

Efficient Computation of Waiting Time Moments for the DBMAP/G/1/N Queue*

Christoph Herrmann

Philips GmbH Research Laboratories

Weisshausstr. 2, P.O. Box 500145, 52085 Aachen

e-mail: herrmann@pfa.research.philips.com

In the recent past, discrete-time queueing models of the DBMAP/G/1/N-type (D[B]MAP: Discrete [Batch] Markovian Arrival Process) were developed, which allow for computation of per-stream loss probabilities, one type of solution based on the method of the unfinished work, and the other using the M/G/1-paradigm. The method of the unfinished work (applicable for deterministic service time only) provides – besides formulae for computing per-stream loss probabilities – also expressions for obtaining per-stream waiting-time probability functions, and therefore arbitrary moments of the waiting-time of each stream [7]. What is missing up to now (see also [3]), is an algorithmic recipe to compute waiting-time moments for DBMAP/G/1/N queueing system. This paper derives the z -transform of the actual waiting-time and derives a fast algorithm to determine the waiting-time probability function for an arbitrary renewal service time. The framework presented allows dealing with a superposition DMAP+DBMAP as an input process. For both streams in the superposition, per-stream waiting-time probability functions are presented for the case of a deterministic service time, which is relevant for modelling of ATM networking (both wired and wireless). Together with findings for the continuous-time case ([2, 12]), this paper completes the insights into finite queueing systems of the M/G/1-type. In comparison with the method of the unfinished work, the M/G/1-paradigm provides much faster algorithms to compute loss probabilities and waiting-time moments, due to the smaller system matrix. Note that the D(B)MAP has proved a versatile stochastic process, which can also be tuned to represent periodic correlation functions [8] and not only geometrically decaying ones.

Keywords DMAP, DBMAP, MAP, BMAP, discrete-time queue, finite buffer, waiting-time probability function, z -transform of the waiting-time probability function.

1. Introduction

So far there is no expression for the z -transform of the waiting-time probability function in discrete-time M/G/1-type queueing systems. [2] presents the solution for the MAP/G/1/N-case, which is continuous-time. The essential trick there is a decomposition of the series $[\mathbf{I}s + (\mathbf{C} + z\mathbf{D})]^{-1}$ into a sum of coefficients in z . However, this is not applicable to the general case of the BMAP/G/1/N

*The author's daily work on wireless ATM networks is supported by: The Federal Ministry of Education, Science, Research, and Technology, Germany

queue, where this sum becomes $[\mathbf{I}s + \sum_{k=0}^{\infty} \mathbf{D}_k z^k]^{-1}$, neither is it helpful for obtaining the z -transform of the waiting-time of the DBMAP/G/1/ N -queue¹. This paper provides an algorithm for both computing the z -transform of the (actual) waiting-time of the above queueing system and for obtaining the waiting-time's probability function. The results are so general that also the superposition of two input streams DMAP+DBMAP can be considered, the superposition of which is again a DBMAP. For this superposition per-stream expressions for the waiting-time probabilities are presented in case of a deterministic service time. Together with [7], which is treating per-stream loss probabilities in the DBMAP/G/1/ N -queue, now the full analysis of such discrete time queues is available. With respect to performance evaluation of *wireless* ATM networking, waiting-time moments are of more interest since they determine the expected delay, and delay jitter, when loss probability is sufficiently reduced due to Forward Error Correction.

2. The Discrete Batch Markovian Arrival Process (DBMAP)

We recall the definition of the DBMAP (Discrete time analogue of the Batch Markovian Arrival process due to Lucantoni [11]) given in [4]. Consider an m -state Markov chain (MC). A transition from state $S(t) = i$ to state $S(t+1) = j$, $t = 0, 1, \dots$, happens according to the transition probabilities $[\mathbf{D}_w]_{ij}$ (\mathbf{D}_w : transition matrix, $[\cdot]_{ij}$: entry in the i -th row and j -th column) $w = 0, 1, 2, \dots, b_{max}$, $\mathbf{D}_w := \mathbf{0}$ for $w > b_{max}$, and a transition is interpreted as an arrival of a w -batch (batch of size w), if $w > 0$ and there is no arrival if $w = 0$. Thus the model generates interarrival times with batch arrivals possible and successive interarrival times are not independent. In order to distinguish from other state random variables, the m states of the DBMAP are usually referred to as *phases*. The MC has *different types* of transitions between two phases i and j , thus extending the conventional definition of a MC. $\widehat{\mathbf{D}} := \sum_{w=0}^{\infty} \mathbf{D}_w$ is a stochastic matrix and represents the (conventional) MC which only considers the transitions between two phases no matter what type of transition. Let $\underline{\pi}$ denote the stationary phase probability vector $\underline{\pi} \widehat{\mathbf{D}} = \underline{\pi}$. It can be proved that the point process of the arrival instants is Semi-Markovian with the following SM kernel (with T_n denoting the discrete-time instant of the n -th arrival)

$$P\{S(T_{n+1}) = j, T_{n+1} - T_n = k | S(T_n) = i\} = \left[\mathbf{D}_0^{k-1} \cdot (\widehat{\mathbf{D}} - \mathbf{D}_0) \right]_{ij}.$$

The embedded MC has the transition matrix $(\mathbf{I} - \mathbf{D}_0)^{-1}(\widehat{\mathbf{D}} - \mathbf{D}_0)$. With $\underline{p} = \underline{\pi}(\widehat{\mathbf{D}} - \mathbf{D}_0)/[\underline{\pi}(\mathbf{I} - \mathbf{D}_0)\underline{e}]$ denoting its stationary phase probability vector at

¹[1] derives the Laplace transform of the waiting-time distribution function, which however leaves open the decomposition of $[\mathbf{I}s + \sum_{k=0}^{\infty} \mathbf{D}_k z^k]^{-1}$, which corresponds to $[\mathbf{R}(z) + s\mathbf{I}]^{-1}$ in the N-process notation used there, and confines to the special case of the Markov Modulated Poisson Process, for which this decomposition can be derived. Note that the differentiation approach proposed here can also be applied to the decomposition of $[\mathbf{I}s + \sum_{k=0}^{\infty} \mathbf{D}_k z^k]^{-1}$. Thus, the general expression for the Laplace transform of the waiting-time distribution function of the BMAP/G/1/ N is now possible.

arrival instants², we get for the mean value of the time between successive arrival instants³ $E[A] = \underline{P} \sum_{k=1}^{\infty} k \mathbf{D}_0^{k-1} (\widehat{\mathbf{D}} - \mathbf{D}_0) \underline{e} = \underline{P} (\mathbf{I} - \mathbf{D}_0)^{-1} \underline{e}$. \underline{e} denotes the column vector of 1's, $\underline{e} = (1, 1, \dots, 1)^T$, m components. Note that $E[A]^{-1} \neq \lambda := \underline{\pi} \sum_{w=1}^{\infty} w \mathbf{D}_w \underline{e}$, where λ is the arrival rate. In the special case of the DMAP ($b_{max} = 1$) it is the custom to write: $\mathbf{C} := \mathbf{D}_0$, and $\mathbf{D} := \mathbf{D}_1$.

2.1. The counting process of a DBMAP

Let $N(0, t]$ denote the number of arrivals in the interval $(0, t]$ and define

$$[\mathbf{P}(w, t_0)]_{ij} := P\{N(0; t_0] = w, S(t_0) = j \mid N(0; 0] = 0, S(0) = i\} \quad .$$

The following relations can be proved⁴ (see e.g. [7]):

$$\mathbf{P}(w, 0) = \mathbf{1}_{w=0} \cdot \mathbf{I} \quad , \quad \mathbf{I} : (m \times m)\text{-matrix of unity} \quad .$$

$$\mathbf{P}(w, t + 1) = \sum_{v=0}^w \mathbf{P}(v, t) \mathbf{D}_{w-v} \quad , \quad (1)$$

$$\sum_{w=0}^{\infty} \mathbf{P}(w, t) z^w = \left[\sum_{w=0}^{b_{max}} \mathbf{D}_w z^w \right]^t =: \mathbf{D}(z) \quad \text{for all } t \in \mathbb{N} \quad . \quad (2)$$

In [8], DMAPs are parameterized in such a way that *periodic* correlation functions for the count process or the interarrival times are obtained. The key idea is using a periodic phase transition matrix for the DMAP.

3. The analysis of the queueing system

We consider a FIFO queue with a buffer of N places (total capacity: $N + 1$). The interarrival times are produced according to a DBMAP, the service times are i.i.d. (renewal) with the probability function $h(t)$. Due to the properties of the DBMAP, the occupancy at successive time instants immediately after a departure results in an *embedded* Markov Renewal process. (It is an embedded MRP, since the state the MRP rests between two departures is not equal to the state of the queueing system in terms of the occupancy between departure epochs, since in between arrivals can happen.)

The resulting Semi-Markov kernel $\tilde{\mathbf{q}}(t)$ is of M/G/1 type [12]. By $\tilde{\mathbf{q}} := \sum_{t=1}^{\infty} \tilde{\mathbf{q}}(t)$ we denote the embedded Markov chain (MC) of the MRP. The stationary state probabilities of this MC are needed for computing the probability function of the occupancy at arbitrary and arrival instants, as well as for the probability function of the waiting-time. Furthermore the MRP characterizes the output process in terms of the interdeparture times completely. The computation of the joint probability functions of successive interdeparture times is done by using standard techniques of Semi-Markov processes [5, 6].

² $\underline{P} \underline{e} = 1$ and $\underline{P} (\mathbf{I} - \mathbf{D}_0)^{-1} (\widehat{\mathbf{D}} - \mathbf{D}_0) = \underline{P}$ are easily verified.

³ $A_n = T_{n+1} - T_n$ denoting the n -th interarrival time, $A = \{A_n, n \in \mathbb{N}_0\}$ the stochastic process of the interarrival times

⁴ $\mathbf{1}_{\text{EXPRESSION}} = 1$, if EXPRESSION is true, and = 0 else.

3.1. The embedded Markov Renewal process (MRP) at departure instants

The Semi-Markov kernel $\tilde{\mathbf{q}}(t)$ of the MRP states about the probability $\left[[\tilde{\mathbf{q}}(t)]_{k\ell} \right]_{ij}$ ⁵ that, given a departure leaves k customers in the system with the input process in phase i , the next departure happens after t time units and leaves ℓ customers in the system with the input process in phase j . Here, "leave behind" means "immediately after a departure" so that an arrival simultaneous with a departure is taken into account.

3.2. Arrival First Policy

Due to the fact that arrivals and departures can happen simultaneously with positive probability in discrete-time queueing systems, one has to define, how a simultaneous arrival and departure are dealt with: If the arriving customer still sees the departing one in case of simultaneity, this is called the "arrival first" (AF) policy, which is considered here. The waiting-time probability function of the other case "departure first" (i.e. the departing customer leaves before the arriving one enters in case of simultaneity) is for further study.

The case AF leads to a Semi-Markov kernel $\tilde{\mathbf{q}}(t)$ similar to that of the continuous-time case [7]. It is described by the following matrices:

- $\mathbf{A}_k(t)$, where $[\mathbf{A}_k(t)]_{ij}$ denotes the probability, that given a departure, which leaves at least one customer in the system with the input process in phase i , there are k arrivals within the following service time lasting t time units with the input process in phase j at the next departure. In the present case it is: $\mathbf{A}_k(t) = \mathbf{P}(k, t)h(t)$.

- $\mathbf{B}_k(t)$, where $[\mathbf{B}_k(t)]_{ij}$ denotes the conditional probability, that given a departure, which leaves the system empty with the input process in phase i , the next departure happens after t time units with the input process in phase j , and there were k arrivals during the service time of that departure. In the present case it is

$$\begin{aligned} \mathbf{B}_k(t) &= \sum_{w=1}^{\infty} \sum_{v=1}^t \mathbf{D}_0^{v-1} \mathbf{D}_w \mathbf{P}(k-w+1, t-v) h(t-v) = \\ &= \sum_{w=1}^{\infty} \sum_{v=1}^t \mathbf{D}_0^{v-1} \mathbf{D}_w \mathbf{A}_{k-w+1}(t-v) \end{aligned}$$

With AF a departure can leave at most N customers in the system (as it is in the continuous-time case) so that the state space of the MRP is $\{0, 1, \dots, N\} \times \{1, 2, \dots, m\}$. The resulting Semi-Markov kernel is shown in Fig. 1. The transition probability matrix of the embedded MC $\tilde{\mathbf{q}}$ is composed of the entries $\mathbf{B}_k := \sum_{t=1}^{\infty} \mathbf{B}_k(t)$ and $\mathbf{A}_k := \sum_{t=1}^{\infty} \mathbf{A}_k(t)$ among which there is the following relation:

⁵ $[[\tilde{\mathbf{q}}(t)]_{k\ell}]_{ij}$ means the submatrix in the k -th main row and the ℓ -th column of $\tilde{\mathbf{q}}(t)$ and in this submatrix the entry in the i -th row and the j -th column.

$$\tilde{\mathbf{q}}(t) = \begin{pmatrix} \mathbf{B}_0(t) & \mathbf{B}_1(t) & \mathbf{B}_2(t) & \cdots & \mathbf{B}_{N-1}(t) & \sum_{k=N}^{\infty} \mathbf{B}_k(t) \\ \mathbf{A}_0(t) & \mathbf{A}_1(t) & \mathbf{A}_2(t) & \cdots & \mathbf{A}_{N-1}(t) & \sum_{k=N}^{\infty} \mathbf{A}_k(t) \\ \mathbf{0} & \mathbf{A}_0(t) & \mathbf{A}_1(t) & \cdots & \mathbf{A}_{N-2}(t) & \sum_{k=N-1}^{\infty} \mathbf{A}_k(t) \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_0(t) & \cdots & \mathbf{A}_{N-3}(t) & \sum_{k=N-2}^{\infty} \mathbf{A}_k(t) \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_0(t) & \sum_{k=1}^{\infty} \mathbf{A}_k(t) \end{pmatrix}$$

Fig. 1 Semi-Markov kernel for arrival first.

$$\mathbf{B}_k = \sum_{w=1}^{\infty} \sum_{v=1}^{\infty} \mathbf{D}_0^{v-1} \mathbf{D}_w \sum_{t=1}^{\infty} \mathbf{A}_{k-w+1}(t-v) = (\mathbf{I} - \mathbf{D}_0)^{-1} \sum_{w=0}^k \mathbf{D}_{w+1} \mathbf{A}_{k-w}. \quad (3)$$

4. The z-transform of the waiting-time

The derivation bases on similar arguments as used for obtaining expressions for the occupancy at arrival instants (for details see e.g. [10, 7, 12]).

Consider Fig. 2. At time instant u there is a departure, which leaves the system empty; v time units later, there is a (μ -batch) arrival (i.e. μ customers arrive simultaneously, which is incorporated in the input process $\underline{\text{DBMAP}}$). The very customer of the batch arrival, who is admitted first, will stay in the server until his service time is elapsed; in Fig. 2 this service time lasts $t-u-v+\tau$ time units. A further arrival at time instant t will see μ customers in the system. The time until these customers will have left the system equals the unfinished work seen at time instant t , and it is composed of the residual service time τ of the customer in service and the sum of the service times of the other customers of the μ -batch.

A second case, where a departure does not leave the system empty, has to be discussed separately and completes the description of the system's evolution in time (Fig 3). Note that as soon as a customer leaves a non-empty system, the service time of the next customer in service starts.

4.1. Waiting time at finite time instants

In the following $\hat{r}_{k\ell}^{ij'}(u)$ means the probability that given a departure at time instant 0, which leaves k customers in the system with the input process in phase i , a departure at time instant u leaves ℓ customers in the system with the input process in phase j' . This probability can be expressed by means of the

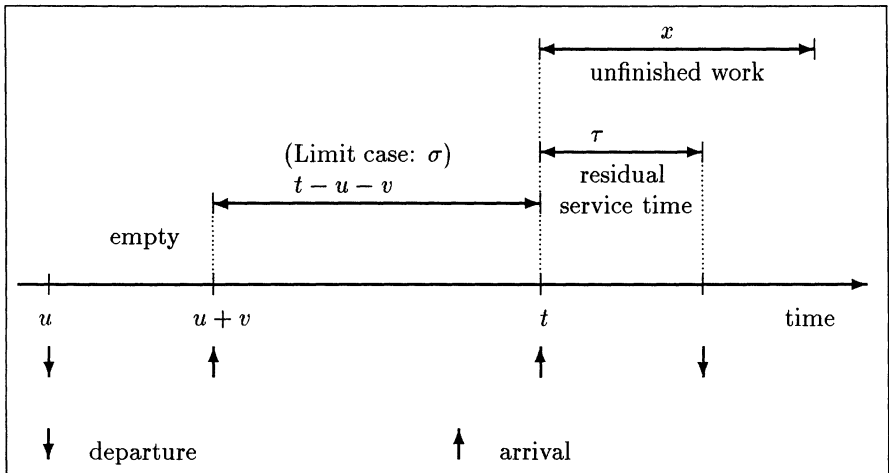


Fig. 2 Time scheme in case that a departure leaves the system empty at u . x represents the unfinished work seen by the arrival at t .

Semi-Markov kernel describing the queuing system by

$$\hat{\mathbf{r}}(t) = \sum_{n=1}^t \tilde{\mathbf{q}}^{*n}(t) \quad .^6$$

For the following discussion, the explicit representation of $\hat{r}_{k\ell}^{ij'}(t) := [[\hat{\mathbf{r}}(t)]_{k\ell}]_{ij'}$, is not of importance, since it will disappear in the limit case due to the limit theorem for Markov renewal processes.

Define $w_{kr}^{ij}(t, x, \mu)$ as the conditional probability that given a departure at 0 which leaves k customers in the system with the input process in phase i , a μ -batch arrival at t sees a waiting-time of x and an occupancy of r customers with the input process in phase j . Using $\hat{r}_{k\ell}^{ij'}(u)$ we get (where \mathcal{N}_t is a normalizing constant depending on the time instant t):

$$\mathcal{N}_t w_{k0}^{ij}(t, x, \mu) = \mathbf{1}_{x=0} \sum_{j' \in \mathcal{E}} \sum_{u=0}^{t-1} \hat{r}_{k0}^{ij'}(u) \left[\mathbf{P}(0, t-u-1) \mathbf{D}_\mu \right]_{j'j} \quad (4)$$

For $1 \leq r \leq N$:

$$\mathcal{N}_t w_{kr}^{ij}(t, x, \mu) =$$

⁶ $*n$ denotes the n -fold convolution.

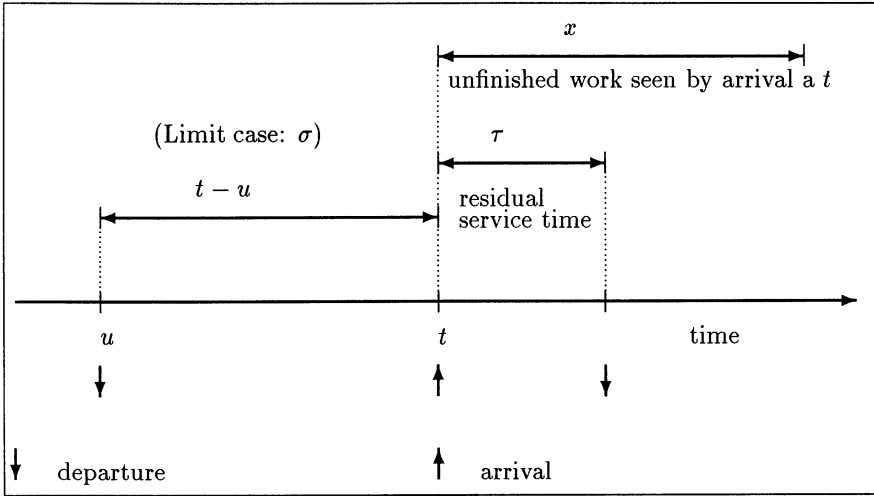


Fig. 3 Time scheme in case that a departure leaves a non-empty system at u . x represents the unfinished work seen by the arrival at t .

$$\begin{aligned}
 & \sum_{j' \in \mathbb{IE}} \sum_{u=0}^{t-2} \hat{r}_{k0}^{ij'}(u) \sum_{v=1}^{t-u-1} \sum_{w=1}^r \left[\mathbf{D}_0^{v-1} \mathbf{D}_w \mathbf{P}(r-w, t-u-v-1) \mathbf{D}_\mu \right]_{j'j} \hat{h}_r(t-u-v, x) + \\
 & + \sum_{j' \in \mathbb{IE}} \sum_{u=0}^{t-1} \sum_{\ell=1}^r \hat{r}_{k\ell}^{ij'}(u) \left[\mathbf{P}(r-\ell, t-u-1) \mathbf{D}_\mu \right]_{j'j} \hat{h}_r(t-u, x) \quad (5)
 \end{aligned}$$

where $\hat{h}_r(t, x)$ is given by:

$$\hat{h}_r(t, x) := \begin{cases} 0 & \text{for } r = 0 \\ h(t+x) & \text{for } r = 1 \\ \sum_{\tau=0}^x h(t+\tau) h^{*(r-1)}(x-\tau) & \text{for } r \geq 2 \end{cases}$$

Here, $h^{*r}(x)$ means the r -fold convolution of $h(x)$. For $r = N + 1$, the arrival is lost. Therefore, this case does not contribute to the waiting-time.

4.2. Waiting time in the limit case

The limit theorem of Markov Renewal functions (see e.g. [13], p. 1251f.) applied to eqs. (4) and (5) yields expressions no longer depending on i and k . Therefore they are dropped in the limit expression defined by

$$[\underline{w}_r(x, \mu)]_j := \lim_{t \rightarrow \infty} w_{kr}^{ij}(t, x, \mu) \quad ,$$

which denotes the probability that in the limit case an arriving μ -batch sees an unfinished work of x with the input process in phase j . Applying the limit

theorem yields

$$\mathcal{N} \underline{w}_0(x, \mu) = \frac{\mathbf{1}_{x=0}}{E^*} \underline{x}_0 (\mathbf{I} - \mathbf{D}_0)^{-1} \mathbf{D}_\mu \quad (6)$$

and for $1 \leq r \leq N$

$$\begin{aligned} \mathcal{N} \underline{w}_r(x, \mu) &= \\ &= \frac{1}{E^*} \sum_{\ell=1}^r \left[\left(\underline{x}_0 (\mathbf{I} - \mathbf{D}_0)^{-1} \mathbf{D}_\ell + \underline{x}_\ell \right) \sum_{\sigma=0}^{\infty} \mathbf{P}(r - \ell, \sigma) \mathbf{D}_\mu \widehat{h}_r(\sigma + 1, x) \right], \quad (7) \end{aligned}$$

where E^* is the mean sojourn time of the underlying Markov Renewal Process, which is here the mean interdeparture time of the queue. E^* need not be computed due to the normalizing constant \mathcal{N} .

4.3. Efficient computation of the z -transform

Define the z -transform of the waiting-time by

$$\underline{W}_r(z_1, \mu) := \sum_{x=0}^{\infty} \underline{w}_r(x, \mu) z_1^x .$$

For $\widehat{h}_r(t, x)$ we get:

$$\widehat{h}_r(t, x) \circ \bullet z^{-t} \left(h_z(z) - \sum_{\nu=0}^{t-1} z^\nu h(\nu) \right) [h_z(z)]^{r-1} \quad \text{for } r \geq 1 ,$$

if z -transformation is defined by $f(k) \circ \bullet \sum_{k=0}^{\infty} f(k) z^k$. $h_z(z) = \sum_{k=0}^{\infty} h(k) z^k$ is the z -transform of the service time.

Theorem 1 For AF:

$$\begin{aligned} \mathcal{N} \cdot \underline{W}_r(z_1, \mu) &= \\ &= \begin{cases} \frac{1}{E^*} \underline{x}_0 (\mathbf{I} - \mathbf{D}_0)^{-1} \mathbf{D}_\mu & \text{for } r = 0 \\ \frac{1}{E^*} \sum_{\ell=1}^r \left[\left(\underline{x}_0 (\mathbf{I} - \mathbf{D}_0)^{-1} \mathbf{D}_\ell + \underline{x}_\ell \right) \mathbf{h}_{r-\ell}(z_1) [h_z(z_1)]^{\ell-1} \mathbf{D}_\mu \right] & \text{for } 1 \leq r \leq N \\ 0 & \text{for } r = N + 1 \end{cases} \quad (8) \end{aligned}$$

\mathcal{N} is determined by a completeness condition, which can e.g. be defined by $\sum_{\mu=1}^{\theta_{\max}} \underline{W}_r(1, \mu) = 1$, and then refers to the unfinished work seen at an arrival instant. It is

$$\mathbf{h}_n(z_1) = h_z(z_1) \mathcal{R}_n(z_1) - \sum_{k=0}^n \mathcal{R}_k(z_1) \mathbf{A}_{n-k} [h_z(z_1)]^{n-k} , \quad (9)$$

where the $\mathcal{R}_n(z_1)$ are coefficients of the series

$$\sum_{n=0}^{\infty} \mathcal{R}_n(z_1) z_2^n = [\mathbf{I} z_1 - \mathbf{D}(z_2 h_z(z_1))]^{-1} , \quad (10)$$

for which the following recursion holds

$$\mathcal{R}_n(z_1) = \left[\sum_{k=0}^{n-1} \mathcal{R}_k(z_1) \mathbf{D}_{n-k} [h_z(z_1)]^{n-k} \right] (\mathbf{I}z_1 - \mathbf{D}_0)^{-1} \quad \text{for } n \geq 1 \quad , \quad (11)$$

$$\mathcal{R}_0(z_1) = [\mathbf{I}z_1 - \mathbf{D}_0]^{-1} \quad . \quad (12)$$

Proof. The central idea is to apply – besides the z -transform with respect to the time variable x – also a z -transform with respect to the count of arrivals n : Let

$$\mathbf{h}_{n,\ell}(z_1) := \sum_{\sigma=0}^{\infty} \mathbf{P}(n, \sigma) \sum_{x=0}^{\infty} z_1^x \widehat{h}_{n+\ell}(\sigma+1, x) \quad .$$

Then we have

$$\sum_{n=0}^{\infty} \mathbf{h}_{n,\ell}(z_1) z_2^n = \sum_{n=0}^{\infty} z_2^n \sum_{\sigma=0}^{\infty} \mathbf{P}(n, \sigma) \sum_{x=0}^{\infty} z_1^x \widehat{h}_{n+\ell}(\sigma+1, x) = \quad (13)$$

$$= \sum_{\sigma=0}^{\infty} \sum_{n=0}^{\infty} \mathbf{P}(n, \sigma) [z_2 h_z(z_1)]^n z_1^{-(\sigma+1)} \left(h_z(z_1) - \sum_{x=0}^{\sigma} z_1^x h(x) \right) h_z(z_1)^{\ell-1} = \quad (14)$$

(because of eq. (2))

$$= \sum_{\sigma=0}^{\infty} [z_1^{-1} \mathbf{D}(z_2 h_z(z_1))]^{\sigma} z_1^{-1} \left(h_z(z_1) - \sum_{x=0}^{\sigma} z_1^x h(x) \right) h_z(z_1)^{\ell-1} = \quad (15)$$

(developing into an infinite series yields)

$$= [\mathbf{I}z_1 - \mathbf{D}(z_2 h_z(z_1))]^{-1} \left(\mathbf{I}h_z(z_1) - \sum_{n=0}^{\infty} \mathbf{A}_n [z_2 h_z(z_1)]^n \right) h_z(z_1)^{\ell-1} \quad . \quad (16)$$

With

$$[\mathbf{I}z_1 - \mathbf{D}(z_2 h_z(z_1))]^{-1} =: \sum_{n=0}^{\infty} \mathcal{R}_n(z_1) z_2^n$$

we get:

$$\begin{aligned} \sum_{n=0}^{\infty} \mathbf{h}_{n,\ell}(z_1) z_2^n &= \\ &= \left(\sum_{n=0}^{\infty} \mathcal{R}_n(z_1) z_2^n \right) \left(\mathbf{I}h_z(z_1) - \sum_{n=0}^{\infty} \mathbf{A}_n [z_2 h_z(z_1)]^n \right) h_z(z_1)^{\ell-1} \quad , \quad (17) \end{aligned}$$

which yields (due to the Cauchy product formula)

$$\mathbf{h}_{n,\ell}(z_1) = \left[\mathcal{R}_n(z_1) h_z(z_1) - \sum_{k=0}^{\infty} \mathcal{R}_k(z_1) \mathbf{A}_{n-k} [h_z(z_1)]^{n-k} \right] h_z(z_1)^{\ell-1} \quad (18)$$

and finally

$$\mathbf{h}_n(z_1) := \frac{\mathbf{h}_{n,\ell}(z_1)}{h_z(z_1)^{\ell-1}} = \mathcal{R}_n(z_1)h_z(z_1) - \sum_{k=0}^{\infty} \mathcal{R}_k(z_1)\mathbf{A}_{n-k}[h_z(z_1)]^{n-k} \quad (19)$$

The expression for $r = 0$ follows directly from eq. (6).

Determining $\mathcal{R}_n(z_1)$

What is left, is the determination of $\mathcal{R}_n(z_1)$, which is the n -th coefficient of the series of $[\mathbf{I}z_1 - \mathbf{D}(z_2)]^{-1}$. The trick of the DMAP-case in [2] does not work, therefore a differentiation is used for determining $\mathcal{R}_n(z_1)$:

$$\sum_{n=0}^{\infty} \mathcal{R}_n(z_1)z_2^n = [\mathbf{I}z_1 - \mathbf{D}(z_2h_z(z_1))]^{-1} =: \mathbf{f}(z_1, z_2) \quad (20)$$

$$\mathbf{f}(z_1, z_2) = [\mathbf{g}(z_1, z_2)]^{-1}, \quad \mathbf{g}(z_1, z_2) := \mathbf{I}z_1 - \mathbf{D}(z_2h_z(z_1))$$

$$n!\mathcal{R}_n(z_1) = \left. \frac{\partial^n}{\partial z_2^n} \mathbf{f}(z_1, z_2) \right|_{z_2=0}, \quad \mathcal{R}_0(z_1) = [\mathbf{I}z_1 - \mathbf{D}_0]^{-1} \quad (21)$$

Leibnitz's Rule

$$\frac{\partial^n}{\partial z_2^n} [\mathbf{f}(z_1, z_2)\mathbf{g}(z_1, z_2)] = \mathbf{0} = \quad (22)$$

$$= \sum_{k=0}^n \binom{n}{k} \frac{\partial^k}{\partial z_2^k} \mathbf{f}(z_1, z_2) \frac{\partial^{n-k}}{\partial z_2^{n-k}} \mathbf{g}(z_1, z_2) \quad (23)$$

yields

$$\mathbf{f}_{z_2}^{(n)}(z_1, z_2) = - \left[\sum_{k=0}^{n-1} \binom{n}{k} \mathbf{f}_{z_2}^{(k)}(z_1, z_2) \mathbf{g}_{z_2}^{(n-k)}(z_1, z_2) \right] \mathbf{g}^{-1}(z_1, z_2) \quad (24)$$

which results in

$$n!\mathcal{R}_n(z_1) = \mathbf{f}_{z_2}^{(n)}(z_1, 0) = - \left[\sum_{k=0}^{n-1} \binom{n}{k} \mathbf{f}_{z_2}^{(k)}(z_1, 0) \mathbf{g}_{z_2}^{(n-k)}(z_1, 0) \right] \mathbf{g}^{-1}(z_1, 0) \quad (25)$$

and because of $\mathbf{g}^{-1}(z_1, 0) = [\mathbf{I}z_1 - \mathbf{D}_0]^{-1}$,

$$\mathbf{f}_{z_2}^{(k)}(z_1, 0) = k!\mathcal{R}_k(z_1), \quad \text{and} \quad \mathbf{g}_{z_2}^{(n-k)}(z_1, 0) = -(n-k)! \cdot \mathbf{D}_{n-k}[h_z(z_1)]^{n-k}$$

it is

$$\mathcal{R}_n(z_1) = \begin{cases} [\mathbf{I}z_1 - \mathbf{D}_0]^{-1} & \text{for } n = 0, \\ \left[\sum_{k=0}^{n-1} \mathcal{R}_k(z_1)\mathbf{D}_{n-k}[h_z(z_1)]^{n-k} \right] (\mathbf{I}z_1 - \mathbf{D}_0)^{-1} & \text{for } n \geq 1. \end{cases} \quad (26)$$

Applying eq. (19) to eq. (7) yields the expressions for $1 \leq r \leq N$. \square

4.4. Determining the probability function of the unfinished work at arrival instants

The above theorem allows the computation of arbitrary moments of the unfinished work at arrival instants. However, for capturing the waiting time seen by a test-cell of stream 1 or 2 in a superposition, it is necessary to find the probability function. This can be done by “simply” differentiating the z -transform often enough and considering the value of the differentiated function at 0. The problem, however, is that you need the derivatives up to the ND -th one. The following Lemma establishes a simple algorithm to determine the value at 0 of all these derivatives.

Lemma 1 *Given the z -transform of the service time $h_z(z) := \sum_{k=0}^{\infty} h(k)z^k$. The probability function of the unfinished work is obtained from*

$$\underline{w}_r(x, \mu) = \frac{1}{x!} \frac{d^x}{dz^x} W_r(z, \mu) \Big|_{z=0} .$$

$$\text{Define } \partial h[m][n] := \frac{1}{m!} \frac{d^m}{dz^m} [h_z(z)]^n \Big|_{z=0} ,$$

$$\partial \mathbf{R}[m][n] := \frac{1}{m!} \frac{d^m}{dz^m} \mathbf{R}_n(z) \Big|_{z=0} ,$$

$$\partial \mathbf{H}[m][n] := \frac{1}{m!} \frac{d^m}{dz^m} \mathbf{h}_n(z) \Big|_{z=0} .$$

Then it is for $0 \leq n \leq N-1$, $0 \leq m \leq ND$:

$$\frac{1}{m!} \frac{d^m}{dz^m} [\mathbf{h}_n(z)[h_z(z)]^{\ell-1}]_{z=0} = \sum_{k=0}^m \partial \mathbf{H}[k][n] \cdot \partial h[m-k][\ell-1]$$

with

$$\begin{aligned} \partial \mathbf{H}[m][n] &= \\ &= \sum_{k=0}^m \partial \mathbf{R}[k][n] \cdot h(m-k) - \sum_{k=0}^m \sum_{\ell=0}^n \partial \mathbf{R}[k][\ell] \cdot \partial h[m-k][n-\ell] \cdot \mathbf{A}_{n-\ell} , \\ \partial \mathbf{R}[m][n] &= \\ &= \left[\sum_{\nu=0}^{n-1} \sum_{k=0}^m \partial \mathbf{R}[k][\nu] \cdot \partial h[m-k][n-\nu] \cdot \mathbf{D}_{n-\nu} - \partial \mathbf{R}[m-1][n] \right] \cdot (-\mathbf{D}_0)^{-1} , \\ \partial h[m][n] &= \sum_{k=0}^{\infty} h(k) \cdot \partial h[m-k][n-1] . \end{aligned}$$

Initial values:

$$\begin{aligned} \partial h[m][0] &= \mathbf{1}_{m=0}, & \partial h[0][n] &= \mathbf{1}_{n=0}, & \partial \mathbf{R}[0][n] &= \mathbf{1}_{n=0}(-\mathbf{D}_0)^{-1}, \\ \partial \mathbf{R}[m][0] &= (-\mathbf{D}_0)^{-(m+1)}, & \partial \mathbf{H}[0][n] &= \mathbf{1}_{n=0} \mathbf{D}_0^{-1} \mathbf{A}_0 . \end{aligned}$$

Proof. Apply Leibnitz' rule to eqs. (9), (11). □

4.5. Numerical stability

This algorithm proves to be numerically stable, if \mathbf{D}_0 is a diagonal matrix. If not, the high powers of the inverse of \mathbf{D}_0 (which has entries significantly greater than 1, since \mathbf{D}_0 is sub-stochastic) soon yield exploding matrix entries in $\partial \mathbf{R}$ and $\partial \mathbf{H}$, which then cause wrong values of the probability function. Eq. (8) can also be used for obtaining moments of the unfinished work at arrival instants by simply considering derivatives at $z = 1$, then also for non-diagonal matrices \mathbf{D}_0 , with very good numerical results. However, then it is not possible to deal with per-stream values as discussed in the next section.

5. Per-stream waiting-time probability functions

Performance evaluation of ATM networks concentrates on deterministic service times due to the transmission of fixed-length packets in ATM networks. We now focuss on a multiplexer facing two input streams 1 and 2, stream 1 represented by a $\text{DMAP}_1^{[X]}$, and stream 2 by a DMAP_2 , both being special cases of a DBMAP .

Let DBMAP_1 be given by $\mathbf{D}_w^{(1)}$, $w \in \mathbb{N}_0$, $b_{max}^{(1)}$, $\widehat{\mathbf{D}}^{(1)} := \sum_{w=0}^{b_{max}^{(1)}} \mathbf{D}_w^{(1)}$ and DMAP_2 by $\mathbf{C}^{(2)}$ and $\mathbf{D}^{(2)}$. The resulting superposition, again a DBMAP , is given by $\widetilde{\mathbf{D}}_w$, $w \in \mathbb{N}_0$, $\sum_{w=0}^{\infty} \widetilde{\mathbf{D}}_w = \widetilde{\mathbf{D}}$; these matrices are obtained by means of the Kronecker product \otimes according to

$$\widetilde{\mathbf{D}}_w = \begin{cases} \mathbf{D}_0^{(1)} \otimes \mathbf{C}^{(2)} & \text{for } w = 0 \\ \mathbf{D}_w^{(1)} \otimes \mathbf{C}^{(2)} + \mathbf{D}_{w-1}^{(1)} \otimes \mathbf{D}^{(2)} & \text{for } 1 \leq w \leq b_{max}^{(1)} \\ \mathbf{D}_{w-1}^{(1)} \otimes \mathbf{D}^{(2)} & \text{for } w = b_{max}^{(1)} + 1 . \end{cases}$$

Here, we have for stream 1 due to the $\text{DMAP}^{[X]}$

$$\mathbf{D}_w^{(1)} = \begin{cases} x(w)\mathbf{D}^{(1)} & w > 0 \\ \mathbf{C}^{(1)} & w = 0 \end{cases} . \quad (27)$$

Now the matrices $\widetilde{\mathbf{D}}_w$ determine the Semi-Markov kernel governed by the matrices \mathbf{A}_n , and \mathbf{B}_n . The z -transform of the waiting-time of each stream ($i = 1, 2$, and 3 representing the full stream) is simply obtained by:

$$\begin{aligned} \mathcal{N}^{(i)} \cdot \underline{W}_r^{(i)}(z_1, \mu) &= \\ &= \begin{cases} \frac{1}{E^*} \underline{x}_0 (\mathbf{I} - \widetilde{\mathbf{D}}_0)^{-1} \widetilde{\mathbf{D}}_\mu^{(i)} & \text{for } r = 0 \\ \frac{1}{E^*} \sum_{\ell=1}^r \left[\left(\underline{x}_0 (\mathbf{I} - \widetilde{\mathbf{D}}_0)^{-1} \widetilde{\mathbf{D}}_\ell + \underline{x}_\ell \right) \mathbf{h}_{r-\ell}(z_1) [h_z(z_1)]^{\ell-1} \widetilde{\mathbf{D}}_\mu^{(i)} \right] & \text{for } 1 \leq r \leq N \\ 0 & \text{for } r = N + 1 \end{cases} \quad (28) \end{aligned}$$

where

$\widetilde{\mathbf{D}}_\mu^{(1)} = \mathbf{D}_\mu^{(1)} \otimes \mathbf{C}^{(2)} + \mathbf{1}_{\mu > 1} \mathbf{D}_{\mu-1}^{(1)} \otimes \mathbf{D}^{(2)}$, i.e. the arriving batch contains at least one cell of stream 1,

$\tilde{\mathbf{D}}_\mu^{(2)} = \mathbf{D}_{\mu-1}^{(1)} \otimes \mathbf{D}^{(2)}$, i.e. the arriving batch contains a cell of stream 2,
 $\tilde{\mathbf{D}}_\mu^{(3)} = \tilde{\mathbf{D}}_w$, i.e the arriving batch contains a cell of stream 1 or 2.

The waiting-time of a cell, which is not lost, is given by the unfinished work immediately before the arrival instant of that cell plus the length of the cells of the same batch, which are admitted before that cell. As a lost cell cannot see any waiting-time, the waiting-time takes values in $\{0, 1, \dots, ND\}$. Since the waiting-time p.f. is related to $u(k, i)$, $0 \leq k \leq ND$, normalizing constants $\mathcal{N}_1, \mathcal{N}_2, \mathcal{N}_3$ are needed. Note that, as usual test-cell admission for stream 2 is completely random. With

$$\underline{w}_r^{(i)}(x, \mu) \overset{x}{\circlearrowleft} \underline{W}_r^{(i)}(z_1, \mu), \quad i = 1, 2, 3$$

it is

$$P\{W_i = \ell\} = \frac{\mathbf{1}_{\ell=0}^{ND}}{\mathcal{N}_i} \left[\sum_{\mu=1}^{b_{max}+1} \frac{1}{\mu} \sum_{\psi=1}^{\mu} \mathbf{1}_{\ell \geq (\psi-1)D} \sum_{r=0}^N \underline{w}_r^{(i)}(\ell - (\psi-1)D, \mu) \underline{e} \right]$$

\mathcal{N}_i given by $\sum_{\ell=0}^{ND} P\{W_i = \ell\} = 1$.

5.1. Formula validation

The above formulae were implemented in C programmes. For diagonal matrices \mathbf{D}_0 the results of the per-stream probability functions obtained were in very good correspondence with those obtained from the method of the unfinished work described in [7, 10]: For **double** numbers the values of the probability function differed first in the 10th figure (of 16). Using derivatives of eq. (8) at $z = 1$, the results were as good also for non-diagonal matrices. Compared with the method of the unfinished work the algorithm presented above is much faster (and is applicable also for non-deterministic service time distributions), since the system matrix grows only with the second power of $N \cdot m$ in contrast to $N \cdot m \cdot D$ for the method of the unfinished work. The latter, however, allows to compute transient loss probabilities and waiting-time probability functions [9]. Table 1 compares the potential of both methods.

6. Numerical Examples

For the numerical examples, the per-stream mean waiting-times and loss probabilities were computed keeping the traffic load of the full stream constant at 0.8, and varying the ratio of the traffic load. Stream 1 is a DMAP with two phases $\mathbf{C} = \begin{pmatrix} q_1 & 0 \\ 0 & q_2 \end{pmatrix}$, $\mathbf{D} = \begin{pmatrix} 1-q_1 & 0 \\ 0 & 1-q_2 \end{pmatrix} \cdot \mathbf{p}$, \mathbf{p} transition matrix of the embedded Markov chain, ($b_{max} = 1$, coefficient of correlation of its embedded Markov chain equals 0.6), stream 2 is a Bernoulli process.

It is easily seen in Fig. 4 and 5 that as soon as the correlated stream 1 gets stronger, the mean waiting-times and loss probabilities of *both* streams in the superposition increase. Loss probabilities were computed according to [7].

Table 1

Comparison of the method of the unfinished work and the M/G/1-paradigm.

	Unfinished work [6, 9]	M/G/1 paradigm
System size	$(N + 1) \cdot D \cdot m$ slower for $D > 1$	$(N + 1) \cdot m$
non-diagonal \mathbf{C}	no problem	rounding errors for per-stream expressions
Waiting Times	x	x
Unfinished Work	x	x
Loss Probabilities	x	x
Conditional Loss Probability	x	-
Transient Quantities	x	-
Departure Process	-	x
Continuous Time Analogue	-	x

7. Conclusions

This paper presents a fast algorithm for computing waiting-time moments for the DBMAP/G/1/N queue with arrival first policy, and generalizes the algorithm in such a way that per-stream waiting-time probability functions of the DBMAP₁+DMAP₂/G/1/N-queue can be computed. Thus, it extends [7], where only the computation of per-stream loss probabilities was solved using the M/G/1-paradigm. Furthermore, it supplements [1], since the decomposition of the matrix expression $[\mathbf{I}z_1 - \mathbf{D}(z_2 h_z(z_1))]^{-1}$ into a series in z_2 , which is the crucial point in finding the algorithm, can also be applied to the composition of the matrix expression $[\mathbf{I}s + \sum_{k=0}^{\infty} \mathbf{D}_k z^k]^{-1}$ into a series of z . The algorithm was validated by comparing the numerical results obtained by C programmes with those, which result from the method of the unfinished work, which - for some input processes - can also compute waiting-time moments of the DBMAP₁+DMAP₂/D/1/N-queue [9]. Very good correspondence was found.

REFERENCES

1. C. Blondia. The N/G/1 finite capacity queue. *Communications in Statistics - Stochastic Models*, 5(2), pp. 273–294, 1989.
2. C. Blondia. Finite Capacity Vacation Models with Non-Renewal Input. *J. Appl. Prob.*, 28, pp. 174–197, 1991.
3. C. Blondia. A discrete-time batch Markovian arrival process as B-ISDN traffic model. *Belgian Journal of Operations Research, Statistics and Computer Science*, 32(3,4), pp. 3–23, 1993.
4. C. Blondia and O. Casals. Statistical multiplexing of VBR sources: A matrix-analytic approach. *Perf. Eval.*, 16(1-3), pp. 5–20, Nov. 1992.
5. E. Cinlar. *Introduction to Stochastic Processes*. Prentice Hall, New Jersey, 1st edition, 1975.

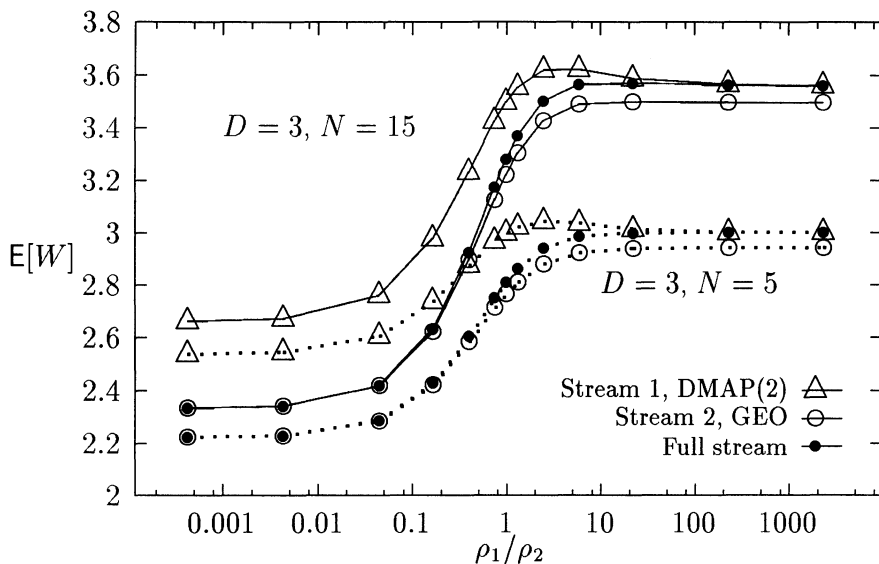


Fig. 4 Per-stream Mean Waiting Times for AF and two different buffer sizes N . $\rho_1 + \rho_2 = 0.8$, ρ_1/ρ_2 varied. Coefficient of correlation of the embedded Markov chain of DMAP(2): 0.6 .

6. C. Herrmann. A Performance Model for Statistical Multiplexing of correlated ATM Traffic Superpositions. In *7. ITG/GI Fachtagung: MMB'93*, Aachen, Sept. 21.-23., 1993, Springer-Verlag.
7. C. Herrmann. Discrete-time Models for Connection Admission Control (CAC) in ATM. In *Performance Modelling and Evaluation of ATM Networks, Vol. 1*, pages 85–104, Chapman & Hall, 1995, ISBN 0-412-71140-0.
8. C. Herrmann. VBR Video in ATM without Frame-Buffering: Influence of a Periodic Correlation Function on QOS Parameters. In *Performance Modelling and Evaluation of ATM Networks, Vol. 1*, pages 14–31, Chapman & Hall, 1995, ISBN 0-412-71140-0.
9. C. Herrmann. Transient Behaviour of QOS Parameters During Connection Admission in ATM. *International Journal of Electronics and Communications*, 50(4), pp. 274–284, July 1996.
10. C. Herrmann. *Stochastische Modelle für ATM-Konzepte*. PhD thesis, Verlag der Augustinus Buchhandlung, Aachen, Germany, ISBN 3-86073-380-X, Jan. 1995.
11. D.M. Lucantoni. New Results on the single Server Queue with a Batch Markovian Arrival Process. *Stochastic Models, Communications in Statistics*, 7(1), pp. 1–46, 1991.
12. Marcel F. Neuts. *Structured Stochastic Matrices of M/G/1 Type and their Applications*. Marcel Dekker, New York, 1989.

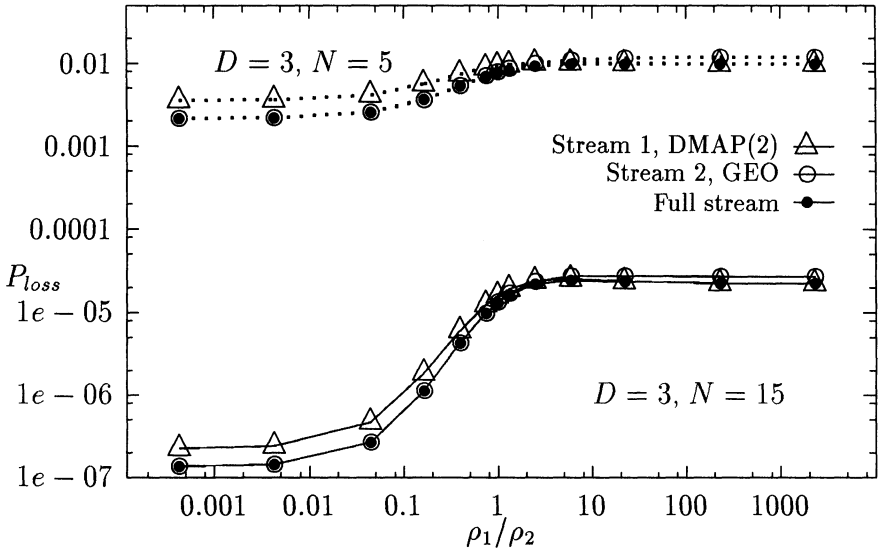


Fig. 5 Per-stream loss probabilities for AF and two different buffer sizes N . $\rho_1 + \rho_2 = 0.8$, ρ_1/ρ_2 varied. Coefficient of correlation of the embedded Markov chain of DMAP(2): 0.6 .

13. R. Pyke. Markov Renewal processes with finitely many states. *Ann. Math. Stat.*, 32, pp. 1243–1259, 1961.