netze so ausgelegt werden können, daß die Mechanismen zur Datenfluß-Steuerung nur zur Abwehr von Überlast wirksam werden müssen.

In den bisherigen Beispielen wurde davon ausgegangen, daß sich der Verkehr so auf die abgehenden Übertragungskanäle verteilt, daß alle Kanäle gleich stark ausgelastet sind. Im Bild 3.18 ist nun anhand eines Vermittlungsknotens mit rein lokalem Verkehr der Einfluß einer ungleichen Belastung der Kanäle auf die Überlaufwahrscheinlichkeit des Paketspeichers dargestellt. Aufge-

tragen ist die Überlaufwahrscheinlichkeit B als Funktion des Paketdurchsatzes D für vier verschiedene Werte des Schiefe-maßes S . Dieses ist in Anlehnung an eine bei der Beurteilung der Verkehrsgüte von Mischungen erfolgreich angewandte Definition [61,62] folgendermaßen festgelegt.

Es wird angenommen, daß das Verkehrsangebot von der ersten bis zur letzten abgehenden Leitung Nr. R in gleichen Stufen ΔA ansteigt. Die Schiefe S des Angebots wird definiert als das Verhältnis des höchsten zum niedrigsten Angebot:

$$S = \frac{A_{2R}}{A_{21}} \quad (3.58)$$

Setzt man der Einfachheit halber voraus, daß die Wiederholungswahrscheinlichkeit für alle Kanäle einheitlich gleich q und die mittlere Übertragungszeit für alle Kanäle gleich b_2 ist, so läßt sich die Schiefe S mit Hilfe von Gleichung (3.6) auch darstellen als:

$$S = \frac{\frac{p_R}{1-q} \lambda b_2}{\frac{p_1}{1-q} \lambda b_2} = \frac{p_R}{p_1} \quad (3.59)$$

Entsprechend wie das Angebot nimmt hiermit auch die Verzweigungswahrscheinlichkeit p_r von der ersten bis zur letzten Leitung um denselben Betrag Δp zu; das heißt, es gilt:

$$p_r = p_1 + (r-1)\Delta p \quad r \in \{1, 2, \dots, R\} \quad (3.60)$$

Mit Hilfe der Bedingung

$$\sum_{r=1}^R p_r = 1 \quad (3.61)$$

erhält man für einen vorgegebenen Wert S der Schiefe und eine vorgegebene Anzahl R abgehender Leitungen die folgenden Ergebnisse für p_1 und Δp :

$$p_1 = \frac{2}{R(S+1)} \quad (3.62)$$

$$\Delta p = \frac{S-1}{R-1} \cdot \frac{2}{R(S+1)} \quad (3.63)$$

Wie aus Bild 3.18 ersichtlich ist, reagiert die Speicherüberlaufwahrscheinlichkeit B äußerst empfindlich auf ein schiefes Ver-

kehrsangebot. Eine wirksame Maßnahme gegen dieses Verhalten ist die künstliche Beschränkung der Warteschlangen vor den einzelnen Übertragungskanälen [40,48]. Hierdurch kann eine Monopolisierung des Speichers durch den Verkehr momentan überlasteter Kanäle verhindert werden. Diese Möglichkeit wird im Abschnitt 4.2 bei der Behandlung von modular aufgebauten Vermittlungsknoten mit zielorientierter Verkehrsaufteilung betrachtet.

Während die bisher gezeigten Ergebnisse für exponentialverteilte Übertragungs- und Prozessorbelegungszeiten gültig sind, zeigen die Bilder 3.19 bis 3.21 den Einfluß andersartiger Belegungszeit-Verteilungen auf die wichtigsten Verkehrsgrößen des Systems.

Zum Vergleich sind neben den Ergebnissen der Näherungsrechnung auch Simulationsergebnisse mit ihren Vertrauensintervallen eingetragen (95% Aussagesicherheit).

Bild 3.19 zeigt den Verlauf der Speicherüberlauf-Wahrscheinlichkeit B über der Paketankunftsrate λ für vier verschiedene Verteilungsfunktionen der Übertragungszeiten:

- konstant (D)
- Erlang-4 (E_4)
- exponentiell (M)
- hyperexponentiell der Ordnung 2 mit quadriertem Variationskoeffizienten $c^2 = \text{Var}[T_b]/E[T_b]^2 = 2$ (H_2).

Wie man sieht, nimmt die Speicherüberlauf-Wahrscheinlichkeit mit zunehmender relativer Streuung der Übertragungszeiten deutlich zu.

Der Grund hierfür ist, daß die Warteschlangen vor den Übertragungskanälen bei stärker streuenden Übertragungszeiten im Mittel länger werden und somit die Anzahl gleichzeitig gespeicherter Pakete zunimmt.

Im Bild 3.20 ist für konstante, exponentiell und hyperexponentiell verteilte Übertragungszeiten die normierte Speicherbelastung Y_{SP}/n als Funktion der Paketankunftsrate λ aufgetragen. Wie dieses Beispiel zeigt, ist diese Verkehrsgröße relativ unempfindlich bezüglich dem Typ der Verteilungsfunktion der Übertragungszeiten.

Dagegen ergeben sich - wie im Bild 3.21 dargestellt - für die mittlere Speicherverweilzeit t_{SP} deutliche Unterschiede abhängig von der Übertragungszeit-Verteilung. Dies rührt wieder daher, daß mit zunehmender Streuung der Übertragungszeiten die Wartezeiten vor den Kanälen ansteigen. Bei größeren Paketankunfts-raten, also höheren Belastungen der Übertragungskanäle, stellen diese Wartezeiten den größten Anteil der gesamten Speicherverweilzeit dar.

Die drei zuletzt diskutierten Diagramme lassen erkennen, daß die wichtigsten Verkehrsgrößen des Systems mit der entwickelten Näherungslösung mit guter Genauigkeit bestimmt werden können.

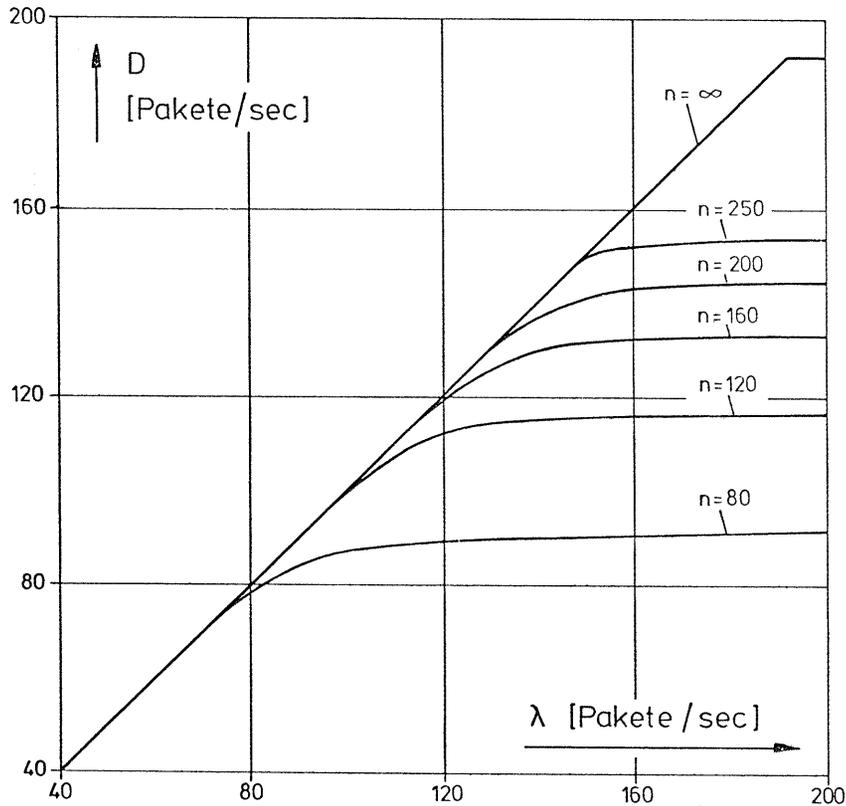


Bild 3.11: Paketdurchsatz D als Funktion der Paketankunftsrate λ für verschiedene Speichergrößen n . Exponentialverteilte Übertragungs- und Prozessorbelegungszeiten.

$R = 42$ abgehende Übertragungskanäle
 mittlere Prozessorbelegungszeit: $b_1 = 5$ msec
 mittlere Übertragungszeiten pro Paket:
 $b_{21} = \dots = b_{2,40} = 417$ msec ; $b_{2,41} = b_{2,42} = 21$ msec
 mittlere Negativ-Quittierzeiten:
 $b_{31} = \dots = b_{3,40} = 1250$ msec ; $b_{3,41} = b_{3,42} = 62$ msec
 mittlere Positiv-Quittierzeiten:
 $b_{41} = \dots = b_{4,40} = 833$ msec ; $b_{4,41} = b_{4,42} = 42$ msec
 Verzweigungswahrscheinlichkeiten:
 $p_1 = \dots = p_{40} = 0.0125$; $p_{41} = p_{42} = 0.25$
 Wiederholungswahrscheinlichkeiten: $q_1 = \dots = q_{42} = 0.01$

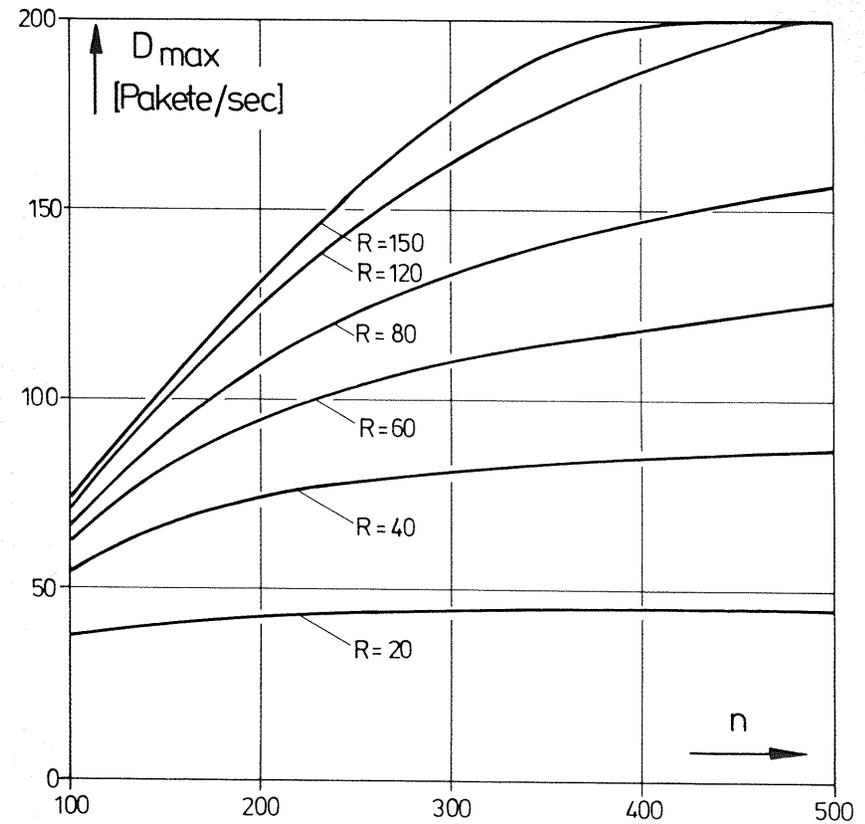


Bild 3.12: Maximaler Paketdurchsatz D_{max} als Funktion der Speichergröße n für verschiedene Leitungszahlen R . Exponentialverteilte Übertragungs- und Prozessorbelegungszeiten.

mittlere Prozessorbelegungszeit: $b_1 = 5$ msec
 mittlere Übertragungszeiten pro Paket: $b_{21} = \dots = b_{2R} = 417$ msec
 mittlere Negativ-Quittierzeiten: $b_{31} = \dots = b_{3R} = 1250$ msec
 mittlere Positiv-Quittierzeiten: $b_{41} = \dots = b_{4R} = 833$ msec
 Verzweigungswahrscheinlichkeiten: $p_1 = \dots = p_R = 1/R$
 Wiederholungswahrscheinlichkeiten: $q_1 = \dots = q_R = 0.01$

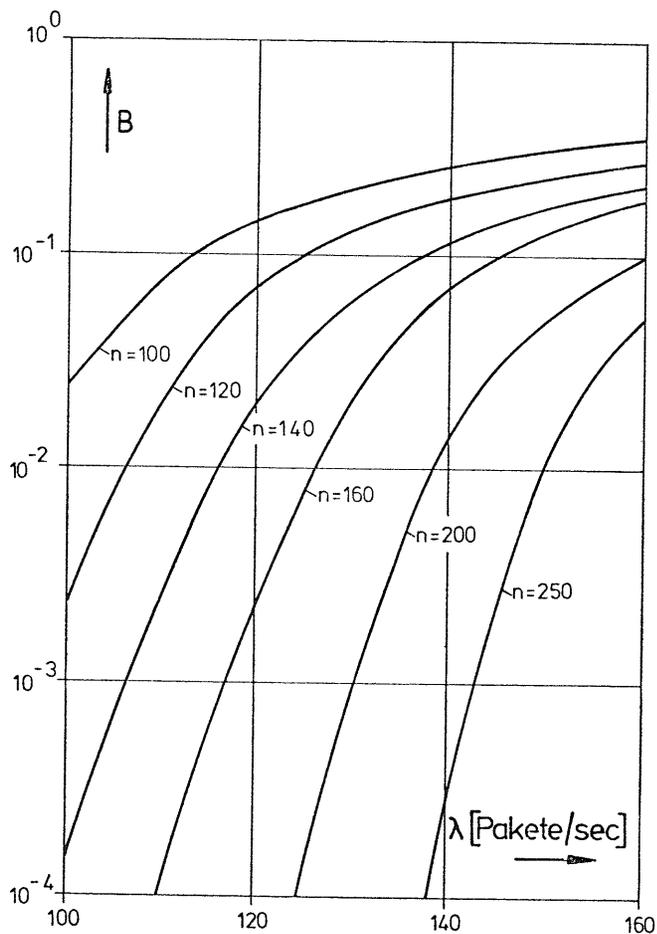


Bild 3.13: Speicherüberlauf-Wahrscheinlichkeit B als Funktion der Paketankunftsrate λ für verschiedene Speichergrößen n. Exponentialverteilte Übertragungs- und Prozessorbelegungszeiten.

R = 42 abgehende Übertragungskanäle
 mittlere Prozessorbelegungszeit: $b_1 = 5$ msec
 mittlere Übertragungszeiten pro Paket:
 $b_{21} = \dots = b_{2,40} = 417$ msec ; $b_{2,41} = b_{2,42} = 21$ msec
 mittlere Negativ-Quittierzeiten:
 $b_{31} = \dots = b_{3,40} = 1250$ msec ; $b_{3,41} = b_{3,42} = 62$ msec
 mittlere Positiv-Quittierzeiten:
 $b_{41} = \dots = b_{4,40} = 833$ msec ; $b_{4,41} = b_{4,42} = 42$ msec
 Verzweigungswahrscheinlichkeiten:
 $p_1 = \dots = p_{40} = 0.0125$; $p_{41} = p_{42} = 0.25$
 Wiederholungswahrscheinlichkeiten:
 $q_1 = \dots = q_{42} = 0.01$

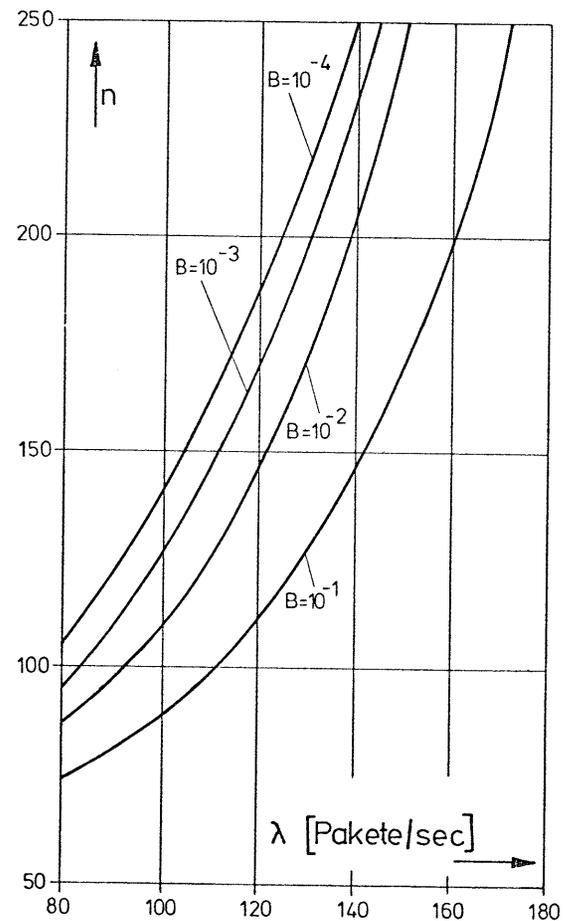


Bild 3.14: Erforderliche Speichergröße n als Funktion der Paketankunftsrate λ für vorgegebene Werte der Speicherüberlauf-Wahrscheinlichkeit B. Exponentialverteilte Übertragungs- und Prozessorbelegungszeiten.

R = 42 abgehende Übertragungskanäle
 mittlere Prozessorbelegungszeit: $b_1 = 5$ msec
 mittlere Übertragungszeiten pro Paket:
 $b_{21} = \dots = b_{2,40} = 417$ msec ; $b_{2,41} = b_{2,42} = 21$ msec
 mittlere Negativ-Quittierzeiten:
 $b_{31} = \dots = b_{3,40} = 1250$ msec ; $b_{3,41} = b_{3,42} = 62$ msec
 mittlere Positiv-Quittierzeiten:
 $b_{41} = \dots = b_{4,40} = 833$ msec ; $b_{4,41} = b_{4,42} = 42$ msec
 Verzweigungswahrscheinlichkeiten:
 $p_1 = \dots = p_{40} = 0.0125$; $p_{41} = p_{42} = 0.25$
 Wiederholungswahrscheinlichkeiten:
 $q_1 = \dots = q_{40} = 0.01$

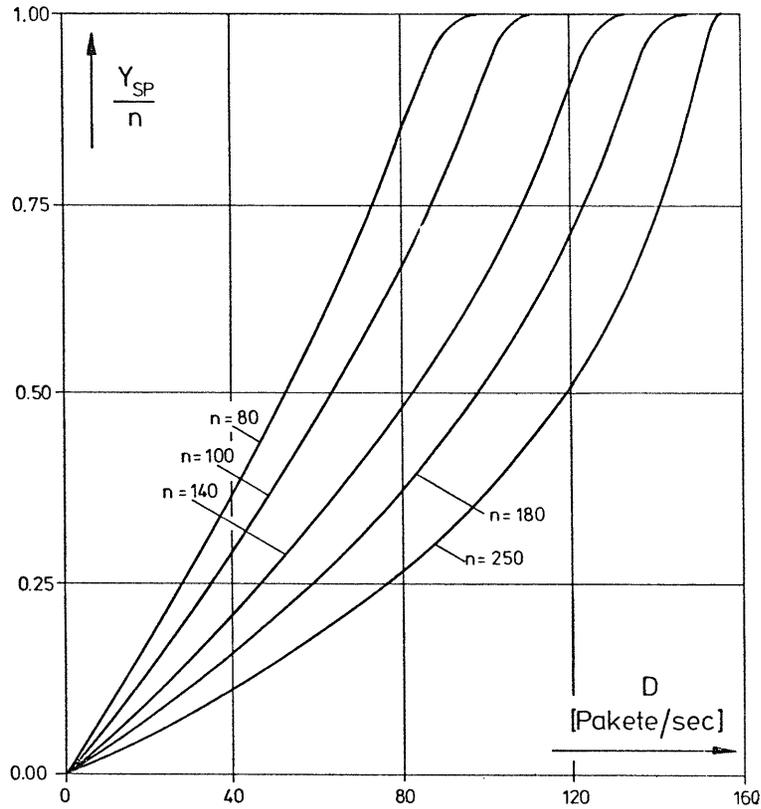


Bild 3.15: Normierte Speicherbelastung Y_{SP}/n als Funktion des Paketdurchsatzes D für verschiedene Speichergrößen n . Exponentialverteilte Übertragungs- und Prozessorbelegungszeiten.

$R = 42$ abgehende Übertragungskanäle
 mittlere Prozessorbelegungszeit pro Paket: $b_1 = 5$ msec
 mittlere Übertragungszeiten pro Paket:
 $b_{21} = \dots = b_{2,40} = 417$ msec ; $b_{2,41} = b_{2,42} = 21$ msec
 mittlere Negativ-Quittierzeiten:
 $b_{31} = \dots = b_{3,40} = 1250$ msec ; $b_{3,41} = b_{3,42} = 62$ msec
 mittlere Positiv-Quittierzeiten:
 $b_{41} = \dots = b_{4,40} = 833$ msec ; $b_{4,41} = b_{4,42} = 42$ msec
 Verzweigungswahrscheinlichkeiten:
 $p_1 = \dots = p_{40} = 0.0125$; $p_{41} = p_{42} = 0.25$
 Wiederholungswahrscheinlichkeiten:
 $q_1 = \dots = q_{42} = 0.01$

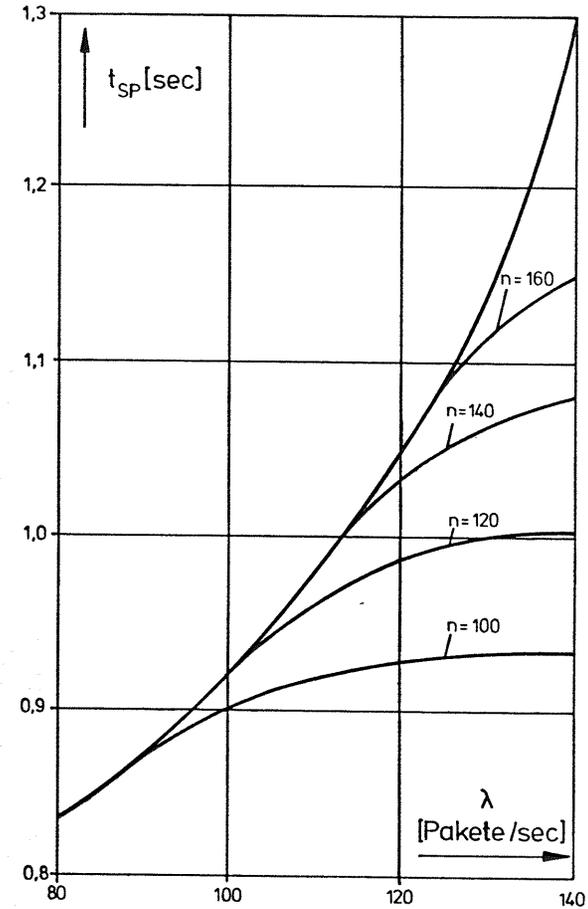


Bild 3.16: Mittlere Speicherverweilzeit der Pakete t_{SP} als Funktion der Paketankunftsrate λ für verschiedene Speichergrößen n . Exponentialverteilte Übertragungs- und Prozessorbelegungszeiten.

$R = 42$ abgehende Übertragungskanäle
 mittlere Prozessorbelegungszeit pro Paket: $b_1 = 5$ msec
 mittlere Übertragungszeiten pro Paket:
 $b_{21} = \dots = b_{2,40} = 417$ msec ; $b_{2,41} = b_{2,42} = 21$ msec
 mittlere Negativ-Quittierzeiten:
 $b_{31} = \dots = b_{3,40} = 1250$ msec ; $b_{3,41} = b_{3,42} = 62$ msec
 mittlere Positiv-Quittierzeiten:
 $b_{41} = \dots = b_{4,40} = 833$ msec ; $b_{4,41} = b_{4,42} = 42$ msec
 Verzweigungswahrscheinlichkeiten:
 $p_1 = \dots = p_{40} = 0.0125$; $p_{41} = p_{42} = 0.25$
 Wiederholungswahrscheinlichkeiten: $q_1 = \dots = q_{42} = 0.01$

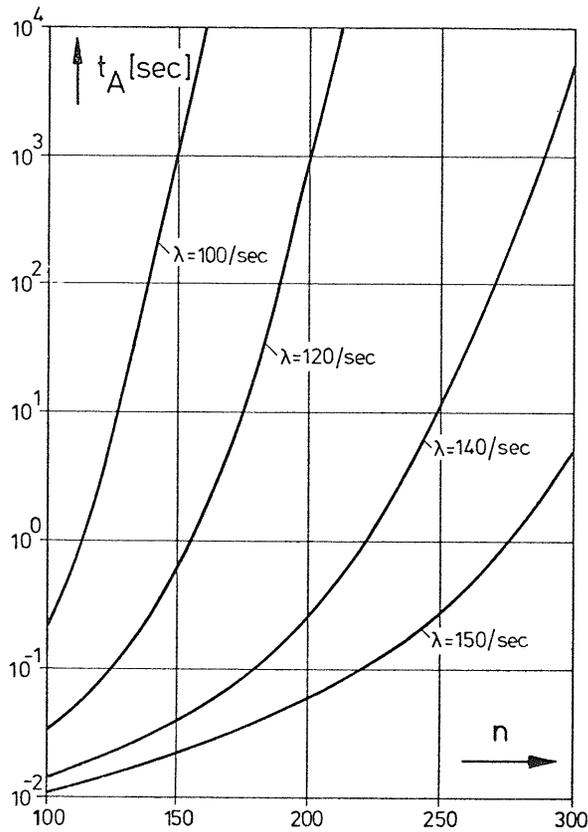


Bild 3.17: Mittlerer Abstand zwischen aufeinanderfolgenden Blockierungszuständen t_A als Funktion der Paketspeichergröße n für verschiedene Paketankunfts-raten λ . Exponentialverteilte Übertragungs- und Prozessorbelegungszeiten.

R = 42 abgehende Übertragungskanäle
 mittlere Prozessorbelegungszeit pro Paket: $b_1 = 5$ msec
 mittlere Übertragungszeiten pro Paket:
 $b_{21} = \dots = b_{2,40} = 417$ msec ; $b_{2,41} = b_{2,42} = 21$ msec
 mittlere Negativ-Quittierzeiten:
 $b_{31} = \dots = b_{3,40} = 1250$ msec ; $b_{3,41} = b_{3,42} = 62$ msec
 mittlere Positiv-Quittierzeiten:
 $b_{41} = \dots = b_{4,40} = 833$ msec ; $b_{4,41} = b_{4,42} = 42$ msec
 Verzweigungswahrscheinlichkeiten:
 $p_1 = \dots = p_{40} = 0.0125$; $p_{41} = p_{42} = 0.25$
 Wiederholungswahrscheinlichkeiten: $q_1 = \dots = q_{42} = 0.01$

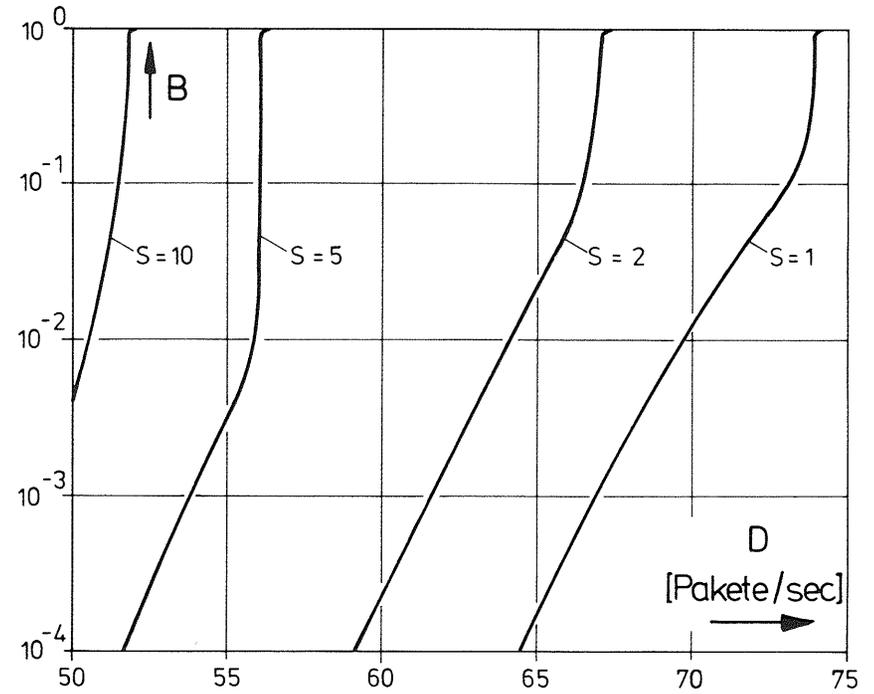


Bild 3.18: Speicherüberlauf-Wahrscheinlichkeit B als Funktion des Paketdurchsatzes D für verschiedene Werte der Schiefe S . Exponentialverteilte Übertragungs- und Prozessorbelegungszeiten.

R = 40 abgehende Übertragungskanäle
 mittlere Prozessorbelegungszeit pro Paket: $b_1 = 5$ msec
 mittlere Übertragungszeiten pro Paket: $b_{21} = \dots = b_{2,40} = 417$ msec
 mittlere Negativ-Quittierzeiten: $b_{31} = \dots = b_{3,40} = 1250$ msec
 mittlere Positiv-Quittierzeiten: $b_{41} = \dots = b_{4,40} = 833$ msec
 Wiederholungswahrscheinlichkeiten: $q_1 = \dots = q_{40} = 0.01$
 $n = 200$ Paketspeicherplätze

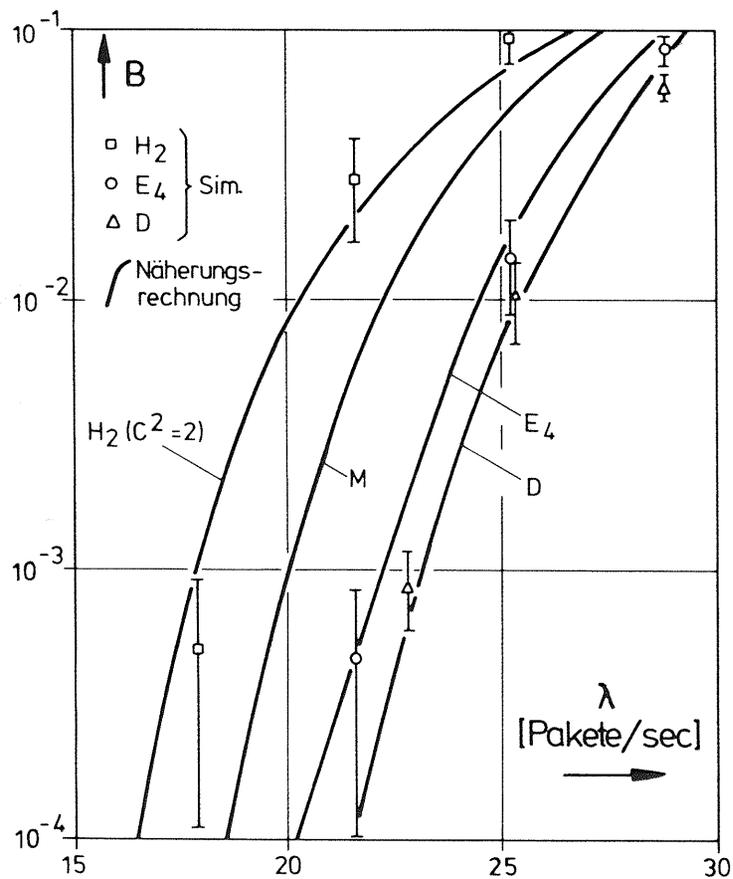


Bild 3.19: Speicherüberlauf-Wahrscheinlichkeit B als Funktion der Paketankunftsrate λ für verschiedene Verteilungsfunktionen der Übertragungszeiten. Exponentialverteilte Prozessorbelegungszeiten.

$R = 15$ abgehende Übertragungskanäle
 $n = 60$ Paketspeicherplätze
 mittlere Prozessorbelegungszeit pro Paket: $b_1 = 5$ msec
 mittlere Übertragungszeiten pro Paket: $b_{21} = \dots = b_{2,15} = 417$ msec
 mittlere Negativ-Quittierzeiten: $b_{31} = \dots = b_{3,15} = 1250$ msec
 mittlere Positiv-Quittierzeiten: $b_{41} = \dots = b_{4,15} = 833$ msec
 Verzweigungswahrscheinlichkeiten: $p_1 = \dots = p_{15} = 0.067$
 Wiederholungswahrscheinlichkeiten: $q_1 = \dots = q_{15} = 0.01$

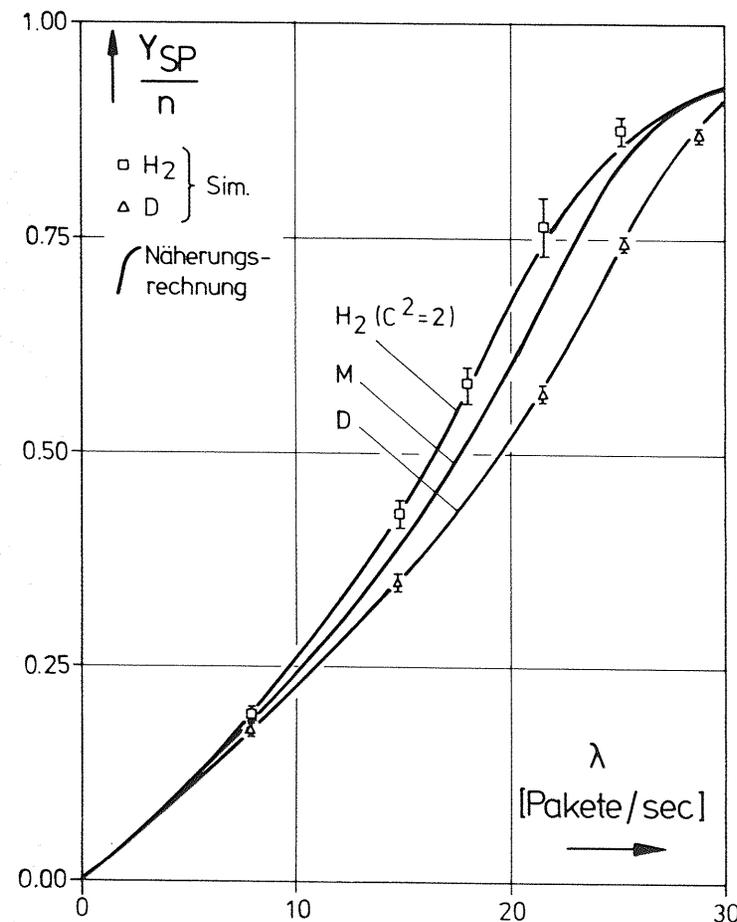


Bild 3.20: Normierte Speicherbelastung $\frac{Y_{SP}}{n}$ als Funktion der Paketankunftsrate λ für verschiedene Verteilungsfunktionen der Übertragungszeiten. Exponentialverteilte Prozessorbelegungszeiten.

$R = 15$ abgehende Übertragungskanäle
 $n = 60$ Paketspeicherplätze
 mittlere Prozessorbelegungszeit pro Paket: $b_1 = 5$ msec
 mittlere Übertragungszeiten pro Paket: $b_{21} = \dots = b_{2,15} = 417$ msec
 mittlere Negativ-Quittierzeiten: $b_{31} = \dots = b_{3,15} = 1250$ msec
 mittlere Positiv-Quittierzeiten: $b_{41} = \dots = b_{4,15} = 833$ msec
 Verzweigungswahrscheinlichkeiten: $p_1 = \dots = p_{15} = 0.067$
 Wiederholungswahrscheinlichkeiten: $q_1 = \dots = q_{15} = 0.01$

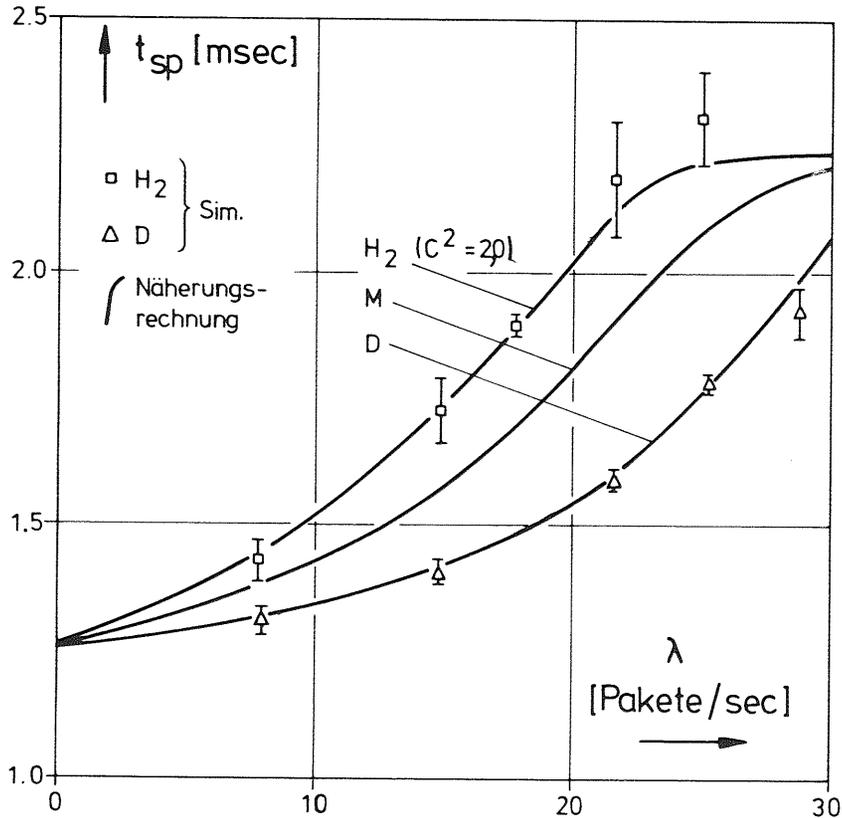


Bild 3.21: Mittlere Speicherverweilzeit der Pakete t_{sp} als Funktion der Paketankunftsrate λ für verschiedene Verteilungsfunktionen der Übertragungszeiten. Exponentialverteilte Prozessorbelegungszeiten.

R = 15 abgehende Übertragungskanäle
 n = 60 Paketspeicherplätze
 mittlere Prozessorbelegungszeit pro Paket: $b_1 = 5$ msec
 mittlere Übertragungszeiten pro Paket: $b_{21} = \dots = b_{2,15} = 417$ msec
 mittlere Negativ-Quittierzeiten: $b_{31} = \dots = b_{3,15} = 1250$ msec
 mittlere Positiv-Quittierzeiten: $b_{41} = \dots = b_{4,15} = 833$ msec
 Verzweigungswahrscheinlichkeiten: $p_1 = \dots = p_{15} = 0.067$
 Wiederholungswahrscheinlichkeiten: $q_1 = \dots = q_{15} = 0.01$

4. VERKEHR IN VERMITTLUNGSKNOTEN MIT MEHREREN VERMITTLUNGSMODULN

4.1 Aufbau und Funktion der Vermittlungsknoten

4.1.1 Gründe für die Modularisierung

Wie in der Fernsprechvermittlungstechnik sprechen auch in der Datenvermittlungstechnik gewichtige Gründe dafür, nicht alle Steuerungsfunktionen an einer einzigen Stelle auszuführen, sondern die Aufgaben in geeigneter Weise auf mehrere Einheiten aufzuteilen. Die Begründung hierfür ist:

- erhöhte Ausfallsicherheit
- verringerte Wirkbreite von Störungen
- verbesserte Ausbaufähigkeit
- Entlastung der zentralen Einheiten
- u.U. kostengünstigere Lösungen.

Für Paketvermittlungen ist in jüngster Zeit eine Reihe von dezentralen Lösungen entwickelt worden [46,73,37, 23,19]. Die im Bild 4.1 gezeigte Darstellung eines solchen Systems wird im folgenden kurz besprochen.

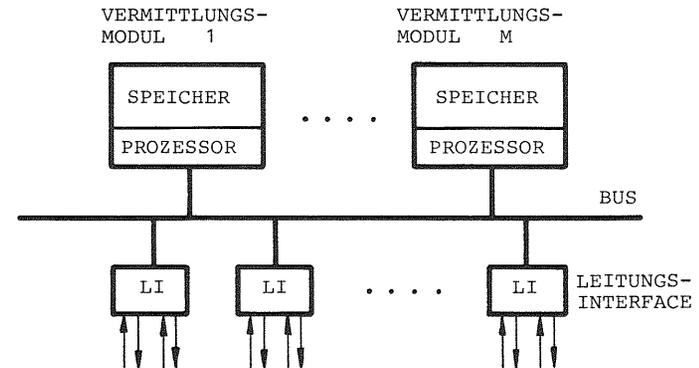


Bild 4.1: Möglicher Aufbau eines Vermittlungsknotens mit mehreren Vermittlungsmoduln

Sowohl Paketspeicher als auch Steuerungsprozessoren sind hier nicht mehr zentralisiert, sondern sie sind - paarweise zu einem Vermittlungsmodul zusammengefaßt - mehrfach vorhanden.

Außerdem sind natürlich, wie in der im Abschnitt 3.1 besprochenen zentralen Lösung, Leitungsinterfaces vorhanden, die im Prinzip dieselben Funktionen wie dort ausüben. Alle Funktionsmoduln sind wiederum über einen Bus verbunden.

4.1.2 Möglichkeiten der Verkehrsaufteilung

Ein entscheidender Punkt bei einem solchen Multiprozessor-System ist die Verteilung der Aufgaben auf die mehrfach vorhandenen Einheiten. Je sauberer eine Trennung der Zuständigkeiten vorgenommen wird, umso geringer wird der Koordinierungsaufwand zwischen den einzelnen Vermittlungsmoduln.

Bei der in [46] entwickelten Lösung wird die Zuständigkeit der Moduln in der folgenden Weise geregelt (vgl. Bild 4.2):

Ein ankommendes Paket wird aufgrund der Leitung, auf der es ankommt, in einem ganz bestimmten Modul Nr. m zwischengespeichert. Dieser Speichermodul wird ausschließlich von dem bei ihm angebrachten Prozessor Nr. m verwaltet. Nur im Falle einer momentanen Speicher- oder Prozessorüberlast ist ein Überlaufen des Pakets auf einen anderen Modul möglich. Diese Möglichkeit wird im weiteren hier nicht betrachtet; ihre verkehrsmäßigen Konsequenzen wurden in [10] untersucht.

Somit ist eine einfache Zuweisung der Pakete auf die einzelnen Moduln des Vermittlungsknotens aufgrund ihrer Herkunft erreicht. Allerdings kann trotzdem in abgehender Richtung ein Konflikt auftreten, wenn mehr als ein Prozessor für eines "seiner" Pakete eine abgehende Leitung belegen möchte. Um diesen Konflikt möglichst aufwandsarm zu beseitigen, wird in [46] vorgesehen, daß jede abgehende Leitung von einem bestimmten Prozessor verwaltet wird. Dieser nimmt die sogenannte Dispatch-Funktion wahr, das heißt er organisiert die geordnete Übertragung der Pakete aller Moduln, welche eine bestimmte abgehende Leitung benutzen wollen.

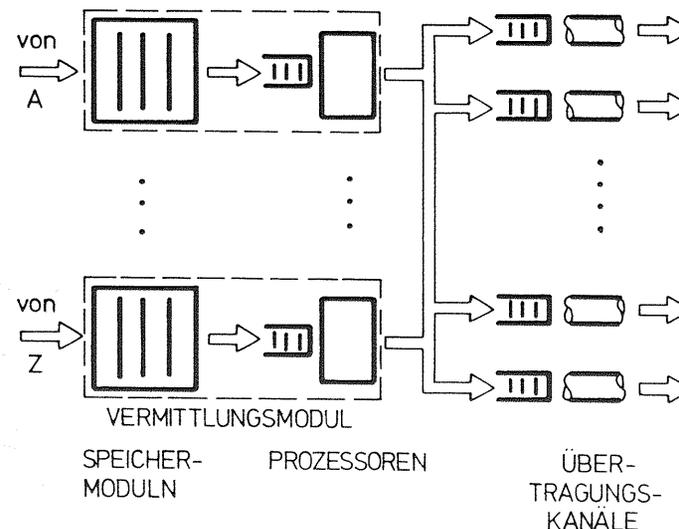


Bild 4.2: Verkehrsfluß durch einen Vermittlungsknoten mit mehreren Vermittlungsmoduln und herkunftsorientierter Verkehrsaufteilung auf die Moduln

Im Gegensatz zu dieser Lösung soll jetzt noch eine weitere Möglichkeit der Verkehrsaufteilung auf die Moduln dargestellt werden (vgl. Bild 4.3). Diese Lösung mit zielorientierter Aufteilung des Verkehrs auf die Vermittlungsmoduln erfordert, sowohl Speicherplatz als auch Intelligenz aus den Vermittlungsmoduln abziehen und dafür die Funktion der Leitungsinterfaces wesentlich zu erweitern. Damit soll erreicht werden, daß für eine bestimmte Gruppe abgehender Leitungen ausschließlich ein einziger Vermittlungsmodul zuständig gemacht wird. Da somit ein Paket, welches in eine bestimmte Richtung weiter geschickt werden soll, in einem ganz bestimmten Speichermodul zwischengespeichert werden muß, folgt zwangsläufig, daß die Vermittlung der Pakete - zumindest grob - bereits vor der eigentlichen Speicherung vorgenommen werden muß. Diese Aufgabe wird von den Prozessormoduln abgezogen

und den Leitungsinterfaces zugewiesen. Diese neuen Einheiten müssen zur kurzzeitigen Speicherung der Pakete eine gewisse Anzahl eigener Speicherplätze besitzen. Die weitaus größere Speicherlast - verursacht durch das Warten auf eine abgehende Leitung und die Quittierung durch den Nachbarknoten - tragen jedoch nach wie vor die dafür vorgesehenen Speichermoduln.

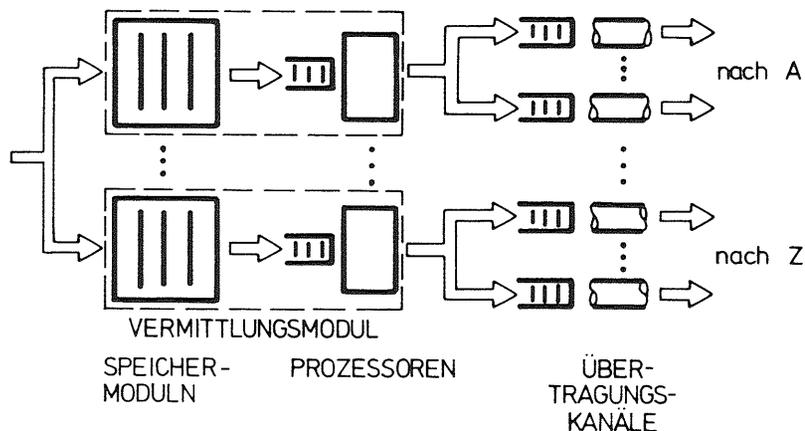


Bild 4.3: Verkehrsfluß durch einen Vermittlungsknoten mit mehreren Vermittlungsmoduln und zielorientierter Verkehrsaufteilung auf die Moduln

Ein Vorteil dieser Lösung mit zielorientierter Aufteilung des Verkehrs auf die Vermittlungsmoduln gegenüber der zuvor beschriebenen herkunftsorientierten Aufteilung besteht darin, daß die Koordinierungsfunktion des Dispatchers überflüssig wird. Als Nachteil ist die notwendige Aufteilung der Paketspeicherplätze auf die Speichermoduln und die Leitungsinterfaces anzusehen.

Die verkehrsmäßigen Eigenschaften beider Lösungen werden im folgenden untersucht. Diese sind im übrigen nicht nur bei den beschriebenen hardwaremäßigen Voraussetzungen der Speichermodularisierung von Interesse. Auch bei Vermittlungsknoten mit

einem einzigen, gemeinsamen Paketspeicher können folgende Gründe dafür sprechen, den Speicher logisch in mehrere Teile zu zerlegen:

1. Durch eine sinnvolle Klasseneinteilung der Pakete, abhängig von ihrem bereits im Netz zurückgelegten Weg, kann erreicht werden, daß Pakete, welche ihrem Ziel schon nahe sind, mit hoher Wahrscheinlichkeit einen Speicherplatz erhalten. Durch einen solchen Mechanismus lassen sich sogenannte "Puffer-Deadlocks" verhindern [35,79,60].
2. Eine Aufteilung des Speichers kann verhindern, daß einzelne Verkehrsbeziehungen in einem Vermittlungsknoten so dominieren, daß alle anderen Beziehungen massiv benachteiligt werden [40, 60]. (Beispiele hierfür siehe Abschnitt 3.5 und 4.2.3.)

Im folgenden wird zuerst auf die zielorientierte und danach auf die herkunftsorientierte Verkehrsaufteilung eingegangen.

4.2 Zielorientierte Verkehrsaufteilung

4.2.1 Verkehrsmodell

Wir gehen davon aus, daß die Zuteilung der ankommenden Pakete auf die vorhandenen Speichermoduln so vorgenommen wird, daß jeweils nur Pakete, die danach über bestimmte abgehende Leitungen übertragen werden, in einem speziellen Speichermodul zwischengespeichert werden (vgl. Bild 4.4).

Im Bild 4.4 werden beispielweise im Vermittlungsmodul Nr. 1 alle Pakete zwischengespeichert, welche später über die Kanäle mit den Richtungsnummern (1;1) bis (1;R₁) weitergesendet werden und entsprechend im Modul Nr. M alle Pakete für die abgehenden Richtungen Nr. (M;1) bis (M;R_M).

Eingetragen im Bild 4.4 sind ferner die Parameter des Modells, die folgendermaßen definiert sind:

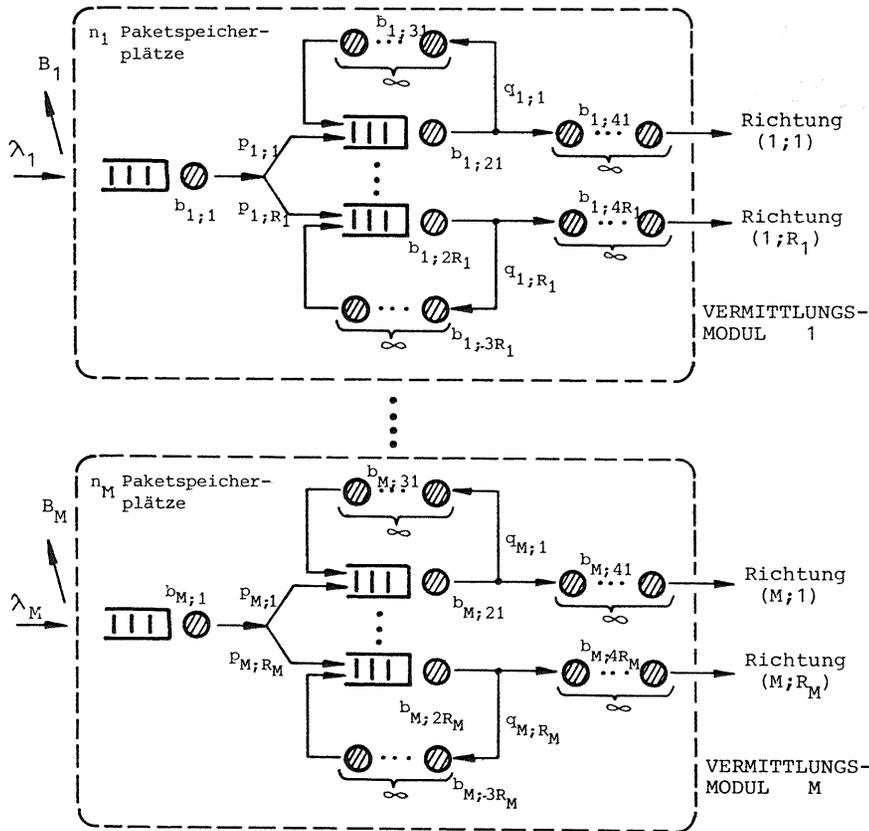


Bild 4.4: Verkehrsmodell eines Vermittlungsknotens mit zielorientierter Verkehrsaufteilung auf M Vermittlungsknoten

- n_m : Anzahl der Paketspeicherplätze im Modul Nr. m
- λ_m : Paketankunftsrate für Modul Nr. m
- $b_{m;1}$: mittlere Belegungszeit des Prozessors im Modul Nr. m
- $b_{m;2r}$: mittlere Belegungszeit des Übertragungskanals Nr. (m;r)
- $b_{m;3r}$: mittlere Verzögerungszeit der negativen Quittungen für Übertragungskanal Nr. (m;r)
- $b_{m;4r}$: mittlere Verzögerungszeit der positiven Quittungen für Übertragungskanal Nr. (m;r)
- $p_{m;r}$: Wahrscheinlichkeit, daß ein Paket des Moduls Nr. m über den Übertragungskanal Nr. (m;r) übertragen wird
- $q_{m;r}$: Wiederholungswahrscheinlichkeit bezüglich Übertragungskanal Nr. (m;r)
- R_m : Anzahl der abgehenden Übertragungskanäle, welche dem Modul Nr. m zugeordnet sind.

4.2.2 Analyse

Wie man bereits aus Bild 4.4 ersieht, spaltet sich das Modell des gesamten Vermittlungsknotens in Teilmodelle auf. Diese umfassen jeweils einen Vermittlungsmodul (Prozessor und Paketspeicher), sowie die diesem Modul zugeordneten abgehenden Übertragungskanäle. Die Analyse des Verkehrs ist somit völlig identisch mit der Analyse des Modells mit nur einem Vermittlungsmodul im Abschnitt 3.

4.2.3 Ergebnisse

In diesem Abschnitt werden numerische Ergebnisse für das Verkehrsmodell eines Vermittlungsknotens mit M Vermittlungsmodulen und zielorientierter Verkehrsaufteilung diskutiert. Gezeigt wird in erster Linie der verkehrsmäßige Einfluß der Modularisierung. Den Ergebnissen liegt ein Modell gemäß Bild 4.4 zugrunde. Es wurde in

den gezeigten Beispielen angenommen, daß der Vermittlungsknoten symmetrisch ausgebaut ist. Das bedeutet, daß die Anzahl der Paketspeicherplätze pro Modul gleich ist, daß alle Prozessoren dieselbe mittlere Belegungszeit besitzen und daß ferner allen Modulen jeweils dieselbe Art und Anzahl von Übertragungskanälen zugeordnet ist.

Analog zu der Untersuchung von Vermittlungsknoten mit einem Vermittlungsmodul im Abschnitt 3.5 wurden gemeinsam für alle Bilder die folgenden Zahlenwerte angenommen:

- 1) Die mittlere Belegungszeit der Prozessoren betrage pro durchgesetztem Paket 5 msec.
- 2) Die mittlere Paketlänge betrage 1000 bit. Bei angenommenen Übertragungsgeschwindigkeiten von 2.4 kbit/sec der Teilnehmeranschlußleitungen bzw. 48 kbit/sec der Verbindungsleitungen ergibt dies mittlere Übertragungszeiten der Pakete von 417msec bzw. 21 msec.
- 3) Die Verzögerungszeiten für die positiven Quittungen seien im Mittel doppelt so lang wie die Übertragungszeiten. Die Verzögerungszeiten für negative Quittungen seien im Mittel dreimal so lang wie die mittleren Übertragungszeiten.
- 4) Für die Wiederholungswahrscheinlichkeiten wird einheitlich ein Wert von 0.01 angenommen.

Variiert werden in den Diagrammen die Gesamtzahl der Paketspeicherplätze im Vermittlungsknoten $n = n_1 + \dots + n_M$, die Anzahl der Modulen M , die gesamte Paketankunftsrate $\lambda = \lambda_1 + \dots + \lambda_M$ sowie die Gesamtzahl der abgehenden Übertragungskanäle $R = R_1 + \dots + R_M$.

Bild 4.5 zeigt am Beispiel eines Vermittlungsknotens mit einem bzw. mit zwei Modulen, wie der maximale Paketdurchsatz durch die Modularisierung gesteigert werden kann. Aufgetragen ist der maximale Gesamtdurchsatz des Vermittlungsknotens D_{max} über der Gesamtzahl der abgehenden Übertragungskanäle R ; Parameter der Kurvenschar ist die Gesamtzahl der Paketspeicherplätze n . Es

ist offensichtlich, daß der Paketdurchsatz mit zunehmender Zahl angeschlossener Leitungen steigt; wie das Bild zeigt, hängt aber das Ausmaß und die Grenze der Durchsatzsteigerung sowohl von der Speicherplatzzahl, als auch insbesondere von der Anzahl der Modulen ab. Zum besseren Verständnis sind für beide Modulzahlen die Grenzen angegeben, welche sich für einen unendlich großen Paketspeicher ergeben. Es ist ersichtlich, daß der maximale Gesamtdurchsatz durch die Summe der Bedienungsraten aller Prozessoren begrenzt ist und daß deshalb - wie zu erwarten - die Modularisierung ein Mittel zur Steigerung des maximalen Durchsatzes darstellt.

Im Bild 4.6 ist der Durchsatz D als Funktion der Paketankunftsrate λ für verschiedene Modulzahlen M dargestellt. Dabei wurde vorausgesetzt, daß unabhängig vom Modularisierungsgrad die Gesamtzahl n aller Paketspeicherplätze im Vermittlungsknoten immer konstant gehalten wird. Wie man sieht, zeigt der hier dargestellte tatsächliche (nicht der maximale !) Durchsatz ein interessantes Verhalten: Bei kleineren und mittleren Ankunfts-raten führt eine Erhöhung des Modularisierungsgrades zu einer Durchsatzminderung, während sich bei extremen Ankunfts-raten genau der umgekehrte Effekt einstellt.

Dieses Verhalten läßt sich folgendermaßen erklären: Bei kleineren Ankunfts-raten ist der auch aus der Fernsprechkverkehrstheorie bekannte Effekt des "Bündelungsgewinns" [63] dafür verantwortlich, daß mit kleinerem Modularisierungsgrad, also größeren Speichern pro Modul, die Speicherüberlauf-Wahrscheinlichkeit abnimmt und deshalb der Durchsatz steigt. Im Überlastfall wirkt sich der Bündelungsgewinn nicht mehr aus; dagegen zieht der folgende Effekt in die entgegengesetzte Richtung. Bei hoher Belastung und großem Paketspeicher ist es häufig der Fall, daß vor einem momentan stark überlasteten Übertragungskanal sehr viele Pakete auf ihre Übertragung warten. Da alle diese Pakete Speicherplatz belegen, kommt es öfters vor, daß momentan ein anderer Übertragungskanal frei steht, weil Pakete, die über diesen Kanal übertragen werden sollten, wegen Speicherplatzmangel abgewiesen werden. Die Aufteilung des Speicherplatzes bei zielorientierter Verkehrsauftei-

lung wirkt nun aber so, daß diese Monopolisierung des Speichers durch überlastete Kanäle vermindert wird. Im Extremfall einer völligen Auftrennung des Speichers für jede Richtung ist sie sogar vollständig ausgeschlossen. Dies erklärt den Effekt des mit zunehmender Modularisierung wachsenden Durchsatzes bei Überlast.

Bild 4.7 zeigt für verschiedene vorgeschriebene Werte B der Speicherüberlauf-Wahrscheinlichkeit die erforderliche Gesamt-speichergröße n als Funktion der Modulzahl M. Die gesamte Paketankunftsrate λ des Vermittlungsknotens ist für die durchgezogenen Kurven konstant gleich 180 Pakete/sec gehalten, für die gestrichelten Kurven konstant gleich 360 Pakete/sec. Man erkennt wiederum den beschriebenen Effekt des Bündelungsgewinns, daß mit kleinerer Modulzahl, also größerem Paketspeicher pro Modul, die Speicherüberlauf-Wahrscheinlichkeit abnimmt.

Die Kurven für 360 Pakete/sec sind bei $M = 2$ abgebrochen, da sich bei einem Modul keiner der drei vorgegebenen Werte für die Speicherüberlauf-Wahrscheinlichkeit erreichen läßt. Der Grund hierfür ist, daß bei $M = 1$ das Angebot an den einen Prozessor $360/\text{sec} \times 5 \text{ msec} = 1,8 \text{ Erl.}$ betragen würde. Aus diesem Beispiel wird nochmals deutlich, daß bei höheren Paketankunftsraten und fester Prozessorgeschwindigkeit eine Verkehrsaufteilung auf mehrere Vermittlungsmoduln unumgänglich ist. Allerdings zeigt das Bild auch, daß die Vorteile der Modularisierung mit einem Mehraufwand an Speicherplatz erkauft werden müssen.

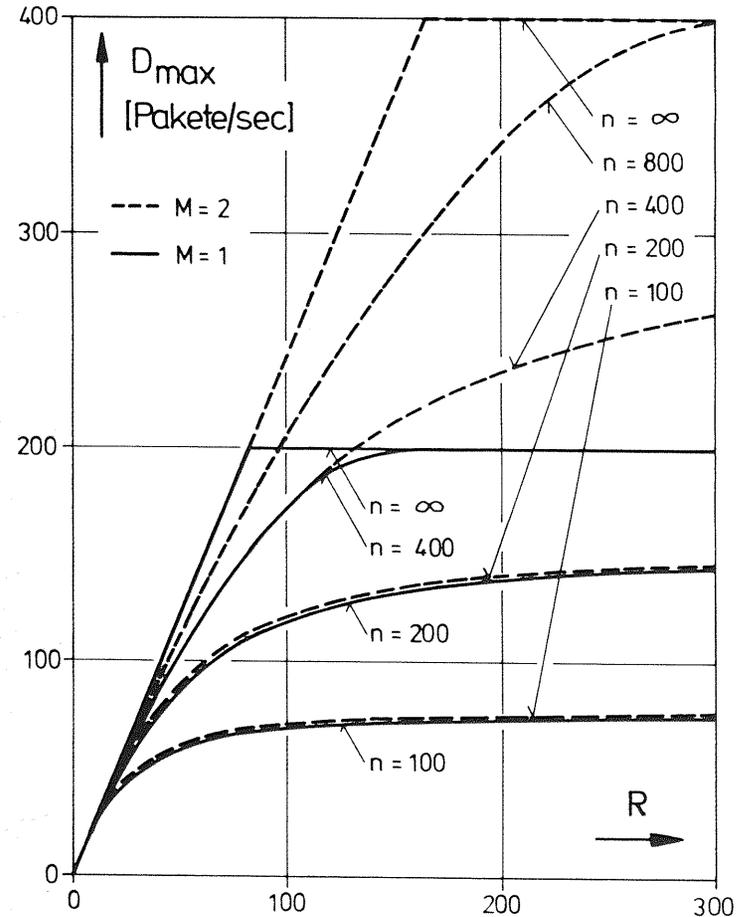


Bild 4.5: Maximaler Gesamtdurchsatz des Vermittlungsknotens D_{\max} als Funktion der Gesamtzahl abgehender Übertragungskanäle R für $M = 1$ und 2 Moduln und verschiedene Gesamt-speichergrößen n. Exponentialverteilte Übertragungs- und Prozessorbelegungszeiten. Zielorientierte Verkehrsaufteilung auf die Vermittlungsmoduln.

$R_m = R/M$ abgehende Übertragungskanäle pro Vermittlungsmodul
 $n_m = n/M$ Paketspeicherplätze pro Vermittlungsmodul
 mittlere Prozessorbelegungszeiten: $b_{m;1} = 5 \text{ msec}$
 mittlere Übertragungszeiten pro Paket: $b_{m;21} = \dots = b_{m;2R_m} = 417 \text{ msec}$
 mittlere Negativ-Quittierzeiten: $b_{m;31} = \dots = b_{m;3R_m} = 1250 \text{ msec}$
 mittlere Positiv-Quittierzeiten: $b_{m;41} = \dots = b_{m;4R_m} = 833 \text{ msec}$
 Verzweigungswahrscheinlichkeiten: $P_{m;1} = \dots = P_{m;R_m} = 1/R_m$
 Wiederholungswahrscheinlichkeiten: $q_{m;1} = \dots = q_{m;R_m} = 0.01$

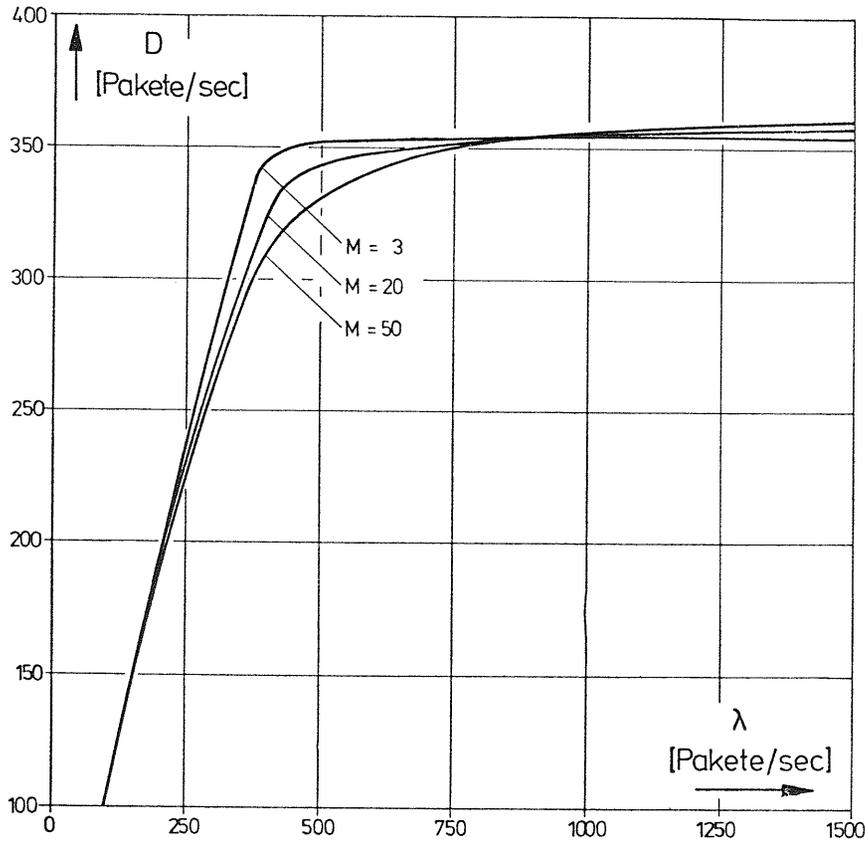


Bild 4.6: Gesamtdurchsatz D des Vermittlungsknotens als Funktion der gesamten Paketankunftsrate λ für verschiedene Modulzahlen M . Exponentialverteilte Übertragungs- und Prozessorbelegungszeiten. Zielorientierte Verkehrsaufteilung auf die Vermittlungsmoduln.

$R = 300$ abgehende Übertragungskanäle insgesamt
 $R_m = 300/M$ abgehende Übertragungskanäle pro Vermittlungsmodul
 $n = 600$ Paketspeicher im gesamten Vermittlungsmodul
 $n_m = 600/M$ Paketspeicherplätze pro Vermittlungsmodul
 Paketankunftsrate pro Vermittlungsmodul: $\lambda_m = \lambda/M$
 mittlere Prozessorbelegungszeiten: $b_{m;1} = 5$ msec
 mittlere Übertragungszeiten pro Paket: $b_{m;21} = \dots = b_{m;2R_m} = 417$ msec
 mittlere Negativ-Quittierzeiten: $b_{m;31} = \dots = b_{m;3R_m} = 1250$ msec
 mittlere Positiv-Quittierzeiten: $b_{m;41} = \dots = b_{m;4R_m} = 833$ msec
 Verzweigungswahrscheinlichkeiten: $P_{m;1} = \dots = P_{m;R_m} = 1/R_m$
 Wiederholungswahrscheinlichkeiten: $q_{m;1} = \dots = q_{m;R_m} = 0.01$

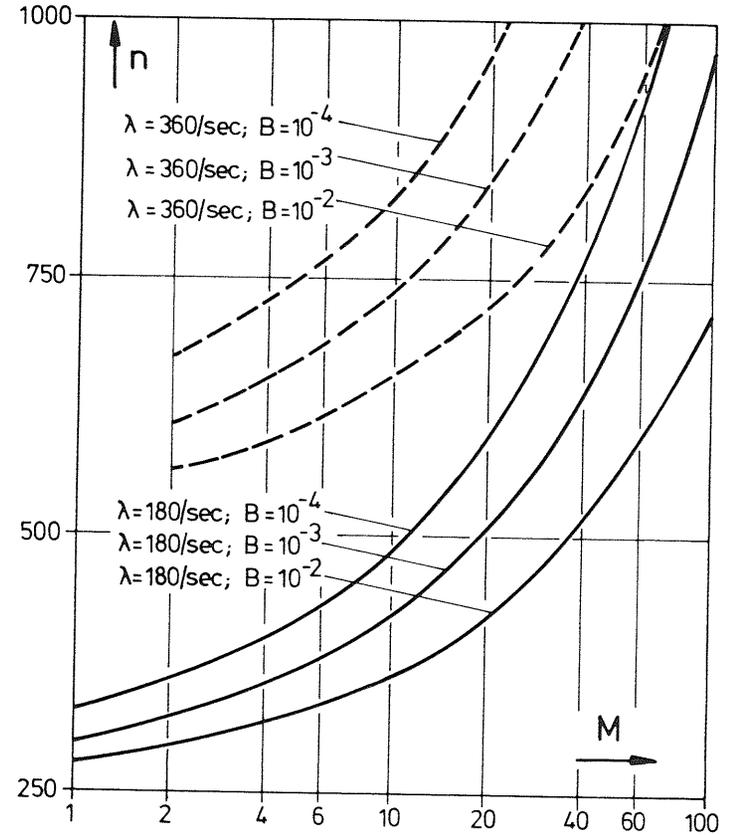


Bild 4.7: Erforderliche Gesamtspeichergröße n als Funktion der Modulzahl M für vorgeschriebene Werte der Speicherüberlauf-Wahrscheinlichkeit B und zwei Werte der gesamten Paketankunftsrate, $\lambda = 180$ /sec und 360 /sec. Exponentialverteilte Übertragungs- und Prozessorbelegungszeiten. Zielorientierte Verkehrsaufteilung auf die Vermittlungsmoduln.

$R = 300$ abgehende Übertragungskanäle insgesamt
 $R_m = 300/M$ abgehende Übertragungskanäle pro Vermittlungsmodul
 $n_m = n/M$ Paketspeicherplätze pro Vermittlungsmodul
 Paketankunftsrate pro Vermittlungsmodul: $\lambda_m = \lambda/M$
 mittlere Prozessorbelegungszeiten: $b_{m;1} = 5$ msec
 mittlere Übertragungszeiten pro Paket: $b_{m;21} = \dots = b_{m;2R_m} = 417$ msec
 mittlere Negativ-Quittierzeiten: $b_{m;31} = \dots = b_{m;3R_m} = 1250$ msec
 mittlere Positiv-Quittierzeiten: $b_{m;41} = \dots = b_{m;4R_m} = 833$ msec
 Verzweigungswahrscheinlichkeiten: $P_{m;1} = \dots = P_{m;R_m} = 1/R_m$
 Wiederholungswahrscheinlichkeiten: $q_{m;1} = \dots = q_{m;R_m} = 0.01$

4.3 Herkunftsorientierte Verkehrsaufteilung

4.3.1 Verkehrsmodell

Alle im betrachteten Vermittlungsknoten ankommenden Pakete seien nun aufgrund der Leitung, auf der sie ankommen, in M verschiedene Klassen eingeteilt (siehe Bild 4.8). Pakete einer Klasse m werden dann und nur dann abgewiesen, wenn alle Speicherplätze des Vermittlungsmoduls Nr. m belegt sind.

Die Pakete der Klasse m werden vom Prozessor Nr. $(m;1)$ bearbeitet. Auf den Übertragungskanälen werden die Pakete aller Klassen gleich behandelt und streng in der Reihenfolge ihres Eintreffens übertragen (Abfertigungsdisziplin "first-come, first-served").

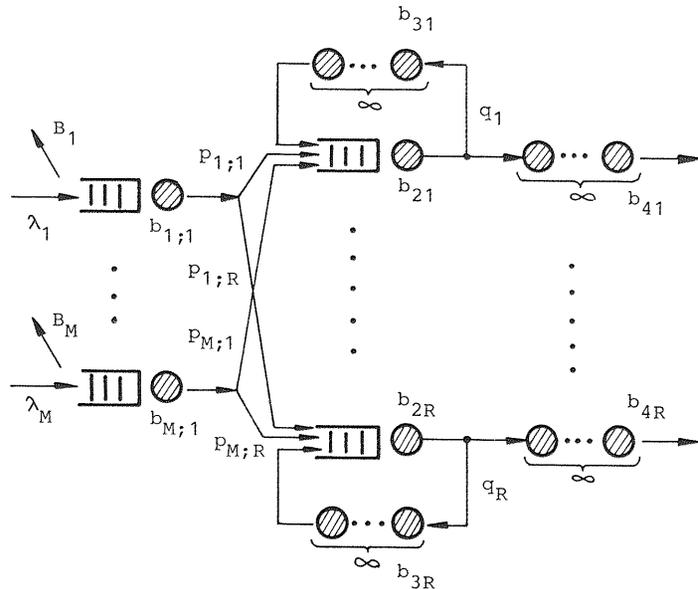


Bild 4.8: Verkehrsmodell eines Vermittlungsknotens mit herkunftsorientierter Verkehrsaufteilung auf M Vermittlungsmoduln

4.3.2 Exakte Analyse bei exponentialverteilten Übertragungs- und Prozessorbelegungszeiten

Das im vorigen Abschnitt beschriebene Verkehrsmodell stellt wiederum ein Warteschlangennetz dar. Da die Pakete hier in M verschiedenen Speichermoduln gespeichert werden, müssen wir die korrespondierenden Anforderungen im Warteschlangennetz dementsprechend unterscheiden. Das bedeutet, daß wir alle ankommenden Anforderungen in M verschiedene Klassen einteilen und diese Klassenzugehörigkeit für jede Anforderung während ihres Wegs durch das Netz beachten.

Im Sinne einer möglichst universellen Analyse wird zugelassen, daß die Anforderungen jeder Klasse sich beliebig auf alle R abgehenden Übertragungskanäle verzweigen können.

Unter der Annahme eines Poisson-Ankunftsprozesses jeder Klasse werden im folgenden verschiedene analytische Untersuchungen beschrieben. Zunächst wird in diesem Abschnitt unter der Voraussetzung exponentialverteilter Übertragungs- und Prozessorbelegungszeiten eine exakte Analyse des Problems durchgeführt. Da es sich herausstellt, daß sich diese Lösung für große Systeme nur schwer numerisch auswerten läßt - obwohl sie explizit vorliegt - , wird danach im Abschnitt 4.3.3 unter denselben Voraussetzungen eine Näherungslösung abgeleitet, welche diesen Nachteil vermeidet. Dieser Näherungsansatz dient schließlich außerdem als Basis einer weiteren approximativen Lösung für die Behandlung allgemein verteilter Übertragungs- und Prozessorbelegungszeiten (siehe Abschnitt 4.3.4).

Wie im Fall eines Vermittlungsmoduls (Abschnitt 3) führt auch hier der Weg zur Bestimmung der charakteristischen Verkehrsgrößen über die Berechnung der Zustandswahrscheinlichkeiten des Systems.

4.3.2.1 Zustandswahrscheinlichkeiten

Für die Analyse des Warteschlangennetzes gemäß Bild 4.8 kann unter der Annahme exponentialverteilter Übertragungs- und Prozessorbelegungszeiten, aber allgemein verteilter Belegungszeiten der Verzögerungsstationen, wiederum das im Anhang 1 beschriebene Theorem herangezogen werden, allerdings jetzt unter Berücksichtigung mehrerer Anforderungsklassen.

Zur Beschreibung des Momentanzustandes des Warteschlangennetzes werden wiederum die Anzahlen von Anforderungen benützt, welche sich gleichzeitig in den verschiedenen Stationen des Netzes gemäß Bild 4.8 befinden. Hierbei muß allerdings - im Unterschied zu Abschnitt 3 - die Klassenzugehörigkeit der Anforderungen beachtet werden. Wir bezeichnen deshalb mit $X_{m;1m}, X_{m;21}, \dots, X_{m;2R}, X_{m;31}, \dots, X_{m;3R}, X_{m;41}, \dots, X_{m;4R}$ die zufällige Anzahl vorhandener Anforderungen in den jeweiligen Bedienstationen Nr. $(m;1), (2,1), \dots, (2,R), (3,1), \dots, (3,R), (4,1), \dots, (4,R)$ gemäß Bild 4.8 und mit den zugehörigen kleinen Buchstaben $x_{m;1m}, x_{m;21}, \dots, x_{m;4R}$ die entsprechenden Realisierungen dieser Zufallsvariablen.

Die Abweisung ankommender Pakete bei vollem Speichermodul Nr. m läßt sich wie früher als zustandsabhängige Ankunftsrate der Anforderungsklasse Nr. m beschreiben:

$$\lambda_m^* = \lambda_m^*(\hat{x}_m) = \begin{cases} \lambda_m & \text{für } \hat{x}_m = x_{m;1,m} + \sum_{s=2}^4 \sum_{r=1}^R x_{m;s,r} < n_m \\ 0 & \text{sonst} \end{cases} \quad (4.1)$$

$$m \in \{1, 2, \dots, M\}$$

Nach dem im Anhang 1 beschriebenen Theorem kann die Lösung für die stationären Zustandswahrscheinlichkeiten in der folgenden Form dargestellt werden:

$$\begin{aligned} & P(X_{1;11} = x_{1;11}, \dots, X_{M;1M} = x_{M;1M}, X_{1;21} = x_{1;21}, \dots, X_{1;2R} = x_{1;2R}, X_{1;31} = x_{1;31}, \dots \\ & \dots X_{1;3R} = x_{1;3R}, X_{1;41} = x_{1;41}, \dots, X_{1;4R} = x_{1;4R}, \dots \\ & \dots X_{M;21} = x_{M;21}, \dots, X_{M;2R} = x_{M;2R}, X_{M;31} = x_{M;31}, \dots, X_{M;3R} = x_{M;3R}, \\ & \dots X_{M;41} = x_{M;41}, \dots, X_{M;4R} = x_{M;4R}) \\ & = p'(x_{1;11}, \dots, x_{M;4R}) \\ & = p'(\vec{x}) \\ & = d'(\vec{x}) g'_{11}(x_{1;11}) \dots g'_{M;1M}(x_{M;1M}) \cdot g'_{21}(x_{1;21}, \dots, x_{M;21}) \dots \\ & \cdot g'_{2R}(x_{1;2R}, \dots, x_{M;2R}) \cdot g'_{31}(x_{1;31}, \dots, x_{M;31}) \dots \cdot g'_{3R}(x_{1;3R}, \dots, x_{M;3R}) \cdot \\ & \cdot g'_{41}(x_{1;41}, \dots, x_{M;41}) \dots \cdot g'_{4R}(x_{1;4R}, \dots, x_{M;4R}) \cdot C' \quad (4.2) \end{aligned}$$

Im folgenden werden die einzelnen Faktoren dieser Lösung für unser Warteschlangennetz bestimmt.

Die Größe $d'(\vec{x})$ in Gl. (4.2) ist abhängig von den Ankunftsraten $\lambda_1, \dots, \lambda_M$ und dem Zustandsvektor \vec{x}' . Sie ist definiert in Gl. (A1.6) im Anhang 1 und ergibt sich wegen der Beziehung (4.1) zu:

$$\begin{aligned} d'(\vec{x}) &= d'(x_{1;1,1}, \dots, x_{M;4,R}) \\ &= \prod_{m=1}^M \prod_{\xi=0}^{\hat{x}_m-1} \lambda_m^*(\xi_m) \\ &= \prod_{m=1}^M \lambda_m^{\hat{x}_m} \quad (4.3a) \end{aligned}$$

$$\hat{x}_m = x_{m;1,m} + \sum_{s=2}^4 \sum_{r=1}^R x_{m;s,r}$$

Die Faktoren $g_{1m}^i, \dots, g_{4R}^i$ in Gl. (4.2) stehen jeweils für eine Station des Warteschlangennetzes. Sie hängen ab von der Anzahl vorhandener Anforderungen aller Klassen in einer Station, vom Typ der Station, ihrer mittleren Bedienungszeit, sowie von der Anzahl der Durchläufe der Anforderungen jeder Klasse durch die Station. Wendet man die Gleichungen (A1.4) und (A1.5) im Anhang 1 auf unser Warteschlangennetz an, so ergibt sich:

$$\begin{aligned}
 g_{1m}^i(x_{m;1m}) &= b_{1m}^{x_{m;1m}} & m \in \{1, \dots, M\} \\
 g_{2r}^i(x_{1;2r}, \dots, x_{M;2r}) &= \binom{x_{1;2r} + \dots + x_{M;2r}}{x_{1;2r}, \dots, x_{M;2r}} \prod_{m=1}^M \left(\frac{p_{m;r}}{1-q_r} b_{2r} \right)^{x_{m;2r}} & r \in \{1, \dots, R\} \\
 g_{3r}^i(x_{1;3r}, \dots, x_{M;3r}) &= \prod_{m=1}^M \left(\frac{p_{m;r} q_r}{1-q_r} b_{3r} \right)^{x_{m;3r}} \frac{1}{x_{m;3r}!} & r \in \{1, \dots, R\} \\
 g_{4r}^i(x_{1;4r}, \dots, x_{M;4r}) &= \prod_{m=1}^M (p_{m;r} b_{4r})^{x_{m;4r}} \frac{1}{x_{m;4r}!} & r \in \{1, \dots, R\}
 \end{aligned} \tag{4.3b}$$

Der Faktor C^i in Gl. (4.2) ist die Normierungskonstante, die so zu bestimmen ist, daß die Summe aller Zustandswahrscheinlichkeiten gleich 1 ist.

Hiermit läßt sich die Gesamtlösung für die stationären Zustandswahrscheinlichkeiten folgendermaßen angeben:

$$\begin{aligned}
 p^i(\vec{x}^i) &= p^i(x_{1;1m}, \dots, x_{M;4R}) = \\
 &= C^i \prod_{m=1}^M A_{m;1m}^{x_{m;1m}} \prod_{r=1}^R \left\{ \binom{x_{1;2r} + \dots + x_{M;2r}}{x_{1;2r}, \dots, x_{M;2r}} \prod_{m=1}^M A_{m;2r}^{x_{m;2r}} \right\} \prod_{m=1}^M \prod_{s=3}^4 \prod_{r=1}^R \frac{A_{m;sr}^{x_{m;sr}}}{x_{m;sr}!} \\
 \sum_{\vec{x}^i \in Z^i} p^i(\vec{x}^i) &= 1 \\
 Z^i &= \left\{ \vec{x}^i \mid x_{1;1m}, \dots, x_{M;4R} \in \mathbb{N}_0 \wedge x_{m;1m} + \sum_{s=2}^4 \sum_{r=1}^R x_{m;sr} \leq n_m, m \in \{1, \dots, M\} \right\}
 \end{aligned} \tag{4.4}$$

$$\left. \begin{aligned}
 A_{m;1m} &= \lambda_m b_{1m} \\
 A_{m;2r} &= \frac{p_{m;r}}{1-q_r} \lambda_m b_{2r} \\
 A_{m;3r} &= \frac{p_{m;r} q_r}{1-q_r} \lambda_m b_{3r} \\
 A_{m;4r} &= p_{m;r} \lambda_m b_{4r}
 \end{aligned} \right\} r \in \{1, \dots, R\} \left. \vphantom{\begin{aligned} A_{m;1m} \\ A_{m;2r} \\ A_{m;3r} \\ A_{m;4r} \end{aligned}} \right\} m \in \{1, \dots, M\}$$

4.3.2.2 Charakteristische Verkehrsgrößen

Die Verkehrsgrößen lassen sich - ähnlich wie im Falle eines einzigen Vermittlungsmoduls - aus den Zustandswahrscheinlichkeiten bestimmen (vgl. Abschnitt 3.3.2). Numerische Ergebnisse für die hier abgeleiteten Verkehrsgrößen werden im Abschnitt 4.3.5 gezeigt.

Speicherüberlauf-Wahrscheinlichkeiten

Wegen der Poisson-Ankunftsprozesse an den einzelnen Vermittlungsmoduln berechnen sich die Speicherüberlauf-Wahrscheinlichkeiten folgendermaßen:

$$\begin{aligned}
 B_m &= P(X_{m;1m} + \sum_{s=2}^4 \sum_{r=1}^R X_{m;sr} = n_m) \\
 &= \sum_{\vec{x}^i \in \Psi_m(n_m)} p^i(\vec{x}^i) & m \in \{1, \dots, M\}
 \end{aligned} \tag{4.5}$$

$$\Psi_m(\hat{x}_m) = \left\{ \vec{x}^i \mid \vec{x}^i \in Z^i \wedge x_{m;1m} + \sum_{s=2}^4 \sum_{r=1}^R x_{m;sr} = \hat{x}_m \right\}$$

Mittlere Abstände zwischen aufeinanderfolgenden Blockierzuständen und mittlere Dauern der Blockierzustände

Der mittlere Abstand zwischen aufeinanderfolgenden Blockierzuständen des Moduls Nr. m berechnet sich analog zu Gl. (3.12) zu:

$$t_{\lambda,m} = \frac{1 - B_m}{P(X_{m,1m} + \sum_{s=2}^4 \sum_{r=1}^R X_{m,sr} = n_m - 1)} \cdot \frac{1}{\lambda_m}$$

$$= \frac{1 - B_m}{\sum_{\vec{x}' \in \Psi_m(n_m - 1)} p'(\vec{x}') \lambda_m} \quad m \in \{1, \dots, M\} \quad (4.6)$$

Die mittlere Dauer eines Blockierzustandes des Speichermoduls Nr. m ergibt sich analog zu Gl. (3.10) zu:

$$t_{B,m} = \frac{B_m}{P(X_{m,1m} + \sum_{s=2}^4 \sum_{r=1}^R X_{m,sr} = n_m - 1)} \cdot \frac{1}{\lambda_m}$$

$$= \frac{B_m}{\sum_{\vec{x}' \in \Psi_m(n_m - 1)} p'(\vec{x}') \lambda_m} \quad (4.7)$$

Speichermodulbelastungen

Die Belastung der Speichermoduln kann analog zu Gl. (3.13) berechnet werden:

$$Y_{SP,m} = \sum_{\hat{x}_m=0}^{n_m} \hat{x}_m P(X_{m,1m} + \sum_{s=2}^4 \sum_{r=1}^R X_{m,sr} = \hat{x}_m)$$

$$= \sum_{\hat{x}_m=0}^{n_m} \hat{x}_m \sum_{\vec{x}' \in \Psi_m(\hat{x}_m)} p'(\vec{x}') \quad (4.8)$$

Mittlere Speicherverweilzeiten

Analog zu Gl. (3.14) ergibt sich:

$$t_{SP,m} = \frac{Y_{SP,m}}{\lambda_m(1 - B_m)} \quad (4.9)$$

Durchsätze

Für die pro Zeiteinheit durch den Vermittlungsmodul Nr. m durchgesetzte Zahl von Paketen gilt:

$$D_m = \lambda_m(1 - B_m) \quad (4.10)$$

Mittlere Warteschlangenlängen und mittlere Wartezeiten vor den Prozessoren

Analog zu Gl. (3.16) gilt:

$$\Omega_{m,1m} = \sum_{\xi=1}^{n_m-1} \xi P(X_{m,1m} = \xi + 1)$$

$$= \sum_{\xi=1}^{n_m-1} \xi \sum_{\vec{x}' \in \Phi_{1m}(\xi+1)} p(\vec{x}') \quad (4.11)$$

$$\Phi_{1m}(\xi) = \{ \vec{x}' \mid \vec{x}' \in Z' \wedge x_{m,1m} = \xi \}$$

$$W_{1m} = \frac{\Omega_{1m}}{\lambda_m(1 - B_m)} \quad (4.12)$$

Belastung der Prozessoren

Es gilt:

$$Y_{m,1m} = \lambda_m(1 - B_m) b_{1m} \quad (4.13)$$

Mittlere Warteschlangenlängen und mittlere Wartezeiten vor den Übertragungskanälen

Für die Praxis ist wegen der für alle Vermittlungsmoduln einheitlichen Paketlängenverteilung und der strengen First-Come-First-Served Abfertigungsdisziplin nur die mittlere Gesamtwarteschlangenlänge vor jedem Übertragungskanal interessant. Sie ergibt sich zu:

$$\Omega_{2r} = \sum_{x=1}^{n_1+\dots+n_m-1} x \cdot P\left(\sum_{m=1}^M X_{m,2r} = x+1\right) = \sum_{x=1}^{n_1+\dots+n_m-1} x \cdot \sum_{\vec{x}' \in \Phi_{2r}(x+1)} p(\vec{x}') \quad r \in \{1, \dots, R\} \quad (4.14)$$

$$\Phi_{2r}(\vec{\xi}) = \left\{ \vec{x}' \mid \vec{x}' \in Z^1 \wedge \sum_{m=1}^M x_{m,2r} = \vec{\xi} \right\}$$

Damit erhält man die mittlere Wartezeit vor den Übertragungskanälen für alle Klassen einheitlich zu:

$$w_{2r} = \Omega_{2r} \left(\sum_{m=1}^M \lambda_m (1-B_m) \frac{\rho_{m,r}}{1-q_r} \right)^{-1} \quad r \in \{1, \dots, R\} \quad (4.15)$$

Belastung der Übertragungskanäle

Die Belastung eines Übertragungskanals errechnet sich aus der Gesamtrate von Paketen, die ihm angeboten wird und seiner mittleren Belegungsdauer:

$$Y_{2r} = \sum_{m=1}^M \lambda_m (1-B_m) \frac{\rho_{m,r}}{1-q_r} b_{2r} \quad r \in \{1, \dots, R\} \quad (4.16)$$

4.3.2.3 Numerische Auswertung

Wie bereits bei der Lösung für Vermittlungsknoten mit einem Modul, so gilt auch hier, daß sich die numerische Bestimmung der Normierungskonstanten und der Verkehrsgrößen durch die einzelne Berechnung der Zustandswahrscheinlichkeiten verbietet, da die Anzahl der Zustände hier sogar noch um Größenordnungen höher ist (vgl. Abschnitt 3.3.3).

Für die numerische Analyse von geschlossenen Warteschlangennetzen mit mehreren Anforderungsklassen ¹⁾ wurden sehr effektive Algorithmen entwickelt und realisiert [81,7]. Diese Verfahren können, wie im folgenden gezeigt wird, auch für die numerische Auswertung unserer Lösung eingesetzt werden.

Um die nachfolgende Darstellung zu vereinfachen, werden - wie im Abschnitt 3.3.3.2 - die Stationen fortlaufend numeriert und folgende Abkürzungen eingeführt:

$$U = 3R + M$$

$$f_u(y_u) = A_{u,1u}^{y_{u,u}} \quad u \in \{1, \dots, M\}$$

$$f_u(y_u) = \left(\frac{y_{1,u} + \dots + y_{M,u}}{y_{1,u} \dots y_{M,u}} \right) A_{1,2,u-M}^{y_{1,u}} \dots A_{M,2,u-M}^{y_{M,u}} \quad u \in \{M+1, \dots, M+R\}$$

$$f_u(y_u) = \prod_{m=1}^M \frac{A_{m,3,u-M-R}^{y_{m,u}}}{y_{m,u}!} \quad u \in \{M+R+1, \dots, M+2R\} \quad (4.17)$$

$$f_u(y_u) = \prod_{m=1}^M \frac{A_{m,4,u-M-2R}^{y_{m,u}}}{y_{m,u}!} \quad u \in \{M+2R+1, \dots, M+3R\}$$

¹⁾ Genauer müßte hier - dem allgemeinen Sprachgebrauch folgend - von "subchains" gesprochen werden. Bei der für unsere Anwendungen gültigen Einschränkung, daß keine Klassenwechsel vorkommen, können jedoch die Bezeichnungen "Anforderungsklasse" und "subchain" synonym verwendet werden.

Mit diesen Abkürzungen läßt sich die Normierungskonstante C' in Gl. (4.4) in der folgenden Form darstellen:

$$C' = \left(\sum_{\vec{y} \in Z_U(n_1, \dots, n_M)} \prod_{u=1}^U f_u(y_u) \right)^{-1}$$

$$\vec{y} = (y_1, y_2, \dots, y_U)$$

$$y_u = (y_{1,u}, y_{2,u}, \dots, y_{M,u})$$
(4.18)

$$Z_U(n_1, \dots, n_M) = \{ \vec{y} \mid y_1, \dots, y_U \in \mathbb{N}_0^M \wedge \sum_{u=1}^U y_{m,u} \leq n_m \text{ für } m \in \{1, \dots, M\} \}$$

Mit Hilfe der Definition

$$D_U(i_1, i_2, \dots, i_M) = \{ \vec{y} \mid y_1, \dots, y_U \in \mathbb{N}_0^M \wedge \sum_{u=1}^U y_{m,u} = i_m \text{ für } m \in \{1, \dots, M\} \}$$
(4.19)

läßt sich C' zerlegen in:

$$C' = \left(\sum_{i_1=0}^{n_1} \dots \sum_{i_M=0}^{n_M} G_U(i_1, \dots, i_M) \right)^{-1}$$
(4.20)

mit

$$G_U(i_1, \dots, i_M) = \sum_{\vec{y} \in D_U(i_1, \dots, i_M)} \prod_{u=1}^U f_u(y_u)$$
(4.21)

Die Größen $G_U(i_1, \dots, i_M)$, auf welche die Normierungskonstante hiermit zurückgeführt ist, lassen sich durch einen Algorithmus, der eine Erweiterung des im Anhang 5 beschriebenen Faltungsalgorithmus darstellt, sehr effektiv berechnen, sofern nicht die Parameter M und n_m große Werte annehmen.

Die Anwendung dieses Algorithmus soll im folgenden an einem Beispiel gezeigt werden, in dem alle Faktoren $f_u(y_u)$ von folgender Form sind:

$$f_u(y_u) = \binom{y_{1,u} + y_{2,u} + \dots + y_{M,u}}{y_{1,u}, y_{2,u}, \dots, y_{M,u}} A_{1,u}^{y_{1,u}} A_{2,u}^{y_{2,u}} \dots A_{M,u}^{y_{M,u}}$$
(4.22)

Dieses Beispiel entspricht dem Fall, daß in einem Warteschlangennetz der beschriebenen Art, ausschließlich Stationen mit einer Bedienungseinheit und first-come, first-served Abfertigungsdisziplin vorkommen. In diesem Fall lassen sich die Größen $G_U(i_1, \dots, i_M)$ mit Hilfe des folgenden Algorithmus berechnen:

(1) Setze $G_1(0, 0, \dots, 0) = 1$ (4.23)

(2) Berechne für $u \in \{1, \dots, U\}$

$$i_1 \in \{1, \dots, n_1\}$$

⋮

$$i_M \in \{1, \dots, n_M\}$$

(4.24)

$$G_u(i_1, i_2, \dots, i_M) = G_{u-1}(i_1, i_2, \dots, i_M) + \sum_{m=1}^M A_{m,u} G_u(i_1, i_2, \dots, i_{m-1}, i_{m-1} - 1, i_{m+1}, \dots, i_M)$$

(Dabei gilt per Definition: $G_0(i_1, i_2, \dots, i_M) = 0$
und $G_u(i_1, i_2, \dots, i_M) = 0$ falls mindestens ein $i_m < 0$.)

Damit ist prinzipiell die Berechnung der Normierungskonstanten auch in diesem Fall durchführbar. Zur Verarbeitung der Faktoren für die Verzögerungsstationen in Gl. (4.4) muß - wie bei der Lösung mit einer Klasse - eine unhandlichere, allgemeine Version des Algorithmus angewendet werden. Dieser Nachteil kann durch geeignete Zusammenfassung der Verzögerungsstationen analog zu der Methode im Abschnitt 3.3.3.1 stark gemindert werden.

Ähnlich wie im Fall mit einer Klasse können auch hier die Verkehrsgrößen des Systems aus den Größen G_u , die bei der Berechnung der Normierungskonstanten anfallen, bestimmt werden. Was jedoch die Anwendbarkeit des oben beschriebenen Algorithmus für unsere Anwendung stark einschränkt, ist der Speicherplatzbedarf zur Ausführung der Rekursion (4.24). Allein für die Größen G_u

sind $n_1 \cdot n_2 \cdot \dots \cdot n_M$ Speicherzellen für Gleitkommazahlen erforderlich. Der gesamte Speicherplatzbedarf liegt noch wesentlich höher. Damit ist von vornherein klar, daß bei den für unsere Anwendungen möglichen Werten der Speichergrößen n_m - zum Beispiel einheitlich 100 - die exakte Lösung nur bei sehr kleiner Modulzahl M ausgewertet werden kann.

Hieraus ergibt sich zwangsläufig die Notwendigkeit, noch einen anderen Lösungsweg zu finden. Im folgenden Abschnitt 4.3.3 wird ein Approximationsverfahren entwickelt, welches sehr genaue Näherungswerte für die interessierenden Verkehrsgrößen liefert und praktisch für eine beliebige Anzahl von Moduln angewendet werden kann.

Trotz ihrer schwierigen Auswertbarkeit ist jedoch die Kenntnis der exakten Lösung äußerst wertvoll, um nämlich die Gültigkeit der Näherungslösung anhand kleinerer Sonderfälle nachprüfen zu können. Zwar kann dafür auch die Simulation herangezogen werden, die Erfahrung zeigt jedoch, daß es für Speicherüberlauf-Wahrscheinlichkeiten kleiner als 10^{-3} äußerst rechenzeitaufwendig ist, statistisch zuverlässige Simulationsergebnisse zu bekommen. Die Auswertung der exakten Lösung erlaubt dagegen, wenigstens für kleinere Beispiele, die Näherungslösung auch in diesem Bereich zu überprüfen (siehe Abschnitt 4.3.5).

4.3.3 Näherungslösung bei exponentialverteilten Übertragungs- und Prozessorbelegungszeiten

4.3.3.1 Motivation

Wie im letzten Abschnitt beschrieben, lassen sich unter der Voraussetzung exponentialverteilter Übertragungs- und Prozessorbelegungszeiten zwar die Verkehrsgrößen unseres Modells explizit bestimmen, aber diese Lösung ist für eine größere Anzahl von Vermittlungsmoduln, wie sie in der Praxis auftreten kann, nicht

mehr numerisch auswertbar. Es wird deshalb in diesem Abschnitt ein Approximationsverfahren entwickelt mit dem Ziel, auch Konfigurationen mit nahezu beliebiger Anzahl von Moduln analysieren zu können. Die Grundidee des Verfahrens besteht darin, Modelle mit mehreren Vermittlungsmoduln in geeigneter Weise auf solche mit nur einem Modul zurückzuspielen und dann diese zu berechnen.

4.3.3.2 Beschreibung des Approximationsverfahrens

In diesem Abschnitt werden obere Schranken für die charakteristischen Größen des Verkehrs hergeleitet, der von einem einzelnen Vermittlungsmodul bei herkunftsorientierter Verkehrsaufteilung bewältigt werden muß.

Die exakte Überlaufwahrscheinlichkeit B_μ eines Speichermoduls Nr. μ gemäß Gl. (4.5) ist sicher nicht größer als der Wert B_μ^* , den man erhielte, falls alle übrigen Moduln einen unendlich großen Speicher besäßen. Die anschauliche Begründung dafür ist, daß die Pakete des Moduls μ bei ihrer Übertragung zum nächsten Vermittlungsknoten umso stärker von Paketen anderer Moduln behindert werden, je mehr solche Pakete gespeichert werden können. Eine stärkere Behinderung drückt sich in höheren Wartezeiten und daher größerer Überlaufwahrscheinlichkeit des Moduls μ aus. Die stärkste Behinderung des Verkehrs der Klasse μ und damit eine obere Schranke für die Überlaufwahrscheinlichkeit des Moduls μ erhält man, wenn alle n_m für $m \neq \mu$ unendlich gesetzt werden.

Wie im Anhang 4 gezeigt wird, erhält man unter dieser Voraussetzung für die Anzahlen der Anforderungen der Klasse μ in allen Stationen des Warteschlangennetzes die folgende Verteilung:

$$\begin{aligned}
 & P(X_{\mu;1\mu} = x_{\mu;1\mu}, X_{\mu;21} = x_{\mu;21}, \dots, X_{\mu;2R} = x_{\mu;2R}, X_{\mu;31} = x_{\mu;31}, \dots \\
 & \quad \dots, X_{\mu;3R} = x_{\mu;3R}, X_{\mu;41} = x_{\mu;41}, \dots, X_{\mu;4R} = x_{\mu;4R}) \\
 &= p_{\mu}(x_{\mu;1\mu}, x_{\mu;21}, \dots, x_{\mu;4R}) \\
 &= p_{\mu}(\vec{x}_{\mu}) \\
 &= C_{\mu} A_{\mu;1\mu}^{x_{\mu;1\mu}} \prod_{r=1}^R \left(\frac{A_{\mu;2r}}{1 - A_{\mu;2r}} \right)^{x_{\mu;2r}} \prod_{s=3}^4 \prod_{r=1}^R \frac{A_{\mu;sr}^{x_{\mu;sr}}}{x_{\mu;sr}!} \quad (4.25)
 \end{aligned}$$

$$\sum_{\vec{x}_{\mu} \in Z_{\mu}} p_{\mu}(\vec{x}_{\mu}) = 1$$

$$Z_{\mu} = \left\{ \vec{x}_{\mu} \mid x_{\mu;1\mu}, \dots, x_{\mu;4R} \in \mathbb{N}_0 \wedge x_{\mu;1\mu} + \sum_{s=2}^4 \sum_{r=1}^R x_{\mu;sr} \leq n_{\mu} \right\}$$

$$A_{\mu;1\mu} = \lambda_{\mu} b_{1\mu} \quad A_{\mu;2r} = \frac{p_{\mu;r}}{1 - q_r} \lambda_{\mu} b_{2r}$$

$$A_{\mu;2r} = \sum_{m \in \{1, \dots, M\} \setminus \{\mu\}} \frac{p_{\mu;r}}{1 - q_r} \lambda_m b_{2r} < 1 \quad r \in \{1, \dots, R\}$$

$$A_{\mu;3r} = \frac{p_{\mu;r} q_r}{1 - q_r} \lambda_{\mu} b_{3r} \quad A_{\mu;4r} = p_{\mu;r} \lambda_{\mu} b_{4r}$$

Diese Lösung läßt sich anschaulich sehr einleuchtend interpretieren:

- Alle Stationen des Warteschlangennetzes, durch die kein Verkehr der Klasse μ fließt, sind in der Lösung nicht enthalten.
- Da in den Verzögerungsstationen (mit unendlich vielen Bedienungseinheiten) keine Behinderung der Klasse μ vorkommen kann, verschwinden alle Anteile der übrigen Verkehre bezüglich diesen Stationen.
- Das Angebot $A_{\mu;2r}$ der Klasse μ an die Station Nr. (2,r), d.h. den Übertragungskanal in die Richtung r, erhöht sich um den Faktor " $(1 - \text{Summe der übrigen Angebote})^{-1}$ ". Diese Erhöhung

spiegelt die Behinderung der Klasse μ durch die übrigen Klassen $m \neq \mu$ wider.

Wie man erkennt, ist Gleichung (4.25) genau gleich aufgebaut wie die Zustandswahrscheinlichkeiten in Gl. (3.6) für den Fall eines einzigen Vermittlungsmoduls. Daher läßt sich die numerische Berechnung der Zustandswahrscheinlichkeiten und der Verkehrsgrößen in genau derselben Weise wie dort durchführen. Es sind einzig allein erhöhte fiktive mittlere Bedienungszeiten b_{2r}^* so anzunehmen, daß die folgende Beziehung erfüllt ist:

$$\frac{A_{\mu;2r}}{1 - A_{\mu;2r}} = \frac{p_{\mu;r}}{1 - q_r} \lambda_{\mu} b_{2r}^* \quad r \in \{1, \dots, R\} \quad (4.26)$$

Dies ist durch die folgende Wahl von b_{2r}^* erfüllt:

$$b_{2r}^* = \frac{b_{2r}}{1 - A_{\mu;2r}} \quad r \in \{1, \dots, R\} \quad (4.27)$$

Dieses Ergebnis ist physikalisch wiederum sehr einleuchtend: Die ursprüngliche (normierte) Kanalkapazität 1 wird reduziert um den resultierenden Verkehrswert $A_{\mu;2r}$, den die übrigen Modulen, wegen ihrer unendlich großen Speicher, über den Übertragungskanal Nr. (2,r) durchsetzen.

Mit diesen fiktiven mittleren Übertragungszeiten b_{2r}^* und den realen Werten $R, n_{\mu}, \lambda_{\mu}, b_{1\mu}, b_{3r}, b_{4r}, p_r, q_r$ lassen sich die Zustandswahrscheinlichkeiten und die interessierenden Verkehrsgrößen (Speicherüberlauf-Wahrscheinlichkeiten, Durchsatz, Speicherbelastungen, Warte- und Verweilzeiten, usw.) genauso wie für einen Vermittlungsknoten mit einem Modul näherungsweise berechnen (siehe Abschnitt 3). Wie aus der Ableitung klar wird, liegen diese Resultate auf der sicheren Seite (numerische Ergebnisse siehe Abschnitt 4.3.5).

4.3.4 Näherungslösung bei allgemein verteilten Übertragungs- und Prozessorbelegungszeiten

Im vorigen Abschnitt wurde gezeigt, wie bei exponentialverteilten Übertragungs- und Prozessorbelegungszeiten die Analyse von Vermittlungsknoten mit mehreren Vermittlungsmoduln auf solche mit einem Modul näherungsweise zurückgeführt werden kann. Da für Vermittlungsknoten mit einem Modul und allgemein verteilten Übertragungs- und Prozessorbelegungszeiten im Abschnitt 3.4 ein Näherungsverfahren entwickelt wurde, liegt es nahe, beide Ansätze zu kombinieren. Diese Kombination liefert somit ein Verfahren zur näherungsweisen Analyse von Vermittlungsknoten mit herkunftsorientierter Verkehrsaufteilung, mehreren Vermittlungsmoduln und allgemein verteilten Übertragungs- und Prozessorbelegungszeiten.

Wie bei der Näherungslösung bei exponentialverteilten Übertragungs- und Prozessorbelegungszeiten betrachten wir die Speicherbelegungen, welche von den Paketen eines bestimmten Moduls herrühren. Analog zu Gl. (3.36) wird angenommen, daß die Verteilung für die Anzahl der Anforderungen der betrachteten Klasse μ in den Stationen des Warteschlangennetzes die folgende Gestalt besitzt:

$$\begin{aligned} P(X_{\mu;1\mu} = x_{\mu;1\mu}, X_{\mu;21} = x_{\mu;21}, \dots, X_{\mu;2R} = x_{\mu;2R}, X_{\mu;31} = x_{\mu;31}, \dots \\ \dots X_{\mu;3R} = x_{\mu;3R}, X_{\mu;41} = x_{\mu;41}, \dots, X_{\mu;4R} = x_{\mu;4R}) \\ = p'_\mu(x_{\mu;1\mu}, x_{\mu;21}, \dots, x_{\mu;2R}, x_{\mu;31}, \dots, x_{\mu;3R}, x_{\mu;41}, \dots, x_{\mu;4R}) \\ = p'_\mu(\vec{x}_\mu) \\ = C'_\mu p_{\mu;1\mu}(x_{\mu;1\mu}) \prod_{r=1}^R p_{\mu;2r}(x_{\mu;2r}) \prod_{s=3}^4 \prod_{r=1}^R \frac{A_{\mu;sr}^{x_{\mu;sr}}}{x_{\mu;sr}!} \end{aligned} \quad (4.28)$$

$$\sum_{\vec{x}_\mu \in Z_\mu} p'_\mu(\vec{x}_\mu) = 1$$

$$Z_\mu = \left\{ \vec{x}_\mu \mid x_{\mu;1\mu}, \dots, x_{\mu;4R} \in \mathbb{N}_0 \wedge x_{\mu;1\mu} + \sum_{s=2}^4 \sum_{r=1}^R x_{\mu;sr} \leq n_\mu \right\}$$

Die Faktoren für die einzelnen Stationen werden - wie im Abschnitt 3.4.3 - durch die Zustandswahrscheinlichkeiten der isolierten Stationen mit angebotenen Poisson-Verkehr bestimmt. Die Angebote an diese isolierten Stationen werden dabei identisch gewählt wie in der Näherungslösung für exponentialverteilte Übertragungs- und Prozessorbelegungszeiten (vgl. Gl. (4.25)):

$$\begin{aligned} A_{\mu;1\mu} &= \lambda_\mu b_{1\mu} & A_{\mu;2r} &= \frac{p_{\mu;r}}{1-q_r} \lambda_\mu b_{2r} \\ A_{0;2r} &= \sum_{m \in \{1, \dots, M\} \setminus \{\mu\}} \frac{p_{m;r}}{1-q_r} \lambda_m b_{2r} & & r \in \{1, \dots, R\} \\ A_{\mu;3r} &= \frac{p_{\mu;r} q_r}{1-q_r} \lambda_\mu b_{3r} \\ A_{\mu;4r} &= p_{\mu;r} \lambda_\mu b_{4r} \end{aligned} \quad (4.29)$$

Aus den Zustandswahrscheinlichkeiten können nun in der üblichen Weise die charakteristischen Verkehrsgrößen berechnet werden.

Wie der Vergleich mit der Simulation zeigt, liefert dieses Näherungsverfahren Ergebnisse von sehr guter Genauigkeit (siehe Abschnitt 4.3.5).

4.3.5 Ergebnisse

Die in diesem Abschnitt vorgestellten Ergebnisse zeigen das grundsätzliche Verhalten des maximalen Durchsatzes und der Speicherüberlauf-Wahrscheinlichkeit von Vermittlungsknoten mit mehreren Vermittlungsmoduln und herkunftsorientierter Verkehrsaufteilung. Sie dienen ferner dazu, die Genauigkeit der beschriebenen Näherungslösungen durch Vergleich mit Simulationsergebnissen und mit exakten Ergebnissen zu demonstrieren. Abschließend werden außerdem Systeme mit herkunfts- und mit zielorientierter Verkehrsaufteilung bezüglich ihrem Speicherplatzbedarf verglichen.

Bild 4.9 zeigt den maximalen Paketdurchsatz D_{\max} als Funktion der Gesamtzahl abgehender Übertragungskanäle R für einen Vermittlungsknoten mit einem bzw. mit zwei Vermittlungsmoduln. Parameter der Kurvenscharen ist die Gesamtzahl n von Paketspeicherplätzen im Vermittlungsknoten. Wie man erkennt, ist der Maximaldurchsatz bei gleicher Speichergröße für kleine Werte von R praktisch unabhängig von der Modulzahl M . Bei größeren Leitungszahlen macht sich dann aber bei $M = 1$ der einzige Prozessor als Engpaß bemerkbar und stellt mit seiner Bedienungsrate von 100 Paketen pro Sekunde eine obere Schranke für den Paketdurchsatz dar. Die obere Schranke für den Durchsatz bei 2 Moduln liegt entsprechend bei 200 Paketen pro Sekunde. Das heißt, auch bei herkunftsorientierter Verkehrsaufteilung stellt die Modularisierung ein Mittel zur Steigerung des Maximaldurchsatzes dar.

Im Bild 4.10 ist die Speicherüberlauf-Wahrscheinlichkeit B über der Paketankunftsrate λ für verschiedene Modulzahlen M aufgetragen. Hierbei ist vorausgesetzt, daß die Gesamtzahl n der Paketspeicherplätze im Vermittlungsknoten konstant gehalten wird, unabhängig von der Modulzahl M . Man erkennt, daß bei größeren Modulzahlen höhere Speicherüberlauf-Wahrscheinlichkeiten auftreten. Der Grund hierfür ist wie bei der zielorientierten Verkehrsaufteilung der sogenannte Bündelungsgewinn, der bei kleinerer Modulzahl, also größerem Speicher pro Modul, zu kleineren Überlaufwahrscheinlichkeiten führt. Aus dem Vergleich der eingezeichneten Näherungslösung und den Simulationsergebnissen wird die sehr gute Genauigkeit dieser Approximation deutlich.

Aus Rechenzeitgründen ist es praktisch unmöglich, mit Hilfe der Simulation Werte kleiner als 10^{-3} für die Speicherüberlauf-Wahrscheinlichkeit zu gewinnen, welche statistisch zuverlässige Aussagen darstellen. Deshalb ist zur Überprüfung der Genauigkeit unserer Näherungslösung in dem Bereich mit kleineren Überlaufwahrscheinlichkeiten die Kenntnis der exakten Lösung sehr wertvoll, auch wenn diese nur für kleine Systeme numerisch ausgewertet werden kann (vgl. Abschnitt 4.3.2.3). Einen solchen Vergleich zeigt Bild 4.11, in dem die Speicherüberlauf-Wahrscheinlichkeit B über der Gesamtzahl der Paketspeicherplätze n aufgetragen ist

und zwar für Vermittlungsknoten mit $M = 1, 2$ und 3 Vermittlungsmoduln. Aufgrund des Bündelungsgewinns nimmt die Speicherüberlauf-Wahrscheinlichkeit bei festgehaltener Gesamtzahl von Speicherplätzen mit wachsendem Modularisierungsgrad zu. Wie man an diesem Beispiel sieht, liefert die Näherung für die praktisch interessanten kleinen Überlaufwahrscheinlichkeiten äußerst präzise, für große Werte von B hinreichend genaue Ergebnisse. In jenem Bereich, in dem die Näherung von der exakten Lösung abweicht, ist jedoch sichergestellt, daß die Approximation auf der sicheren Seite liegt, da sie den Einfluß der übrigen Moduln auf die Überlaufwahrscheinlichkeit des betrachteten Moduls stets überschätzt (vgl. Abschnitt 4.3.3.2).

Während in den bisherigen Beispielen exponentialverteilte Übertragungs- und Prozessorbelegungszeiten angenommen wurden, ist im Bild 4.12 der Einfluß anderer Belegungszeit-Verteilungen dargestellt. Aufgetragen ist die Speicherüberlauf-Wahrscheinlichkeit B über der gesamten Paketankunftsrate λ für einen Vermittlungsknoten mit $M = 4$ Vermittlungsmoduln und herkunftsorientierter Verkehrsaufteilung. Variiert wurde in diesem Diagramm die Verteilungsfunktion der Übertragungszeiten (konstant, Erlang-4, exponentiell, hyperexponentiell), während alle übrigen Parameter festgehalten wurden. Es zeigt sich auch hier, wie schon im Fall eines Moduls, der Effekt, daß stärker streuende Übertragungszeiten zu längeren Warteschlangen und deshalb zu höheren Speicherüberlauf-Wahrscheinlichkeiten führen. Der Vergleich mit den Simulationsergebnissen zeigt die gute Genauigkeit der im Abschnitt 4.3.4 beschriebenen Näherungslösung. Es ist zu erwarten, daß die Güte der Approximation bei größeren Systemen sogar noch höher ist, da dort die im Näherungsansatz nicht berücksichtigten Abhängigkeiten weniger Einfluß besitzen als bei dem hier gezeigten System, welches wegen der Simulation relativ klein gehalten werden mußte.

Im Bild 4.13 wird nun noch herkunfts- und zielorientierte Verkehrsaufteilung bezüglich ihrem Speicherbedarf verglichen. Dargestellt ist für beide Aufteilungsarten die Speicherüberlauf-Wahrscheinlichkeit B als Funktion der Gesamtzahl von Paketspeicherplätzen n für drei Modulzahlen: $M = 2, 10$ und 50 . Der Ver-

gleich der beiden Aufteilungsarten fällt bei großen Überlaufwahrscheinlichkeiten leicht zu Gunsten der zielorientierten Verkehrsaufteilung aus, was jedoch von geringem praktischem Interesse ist. Bei kleineren Überlaufwahrscheinlichkeiten liefert dagegen die herkunftsorientierte Verkehrsaufteilung - insbesondere bei größeren Modulzahlen - deutlich kleinere Speicherüberlauf-Wahrscheinlichkeiten. Anders ausgedrückt, das zielorientierte Verfahren benötigt bei gleicher Verkehrsgüte mehr Speicherplatz als das herkunftsorientierte. Der Grund hierfür ist, daß bei zielorientierter Aufteilung für eine relativ kleine Anzahl abgehender Kanäle ein gemeinsamer Speicherpool vorhanden ist, so daß sich momentane Verkehrsschwankungen eines Kanals relativ stark bemerkbar machen. Bei herkunftsorientierter Verkehrsaufteilung kann dagegen ein Ausgleich zwischen dem Verkehr aller abgehenden Kanäle und deren Warteschlangen stattfinden.

Abschließend soll jedoch bemerkt werden, daß dieses Verkehrsverhalten beider Strategien ein zwar wichtiges, aber nicht das einzige Kriterium zur Auswahl einer günstigen Aufteilung des Verkehrs auf die Vermittlungsmoduln darstellt. Wichtige Faktoren sind - wie schon im Abschnitt 4.1.2 bemerkt wurde - der Mehraufwand für die Ausstattung der Leitungsinterfaces mit mehr Intelligenz und Speicher bei der zielorientierten Verkehrsaufteilung, dem jedoch die aufwendigere Abwicklung des abgehenden Verkehrs bei der herkunftsorientierten Aufteilung gegenübersteht.

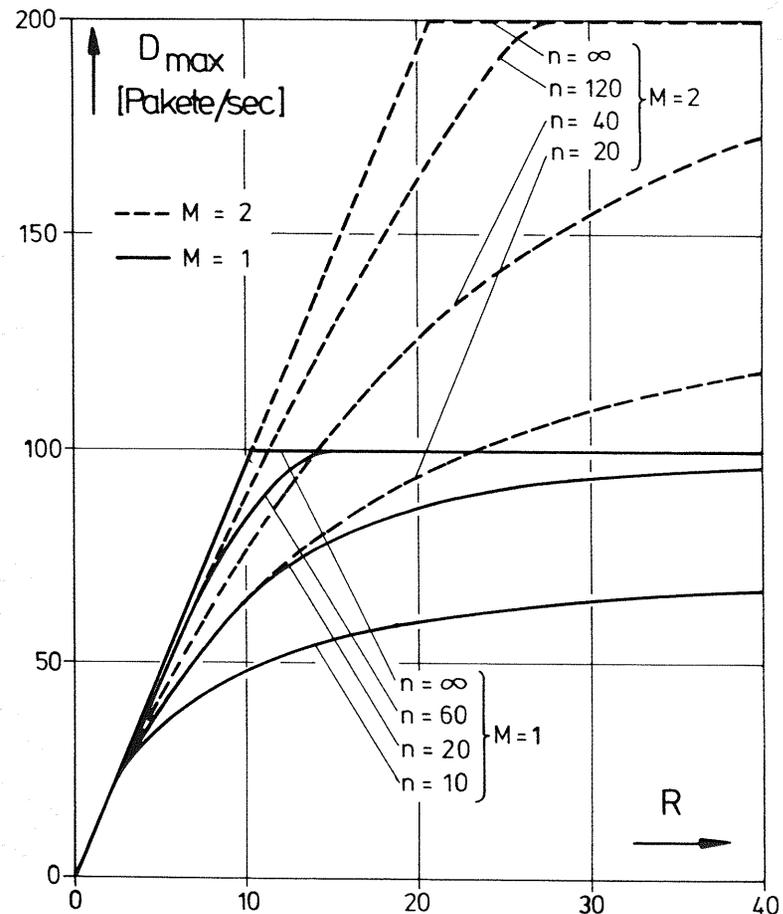


Bild 4.9: Maximaler Gesamtdurchsatz des Vermittlungsknotens D_{max} als Funktion der Gesamtzahl abgehender Übertragungskanäle R für $M = 1$ und 2 Moduln und verschiedene Gesamtspeichergrößen n . Exponentialverteilte Übertragungs- und Prozessorbelegungszeiten. Herkunftsorientierte Verkehrsaufteilung auf die Vermittlungsmoduln.

$n_m = n/M$ Paketspeicherplätze pro Vermittlungsmodul
 mittlere Prozessorbelegungszeiten: $b_{1,1} = \dots = b_{M,1} = 10$ msec
 mittlere Übertragungszeiten pro Paket: $b_{21} = \dots = b_{2R} = 104$ msec
 mittlere Negativ-Quittierzeiten: $b_{31} = \dots = b_{3R} = 0$ msec ¹⁾
 mittlere Positiv-Quittierzeiten: $b_{41} = \dots = b_{4R} = 0$ msec ¹⁾
 Verzweigungswahrscheinlichkeiten: $p_1 = \dots = p_R = 1/R$
 Wiederholungswahrscheinlichkeiten: $q_1 = \dots = q_R = 0.01$

¹⁾ idealisierte Annahme mit Rücksicht auf Rechenzeit- und Speicherplatzbedarf

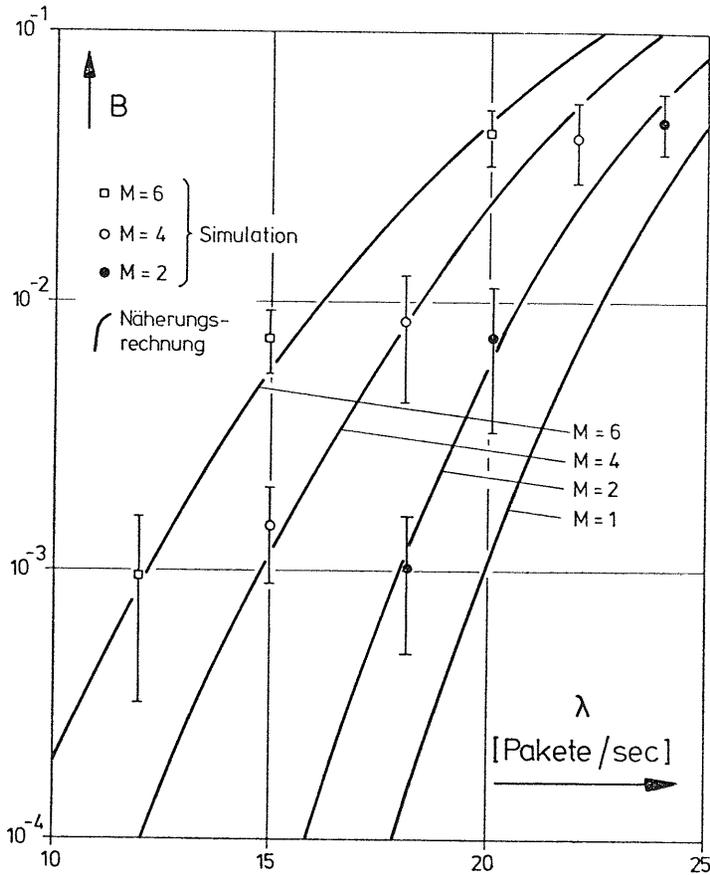


Bild 4.10: Speicherüberlauf-Wahrscheinlichkeit B als Funktion der gesamten Paketankunftsrate λ für verschiedene Modulzahlen M. Exponentialverteilte Übertragungs- und Prozessorbelegungszeiten. Herkunftsorientierte Verkehrsaufteilung auf die Vermittlungsmoduln.

R = 15 abgehende Übertragungskanäle
 n = 60 Paketspeicherplätze im gesamten Vermittlungsknoten
 $n_m = 60/M$ Paketspeicherplätze pro Vermittlungsmodul
 Paketankunftsrate pro Vermittlungsmodul: $\lambda_m = \lambda/M$
 mittlere Prozessorbelegungszeiten: $b_{1,1} = \dots = b_{M,1} = 0 \text{ msec}$ ¹⁾
 mittlere Übertragungszeiten pro Paket: $b_{2,1} = \dots = b_{2,15} = 417 \text{ msec}$
 mittlere Negativ-Quittierzeiten: $b_{3,1} = \dots = b_{3,15} = 1250 \text{ msec}$
 mittlere Positiv-Quittierzeiten: $b_{4,1} = \dots = b_{4,15} = 833 \text{ msec}$
 Verzweigungswahrscheinlichkeiten: $p_1 = \dots = p_{15} = 0.067$
 Wiederholungswahrscheinlichkeiten: $q_1 = \dots = q_{15} = 0.01$

¹⁾ idealisierte Annahme mit Rücksicht auf Speicherbedarf der Simulation

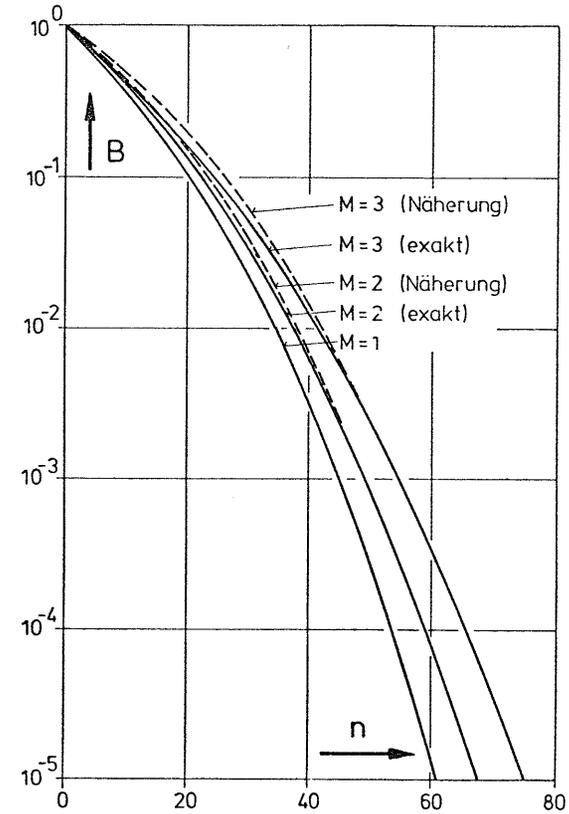


Bild 4.11: Speicherüberlauf-Wahrscheinlichkeit B als Funktion der Gesamt-speichergröße n für verschiedene Modulzahlen M. Exponentialverteilte Übertragungs- und Prozessorbelegungszeiten. Herkunftsorientierte Verkehrsaufteilung auf die Vermittlungsmoduln.

R = 12 abgehende Übertragungskanäle
 $n_m = n/M$ Paketspeicherplätze pro Vermittlungsmodul
 gesamte Paketankunftsrate des Vermittlungsknotens: $\lambda = 69.12/\text{sec}$
 Paketankunftsrate pro Vermittlungsmodul: $\lambda_m = \lambda/M$
 mittlere Prozessorbelegungszeiten: $b_{1,1} = \dots = b_{M,1} = 10 \text{ msec}$
 mittlere Übertragungszeiten pro Paket: $b_{2,1} = \dots = b_{2,12} = 104 \text{ msec}$ ¹⁾
 mittlere Negativ-Quittierzeiten: $b_{3,1} = \dots = b_{3,12} = 0 \text{ msec}$ ¹⁾
 mittlere Positiv-Quittierzeiten: $b_{4,1} = \dots = b_{4,12} = 0 \text{ msec}$ ¹⁾
 Verzweigungswahrscheinlichkeiten: $p_1 = \dots = p_{12} = 0.083$
 Wiederholungswahrscheinlichkeiten: $q_1 = \dots = q_{12} = 0.01$

¹⁾ idealisierte Annahme mit Rücksicht auf Rechenzeit- und Speicherplatzbedarf der exakten Lösung

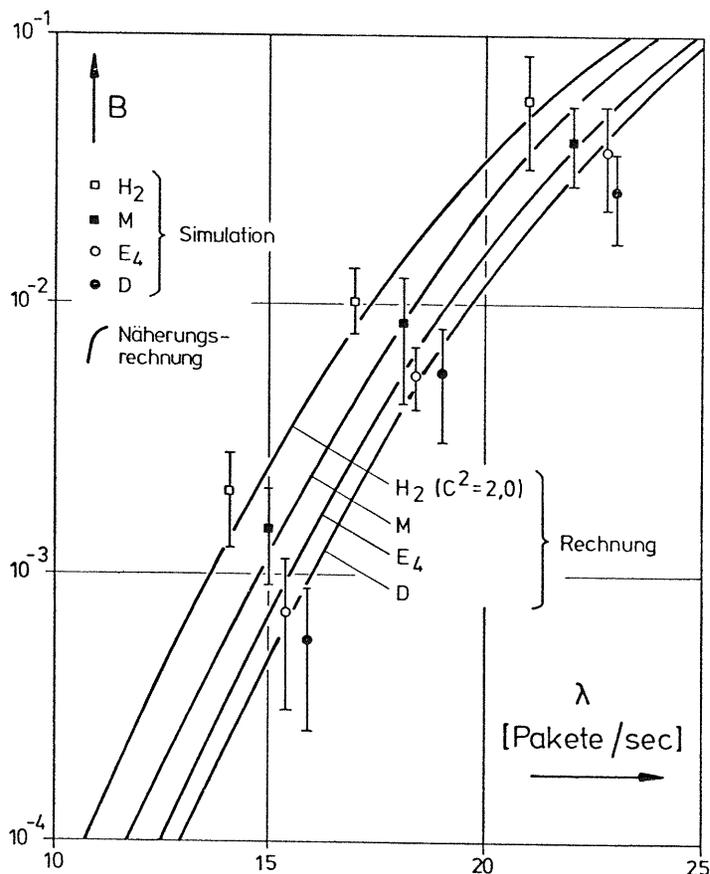


Bild 4.12: Speicherüberlauf-Wahrscheinlichkeit B als Funktion der gesamten Paketankunftsrate λ für verschiedene Verteilungsfunktionen der Übertragungszeiten. Exponentialverteilte Prozessorbelegungszeiten. Herkunftsorientierte Verkehrsaufteilung auf die Vermittlungsmoduln.

$M = 4$ Vermittlungsmoduln
 $R = 15$ abgehende Übertragungskanäle
 $n_m = 15$ Paketspeicherplätze pro Modul
 Paketankunftsrate pro Vermittlungsmodul: $\lambda_m = \lambda/M$
 mittlere Prozessorbelegungszeiten: $b_{1,1} = \dots = b_{4,1} = 0 \text{ msec}$ ¹⁾
 mittlere Übertragungszeiten pro Paket: $b_{21} = \dots = b_{2,15} = 417 \text{ msec}$
 mittlere Negativ-Quittierzeiten: $b_{31} = \dots = b_{3,15} = 1250 \text{ msec}$
 mittlere Positiv-Quittierzeiten: $b_{41} = \dots = b_{4,15} = 833 \text{ msec}$
 Verzweigungswahrscheinlichkeiten: $p_1 = \dots = p_{15} = 0.067$
 Wiederholungswahrscheinlichkeiten: $q_1 = \dots = q_{15} = 0.01$

¹⁾ idealisierte Annahme mit Rücksicht auf Speicherbedarf der Simulation

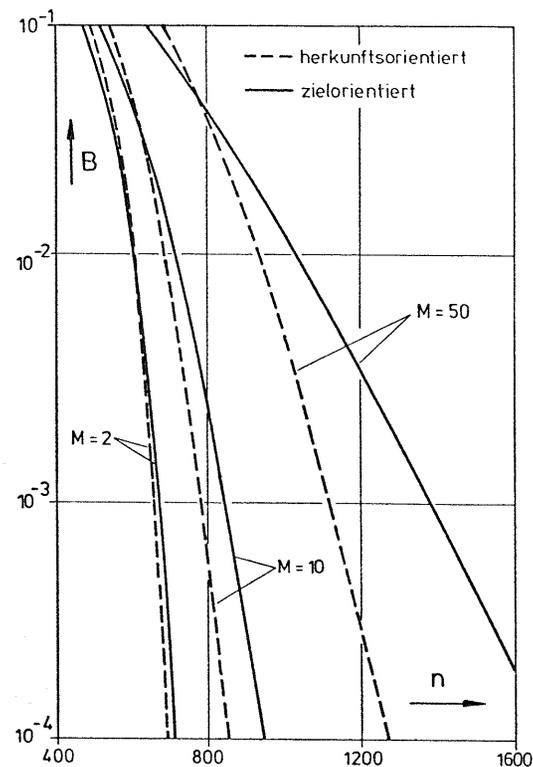


Bild 4.13: Speicherüberlauf-Wahrscheinlichkeit B als Funktion der Gesamt-speichergröße n für verschiedene Modulzahlen M . Exponentialverteilte Übertragungs- und Prozessorbelegungszeiten. Vergleich von herkunfts- und zielorientierter Verkehrsaufteilung auf die Vermittlungsmoduln.

$R = 200$ abgehende Übertragungskanäle
 $n_m = n/M$ Paketspeicherplätze pro Vermittlungsmodul
 gesamte Paketankunftsrate des Vermittlungsknotens: $\lambda = 288/\text{sec}$
 Paketankunftsrate pro Vermittlungsmodul: $\lambda_m = \lambda/M$
 mittlere Prozessorbelegungszeiten: $b_{1,1} = \dots = b_{M,1} = 5 \text{ msec}$
 mittlere Übertragungszeiten pro Paket:
 $b_{21} = \dots = b_{2,200} = b_{m;21} = \dots = b_{m;2,R_m} = 417 \text{ msec}$
 mittlere Negativ-Quittierzeiten:
 $b_{31} = \dots = b_{3,200} = b_{m;31} = \dots = b_{m;3,R_m} = 1250 \text{ msec}$
 mittlere Positiv-Quittierzeiten:
 $b_{41} = \dots = b_{4,200} = b_{m;41} = \dots = b_{m;4,R_m} = 833 \text{ msec}$
 Verzweigungswahrscheinlichkeiten:
 $p_1 = \dots = p_{200} = 0.005$; $p_{m;1} = \dots = p_{m;R_m} = M/200$
 Wiederholungswahrscheinlichkeiten:
 $q_1 = \dots = q_{200} = q_{m;1} = \dots = q_{m;R_m} = 0.01$

5. ZUSAMMENFASSUNG

Die in dieser Arbeit durchgeführten Untersuchungen haben das Ziel, Verfahren zur verkehrsgerechten Dimensionierung der Vermittlungsknoten in Daten- und Rechnernetzen mit Teilstreckenvermittlung bereitzustellen. Verkehrsgerechte Dimensionierung bedeutet in diesem Zusammenhang, die Baugruppen der Vermittlungsknoten und die Übertragungswege so zu bemessen, daß unter Einhaltung einer vorgeschriebenen Verkehrsgüte ein vorgegebener Datenverkehr vermittelt und übertragen werden kann.

Für diese Aufgabenstellung wurden Verkehrsmodelle entwickelt, welche das Ablaufgeschehen in den Vermittlungsknoten mit Hilfe von Warteschlangennetzen beschreiben. Komponenten der Modelle sind die endlich großen Paketspeicher und die Vermittlungsprozessoren, sowie die angeschlossenen Übertragungskanäle mit ihren Übertragungsprozeduren. Ziel der Analyse dieser Modelle ist es, die interessierenden Verkehrsgrößen (Speicherüberlauf-Wahrscheinlichkeit, Durchsatz, Wartezeiten, Speicherverweilzeiten, Belastungen) zu bestimmen. Diese Größen werden hauptsächlich beeinflußt von der Größe des Paketspeichers, von der Bearbeitungszeit der Pakete durch den Prozessor, von der Übertragungszeit der Pakete, sowie von den Eigenschaften der Übertragungskanäle (Laufzeiten, Fehlerwahrscheinlichkeiten) und der Übertragungsprozeduren.

Außerdem besitzt der Aufbau und die Organisation der Vermittlungsknoten einen gewichtigen Einfluß auf die Verkehrsleistung. Zwei unterschiedliche Strukturen von Vermittlungsknoten wurden in dieser Arbeit untersucht: zentralisierte Systeme mit einem Vermittlungsmodul (d.h. einem Paketspeicher und einem Vermittlungsprozessor) und dezentralisierte Systeme mit mehreren Vermittlungsmoduln (d.h. mehreren Paketspeichern und Vermittlungsprozessoren). Die dezentralisierten Systeme besitzen durch die Möglichkeit, mit wachsender Verkehrsbelastung die Last auf eine wachsende Anzahl von Moduln aufzuteilen, grundsätzliche Vorteile gegenüber zentralisierten Systemen. Zwei Arten dieser

Verkehrsaufteilung wurden in der Arbeit untersucht: zielorientierte und herkunftsorientierte Aufteilung des Verkehrs auf die Vermittlungsmoduln.

Für alle entwickelten Verkehrsmodelle wurden zunächst unter der Voraussetzung eines Poisson-Ankunftsprozesses der Pakete und exponentialverteilter Übertragungs- und Prozessorbelegungszeiten exakte Lösungen hergeleitet. Für allgemein verteilte Übertragungs- und Prozessorbelegungszeiten wurden Näherungsverfahren entwickelt. Der Vergleich dieser Näherungslösungen mit Simulationsergebnissen zeigt, daß die Approximationen über ein breites Spektrum von Parametern sehr zuverlässige Resultate liefern.

Ein wichtiges Ziel der Arbeit war es, die Analyseverfahren so weit zu entwickeln, daß sie bei der praktischen Dimensionierung von Vermittlungsknoten eingesetzt werden können, ohne daß sich der Anwender mit der dahinter stehenden Theorie auseinandersetzen muß.

ANHANG

Anhang 1 : Theorem über Warteschlangennetze

Dieser Anhang faßt diejenigen Ergebnisse aus der Theorie der Warteschlangennetze zusammen, die für die Untersuchungen in dieser Arbeit benötigt werden. Er stützt sich im wesentlichen auf die Darstellung in [2] ab, wobei jedoch Ergebnisse und Verallgemeinerungen, welche für unsere Untersuchungen nicht wesentlich sind, unerwähnt bleiben.

Beschreibung des Systems

Die betrachtete Art von Wartesystemen enthält eine beliebige endliche Anzahl U von Bedienstationsstationen. Die zu bedienenden Anforderungen sind in eine beliebige endliche Anzahl K von Anforderungsklassen eingeteilt. Die Anforderungen durchlaufen das Warteschlangennetz gemäß vorgegebenen Übergangswahrscheinlichkeiten: Eine Anforderung der Klasse k , welche gerade in der Station i bedient worden ist, wird als nächstes mit der Wahrscheinlichkeit $q_{k;ij}$ zur Station j gelangen.

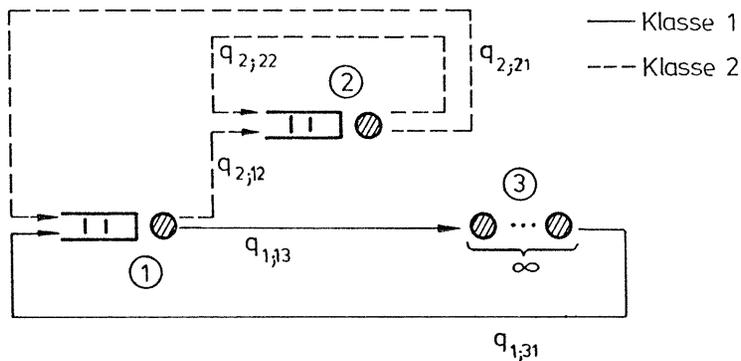


Bild A1: Beispiel eines geschlossenen Warteschlangennetzes mit $U = 3$ Bedienstationsstationen und $K = 2$ Anforderungsklassen

Wir unterscheiden zwischen zwei Arten von Warteschlangennetzen: Bei geschlossenen Netzen ist die Gesamtzahl der Anforderungen jeder Klasse im Netz zu jedem Zeitpunkt gleich groß (vgl. Bild A1). Dagegen treffen bei offenen Warteschlangennetzen Anforderungen von externen Verkehrsquellen ein (vgl. Bild A2). Voraussetzung der im folgenden beschriebenen Lösung ist, daß die Ankunftsprozesse bei offenen Netzen für alle Anforderungsklassen Poisson-Prozesse sind. Allerdings darf die momentane Ankunftsrate λ_k der Klasse k abhängen von der Gesamtzahl ξ_k der Anforderungen dieser Klasse, welche sich momentan im Netz befinden: $\lambda_k = \lambda_k(\xi_k)$. Eine extern ankommende Anforderung der Klasse k gelangt als erstes zur Station i mit der Wahrscheinlichkeit $q_{k;i}$.

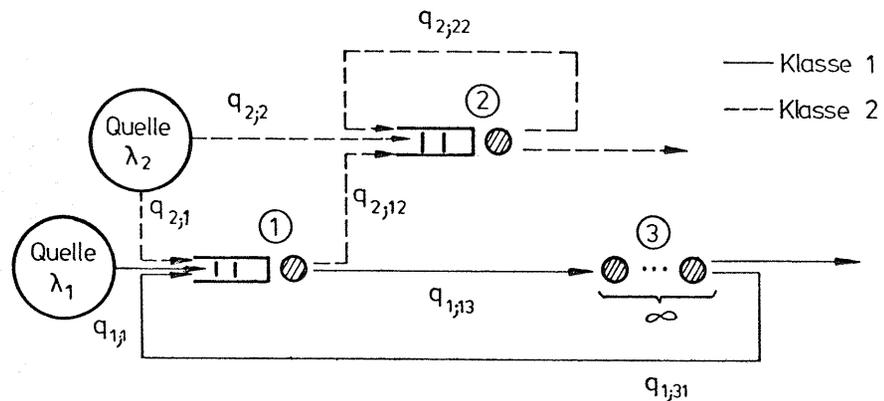


Bild A2: Beispiel eines offenen Warteschlangennetzes mit $U = 3$ Bedienstationsstationen und $K = 2$ Anforderungsklassen

Wir unterscheiden ferner zwischen zwei Typen von Bedienstationsstationen ¹⁾:

Typ 1: Alle Anforderungen besitzen dieselbe exponentialverteilte Bedienungszeit-Verteilung. Die Bedienungsrate

¹⁾ Die hier beschriebene Lösung läßt sich auf weitere zwei Typen von Bedienstationsstationen ausdehnen, welche aber für unsere Anwendungen ohne Interesse sind.

darf von der Anzahl der Anforderungen in der Station abhängen ¹⁾. Die Abfertigung geschieht in der Reihenfolge der Ankunft (first-come, first-served).

Typ 2: Die Anzahl der Bedienungseinheiten ist mindestens so groß wie die maximale Anzahl der Anforderungen, die sich gleichzeitig in der Bedienstation aufhalten können. Die Bedienungszeiten der Anforderungen dürfen für jede Klasse verschieden verteilt sein. Diese Verteilungen müssen eine rationale Laplace-Transformierte besitzen ²⁾.

Zustandswahrscheinlichkeiten

Wir beschreiben einen Zustand \vec{y} des Systems durch das U-Tupel

$$\vec{y} = (y_1, y_2, \dots, y_U) \tag{A1.1}$$

dessen Komponenten y_i selbst Vektoren sind entsprechend:

$$y_i = (y_{1,i}, y_{2,i}, \dots, y_{K,i}) \tag{A1.2}$$

Dabei bedeutet $y_{k,i}$ die Anzahl der Anforderungen der Klasse k in der Station i.

Wenn wir mit v_i die Gesamtzahl der Anforderungen in der Station i bezeichnen und mit $(\mu_{k,i})^{-1}$ die mittlere Bedienungszeit der Klasse k in der Station i, so gilt für die stationären Zustandswahrscheinlichkeiten :

$$p(\vec{y}) = p(y_1, \dots, y_U) = C d(\vec{y}) g_1(y_1) g_2(y_2) \dots g_U(y_U) \tag{A1.3}$$

¹⁾ Damit sind insbesondere Wartesysteme mit mehr als einer Bedienungseinheit erfaßt. Im folgenden beschränken wir uns jedoch auf zustandsunabhängige Bedienungsraten.

²⁾ Verteilungsfunktionen mit rationaler Laplace-Transformierter sind sog. "Phasentyp-Verteilungen", da sie sich immer aus exponentialverteilten Zeitabschnitten (Phasen) erzeugen lassen. Mit Phasentyp-Verteilungen läßt sich jede Verteilung beliebig genau approximieren | 21,9 |.

Dabei gilt:

$$1) \quad g_i(y_i) = \binom{v_i}{y_{1,i}, y_{2,i}, \dots, y_{K,i}} \left(\frac{1}{\mu_i}\right)^{y_i} \prod_{k=1}^K e_{k,i}^{y_{k,i}} \tag{A1.4}$$

$(\mu_i = \mu_{1,i} = \dots = \mu_{K,i})$
falls die Station i vom Typ 1 ist,

$$2) \quad g_i(y_i) = \prod_{k=1}^K \frac{1}{y_{k,i}!} \left(\frac{e_{k,i}}{\mu_{k,i}}\right)^{y_{k,i}} \tag{A1.5}$$

falls die Station i vom Typ 2 ist.

3) Die Größe $d(\vec{y})$ ist bei offenen Netzen eine Funktion der Ankunftsrate gemäß:

$$d(\vec{y}) = \prod_{k=1}^K \frac{y_{k,1} + \dots + y_{k,U} - 1}{y_{k,1}! \dots y_{k,U}!} \lambda_k \left(\frac{\xi_k}{\delta_k}\right) \tag{A1.6}$$

Für geschlossene Netze ist $d(\vec{y}) = 1$.

4) Die Größen $e_{k,i}$ in der Lösung (A1.3) lassen sich interpretieren als die Erwartungswerte für die Anzahl der Durchläufe einer Anforderung der Klasse k durch die Station i. Sie sind durch die folgenden Gleichungssysteme definiert ¹⁾:

$$\sum_{i=1}^U e_{k,i} q_{k,ij} + q_{k,j} = e_{k,j} \quad k \in \{1, \dots, K\}; j \in \{1, \dots, U\} \tag{A1.7}$$

5) Der Wert der Normierungskonstanten C ergibt sich aus der Bedingung

$$\sum_{\text{alle Zustände}} p(\vec{y}) = 1 \tag{A1.8}$$

¹⁾ Für geschlossene Netze ist $q_{k,j} = 0$; die dadurch entstehenden homogenen Gleichungssysteme für die $e_{k,j}$ besitzen eine einparametrische Lösungsschar. Trotzdem ergibt sich aufgrund der Normierung eine eindeutige Lösung für die Zustandswahrscheinlichkeiten.

Anhang 2: Zusammenfassung der Verzögerungsstationen

Hier wird gezeigt, daß - wie im Abschnitt 3.3.3.1 beschrieben - alle Verzögerungsstationen (Stationen des Typs 2) des Warteschlangennetzes gemäß Bild 3.5 in einer einzigen fiktiven Station zusammengefaßt werden können.

Nach Gleichung (3.6) gilt für die Zustandswahrscheinlichkeiten des Netzes:

$$p_e(x_1, x_{21}, \dots, x_{2R}, x_{31}, \dots, x_{3R}, x_{41}, \dots, x_{4R}) = C_e A_1^{x_1} \prod_{r=1}^R A_{2r}^{x_{2r}} \frac{A_{3r}^{x_{3r}}}{x_{3r}!} \frac{A_{4r}^{x_{4r}}}{x_{4r}!} \quad (A2.1)$$

Faßt man alle Zustände zusammen, für welche die Gesamtzahl der Anforderungen in allen Verzögerungsstationen gleich x_0 ist, so erhält man:

$$\begin{aligned} & P(X_1 = x_1, X_{21} = x_{21}, \dots, X_{2R} = x_{2R}, \sum_{S=3}^4 \sum_{r=1}^R X_{Sr} = x_0) \\ &= C_e A_1^{x_1} \prod_{r=1}^R A_{2r}^{x_{2r}} \sum_{x_{31} + \dots + x_{4R} = x_0} \prod_{r=1}^R \frac{A_{3r}^{x_{3r}}}{x_{3r}!} \frac{A_{4r}^{x_{4r}}}{x_{4r}!} \\ &= C_e A_1^{x_1} \prod_{r=1}^R A_{2r}^{x_{2r}} \frac{1}{x_0!} \sum_{x_{31} + \dots + x_{4R} = x_0} \binom{x_0}{x_{31}, x_{32}, \dots, x_{4R}} A_{31}^{x_{31}} A_{32}^{x_{32}} \dots A_{4R}^{x_{4R}} \\ &= C_e A_1^{x_1} \prod_{r=1}^R A_{2r}^{x_{2r}} \frac{1}{x_0!} (A_{31} + A_{32} + \dots + A_{4R})^{x_0} \\ &= C_e \frac{A_0^{x_0}}{x_0!} \prod_{r=1}^R A_{2r}^{x_{2r}} \quad (A_0 = \sum_{S=3}^4 \sum_{r=1}^R A_{Sr}) \quad (A2.2) \end{aligned}$$

Anhang 3: Umformung der Rekursion für die Zustandswahrscheinlichkeiten des Systems M/G/1,s

Gezeigt werden soll die Äquivalenz zwischen Gleichung (3.45) und Gleichung (3.43) im Abschnitt 3.4.3.2 .

a) $x = 0$

Gl.(3.45) ergibt mit der Definition in Gl.(3.44) für $x = 0$:

$$\begin{aligned} p^*(1) &= \frac{1}{a_0} \sigma_1 p^*(0) \\ &= \frac{1}{a_0} \sum_{j=1}^{\infty} a_j p^*(0) \\ &= \frac{1}{a_0} (1 - a_0) p^*(0) \end{aligned} \quad (A3.1)$$

Dies entspricht der ersten Beziehung in Gl.(3.43).

b) $x > 0$

Wir formen Gl.(3.45) mit Hilfe der Definition (3.44) sukzessive um:

$$\begin{aligned} p^*(x+1) &= \frac{1}{a_0} \left[\sigma_{x+1} p^*(0) + \sum_{j=1}^x \sigma_{x-j+2} p^*(j) \right] \\ &= \frac{1}{a_0} \left[\sum_{i=x+1}^{\infty} a_i p^*(0) + \sum_{j=1}^x \sum_{i=x-j+2}^{\infty} a_i p^*(j) \right] \\ &= \frac{1}{a_0} \left[\left(1 - \sum_{i=0}^x a_i\right) p^*(0) + \sum_{j=1}^x \left(1 - \sum_{i=0}^{x-j+1} a_i\right) p^*(j) \right] \\ &= \frac{1}{a_0} \left[(1 - a_1) p^*(x) - a_x p^*(0) - \sum_{j=1}^{x-1} a_{x-j+1} p^*(j) \right] \\ &\quad - \frac{1}{a_0} \underbrace{\left[a_0 p^*(x) + (1 - a_1) p^*(x-1) - a_{x-1} p^*(0) - \sum_{j=1}^{x-2} a_{x-1-j+1} p^*(j) \right]}_{= 0} \\ &\quad - \frac{1}{a_0} \underbrace{\left[a_0 p^*(x-1) + (1 - a_1) p^*(x-2) - a_{x-2} p^*(0) - \sum_{j=1}^{x-3} a_{x-2-j+1} p^*(j) \right]}_{= 0} \\ &\quad \vdots \end{aligned}$$

$$\begin{aligned}
 & - \frac{1}{a_0} \underbrace{[-a_0 p^*(2) + (1-a_1) p^*(1) - a_1 p^*(0)]}_{=0} \\
 & - \frac{1}{a_0} \underbrace{[-a_0 p^*(1) + (1-a_0) p^*(0)]}_{=0}
 \end{aligned}$$

$$p^*(x+1) = \frac{1}{a_0} [(1-a_1) p^*(x) - a_x p^*(0) - \sum_{j=1}^{x-1} a_{x-j+1} p^*(j)] \quad (A3.2)$$

Dies entspricht der zweiten Beziehung in Gl.(3.43).

Anhang 4: Marginalverteilung für einen endlich großen Speichermodul bei endlich großen Speichern der übrigen Moduln

Es wird hier der Beweis für Gl.(4.25) im Abschnitt 4.3.3.2 geführt. Wir nehmen an, daß als einziger der Modul Nr. μ einen endlich großen Speicher der Größe n_μ besitzt, während sämtliche anderen Moduln unendlich große Speicher besitzen. Somit können wir ein Warteschlangennetz mit zwei Anforderungsklassen betrachten:

- die Klasse Nr. μ mit maximal n_μ Anforderungen im gesamten Netz und
- die Klasse Nr. 0, welche alle Anforderungen der übrigen Klassen $m \in \{1, \dots, M\} \setminus \{\mu\}$ umfassen soll.

Mit den Bezeichnungen

$$\begin{aligned}
 X_{0;1m} &= X_{m;1m} \quad \text{für } m \in \{1, \dots, M\} \setminus \{\mu\} \\
 X_{0;sr} &= \sum_{m \in \{1, \dots, M\} \setminus \{\mu\}} X_{m;sr}
 \end{aligned} \quad (A4.1)$$

erhält man mit Hilfe des Theorems im Anhang 1 die folgenden Zustandswahrscheinlichkeiten:

$$\begin{aligned}
 & P(X_{0;11} = x_{0;11}, \dots, X_{\mu;1\mu} = x_{\mu;1\mu}, \dots, X_{0;1M} = x_{0;1M}, X_{0;21} = x_{0;21}, X_{\mu;21} = x_{\mu;21}, \dots \\
 & \dots, X_{0;2R} = x_{0;2R}, X_{\mu;2R} = x_{\mu;2R}, X_{0;31} = x_{0;31}, X_{\mu;31} = x_{\mu;31}, \dots, X_{0;3R} = x_{0;3R}, \\
 & X_{\mu;3R} = x_{\mu;3R}, X_{0;41} = x_{0;41}, X_{\mu;41} = x_{\mu;41}, \dots, X_{0;4R} = x_{0;4R}, X_{\mu;4R} = x_{\mu;4R}) \\
 &= p''(x_{0;11}, \dots, x_{\mu;1\mu}, \dots, x_{0;1M}, x_{0;21}, x_{\mu;21}, \dots, x_{0;2R}, x_{\mu;2R}, x_{0;31}, x_{\mu;31}, \dots \\
 & \dots, x_{0;3R}, x_{\mu;3R}, x_{0;41}, x_{\mu;41}, \dots, x_{0;4R}, x_{\mu;4R}) \\
 &= p''(\vec{x}'') \\
 &= C'' \prod_{m \in \{1, \dots, M\} \setminus \{\mu\}} A_{\mu;1m}^{x_{\mu;1m}} \prod_{r=1}^R \binom{x_{0;2r} + x_{\mu;2r}}{x_{0;2r}} A_{0;2r}^{x_{0;2r}} A_{\mu;2r}^{x_{\mu;2r}} \prod_{s=3}^4 \prod_{r=1}^R \frac{A_{0;sr}^{x_{0;sr}} A_{\mu;sr}^{x_{\mu;sr}}}{x_{0;sr}! x_{\mu;sr}!} \\
 & \sum_{\vec{x}'' \in Z''} p''(\vec{x}'') = 1
 \end{aligned} \quad (A4.2)$$

$$Z^\mu = \{ \vec{x}^\mu \mid x_{0;11}, \dots, x_{\mu;4R} \in \mathbb{N}_0 \wedge x_{\mu;1\mu} + \sum_{s=2}^4 \sum_{r=1}^R x_{\mu;sr} \leq n_\mu \}$$

$$A_{\mu;1\mu} = \lambda_\mu b_{1\mu} \quad A_{0;1m} = A_{m;1m} = \lambda_m b_{1m} \quad m \in \{1, \dots, M\} \setminus \{\mu\}$$

$$\left. \begin{aligned} A_{\mu;2r} &= \frac{p_{m;r}}{1-q_r} \lambda_\mu b_{2r} & A_{0;2r} &= \sum_{m \in \{1, \dots, M\} \setminus \{\mu\}} \frac{p_{m;r}}{1-q_r} \lambda_m b_{2r} \\ A_{\mu;3r} &= \frac{p_{m;r} q_r}{1-q_r} \lambda_\mu b_{3r} & A_{0;3r} &= \sum_{m \in \{1, \dots, M\} \setminus \{\mu\}} \frac{p_{m;r} q_r}{1-q_r} \lambda_m b_{3r} \\ A_{\mu;4r} &= p_{m;r} \lambda_\mu b_{4r} & A_{0;4r} &= \sum_{m \in \{1, \dots, M\} \setminus \{\mu\}} p_{m;r} \lambda_m b_{4r} \end{aligned} \right\} r \in \{1, \dots, R\}$$

Hieraus berechnet man die für die Klasse μ geltende Verteilung der Anforderungszahlen in der folgenden Weise:

$$\begin{aligned} & P(x_{\mu;1\mu} = x_{\mu;1\mu}, x_{\mu;21} = x_{\mu;21}, \dots, x_{\mu;2R} = x_{\mu;2R}, x_{\mu;31} = x_{\mu;31}, \dots \\ & \quad \dots, x_{\mu;3R} = x_{\mu;3R}, x_{\mu;41} = x_{\mu;41}, \dots, x_{\mu;4R} = x_{\mu;4R}) \\ &= P_\mu(x_{\mu;1\mu}, x_{\mu;21}, \dots, x_{\mu;2R}, x_{\mu;31}, \dots, x_{\mu;3R}, x_{\mu;41}, \dots, x_{\mu;4R}) \\ &= p_\mu(\vec{x}_\mu) \\ &= C^\mu A_{\mu;1\mu}^{x_{\mu;1\mu}} \sum_{x_{0;11}=0}^{\infty} A_{0;11}^{x_{0;11}} \dots \sum_{x_{0;1M}=0}^{\infty} A_{0;1M}^{x_{0;1M}} \prod_{r=1}^R [A_{\mu;2r}^{x_{\mu;2r}} \sum_{x_{0;2r}=0}^{\infty} \binom{x_{0;2r} + x_{\mu;2r}}{x_{0;2r}} A_{0;2r}^{x_{0;2r}}] \cdot \\ & \quad \cdot \prod_{s=3}^4 \prod_{r=1}^R \frac{A_{\mu;sr}^{x_{\mu;sr}}}{x_{\mu;sr}!} \prod_{s=3}^4 \prod_{r=1}^R \sum_{x_{0;sr}=0}^{\infty} \frac{A_{0;sr}^{x_{0;sr}}}{x_{0;sr}!} \\ &= C^\mu A_{\mu;1\mu}^{x_{\mu;1\mu}} \prod_{m \in \{1, \dots, M\} \setminus \{\mu\}} (1 - A_{0;1m})^{-1} \prod_{r=1}^R A_{\mu;2r}^{x_{\mu;2r}} (1 - A_{0;2r})^{-x_{\mu;2r}-1} \cdot \\ & \quad \cdot \prod_{s=3}^4 \prod_{r=1}^R \frac{A_{\mu;sr}^{x_{\mu;sr}}}{x_{\mu;sr}!} \prod_{s=3}^4 \prod_{r=1}^R e^{A_{0;sr}} \end{aligned} \quad (A4.3)$$

$$A_{0;1m} < 1 \quad (m \in \{1, \dots, M\})$$

$$A_{0;2r} < 1 \quad (r \in \{1, \dots, R\})$$

$$p_\mu(\vec{x}_\mu) = C_\mu A_{\mu;1\mu}^{x_{\mu;1\mu}} \prod_{r=1}^R \left(\frac{A_{\mu;2r}}{1 - A_{0;2r}} \right)^{x_{\mu;2r}} \prod_{s=3}^4 \prod_{r=1}^R \frac{A_{\mu;sr}^{x_{\mu;sr}}}{x_{\mu;sr}!} \quad (A4.4)$$

$$\sum_{\vec{x}_\mu \in Z_\mu} p_\mu(\vec{x}_\mu) = 1$$

$$Z_\mu = \{ \vec{x}_\mu \mid x_{\mu;1\mu}, \dots, x_{\mu;4R} \in \mathbb{N}_0 \wedge x_{\mu;1\mu} + \sum_{s=2}^4 \sum_{r=1}^R x_{\mu;sr} \leq n_\mu \}$$

$$A_{0;2r} < 1 \quad (r \in \{1, \dots, R\})$$

Anhang 5: Faltungsalgorithmus

Hier soll der an verschiedenen Stellen dieser Arbeit angewandte Faltungsalgorithmus kurz beschrieben werden [13,80].

Der Algorithmus dient zur numerischen Bestimmung von Ausdrücken der Form:

$$G(u,n) = \sum_{\vec{y} \in D(u,n)} \prod_{u=1}^u f_u(y_u)$$

mit $\vec{y} = (y_1, \dots, y_u)$ (A5.1)

$$D(u,n) = \{ \vec{y} \mid y_1, \dots, y_u \in \mathbb{N}_0 \wedge \sum_{u=1}^u y_u = n \}$$

Wie man sieht, gilt folgende rekursive Beziehung:

$$G(u,i) = \sum_{j=0}^i G(u-1, i-j) f_u(j) \quad (A5.2)$$

Der Algorithmus lautet daher in seiner allgemeinen Form:

(1) Setze für $i \in \{0, 1, \dots, n\}$

$$G(1,i) = f_1(i) \quad (A5.3)$$

(2) Berechne für $u \in \{2, 3, \dots, U\}$

und $i \in \{0, 1, \dots, n\}$ rekursiv

$$G(u,i) = \sum_{j=0}^i G(u-1, i-j) f_u(j) \quad (A5.4)$$

Liegen die Funktionen $f_u(y_u)$ in der Form

$$f_u(y_u) = \prod_{j=0}^{y_u} h_{uj} \quad (A5.5)$$

vor, so erfolgt die Berechnung des Schrittes (2) am besten mit Hilfe des Hornerschemas:

$$(2') \quad G(u,i) = h_{u0} [G(u-1,i) + h_{u1} [G(u-1,i-1) + h_{u2} [G(u-1,i-2) + h_{u3} [\dots \dots + h_{u,i-1} [G(u-1,1) + h_{u,i} G(u-1,0) \dots]]]] \quad (A5.6)$$

Besonders effektiv kann Schritt (2) durchgeführt werden, wenn

die Funktionen $f_u(y_u)$ die folgende Form besitzen ¹⁾:

$$f_u(y_u) = h_u^{y_u} \quad (A5.7)$$

Dann vereinfacht sich Schritt (2) folgendermaßen:

$$(2'') \quad G(u,i) = G(u-1,i) + h_u G(u,i-1) \quad (A5.8)$$

$$(G(u,-1) = 0)$$

¹⁾ Dies ist zum Beispiel der Fall für eine Station vom Typ 1 mit einer Bedienungseinheit und einer Anforderungsklasse in einem Warteschlangennetz gemäß Anhang 1.

