University of Würzburg
Institute of Computer Science
Research Report Series

A Numerical Analysis of the Geo/D/N Queueing System

N. Vicari and P. Tran-Gia

Report No. 151 September 1996

Institute of Computer Science, University of Würzburg
Am Hubland, 97074 Würzburg, Germany
Tel.: +49 931 888 5505, Fax: +49 931 888 4601
e-mail: vicari@informatik.uni-wuerzburg.de

Abstract

In this paper we present an analysis of the discrete-time $Geo/D/N - \infty$ *queue. This model has been recently applied as an asymptotic model in performance studies of various modern telecommunication systems. The distribution of the number of customers at arbitrary discrete time instants and the corresponding arrival distribution are derived. Using these results the waiting time distribution is obtained. The results are of an exact nature and the presented algorithm is of low complexity with respect to computation time and memory usage.*

1 Introduction

In the performance evaluation of modern telecommunication systems basic discrete-time queueing systems are frequently applied for system modeling. A special class of queueing systems used to investigate the burst scale performance of ATM multiplexers is formed by queueing models with multiple deterministic servers ([Kam96], [GBSM96]). In this paper we focus on the discrete-time $Geo/D/N-\infty$ queueing system.

While analytical approaches for single-server systems are well-known, solutions for multi-server systems are rarely published. Performance measures for single-server systems of the $GI/G/1$-class can be derived by applying the Lindley Integral method ([Lin52], [Coh69]). This method, however, is not extensible to multi-server systems since the number of states required to keep track of the remaining work grows exponentially with the number of servers. As a special case, the $Geo/D/1-\infty$ queue is analyzed in [TG96] applying the principle of the embedded Markov chain to a model with geometric inter-arrival time distribution.

Multi-server systems with deterministic service time are investigated e.g. in [Kam96]. The steady-state probabilities for $G^{[X]}/D/C/K$-type queues with synchronous service are derived using a matrix geometrical approach. The assumption of synchronous service reduces the number of states to describe the system considerably, allowing efficient computation. The relation between the buffer content and the waiting time in multi-server systems is investigated in [XBS96]. For systems having a deterministic service time equal to one time unit explicit relations between the delay and buffer content are derived.

In this paper we adopt Crommelin's [Cro33] approach for the $M/D/N-\infty$ queue to analyze the system state and waiting time. A survey of this method and approximations for similar queueing systems are given in [Tij86] and [VH83]. The paper is organized as follows: Section 2 provides a description of the $Geo/D/N-\infty$ queueing system and the notation used. Section 3 describes our analytical approach for this system. In Section 4 we present some numerical examples and the paper is summarized in Section 5.

2 The $\mathrm{Geo/D/N}-\infty$ Queueing Model

We first describe the $Geo/D/N-\infty$ queueing model under consideration, as shown in Fig. 1. All events in the model, such as arrivals or service completions, are assumed to occur at discrete time instants. The constant duration between two time instants will be referred to as time unit. Due to the discrete-time nature of the model events can occur simultaneously. Thus, in order to deal with simultaneous events in a discretized time instant priorities have to be defined. We assume that service completions always occur before arrivals, and the service is started immediately after an arrival if a server is idle.

In the following we denote the random variables used to describe the system with capital letters and the corresponding distributions with small letters. The system consists of N servers with a deterministic service time of D time units and an infinite queue. The servers are not synchronized, i.e. an idle server starts with the service immediately and independently of the other servers upon a customer arrival. Customers which arrive when

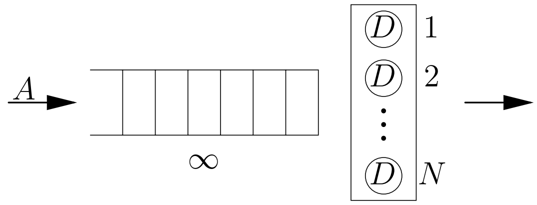


Figure 1: $Geo/D/N-\infty$ queueing model

all servers are busy are queued in a FIFO order. The inter-arrival time A is geometrically distributed:

$$a(k) = p(1-p)^{k-m}, \qquad k \geq m, \tag{1}$$

which will be referred to as $Geo(m)$. According to the arguments in the next section, the two cases $m=0$ and $m=1$ are taken into account:

- $m=0$: For $m=0$ the arrival process will be referred to as $Geo(0)$. At a discrete-time instant a customer arrives with probability p and with probability $1-p$ no arrival occurs. The time to the next arrival can be 0. Therefore, the $Geo(0)$ process includes batch arrivals.
- $m=1$: The arrival process with parameter $m=1$ will be referred to as $Geo(1)$ or Bernoulli process. Due to the shift of $m=1$ a minimal distance of one time unit between two arrivals is given. The process allows no batch arrivals.

Thus, the analysis below will deal with two delay systems: $Geo(0)/D/N$ and $Geo(1)/D/N$.

3 Performance Analysis

3.1 Memoryless Property in Discrete-Time

Before presenting the analysis we describe properties of the geometric arrival process, which are preliminary to the analysis. For the analysis of the $Geo/D/N-\infty$ system we exploit the memoryless property of the discrete-time arrival process for the derivation of the system state immediately before an arrival. Let the random variable X denote the time until the next customer arrives. The equation

$$P\{X \leq t, X > t_0\} = P\{X \leq (t-t_0)\} \tag{2}$$

states the memoryless property since the probability that the next arrival falls into the interval (t_0, t) is independent of t_0 (see [Pap84]), assuming that the arrival does not occur prior to t_0. This property is only fulfilled by processes where the recurrence time distribution equals the inter-arrival time distribution.

In the continuous-time domain the Markov process is memoryless. In the discrete-time domain, however, two processes have the memoryless property, depending on the observation instant [TG96]:

- The $Geo(0)$ process has the memoryless property for observation instants immediately before the discrete-time instants.
- Given observation instants immediately after the discrete time points, the $Geo(1)$ process has the memoryless property.

3.2 Customer Arrivals During a Random Period

Before presenting the computation of the system state and the waiting time distributions we first derive the number of customer arrivals $\Gamma(u)$ during a given time interval with the corresponding distribution $\gamma(u, k)$. Again the two arrival processes considered have to be treated separately.

3.2.1 Geometric Inter-Arrival Distribution

For the $Geo(0)$ process a closed form of the distribution of arrivals in a random period is not available. Therefore, we derive this distribution numerically.

The number of customer arrivals within a given interval is derived in [TG96] for arbitrary distributed inter-arrival times. In an interval of u time units we observe k arrivals if the sum of the first k inter-arrival times is at most u and the sum of the first $(k+1)$ inter-arrival times is larger than u, cf. Fig. 2.

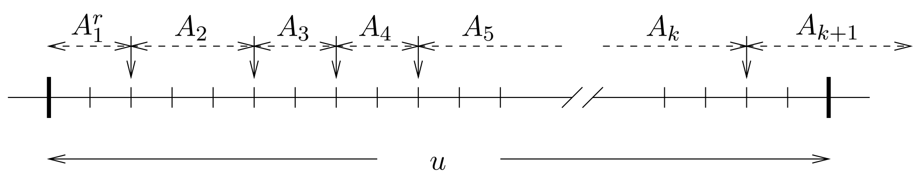


Figure 2: Arrivals during an interval of u time units

The first time interval A_1^r is equal to the time from the beginning of the interval until the first arrival occurs. The distribution of A_1^r is the forward recurrence time of the inter-arrival time distribution A with an observation instant immediately after the event. According to [TG96], this recurrence time distribution of the $Geo(0)$ process is given by:

$$a^r(k) = a(k-1), \qquad k = 1, 2, \ldots \tag{3}$$

The distribution corresponding the sum of k inter-arrival times is expressed by the k-fold convolution of the inter-arrival time distribution and is denoted by $a^{*k}(i)$. Thus the distribution of the number $\Gamma(u)$ of arrivals in an interval of u time units is given by:

$$\gamma(0, k) = \delta(k) = \begin{cases} 1, & k = 0 \\ 0, & k = 1, 2, \ldots \end{cases}, \tag{4}$$

for $u = 0$ and for $u > 0$ by:

$$\gamma(u,0) = 1 - \sum_{j=0}^{u} a^r(j), \tag{5a}$$

$$\gamma(u,k) = \sum_{i=0}^{u} \left[\left(a^r(i) * a^{*(k-1)}(i) \right) \cdot \left(1 - \sum_{j=0}^{u-i} a(j) \right) \right], \qquad k = 1, 2, \ldots. \tag{5b}$$

3.2.2 Bernoulli Arrival Process

In the case of the Bernoulli arrival process the distribution of the number of arrivals during u time units can be expressed in closed-form. With the mean $E[A]$ of the $Geo(1)$ process, the probability p for an arrival in one time unit is:

$$p = \frac{1}{E[A]}. \tag{6}$$

The number of arrivals during u time units follows a binomial distribution with parameters u and p:

$$\gamma(u,k) = \binom{u}{k} p^k (1-p)^{u-k}, \qquad k = 0, 1, \ldots, u. \tag{7}$$

3.3 Computation of the System State Distribution

Let the random variable Z describe the number of customers in the system at an arbitrary time t. The system state Z' at time $t + D$ is derived by the following considerations (according Crommelin [Cro33]). The number of customers in service at time t will have left the system at time $t + D$ while during the D time units $\Gamma(D)$ new customers will arrive. Therefore, the system state Z' at time $t + D$ is given by:

$$Z' = \begin{cases} \Gamma(D), & Z < N \\ Z - N + \Gamma(D), & Z \geq N \end{cases}. \tag{8}$$

For the corresponding distributions we obtain the following equation:

$$z'(k) = w \cdot \gamma(D,k) + (1-w) \cdot (\gamma(D,k) * \tilde{z}(k)). \tag{9}$$

The probability w and the distribution $\tilde{z}(k)$ are given by:

$$w = \sum_{i=0}^{N} z(i), \tag{10a}$$

$$\tilde{z}(k) = \begin{cases} 0, & k = 0 \\ \frac{z(N+k)}{1-w}, & k > 0 \end{cases}. \tag{10b}$$

Using this result the steady-state distribution $z(k)$ of the number of customers in the system at an arbitrary time instant t can be derived iteratively. In the following we describe the relation between the system state Z at an arbitrary time and at observation instants Z^* immediately before customer arrivals. The system state immediately before customer arrivals is used for the derivation of the waiting time distribution. Due to the different properties of the $Geo(0)$ and $Geo(1)$ arrival process, we consider two cases.

3.3.1 Geometric Inter-Arrival Distribution

Since the $Geo(0)$ process with observation instants immediately before customer arrivals is the discrete-time equivalent to the memoryless Poisson process, the BASTA (Bernoulli Arrivals See Time Averages) principle (cf. [BP88], [Hal83]) applies. Therefore, the system state distribution at an arbitrary time instant $z(k)$ is equal to the distribution $z^*(k)$ of the number of customers in the system immediately before a customer arrival.

3.3.2 Bernoulli Arrival Process

As discussed above the Bernoulli arrival process has the memoryless property for observation instants immediately after the discrete-time instants. Applying the BASTA principle, the system state at an arbitrary time instant is identical to the state immediately after an observation instant. At an observation instant the number of customers is increased by one with probability p (cf. Eqn. (6)), while with probability $(1-p)$ the number of customers keeps constant. Thus a simple relationship between the system state before and after observation instants is given. The state probabilities for an observation instant before an arrival are derived by:

$$\begin{aligned} z^*(0) &= \frac{z(0)}{1-p}, \\ z^*(k) &= \frac{1}{1-p} \cdot \left(z(k) - p \cdot z^*(k-1) \right), \qquad k = 1, 2, 3, \ldots . \end{aligned} \tag{11}$$

3.4 Waiting Time Distribution

A waiting time of l time units can be expressed, using modulo D arithmetic, by:

$$l = m \cdot D + r, \qquad 0 \le r < D. \tag{12}$$

The number of customers in the system before the arrival of a marked customer is denoted by $Z^* = Z_0^*$. The amount of those customers remaining in the system t time units later is denoted by Z_t^* (not taking into account arrivals during this interval). If all servers are busy and work continuously $m \cdot N$ customers are served in $m \cdot D$ time units.

Assume that the marked customer has a waiting time of at most x time units. Thus, l time units upon the arrival of this customer at most $N-1$ of the customers in the system prior to the arrival of the marked customer remain in the system. Taking in account continuous service we notice that r time units upon the arrival of the marked customer at most $Z_r^* = mN + N - 1$ of the Z_0^* customers in the system before the arrival remain in the system.

$$\begin{aligned} W \le l &\Leftrightarrow Z_l^* \le N - 1 \\ &\Leftrightarrow Z_r^* \le mN + N - 1. \end{aligned} \tag{13}$$

We define the probability $b_\nu(r)$ that r time units upon the arrival of a customer at most ν of the customers in the system before the arrival will remain in the system:

$$b_\nu(r) = P\{Z_r^* \le \nu\}. \tag{14}$$

The waiting time distribution function is given by:

$$P\{W \le l\} = b_{mN+N-1}(r), \qquad l = mD + r, \quad 0 \le r < D. \tag{15}$$

For the computation of the waiting time distribution the probabilities $b_\nu(r)$ have to be derived. Therefore the following relation is exploited:

$$\begin{aligned} &P\{\text{at most j customers are in the system at time } t\} \\ &= \sum_{k=0}^{j} P\{\text{at most } (j-k) \text{ customers from time } (t-r) \text{ remain at} \\ &\qquad \text{time } t \text{ in the system and } k \text{ arrivals occur in the interval } (t-r, t]\}. \end{aligned} \tag{16}$$

With $t \to \infty$ and the memoryless property of the arrival process the following equation is obtained:

$$P\{Z^* \le j\} = \sum_{k=0}^{j} z^*(k) = \sum_{k=0}^{j} b_{j-k}(r) \cdot \gamma(r,k), \qquad j = 0, 1, 2, \ldots \tag{17}$$

Solving this equation in a recursive manner (for $0 \le r < D$) we arrive at:

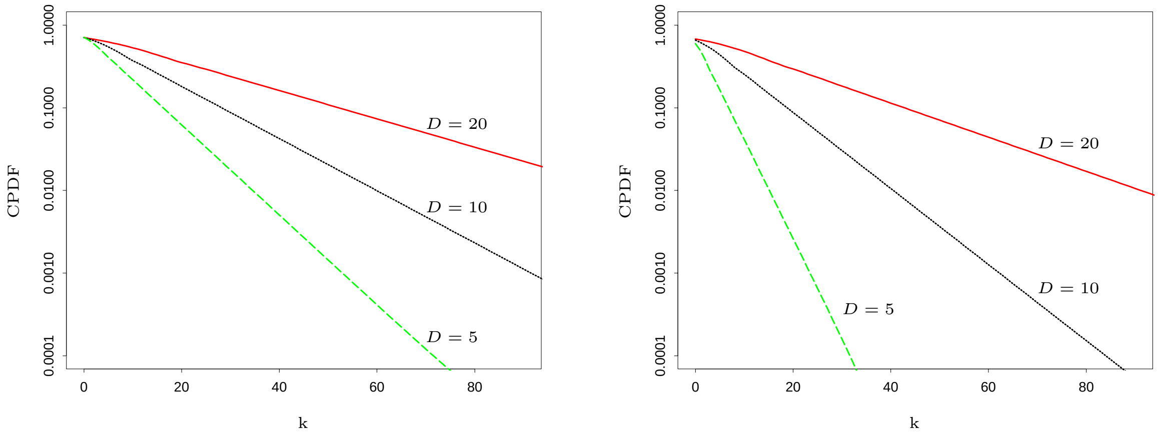
$$b_0(r) = \gamma(r,0)^{-1} \cdot z^*(0), \qquad j = 0, \tag{18a}$$

$$b_j(r) = \gamma(r,0)^{-1} \cdot \left(\sum_{k=0}^{j} z^*(k) - \sum_{k=1}^{j} b_{j-k}(r) \cdot \gamma(r,k), \right), \qquad j = 1, 2, 3, \ldots \tag{18b}$$

4 Numerical Examples

In this section we present numerical examples for the $Geo(0)/D/N$ and the $Geo(1)/D/N$ delay systems. Both systems are stable for load ρ:

$$\rho = \frac{D}{N \cdot E[A]} \le 1. \tag{19}$$

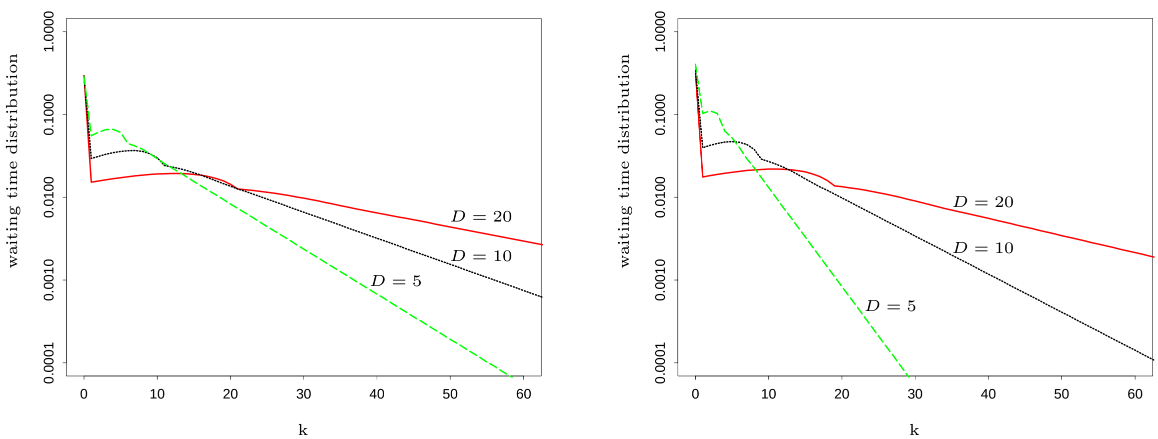


(a) *Geo*(0) inter-arrival times

(b) Bernoulli arrival process

Figure 3: Complementary waiting time distribution functions (CPDF)

In Fig. 3 the complementary waiting time distribution function for systems with $N = 2$ servers and a load of $\rho = 0.8$ is depicted. The service time D is 5, 10 and 20 time units. The arrival rate is derived according to Eqn. (19). With higher service rate we obtain a steeper descend of the distributions due to the higher granularity of the system. The waiting time distributions of the system with Bernoulli arrival process decline faster than those of the system with $Geo(0)$ arrival process. This effect is caused by the batch arrivals of the $Geo(0)$ process.

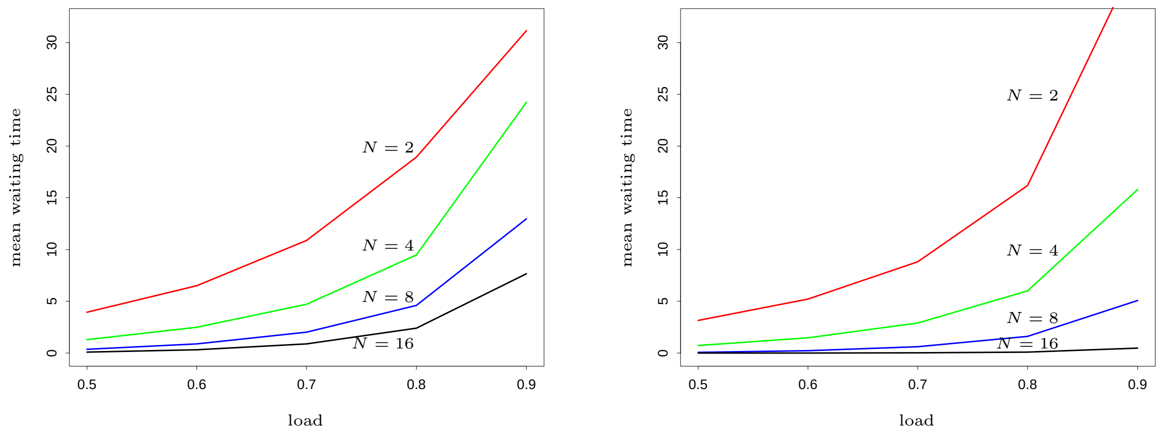


(a) *Geo*(0) inter-arrival times

(b) Bernoulli arrival process

Figure 4: Waiting time distributions

The curves in Fig. 4 render the waiting time distribution for the above mentioned parameter set. The probabilities for a customer waiting between 1 and D time units show a local increase, which could not be modeled using combinatory methods. Only the computation using the modulo D remainder of the waiting time (Eqn. (15)) reflects this property of the waiting time distribution.



(a) *Geo*(0) inter-arrival times

(b) Bernoulli arrival process

Figure 5: Mean waiting time

Figure 5 shows the mean waiting time as a function of the load. We consider systems with $N = 2$, 4, 8, and 16 servers. The service time is set to $D = 20$ time units. Again, the inter-arrival time is derived according to Eqn. (19). For both of the considered arrival processes the waiting time decreases if the number of servers is increased. This is the well-known effect of economy of scale. For the Bernoulli arrival process, the mean waiting time starts to increase faster at higher load than the system with the $Geo(0)$ distributed inter-arrival times.

5 Conclusion

Algorithms to analyze delay systems of the type $Geo(0)/D/N$ and $Geo(1)/D/N$ have been presented. The algorithm is an discrete-time extension of the method of Crommelin [Cro33] for the analysis of the $M/D/N - \infty$ queueing system. The iterative algorithm is numerically stable and converges fast. It allows to compute waiting time probabilities in the range of 10^{-10} with a low computational complexity.

References

[BP88] O. J. Boxma and Groenendijk W. P. Waiting Times in Discrete-Time Cyclic-Service Systems. *IEEE Transactions on Communications*, 36(2):164 –170, February 1988.

[Coh69] C. W. Cohen. *The Single Server Queue*. North Holland, 1969.

[Cro33] C. D. Crommelin. Delay Probability Formulae. *Post Office Electrical Engeneer's Journal*, 26:266–274, 1933.

[GBSM96] A. Gravey, J. Boyer, K. Sevilla, and J. Mignault. Resource allocation for Worst Case Traffic in ATM networks. *to be published in Performance Evaluation*, 1996.

[Hal83] S. Halfin. Batch Delays Versus Customer Delays. *The Bell Sysyem Technical Journal*, 62(7):2011–2015, September 1983.

[Kam96] A. Kamal. Efficient Solution of Multiple Server Queues with Application to the Modeling of ATM Concentrators. *Proceedings of the IEEE Infocom '96, San Francisco, CA*, pages 248–254, March 1996.

[Lin52] D.V. Lindley. The Theory of Queues with a Single Server. *Proc. of the Cambridge Philosophical Society*, 48:277–289, 1952.

[Pap84] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1984.

[TG96] P. Tran-Gia. *Analytische Leistungsbewertung verteilter Systeme*. Springer, 1996.

[Tij86] H. C. Tijms. *Stochastic Modelling and Analysis, a computational approach*. Wiley, 1986.

[VH83] M. H. Van Hoorn. *Algorithms and Approximations for Queueing Models*. PhD thesis, University of Amsterdam, 1983.

[XBS96] Y. Xiong, H. Bruneel, and B. Steyaert. Deriving Delay Characteristics from Queue Length Statistics in Discrete Time Queues with Multiple Servers. *Performance Evaluation*, (24):189–204, 1996.

Verantwortlich: Die Vorstände des Institutes für Informatik.

[100] U. Hertrampf, H. Vollmer und K. W. Wagner. *On the Power of Number-Theoretic Operations with Respect to Counting.* Januar 1995.

[101] O. Rose. *Statistical Properties of MPEG Video Traffic and their Impact on Traffic Modeling in ATM Systems.* Februar 1995.

[102] M. Mittler und R. Müller. *Moment Approximation in Product Form Queueing Networks.* Februar 1995.

[103] D. Rooß und K. W. Wagner. *On the Power of Bio-Computers.* Februar 1995.

[104] N. Gerlich und M. Tangemann. *Towards a Channel Allocation Scheme for SDMA-based Mobile Communication Systems.* Februar 1995.

[105] A. Schömig und M. Kahnt. *Vergleich zweier Analysemethoden zur Leistungsbewertung von Kanban Systemen.* Februar 1995.

[106] M. Mittler, M. Purm und O. Gihr. *Set Management: Synchronization of Prefabricated Parts before Assembly.* März 1995.

[107] A. Schömig und M. Mittler. *Autocorrelation of Cycle Times in Semiconductor Manufacturing Systems.* März 1995.

[108] A. Schömig und M. Kahnt. *Performance Modelling of Pull Manufacturing Systems with Batch Servers and Assembly-like Structure.* März 1995.

[109] M. Mittler, N. Gerlich und A. Schömig. *Reducing the Variance of Cycle Times in Semiconductor Manufacturing Systems.* April 1995.

[110] A. Schömig und M. Kahnt. *A note on the Application of Marie's Method for Queueing Networks with Batch Servers.* April 1995.

[111] F. Puppe, M. Daniel und G. Seidel. *Qualifizierende Arbeitsgestaltung mit tutoriellen Expertensystemen für technische Diagnoseaufgaben.* April 1995.

[112] G. Buntrock, und G. Niemann. *Weak Growing Context-Sensitive Grammars.* Mai 1995.

[113] J. García and M. Ritter. *Determination of Traffic Parameters for VPs Carrying Delay-Sensitive Traffic.* Mai 1995.

[114] M. Ritter. *Steady-State Analysis of the Rate-Based Congestion Control Mechanism for ABR Services in ATM Networks.* Mai 1995.

[115] H. Graefe. *Konzepte für ein zuverlässiges Message-Passing-System auf der Basis von UDP.* Mai 1995.

[116] A. Schömig und H. Rau. *A Petri Net Approach for the Performance Analysis of Business Processes.* Mai 1995.

[117] K. Verbarg. *Approximate Center Points in Dense Point Sets.* Mai 1995.

[118] K. Tutschku. *Recurrent Multilayer Perceptrons for Identification and Control: The Road to Applications.* Juni 1995.

[119] U. Rhein-Desel. *Eine „Übersicht“ über medizinische Informationssysteme: Krankenhausinformationssysteme, Patientenaktensysteme und Kritiksysteme.* Juli 1995.

[120] O. Rose. *Simple and Efficient Models for Variable Bit Rate MPEG Video Traffic.* Juli 1995.

[121] A. Schömig. *On Transfer Blocking and Minimal Blocking in Serial Manufacturing Systems — The Impact of Buffer Allocation.* Juli 1995.

[122] Th. Fritsch, K. Tutschku und K. Leibnitz. *Field Strength Prediction by Ray-Tracing for Adaptive Base Station Positioning in Mobile Communication Networks.* August 1995.

[123] R. V. Book, H. Vollmer und K. W. Wagner. *On Type-2 Probabilistic Quantifiers.* August 1995.

[124] M. Mittler, N. Gerlich, A. Schömig. *On Cycle Times and Interdeparture Times in Semiconductor Manufacturing.* September 1995.

[125] J. Wolff von Gudenberg. *Hardware Support for Interval Arithmetic - Extended Version.* Oktober 1995.

[126] M. Mittler, T. Ono-Tesfaye, A. Schömig. *On the Approximation of Higher Moments in Open and Closed Fork/Join Primitives with Limited Buffers.* November 1995.

[127] M. Mittler, C. Kern. *Discrete-Time Approximation of the Machine Repairman Model with Generally Distributed Failure, Repair, and Walking Times.* November 1995.

[128] N. Gerlich. *A Toolkit of Octave Functions for Discrete-Time Analysis of Queuing Systems.* Dezember 1995.

[129] M. Ritter. *Network Buffer Requirements of the Rate-Based Control Mechanism for ABR Services.* Dezember 1995.

[130] M. Wolfrath. *Results on Fat Objects with a Low Intersection Proportion.* Dezember 1995.

[131] S. O. Krumke and J. Valenta. *Finding Tree–2–Spanners.* Dezember 1995.

[132] U. Hafner. *Asymmetric Coding in (m)-WFA Image Compression.* Dezember 1995.

[133] M. Ritter. *Analysis of a Rate-Based Control Policy with Delayed Feedback and Variable Bandwidth Availability.* January 1996.

[134] K. Tutschku and K. Leibnitz. *Fast Ray-Tracing for Field Strength Prediction in Cellular Mobile Network Planning.* January 1996.

[135] K. Verbarg and A. Hensel. *Hierarchical Motion Planning Using a Spatial Index.* January 1996.

[136] Y. Luo. *Distributed Implementation of PROLOG on Workstation Clusters.* February 1996.

[137] O. Rose. *Estimation of the Hurst Parameter of Long-Range Dependent Time Series.* February 1996.

[138] J. Albert, F. Räther, K. Patzner, J. Schoof, J. Zimmer. *Concepts For Optimizing Sinter Processes Using Evolutionary Algorithms.* February 1996.

[139] O. Karch. *A Sharper Complexity Bound for the Robot Localization Problem.* June 1996.

[140] H. Vollmer. *A Note on the Power of Quasipolynomial Size Circuits.* June 1996.

[141] M. Mittler. *Two-Moment Analysis of Alternative Tool Models with Random Breakdowns.* July 1996.

[142] P. Tran-Gia, M. Mandjes. *Modeling of customer retrial phenomenon in cellular mobile networks.* July 1996.

[143] P. Tran-Gia, N. Gerlich. *Impact of Customer Clustering on Mobile Network Performance.* July 1996.

[144] M. Mandjes, K. Tutschku. *Efficient call handling procedures in cellular mobile networks.* July 1996.

[145] N. Gerlich, P. Tran-Gia, K. Elsayed. *Performance Analysis of Link Carrying Capacity in CDMA Systems.* July 1996.

[146] K. Leibnitz, K. Tutschku, U. Rothaug. *Künstliche Neuronale Netze für die Wegoptimierung in ATG Leiterplattentestern.* Juli 1996.

[147] M. Ritter. *Congestion Detection Methods and their Impact on the Performance of the ABR Flow Control Mechanism.* August 1996.

[148] H. Baier Saip, K.W. Wagner. *The Analytic Polynomial Time Hierarchy.* September 1996.

[149] H. Vollmer, K.W. Wagner. *Measure One Results in Computational Complexity Theory.* September 1996.

[150] O. Rose. *Discrete-time Analysis of a Finite Buffer with VBR MPEG Video Traffic Input.* September 1996.

[151] N. Vicari, P. Tran-Gia. *A Numerical Analysis of the $Geo/D/N$ Queueing System.* September 1996.