

Analysis of congestion periods of an M/M/∞-queue

Frank Roijers^{a,b}, Michel Mandjes^{b,c}, Hans van den Berg^{a,d}

^a TNO Information and Communication Technology, P.O. Box 5050, 2600 GB Delft, The Netherlands

^b Center for Mathematics and Computer Science (CWI), The Netherlands

^c University of Amsterdam, Korteweg-de Vries Institute for Mathematics, The Netherlands

^d University of Twente, Department of Design and Analysis of Communication Systems (DACS), The Netherlands

Received 29 July 2005; received in revised form 23 June 2006; accepted 16 December 2006

Available online 31 January 2007

Abstract

A C-congestion period of an M/M/∞-queue is a period during which the number of customers in the system is continuously above level C. Interesting quantities related to a C-congestion period are, besides its duration D_C , the total area A_C above C, and the number of arrived customers N_C . In the literature Laplace transforms for these quantities have been derived, as well as explicit formulae for their means. Explicit expressions for higher moments and covariances (between D_C , N_C and A_C), however, have not been found so far.

This paper presents recursive relations through which all moments and covariances can be obtained. Up to a starting condition, we explicitly solve these equations; for instance, we write $\mathbb{E}D_C^2$ explicitly in terms of $\mathbb{E}D_0^2$. We then find formulae for these starting conditions (which directly relate to the busy period in the M/M/∞ queue).

Finally, a C-intercongestion period is defined as the period during which the number of customers is continuously below level C. Also for this situation a recursive scheme allows us to explicitly compute higher moments and covariances. Additionally we present the Laplace transform of a so-called *intercongestion* triple of the three performance quantities. It is also shown that expressions for the quantities of a C-intercongestion period can be used in an approximation for the C-congestion period. This is especially useful as the expressions for the C-intercongestion period are numerically more stable than those for the C-congestion period.

© 2007 Elsevier B.V. All rights reserved.

Keywords: M/M/∞; Transient behavior; C-congestion period; Busy period

1. Introduction

In this paper we consider an M/M/∞ queueing system. Customers arrive according to a Poisson process with arrival rate λ and have an exponential service requirement with mean μ^{-1} . There are an infinite number of identical servers and customers start service immediately upon arrival.

The M/M/∞ queueing system can be used as a flow-level model for the occupancy of a link in a communication network, see e.g. [1]. In order to avoid a degradation of the Quality of Service (QoS) of the underlying applications, a network operator should dimension the network links such that the fraction of time that the link occupancy exceeds a certain critical level (close to the link capacity) is kept small, i.e., below a target value ϵ (which can be evaluated

E-mail addresses: frank.roijers@tno.nl (F. Roijers), mmandjes@science.uva.nl (M. Mandjes), j.l.vandenberg@tno.nl (H. van den Berg).

easily; the number of flows has a Poisson distribution). However, QoS as perceived by the users of the network is not only affected by the frequency of congestion periods, but also by their duration. This motivates the interest in so-called *C-congestion periods* in a $M/M/\infty$ system.

A C -congestion period is defined as the period during which the number of users present is continuously above level C . In other words: a C -congestion period is the period starting at the epoch that an arriving customer finds C customers in the system, until the first time that a departing customer leaves behind C customers. The duration of a C -congestion period is denoted by D_C . Other interesting quantities which are related to a C -congestion period are the number of users that arrive during the congestion period, denoted by N_C , and the total amount of work in excess of level C during the C -congestion period, which is the so-called area A_C above level C .

First, in the following section we provide an overview of the existing literature related to congestion periods. This allows us to clarify our contribution in the subsequent section.

1.1. Literature

Keilson [4] studied passage times of birth–death process by decomposing a passage time into the convolution of congestion periods; due to the general nature of birth–death processes the results are rather implicit. There are several papers that have studied the congestion period in $M/M/\infty$ -queueing systems. Guillemin and Simonian [3] present closed-form expressions for the means of D_C , N_C and A_C . They also obtained the Laplace transforms (LTs) for the above-mentioned quantities, and analyzed the first passage time of level C starting in steady-state. Preater [7] elaborates on the results of Guillemin and Simonian; by using an alternative derivation he finds a more attractive form of the LT of the congestion period. He also presents the *joint* LT of the congestion period triple $\Theta_C(D_C, N_C, A_C)$ of the duration, number of arrivals and the area. In another paper [8] Preater examines the height of a congestion period, e.g., the maximum level that is reached during a congestion period. Knessl and Yang [5] study $\mathbb{P}(D_C > t)$ in several asymptotic regimes. Both Guillemin and Simonian [3] and Preater [7] observe that, when C grows large, a C -congestion period of an $M/M/\infty$ -queue behaves similarly to the busy period of the $M/M/1$ -queue. The LT of the duration and number of arriving customers in the busy period of an $M/M/1$ -queue can easily be obtained, and see [2] for an analysis of the area of a busy period. Robert [9] presents an approximation of the order of the mean passage time from level n to level 0 for large n .

Another related subject of frequent study is the busy period of the $M/G/\infty$ queueing system, which in fact coincides with the congestion period of level 0 (i.e., the 0-congestion period), with generally distributed service times. One of the earliest works on the busy period is by Takács [13]. He presents the LST of the busy cycle duration of a so-called type II counter, which is similar to an $M/G/\infty$ -queue. This result is used by others, e.g. Stadje [12] and Liu and Shi [6]. Liu and Shi [6] consider the busy period in $GI^X/G/\infty$ -queueing systems with batch arrivals and for several special cases they obtain expressions for the first and second moment of both the busy period and busy cycle. A joint LT for both the duration and number of arrivals was already presented by Shanbhag [11].

Although the Laplace transforms of D_C , N_C and A_C are known [3,7], differentiating these is fairly non-straightforward due to the rather implicit nature of the functions involved. This explains the absence of explicit formulae for higher moments (the means are known) and covariances (between D_C , N_C and A_C). Also, so far no attention was paid to *C-intercongestion periods*, which are the periods during which the number of customers in the system is continuously *below* C .

1.2. Contribution

This paper studies the duration, number of arrivals and area swept above C (i.e., D_C , N_C and A_C) for C -congestion periods in an $M/M/\infty$ -queue. Recursive relations are derived through which all the moments of the above-mentioned values can be obtained. In particular it is demonstrated that there is a recursive relation between the congestion periods of two adjacent levels, e.g., level C and level $C - 1$: any quantity of level C can be expressed in terms of the same quantity of a $(C - 1)$ -congestion period. Iterating these, we can express the quantities related to a C -congestion period in terms of the quantities related to a 0-congestion period (which is, as observed above, a busy period of the $M/M/\infty$ -queue). For instance, we write $\mathbb{E}D_C^2$ explicitly in terms of $\mathbb{E}D_0^2$. Furthermore, similar recursions are derived for the covariances between the quantities D_C , N_C and A_C .

Thus, in order to solve for the higher moments, we have to find the starting values for our recursion; in our example: to find an expression for $\mathbb{E}D_C^2$ we have to find an explicit formula for $\mathbb{E}D_0^2$. The derivation of these starting values can be done through the differentiation of the LT of these busy-period related quantities. In particular, explicit expressions for the first and second moments are presented. In addition to this, we find the covariances $\text{Cov}(D_C, N_C)$, $\text{Cov}(D_C, A_C)$ and $\text{Cov}(N_C, A_C)$. With $\mathbb{E}D_C$, $\mathbb{E}N_C$ and $\mathbb{E}A_C$ being known, this reduces to finding the ‘joint expectations’ $\mathbb{E}[D_C N_C]$, $\mathbb{E}[D_C A_C]$ and $\mathbb{E}[N_C A_C]$. Again, we first express these in terms of the busy-period quantities (for example, $\mathbb{E}[D_C N_C]$ is phrased in terms of $\mathbb{E}[D_0 N_0]$), and then the busy-period related starting condition is solved. Theoretically, all moments (joint expectations) of the quantities of a C-congestion period can be obtained by differentiating the LT of the quantities (from Preater’s [7] congestion triple), but practically this is far from trivial. It is considerably easier to obtain the moments (and joint expectations) of the busy-period quantities and to insert these as the starting conditions into the recursive relations.

Analogously to a C-congestion period, a C-intercongestion period is defined as the period that the number of users is continuously *below* level C. The analysis and results for the quantities duration, number of arrivals, and the area below C are presented, which are also recursive relations for the moments and covariances. Again, the recursion can be solved in terms of the quantities of level 0. Importantly, these relate to the period that the system has *less* than 0 customers; hence, all moments and joint expectations of the quantities are 0. The recursion has attractive numerical properties: it is more stable than those of the C-congestion periods. In addition, similarly to Preater’s derivation of the LT of a congestion triple [7], the LT of the intercongestion triple is derived.

Guillemin and Simonian [3] and Preater [7] already observed that, for large C, the busy period of an M/M/1-queue can be used to approximate the behavior of a C-congestion period of an M/M/∞-queue. The approximation works well for large C but not for C close to the average number of users in the system ρ . Results in this paper indicate that the quantities of a C-congestion period can be approximated accurately by a $\rho - (C - \rho)$ -intercongestion period (which has, as indicated above, favorable numerical properties). The approximation works particularly well for C close to ρ , and can consequently be used complementary to the above-mentioned M/M/1-based approximation.

1.3. Outline

The outline of this paper is as follows. Section 2 introduces the notation and illustrates how a transient period of an M/M/∞-queue can be subdivided into C-congestion periods. Section 3 presents the recursion schemes for the first and second moment of the D_C , N_C and A_C . The recursions are solved resulting in closed-form expressions which still contain the starting condition: for instance, $\mathbb{E}D_C^2$ is explicitly written in terms of $\mathbb{E}D_0^2$. Similarly, Section 4 yields the derivation of the covariances of the quantities in terms of the covariances relating to the busy period: $\mathbb{E}[D_C N_C]$ is presented in terms of $\mathbb{E}[D_0 N_0]$. In Section 5 the first and second moments of D_0 as well as the joint expectation $\mathbb{E}[D_0 N_0]$ are obtained. These busy-period quantities are then the ‘starting conditions’ of the recursions of Sections 3 and 4. For the first and second moments of N_0 and A_0 and for the joint expectations $\mathbb{E}[D_0 A_0]$ and $\mathbb{E}[N_0 A_0]$ we refer the reader to Section 5 of [10]. Section 6 presents the definition, analysis and results for C-intercongestion periods. Section 7 provides some numerical results and illustrates that a $\rho - (C - \rho)$ -intercongestion period can be used as an accurate approximation of a C-congestion period when C is close to ρ . Section 8 concludes this paper.

2. Preliminaries

2.1. Definitions

Consider an M/M/∞-queue with arrival rate λ and mean service requirement μ^{-1} . For convenience we introduce the notation $v_n := \lambda + n\mu$. The average workload of the system is denoted by $\rho = \lambda/\mu$. Let the Markov process $A_t \in \{0, 1, 2, \dots\}$ denote the number of customers in the system at time t . Let

$$D_j(i) := \inf\{t > 0 : A_t = j \mid A_0 = i\}, \quad i > j, \quad (1)$$

$$N_j(i) := \#\{t : A_t - A_{t-} = 1, 0 < t \leq D_j(i)\}, \quad i > j,$$

$$A_j(i) := \int_{t=0}^{D_j(i)} (A_t - j) dt, \quad i > j, \quad (2)$$

where $A_{t-} := \lim_{\epsilon \downarrow 0} A_{t-\epsilon}$. Then, $D_j(i)$ is the first passage time of state j from state i , $N_j(i)$ the number of arrivals during this first passage time $D_j(i)$, and $A_j(i)$ is the area above j during the same period of time. Note that Guillemin and Simonian (GS) [3] have a slightly different interpretation of the number of arrivals.¹

An important sub-class of these passage times is the class of C-congestion periods. A C-congestion period is the duration until the first return to level C after an arriving customer raised the number of users above level C. So, a C-congestion period is the period that the system is continuously above level C. Duration D_C is defined by (1) where $i = C + 1$ and $j = C$. For short-hand notation we introduce $D_C := D_C(C + 1)$, $N_C := N_C(C + 1)$ and $A_C := A_C(C + 1)$. The special case where $C = 0$ is called the ‘busy period’.

2.2. Decomposition of a passage time into congestion periods

By its definition $D_j(i)$ is a stopping time of the Markov process A_t . It can be decomposed as the sum of the hitting times D_{i-1} and $D_j(i - 1)$. The strong Markov property states that these hitting times are independent. The first component is already a congestion period and the second term can be decomposed repeatedly in a similar way and finally results in the following equality in distribution:

$$D_j(i) = \sum_{k=j}^{i-1} D_k, \quad (3)$$

where the D_k for $k = j, \dots, i - 1$ are independent. Expression (3) resembles expression (5.1.1) of Keilson [4].

The number of arrivals $N_j(i)$ and the area $A_j(i)$ can also be decomposed, based on the decomposition of the duration $D_j(i)$, resulting in

$$N_j(i) = \sum_{k=j}^{i-1} N_k, \quad (4)$$

$$A_j(i) = \sum_{k=j}^{i-1} (A_k + (k - j)D_k). \quad (5)$$

Proof. Eq. (4) follows directly due to (3). Eq. (5) is obtained because area $A_j(i)$ can be decomposed in a similar way as $D_j(i)$ and $N_j(i)$, but caution is required because of the definition of the area. $A_j(i)$ can be decomposed into the terms A_{i-1} and $A_j(i - 1)$, but A_{i-1} only consists of the area above level $i - 1$, ignoring the area between $i - 1$ and j for the duration D_{i-1} . The missing area for A_{i-1} is $(i - 1 - j)D_{i-1}$ and correction of all terms A_k leads to (5). \square

This subdivision of the passage times into the sum of independent congestion periods simplifies the analysis. All the moments can be directly derived from the moments of the individual congestion periods, e.g., for the duration it yields

$$\mathbb{E}D_j(i) = \sum_{k=j}^{i-1} \mathbb{E}D_k \quad \text{and} \quad \mathbb{E}D_j^2(i) = \sum_{k=j}^{i-1} \mathbb{E}D_k^2 + 2 \sum_{k=j}^{i-2} \sum_{l=k+1}^{i-1} \mathbb{E}D_k \mathbb{E}D_l.$$

2.3. Analysis of a C-congestion period

In this section a recursive relation for the quantities of a C-congestion period are derived using straightforward analysis.

A C-congestion period is initiated by a customer who finds C other customers in the system upon arrival. The number of customers is increased to $C + 1$ and the system will remain at this level for an exponentially ν_{C+1} distributed time, as both the interarrival time and the service times are exponentially distributed. The next transition of

¹ GS [3] include the arrival that starts a C-congestion period. Formally this arrival did not occur within the C-congestion period as the customer entered the system when only C customers were present. Preater [7] also ignores the arrival that initiates the congestion period.

the system is caused either by the arrival of a new customer or by the departure of one of the $C + 1$ customers present. With probability $(C + 1)\mu/\nu_{C+1}$ the next transition is a departure, which immediately ends the currently ongoing C-congestion period. With probability λ/ν_{C+1} the next transition is initiated by an arrival, which increases the number of customers to $C + 2$; then the remaining duration of the C-congestion period is the duration of a transient period $D_C(C + 2)$.

Let T_C be the duration that the system remains at level C, which is exponentially ν_{C+1} distributed, and define the random variable P_C as

$$P_C = \begin{cases} 1 & \text{with probability } \lambda/\nu_C \\ 0 & \text{with probability } C\mu/\nu_C. \end{cases}$$

Notice that, as P_C is alternatively distributed all moments are the same: $\mathbb{E}P_C^k = \lambda/\nu_C$ for all k . Now, for the duration of a C-congestion period D_C the above reasoning leads to:

$$D_C = T_{C+1} + P_{C+1}D_C(C + 2) = T_{C+1} + P_{C+1}(D_{C+1} + D'_C). \quad (6)$$

Here X' denotes an independent, statistically identical copy of X . By the memoryless property of the exponential distribution all the random variables, e.g., T_{C+1} , P_{C+1} and $D_C(C + 2)$, are mutually independent. Expression (6) is a recursive relation which illustrates that the duration of a C-congestion period can be expressed in terms of the duration of a $(C - 1)$ -congestion period. By repeated iteration the duration can be expressed in terms of D_0 , which is the duration of a busy period.

For the quantities N_C and A_C the following similar relations can be derived:

$$N_C = P_{C+1}(1 + N_C(C + 2)) = P_{C+1}(1 + N_{C+1} + N'_C). \quad (7)$$

$$A_C = T_{C+1} + P_{C+1}A_C(C + 2) = T_{C+1} + P_{C+1}(A_{C+1} + D_{C+1} + A'_C). \quad (8)$$

Note that by definition of N_C the (possible) arrival that ends T_{C+1} and initiates the passage time $D_C(C + 2)$ is not accounted for in $N_C(C + 2)$ and has to be accounted for separately. The second equality of (8) follows directly from (5).

3. Quantities of a C-congestion period

In this section we present the mean and second moment of the duration, number of arrivals and the area swept above C. For the quantity duration the mean, the second moment and also higher moments are written out. Next, the moments of the number of arrivals and the mean of the area are rather trivial, the second moment of the area is more complicated as definition (8) includes a term D_C which requires the joint expectation of the quantities D_C and A_C .

3.1. Duration of a C-congestion period

For the derivations of the moments of the duration we use result (6) of Section 2.3. Although the expected duration of a congestion period is already given in Guillemin and Simonian [3], the derivation of the mean duration is presented to become acquainted with the methodology of the recursions.

3.1.1. Mean duration of a C-congestion period

Taking the expectation on both sides of expression (6) yields

$$\mathbb{E}D_C = \mathbb{E}[T_{C+1} + P_{C+1}(D_{C+1} + D'_C)] = \frac{1}{\nu_{C+1}} + \frac{\lambda}{\nu_{C+1}}(\mathbb{E}D_{C+1} + \mathbb{E}D_C).$$

By isolating $\mathbb{E}D_{C+1}$ at the left side, we obtain the following expression:

$$\mathbb{E}D_{C+1} = \frac{(C + 1)\mu\mathbb{E}D_C - 1}{\lambda}. \quad (9)$$

Expression (9) is a difference equation and illustrates that the mean duration of a $(C + 1)$ -congestion period depends on the mean duration of C-congestion period. By iteration $\mathbb{E}D_{C+1}$ (or preferably $\mathbb{E}D_C$) can be expressed in terms of

$\mathbb{E}D_0$, which is the expected duration of a busy period. This yields the following closed-form expression

$$\mathbb{E}D_C = \frac{C!}{\rho^C} \mathbb{E}D_0 - \frac{C!}{\lambda \rho^C} \sum_{j=1}^C \frac{\rho^j}{j!} = \frac{C!}{\lambda \rho^C} \sum_{j=C+1}^{\infty} \frac{\rho^j}{j!}, \quad (10)$$

where $\mathbb{E}D_0$ is obtained via renewal arguments. Let π_0 denote the fraction of time that the system is empty, T_{idle} the duration that the system is empty, and T_{busy} the duration that the system is busy. As $\pi_0 = e^{-\rho}$, $\mathbb{E}T_{\text{idle}} = 1/\lambda$, $\mathbb{E}D_0 = \mathbb{E}T_{\text{busy}}$ and $\pi_0 = \mathbb{E}T_{\text{idle}}/(\mathbb{E}T_{\text{busy}} + \mathbb{E}T_{\text{idle}})$ it follows that $\mathbb{E}D_0 = (e^{\rho} - 1)/\lambda$.

3.1.2. Second moment of duration of a C-congestion period

The second moment of the duration can also be obtained by taking the second moment of expression (6). Then we obtain

$$\begin{aligned} \mathbb{E}D_C^2 &= \mathbb{E}[T_{C+1} + P_{C+1}(D_{C+1} + D_C)]^2 \\ &= \mathbb{E}[T_{C+1}^2 + 2T_{C+1}P_{C+1}(D_{C+1} + D_C) + P_{C+1}^2(D_{C+1}^2 + 2D_{C+1}D_C + D_C^2)] \\ &= \frac{2}{v_{C+1}^2} + \frac{2\lambda}{v_{C+1}^2}(\mathbb{E}D_{C+1} + \mathbb{E}D_C) + \frac{\lambda}{v_{C+1}}(\mathbb{E}D_{C+1}^2 + 2\mathbb{E}D_{C+1}\mathbb{E}D_C + \mathbb{E}D_C^2) \end{aligned}$$

as $\mathbb{E}[D_{C+1}D_C] = \mathbb{E}D_{C+1}\mathbb{E}D_C$ by the strong Markov property. Rearranging leads to the following difference equation:

$$\mathbb{E}D_{C+1}^2 = \frac{C+1}{\rho} \mathbb{E}D_C^2 - \frac{2}{\lambda v_{C+1}} - \frac{2}{v_{C+1}}(\mathbb{E}D_{C+1} + \mathbb{E}D_C) - 2\mathbb{E}D_{C+1}\mathbb{E}D_C.$$

This equation can be solved in terms of $\mathbb{E}D_0^2$, the second moment of the duration of a busy period which is treated in Section 5.2, and yields

$$\mathbb{E}D_C^2 = \frac{C!}{\rho^C} \mathbb{E}D_0^2 - 2 \frac{C!}{\rho^C} \sum_{j=1}^C \frac{\rho^j}{j!} \frac{1}{v_j} [\mathbb{E}D_{j-1} + \mathbb{E}D_j] - 2 \frac{C!}{\rho^C} \sum_{j=1}^C \frac{\rho^j}{j!} \mathbb{E}D_{j-1} \mathbb{E}D_j - \frac{2}{\lambda} \frac{C!}{\rho^C} \sum_{j=1}^C \frac{\rho^j}{j!} \frac{1}{v_j}. \quad (11)$$

3.1.3. Higher moments of the duration of a C-congestion period

Higher moments can also be obtained using the recursive relation (6), although calculations are more tedious. By using the binomial theorem for both $\mathbb{E}D_C^n$ and $\mathbb{E}D_C^n(C+2)$ we obtain:

$$\begin{aligned} \mathbb{E}D_C^n &= \sum_{l=0}^n \binom{n}{l} \mathbb{E}T_{C+1}^{n-l} \mathbb{E}[P_{C+1}D_C(C+2)]^l \\ &= \frac{n!}{v_{C+1}^n} + \frac{\lambda}{v_{C+1}} \sum_{l=1}^{n-1} \binom{n}{l} \frac{(n-l)!}{v_{C+1}^{n-l}} \sum_{k=0}^l \binom{l}{k} \mathbb{E}D_{C+1}^k \mathbb{E}D_C^{l-k} + \frac{\lambda}{v_{C+1}} \sum_{l=0}^n \binom{n}{l} \mathbb{E}D_{C+1}^l \mathbb{E}D_C^{n-l}. \end{aligned}$$

Rearranging leads to a difference equation which can be solved in terms of $\mathbb{E}D_0^n$:

$$\begin{aligned} \mathbb{E}D_C^n &= \frac{C!}{\rho^C} \mathbb{E}D_0^n - \frac{C!}{\rho^C} \sum_{j=1}^C \frac{\rho^j}{j!} \sum_{l=1}^{n-1} \binom{n}{l} \frac{(n-l)!}{v_j^{n-l}} \sum_{k=0}^l \binom{l}{k} \mathbb{E}D_j^k \mathbb{E}D_{j-1}^{l-k} \\ &\quad - \frac{C!}{\rho^C} \sum_{j=1}^C \frac{\rho^j}{j!} \sum_{l=1}^{n-1} \binom{n}{l} \mathbb{E}D_j^l \mathbb{E}D_{j-1}^{n-l} - \frac{n!}{\lambda} \frac{C!}{\rho^C} \sum_{j=1}^C \frac{\rho^j}{j!} \frac{1}{v_j^{n-1}}. \end{aligned} \quad (12)$$

From expression (12) it can be observed that the n -th moment of level C depends on all moments $\mathbb{E}D_C^m$ for $m < n$ and $\mathbb{E}D_k^m$ for $k < C$, $m \leq n$. This illustrates that for $\mathbb{E}D_C^n$ all moments $\mathbb{E}D_0^m$ for $m = 1, \dots, n$ have to be known. This is a drawback as closed-form expressions for the second and higher moments are not presented in literature. An expression for $\mathbb{E}D_0^2$ will be derived in Section 5.2. The method can also be used for higher moments, but the calculations become substantially more tedious.

3.2. Number of arriving customers during a C-congestion period

The mean and second moment are obtained by taking the expectation of expression (7) and the square of expression (7) respectively.

3.2.1. Mean number of arriving customers in a C-congestion period

Taking the expectation of expression (7) and rearranging leads to a difference equation in terms of $\mathbb{E}N_0$, which is the number of arrivals during a busy period. $\mathbb{E}N_0$ is easily obtained as $\mathbb{E}N_0 = \lambda \mathbb{E}D_0 = e^\rho - 1$ and the solution of the difference equation is the following closed-form expression:

$$\mathbb{E}N_C = \frac{C!}{\rho^C} \mathbb{E}N_0 - \frac{C!}{\rho^C} \sum_{j=0}^{C-1} \frac{\rho^j}{j!} = \frac{C!}{\rho^C} \sum_{j=C+1}^{\infty} \frac{\rho^j}{j!}. \quad (13)$$

3.2.2. Second moment of the number of arriving customers

The second moment is derived in terms of $\mathbb{E}N_0^2$ in a similar manner and yields

$$\mathbb{E}N_C^2 = \frac{C!}{\rho^C} \mathbb{E}N_0^2 - \frac{C!}{\rho^C} \sum_{j=1}^C \frac{\rho^j}{j!} (1 + 2\mathbb{E}N_j \mathbb{E}N_{j-1} + 2\mathbb{E}N_j + 2\mathbb{E}N_{j-1}). \quad (14)$$

For the derivation of $\mathbb{E}N_0^2$ we refer to Section 5.3 of [10].

3.3. Area swept above C during a C-congestion period

The mean and second moment can be obtained by using expression (8).

3.3.1. Mean area swept above C

Taking the expectation of (8) leads to a difference equation that can be solved iteratively in terms of $\mathbb{E}A_0$. $\mathbb{E}A_0$, the area above 0 during a busy period, can be obtained by considering that the system is a renewal process of cycles consisting of busy and idle periods. The average workload ρ during a cycle should all be obtained during a busy period. Then $\rho = \mathbb{E}A_0 / (\mathbb{E}D_0 + 1/\lambda)$ and thus $\mathbb{E}A_0 = \rho e^\rho$. Finally, we obtain the following closed-form expression for $\mathbb{E}A_C$:

$$\mathbb{E}A_C = \frac{C!}{\rho^C} \mathbb{E}A_0 - \sum_{j=1}^C \frac{C!}{j!} \left(\mathbb{E}D_j + \frac{1}{\lambda} \right) = \frac{1}{\lambda} \frac{C!}{\rho^C} \sum_{j=C+1}^{\infty} \frac{\rho^j}{j!}. \quad (15)$$

3.3.2. Second moment of the area swept above C

Taking the second moment of (8) and isolating $\mathbb{E}A_{C+1}^2$ leads to a difference equation. The difference equation includes the ‘joint expectation’ $\mathbb{E}[D_{C+1}A_{C+1}]$, which results from the term $\mathbb{E}[A_C(C+2)]^2$. By definition (2) A_{C+1} is dependent on D_{C+1} , hence $\mathbb{E}[D_{C+1}A_{C+1}] \neq \mathbb{E}D_{C+1}\mathbb{E}A_{C+1}$; an expression for $\mathbb{E}[D_{C+1}A_{C+1}]$ is required and will be derived in Section 4. Then, the difference equation can be solved in terms of $\mathbb{E}A_0^2$ (see Section 5.4 of [10]) and the solution yields

$$\begin{aligned} \mathbb{E}A_C^2 &= \frac{C!}{\rho^C} \mathbb{E}A_0^2 - \frac{C!}{\rho^C} \sum_{j=1}^C \frac{\rho^j}{j!} (\mathbb{E}D_j^2 + 2\mathbb{E}A_j \mathbb{E}A_{j-1} + 2\mathbb{E}[D_j A_j] + 2\mathbb{E}D_j \mathbb{E}A_{j-1}) \\ &\quad - 2 \frac{C!}{\rho^C} \sum_{j=1}^C \frac{\rho^j}{j!} \frac{1}{v_j} (\mathbb{E}A_j + \mathbb{E}D_j + \mathbb{E}A_{j-1}) - \frac{2}{\lambda} \frac{C!}{\rho^C} \sum_{j=1}^C \frac{\rho^j}{j!} \frac{1}{v_j}. \end{aligned} \quad (16)$$

Observe that expression (16) requires, besides $\mathbb{E}A_0^2$, the terms $\mathbb{E}D_j^2$ and $\mathbb{E}[D_j A_j]$ for $1 \leq j \leq C$. Recall that $\mathbb{E}D_j^2$ is given by (12), and $\mathbb{E}[D_j A_j]$ can be obtained from Section 4, so expressions are available for all the required terms.

4. Joint expectations of the C-congestion period quantities

In this section the joint expectations $\mathbb{E}[D_C N_C]$, $\mathbb{E}[D_C A_C]$ and $\mathbb{E}[N_C A_C]$ are derived. The covariances between the quantities can easily be found as, e.g., $\text{Cov}(D_C, N_C) = \mathbb{E}[D_C N_C] - \mathbb{E}D_C \mathbb{E}N_C$. Furthermore, the joint expectation $\mathbb{E}[D_C A_C]$ is required to determine the second moment of the area swept above k for all $k \geq C$, see Section 3.3.

4.1. Joint expectation of the duration and number of arrivals

By (6) and (7) we have

$$\begin{aligned}\mathbb{E}[D_C N_C] &= \mathbb{E}[(T_{C+1} + P_{C+1} D_C(C+2)) P_{C+1}(1 + N_C(C+2))] \\ &= \frac{\lambda}{v_{C+1}} (\mathbb{E}T_{C+1} + \mathbb{E}T_{C+1} \mathbb{E}N_{C+1} + \mathbb{E}T_{C+1} \mathbb{E}N_C + \mathbb{E}[D_{C+1} N_{C+1}] \\ &\quad + \mathbb{E}D_{C+1} \mathbb{E}N_C + \mathbb{E}D_C \mathbb{E}N_{C+1} + \mathbb{E}[D_C N_C] + \mathbb{E}D_{C+1} + \mathbb{E}D_C).\end{aligned}$$

This difference equation can be solved in terms of $\mathbb{E}[D_0 N_0]$, the derivation of which is presented in Section 5.3, and yields

$$\begin{aligned}\mathbb{E}[D_C N_C] &= \frac{C!}{\rho^C} \mathbb{E}[D_0 N_0] - \frac{C!}{\rho^C} \sum_{j=1}^C \frac{\rho^j}{j!} (\mathbb{E}D_j \mathbb{E}N_{j-1} + \mathbb{E}D_{j-1} \mathbb{E}N_j + \mathbb{E}D_j + \mathbb{E}D_{j-1}) \\ &\quad - \frac{C!}{\rho^C} \sum_{j=1}^C \frac{\rho^j}{j!} \frac{1}{v_j} (1 + \mathbb{E}N_j + \mathbb{E}N_{j-1}).\end{aligned}\tag{17}$$

4.2. Joint expectation of the duration and the area swept above C

By (6) and (8) we have

$$\mathbb{E}[D_C A_C] = \mathbb{E}[(T_{C+1} + P_{C+1} D_C(C+2))(T_{C+1} + P_{C+1} A_C(C+2))].$$

Isolating $\mathbb{E}[D_{C+1} A_{C+1}]$ yields

$$\begin{aligned}\mathbb{E}[D_{C+1} A_{C+1}] &= \frac{C+1}{\rho} \mathbb{E}[D_C A_C] - (\mathbb{E}D_{C+1}^2 + \mathbb{E}D_C \mathbb{E}A_{C+1} + \mathbb{E}D_{C+1} \mathbb{E}A_C + \mathbb{E}D_{C+1} \mathbb{E}D_C) \\ &\quad - \frac{1}{v_{C+1}} (2\mathbb{E}D_{C+1} + \mathbb{E}D_C + \mathbb{E}A_{C+1} + \mathbb{E}A_C) - \frac{2}{\lambda v_{C+1}}.\end{aligned}$$

Notice that this expression includes a term $\mathbb{E}D_{C+1}^2$ which results from the decompositions of $D_C(C+2)$ and $A_C(C+2)$ that both consist of a term D_{C+1} . The difference equation can be solved in terms of $\mathbb{E}[D_0 A_0]$, which is deduced in Section 5.6 of [10], and yields

$$\begin{aligned}\mathbb{E}[D_C A_C] &= \frac{C!}{\rho^C} \mathbb{E}[D_0 A_0] - \frac{C!}{\rho^C} \sum_{j=1}^C \frac{\rho^j}{j!} (\mathbb{E}D_j^2 + \mathbb{E}D_{j-1} \mathbb{E}A_j + \mathbb{E}D_{j-1} \mathbb{E}A_j + \mathbb{E}D_j \mathbb{E}D_{j-1}) \\ &\quad - \frac{C!}{\rho^C} \sum_{j=1}^C \frac{\rho^j}{j!} \frac{1}{v_j} (2\mathbb{E}D_j + \mathbb{E}D_{j-1} + \mathbb{E}A_j + \mathbb{E}A_{j-1}) - \frac{2}{\lambda} \frac{C!}{\rho^C} \sum_{j=1}^C \frac{\rho^j}{j!} \frac{1}{v_j}.\end{aligned}\tag{18}$$

Observe that the solution requires the second moments $\mathbb{E}D_j^2$ for $1 \leq j \leq C$, which are given by (11).

4.3. Joint expectation of the number of arrivals and the area swept above C

By (7) and (8) we have

$$\mathbb{E}[N_C A_C] = \mathbb{E}[P_{C+1}(1 + N_C(C+2))(T_{C+1} + P_{C+1} A_C(C+2))].$$

The solution in terms of $\mathbb{E}[N_0 A_0]$, see Section 5.7 of [10], yields

$$\begin{aligned} \mathbb{E}N_c A_c &= \frac{c!}{\rho^c} \mathbb{E}[N_0 A_0] - \frac{c!}{\rho^c} \sum_{j=1}^c \frac{\rho^j}{j!} (\mathbb{E}[D_j N_j] + \mathbb{E}N_j \mathbb{E}A_{j-1} + \mathbb{E}N_{j-1} \mathbb{E}A_j + \mathbb{E}D_j \mathbb{E}N_{j-1} \\ &\quad + \mathbb{E}A_j + \mathbb{E}D_j + \mathbb{E}A_{j-1}) - \frac{c!}{\rho^c} \sum_{j=1}^c \frac{\rho^j}{j!} \frac{1}{v_j} (1 + \mathbb{E}N_j + \mathbb{E}N_{j-1}). \end{aligned} \quad (19)$$

5. Moments and joint expectations of the busy-period quantities

In Sections 3 and 4 expressions were obtained for the moments and joint expectations for the quantities of a C-congestion period. The expressions are all solved in terms of the busy-period quantities (i.e., 0-congestion period quantities). The goal of this section is to derive these busy-period quantities. This section only presents the first and second moments of the duration and the joint expectation of the duration and number of arrivals; the derivations of the other busy-period quantities $\mathbb{E}N_0$, $\mathbb{E}A_0$, $\mathbb{E}[D_0 A_0]$ and $\mathbb{E}[N_0 A_0]$ are presented Section 5 of [10].

The moments of the quantities are obtained by differentiating the Laplace transform (LT) of the congestion triple (D_0, N_0, A_0) that was obtained by Preater [7]. Section 5.1 presents Preater's LT and additionally a lemma that simplifies the calculations that are presented in the succeeding subsections.

Theoretically, all moments and joint expectations of the quantities of level C can be obtained by differentiating Preater's LT of the congestion triple, but this task appeared to be far from trivial. Therefore we decided to first express them in terms of moments and joint expectations of the 0-congestion period; subsequently, we derive these 0-congestion period quantities through (relatively easy, but still tedious) differentiations.

5.1. Preater's LT of the 0-congestion triple (D_0, N_0, A_0)

Analogously to Guillemin and Simonian [3], Preater uses $\mu = 1$, and so $\lambda = \rho$. To obtain the Laplace transform of the C-congestion triple, Preater first considers the LT of the duration of a C-congestion period. By two different derivations he obtains the LT in two different expressions: the first is a continued fraction, the second is a fraction of the functions I_{C+1} and I_C (see (21)). The equality of these two expressions is the most important result of his Proposition 2.2. In his Theorem 3.1 he derives the (joint) LT of the congestion triple by the first derivation and the result is also in the form of a continued fraction. Using the equality of his Proposition 2.2, the continued fraction can be rewritten as a fraction of I_{C+1} and I_C . The LT for $C = 0$ resulting from his Proposition 2.2 and Theorem 3.1 is stated below.

Preater's Theorem 3.1 and Proposition 2.2 combined for $C=0$. The vector (D_0, N_0, A_0) has Laplace transform

$$\Theta_0^*(s, t, u) := \mathbb{E} \exp(-s D_0 - t N_0 - u A_0) = \frac{1}{u+1} \frac{I_1(a-b, b)}{I_0(a-b, b)} \quad (20)$$

where

$$\begin{aligned} a &:= a(s, t, u) = \frac{s + \rho}{u + 1}, & b &:= b(s, t, u) = \frac{\rho e^{-t}}{(u + 1)^2}, \\ I_C(a, b) &:= \int_0^1 e^{-bx} (1-x)^{a-1} x^C dx. \end{aligned} \quad (21)$$

Differentiating (21) is a tedious job, but can be simplified considerably by the next lemma.

Lemma 1.

$$I_0(a, b) = e^{-b} \sum_{k=0}^{\infty} \frac{1}{a+k} \frac{b^k}{k!}, \quad I_1(a, b) = I_0(a, b) - I_0(a+1, b). \quad (22)$$

Proof of (22).

$$\begin{aligned}
I_0(a, b) &= \int_0^1 e^{-bx} (1-x)^{a-1} dx = e^{-b} \int_0^1 e^{bx} x^{a-1} dx \\
&= e^{-b} \sum_{k=0}^{\infty} \frac{b^k}{k!} \int_0^1 x^{k+a-1} dx = e^{-b} \sum_{k=0}^{\infty} \frac{1}{a+k} \frac{b^k}{k!}. \\
I_1(a, b) &= \int_0^1 e^{-bx} (1-x)^{a-1} x dx = e^{-b} \int_0^1 e^{bx} x^{a-1} (1-x) dx \\
&= e^{-b} \sum_{k=0}^{\infty} \frac{b^k}{k!} \int_0^1 (x^{k+a-1} - x^{k+a}) dx \stackrel{\text{by (22)}}{=} I_0(a, b) - I_0(a+1, b). \quad \square
\end{aligned}$$

Furthermore, we introduce the following notation:

$$\xi(\rho) := \sum_{k=0}^{\infty} \frac{1}{(k+1)^2} \frac{\rho^k}{k!}.$$

Notice that $\xi(\rho) < \infty$.

5.2. Moments of the duration of the busy period

By (20) and using (22) we have

$$D_0^*(s) = \Theta_0^*(s, 0, 0) = 1 - \frac{f(s)}{n(s)}, \quad (23)$$

where

$$f(s) := \sum_{k=0}^{\infty} \frac{1}{s+k+1} \frac{\rho^k}{k!} \quad \text{and} \quad n(s) := \sum_{k=0}^{\infty} \frac{1}{s+k} \frac{\rho^k}{k!}.$$

Let $n^{(m)}(s)$ denote the m -th derivative of $n(s)$ (hence $n(s) = n^{(0)}(s)$). Then it can be shown that

$$n^{(m)}(s) = \sum_{k=0}^{\infty} \frac{(-1)^m \cdot m!}{(s+k)^{m+1}} \frac{\rho^k}{k!} \quad \text{and} \quad \lim_{s \rightarrow 0} n^{(m)}(s) \sim \frac{(-1)^m m!}{s^{m+1}}.$$

The first equation is obtained by repeated derivation of $n(s)$. The second statement is obtained by proving that the $(k \geq 1)$ -terms can be bounded by a finite term as follows:

$$\sum_{k=1}^{\infty} \frac{\rho^k}{k^m \cdot k!} \leq \sum_{k=1}^{\infty} \frac{\rho^k}{k!} < \sum_{k=0}^{\infty} \frac{\rho^k}{k!} = e^\rho.$$

Then the second statement is proven by observing that the second statement is exactly the $(k = 0)$ -term which goes to infinity for s close to 0.

5.2.1. First moment

Although the first moment is already obtained in Section 3.1, we also present its derivation for the sake of completeness. It is well known that $\mathbb{E}D_0 = -(D_0^*)'(0)$. Differentiation of (23) yields

$$(D_0^*)'(s) = \frac{d}{ds} \left(1 - \frac{f(s)}{n(s)} \right) = \frac{n'(s)f(s)}{n^2(s)} - \frac{f'(s)}{n(s)}.$$

We conclude that $\mathbb{E}D_0 = f(0) - 0 = (e^\rho - 1)/\rho$, which coincides with the results earlier obtained in Section 3.1 for $\mu = 1$.

5.2.2. Second moment

Now $\mathbb{E}D_0^2 = (D_0^*)''(0)$. The second derivative is

$$(D_0^*)''(s) = \frac{d}{ds} \left(\frac{n'(s)f(s)}{n^2(s)} - \frac{f'(s)}{n(s)} \right) = -\frac{f''(s)}{n(s)} + 2\frac{n'(s)f'(s)}{n^2(s)} - 2\frac{(n'(s))^2 f(s)}{n^3(s)} + \frac{n''(s)f(s)}{n^2(s)}.$$

The first of these four terms goes to 0, and the second to $-2f'(0)$. The third term goes to $-\infty$, and, as $n''(s) \sim 2/s^3$, the fourth term goes to $+\infty$. Define for ease

$$g_n(s) := \sum_{k=1}^{\infty} \frac{1}{(k+s)^n} \frac{\rho^k}{k!};$$

for any $n \in \mathbb{N}$, it holds that $g_n(0) < e^\rho < \infty$. Simple manipulations yield

$$\lim_{s \downarrow 0} \left(\frac{n''(s)}{n^2(s)} - 2\frac{(n'(s))^2}{n^3(s)} \right) = \lim_{s \downarrow 0} \frac{(s^{-1} + g_1(s))(2s^{-3} + 2g_3(s)) - 2(s^{-2} + g_2(s))^2}{(s^{-1} + g_1(s))^3} = 2g_1(0).$$

Thus

$$\mathbb{E}D_0^2 = 2g_1(0)f(0) - 2f'(0) = 2 \left(\sum_{k=1}^{\infty} \frac{1}{k} \frac{\rho^k}{k!} \right) \frac{e^\rho - 1}{\rho} + 2 \sum_{k=0}^{\infty} \frac{1}{(k+1)^2} \frac{\rho^k}{k!} = 2e^\rho \xi(\rho). \quad (24)$$

5.2.3. Relation between (24) and the results of Liu and Shi [6]

Liu and Shi [6] obtained the following expression for the second moment of the busy period of an M/G/∞-queue:

$$\mathbb{E}D_0^2 = \frac{2}{\lambda P_0^2} \int_0^\infty [P_0(t) - P_0] dt$$

where $P_0(t) = \exp\{-\rho \int_0^t e^{-x} dx\} = \exp\{-\rho(1-e^{-t})\}$ and P_0 is the probability that the system is idle, thus $P_0 = e^{-\rho}$. Then, by using that $\exp\{\rho e^{-t}\} = \sum_{k=0}^{\infty} (\rho e^{-t})^k / k!$, we have

$$\begin{aligned} \frac{2}{\rho P_0^2} \int_0^\infty [P_0(t) - P_0] dt &= \frac{2e^{2\rho}}{\rho} \int_0^\infty e^{-\rho} [e^{\rho e^{-t}} - 1] dt = \frac{2e^\rho}{\rho} \int_0^\infty \sum_{k=1}^{\infty} \frac{(\rho e^{-t})^k}{k!} dt \\ &= \frac{2e^\rho}{\rho} \sum_{k=1}^{\infty} \frac{\rho^k}{k!} \int_0^\infty e^{-kt} dt = \frac{2e^\rho}{\rho} \sum_{k=0}^{\infty} \frac{\rho^k}{(k+1)^2 k!} = 2e^\rho \xi(\rho). \end{aligned}$$

We conclude that expression (24) and the result of Liu and Shi [6] coincide.

5.3. Joint expectation $\mathbb{E}[D_0 N_0]$ of the busy period

The joint expectation $\mathbb{E}[D_0 N_0]$ can be obtained by differentiating the Laplace transform (20) by both s and t :

$$\mathbb{E}[D_0 N_0] = \lim_{s \downarrow 0, t \downarrow 0} \frac{d^2}{ds dt} \mathbb{E}e^{-s D_0 - t N_0}.$$

By (20) and (22) we have

$$\mathbb{E}e^{-s D_0 - t N_0} = \Theta_0^*(s, t, u) = 1 - \frac{I_0(a - b + 1, b)}{I_0(a - b, b)} = 1 - \frac{f(s, t)}{n(s, t)},$$

where $a = a(s, t) = s + \rho$ and $b = b(s, t) = \rho e^{-t}$ and by the definition we have

$$n(s, t) := \sum_{k=0}^{\infty} \frac{\rho^k}{k!} \frac{e^{-kt}}{s + \rho(1 - e^{-t}) + k}; \quad f(s, t) := \sum_{k=0}^{\infty} \frac{\rho^k}{k!} \frac{e^{-kt}}{s + \rho(1 - e^{-t}) + k + 1}.$$

Define $n'_s := dn(s, t)/ds$, $n'_t := dn(s, t)/dt$ and $n''_{st} := d^2n(s, t)/dtds$. Analogously, the derivatives f'_s , f'_t and f''_{st} are defined. Then

$$\frac{d^2}{dtds} \mathbb{E} e^{-sD_0-tN_0} = \frac{d}{dt} \left[\frac{fn'_s}{n^2} - \frac{f'_s}{n} \right] = -\frac{f''_{st}}{n} + \frac{f'_s n'_t + f'_t n'_s}{n^2} + \frac{fn''_{st}}{n^2} - \frac{2fn'_s n'_t}{n^3}.$$

For $s, t \rightarrow 0$, the first term goes to 0, and the second term yields $-\rho f'_s(0, 0) - f'_t(0, 0)$. The third and fourth terms result in $(\infty - \infty)$, and require careful analysis, similar as for $\mathbb{E}D_0^2$. This eventually yields $2\rho^2 \xi(\rho) f(0, 0)$. Then we obtain the following expression for $\mathbb{E}[D_0 N_0]$:

$$\mathbb{E}[D_0 N_0] = -\rho f'_s(0, 0) - f'_t(0, 0) + 2\rho^2 \xi(\rho) f(0, 0). \quad (25)$$

Notice that $\mathbb{E}[D_0 N_0]$ is bounded for finite ρ ; $f(0, 0) \leq \sum_{k=0}^{\infty} \rho^k / k! = e^\rho$ and similar bounds can be obtained for $f'_s(0, 0)$ and $f'_t(0, 0)$.

5.4. Moments for service times other than 1

In this section expressions are derived for the moments of D_0 and the joint expectation of $D_0 N_0$ with mean service time $\mu = 1$. To adapt the derived expressions for service times $\mu \neq 1$, it suffices to see that varying λ or μ for fixed ρ is only a scaling of time. Time scaling does not influence the number of arrivals, but it does influence the duration and area. The expressions derived in this section can be adapted to $\mu \neq 1$ by a factor $(\mu^{-1})^n$ where n is the order of the moment, e.g., in self-evident notation:

$$\mathbb{E}D_C^n = \mu^{-n} \mathbb{E}D_{C|\{\mu=1\}}^n, \quad \mathbb{E}D_C N_C = \mu^{-1} \mathbb{E}D_C N_{C|\{\mu=1\}}.$$

6. C-intercongestion periods

Besides the duration of a C-congestion period, we are also interested in the time that the system is below level C, a so-called C-intercongestion period. This section consists of the definitions of a C-intercongestion period, the derivation of a LT of the intercongestion triple, and the derivation of the first and second moments of the quantities and the covariances between the quantities.

6.1. Definitions

Analogously to the definitions of a C-congestion period in Section 2.1 we define

$$\begin{aligned} \mathcal{D}_j(i) &:= \inf\{t > 0 : A_t = j \mid A_0 = i\}, \quad i < j, \\ \mathcal{N}_j(i) &:= \#\{t : A_t - A_{t-} = 1, 0 < t \leq \mathcal{D}_j(i)\}, \quad i < j, \\ \mathcal{A}_j(i) &:= \int_{t=0}^{\mathcal{D}_j(i)} (j - A_t) dt, \quad i < j, \end{aligned}$$

where $\mathcal{D}_j(i)$ is the duration of the transient period to go from state i to state j for $i < j$, $\mathcal{N}_j(i)$ is the number of arrivals during $\mathcal{D}_j(i)$, and $\mathcal{A}_j(i)$ is the area under C during $\mathcal{D}_j(i)$. For convenience we use the following short notation: $\mathcal{D}_C := \mathcal{D}_C(C-1)$, $\mathcal{N}_C := \mathcal{N}_C(C-1)$ and $\mathcal{A}_C := \mathcal{A}_C(C-1)$. $\mathcal{D}_j(i)$ is a hitting time which allows for the following decomposition $\mathcal{D}_j(i) = \sum_{k=i+1}^j \mathcal{D}_k$. Due to the definition of $\mathcal{N}_j(i)$ and $\mathcal{A}_j(i)$ these can be similarly decomposed $\mathcal{N}_j(i) = \sum_{k=i+1}^j \mathcal{N}_k$ and $\mathcal{A}_j(i) = \sum_{k=i+1}^j (\mathcal{A}_k + (j-k)\mathcal{D}_k)$.

Finally, we derive a recursive structure for a C-intercongestion period. Using random variables T_C and P_C , which have the same definition as in Section 2.1, the duration \mathcal{D}_C ($C \geq 2$) can be subdivided into the independent durations T_{C-1} and $\mathcal{D}_C(C-2)$ as follows:

$$\mathcal{D}_C = T_{C-1} + (1 - P_{C-1})\mathcal{D}_C(C-2) \quad \text{for } C \geq 2.$$

Similarly decompositions of \mathcal{N}_C and \mathcal{A}_C give the following recursive equations:

$$\begin{aligned} \mathcal{N}_C &= P_{C-1} + (1 - P_{C-1})\mathcal{N}_C(C-2) \quad \text{for } C \geq 2, \\ \mathcal{A}_C &= T_{C-1} + (1 - P_{C-1})\mathcal{A}_C(C-2) \quad \text{for } C \geq 2. \end{aligned}$$

This set of equations again gives rise to recursions for D_C , \mathcal{N}_C and \mathcal{A}_C , which can be solved in terms of D_0 , \mathcal{N}_0 and \mathcal{A}_0 .

6.2. Laplace transforms of the duration and the intercongestion triple

The derivation of the Laplace transforms is done analogously to the derivation of the LTs of the congestion period done by Preater [7]. First, the LT of the duration will be derived in two different ways which result in two different forms. The equality of these two forms is exploited in the derivation of the LT of the intercongestion triple. In this section we follow Preater's assumption that $\mu = 1$.

6.2.1. Laplace transform of the intercongestion period duration

Lemma 2. Let x_n be a non-negative, bounded sequence satisfying

$$x_n := \frac{a + bn}{n + c - x_{n-1}}, \quad n \geq 1,$$

where $a, c > 0, b \geq 0$. Then

$$x_0 = \mathcal{L}(a, b, c) := -1 - c + \frac{a + b}{-2 - c + \frac{a+2b}{-3-c+\frac{a+3b}{-4-c+\dots}}}. \quad (26)$$

Proof. The proof is derived by mimicking Lemma 2.1 of [7] for

$$x_{n-1} := -n - c + \frac{a + bn}{x_n}, \quad n \geq 1.$$

Writing x_0 as a continued fraction yields (26). \square

Proposition 1. The Laplace transform of the duration \mathcal{D}_C is

$$\begin{aligned} \mathcal{D}_C^*(s) &= c^{-1} \mathcal{L}(\lambda c, \lambda, \lambda + s + c - 1) \\ &= \frac{\lambda}{c} \frac{I_C(s, \lambda)}{I_{C-1}(s, \lambda)} \frac{\sum_{k=0}^{c-1} \binom{c-1}{k} \frac{\lambda^k}{k!} I_{2k}(s + c - 1 - k, \lambda)}{\sum_{k=0}^c \binom{c}{k} \frac{\lambda^k}{k!} I_{2k}(s + c - k, \lambda)}. \end{aligned} \quad (27)$$

Proof. (a) An n -intercongestion period starts with a sojourn time T_{n-1} at level $n - 1$. At the end of T_{n-1} with probability $(n - 1)/(\lambda + n - 1)$ a customer departs, starting a $(n - 1)$ -intercongestion period followed by another sojourn at level $n - 1$. At the end of each sojourn time T_{n-1} a new $(n - 1)$ -intercongestion period can be started by a departure or the n -intercongestion can be ended by the arrival of a new customer. With obvious notation, we find the following equality in distribution:

$$\mathcal{D}_n = T_{n-1}^{(0)} + \sum_{i=1}^{G_{n-1}-1} (\mathcal{D}_{n-1}^{(i)} + T_{n-1}^{(i)}), \quad n \geq 0,$$

where all variables on the right are independent, T_n is exponentially $(\lambda + n)$ distributed and G_n is geometrically $(p_n := \lambda/(\lambda + n))$ distributed. Then,

$$T_n^*(s) := \mathbb{E}e^{-sT_n} = \frac{\lambda + n}{s + \lambda + n},$$

and

$$\mathcal{D}_n^*(s) = \frac{p_{n-1} T_{n-1}^*(s)}{1 - (1 - p_{n-1}) T_{n-1}^*(s) \mathcal{D}_{n-1}^*(s)} = \frac{\lambda}{n - 1 + s + \lambda - (n - 1) \mathcal{D}_{n-1}^*(s)}. \quad (28)$$

Let $x_n = (n + c)\mathcal{D}_{n+c}^*(s)$. Then (28) fulfils the setting of Lemma 2 with $a = \lambda c$, $b = \lambda$ and $c = \lambda + s + c - 1$. Hence, the first equality in (27) follows from (26).

(b) We follow the lines of the proof of Proposition 2.2 of Preater [7]. Let X_t be a stationary version of the M/M/ ∞ occupation process, so X_t is Poisson (ρ) distributed ($\rho = \lambda$ as $\mu = 1$). Preater defined $\pi_n(t) := \mathbb{P}(X_t = n \mid X_0 = 0)$, which is Poisson ($\lambda(1 - e^{-t})$) distributed and has LT $\pi_n^*(s) = (\lambda^n/n!)I_n(s, \lambda)$. Additionally we define $\chi_n(t) := \mathbb{P}(X_t = n \mid X_0 = n)$ and denote its LT by $\chi_n^*(s)$. By conditioning on the number $k \leq n$ of the initial n customers that were present at epoch 0 and that are still present at epoch t , we obtain

$$\chi_n(t) = \sum_{k=0}^n \binom{n}{k} (1 - e^{-t})^{n-k} (e^{-t})^k \pi_k(t).$$

Then its LT can be obtained as follows:

$$\begin{aligned} \chi_n^*(s) &= \int_0^\infty e^{-st} \sum_{k=0}^n \binom{n}{k} (1 - e^{-t})^{n-k} (e^{-t})^k \frac{(\lambda(1 - e^{-t}))^k}{k!} e^{-\lambda(1 - e^{-t})} dt \\ &= \sum_{k=0}^n \binom{n}{k} \frac{\lambda^k}{k!} \int_0^1 (1 - u)^{s+n-k-1} u^{2k} e^{-\lambda u} du = \sum_{k=0}^n \binom{n}{k} \frac{\lambda^k}{k!} I_{2k}(s + n - k, \lambda), \end{aligned}$$

by using the substitution $u := 1 - e^{-t}$ in the second step.

Next, we introduce the first passage time $\tau_n := \inf\{t \geq 0 : X_t = n \mid X_0 = 0\}$. Then for $n \geq 0$

$$\int_0^t \chi_n(t - x) \mathbb{P}(\tau_n \in dx) = \pi_n(t).$$

Taking Laplace transforms on both sides results in $\mathbb{E}e^{-s\tau_n} = \pi_n^*(s)/\chi_n^*(s)$. We thus obtain

$$\mathcal{D}_c^*(s) = \frac{\mathbb{E}e^{-s\tau_c}}{\mathbb{E}e^{-s\tau_{c-1}}} = \frac{\lambda}{c} \frac{I_c(s, \lambda)}{I_{c-1}(s, \lambda)} \frac{\sum_{k=0}^{c-1} \binom{c-1}{k} \frac{\lambda^k}{k!} I_{2k}(s + c - 1 - k, \lambda)}{\sum_{k=0}^c \binom{c}{k} \frac{\lambda^k}{k!} I_{2k}(s + c - k, \lambda)}$$

which proves the second equality in (27). \square

A ‘sanity check’ of (27) is the special case $c = 1$; for $c = 1$ the intercongestion period reduces to an exponentially (λ) distributed idle period. In Appendix B of [10] it is shown that then (27) indeed reduces to $\lambda/(\lambda + s)$.

6.2.2. Laplace transform of c -intercongestion triple $(\mathcal{D}_c, \mathcal{N}_c, \mathcal{A}_c)$

Theorem 1. Let $c \in \mathbb{N}$. The vector $(\mathcal{D}_c, \mathcal{N}_c, \mathcal{A}_c)$ has LT

$$\Omega_c^*(s, t, u) := \mathbb{E} \exp(-s\mathcal{D}_n - t\mathcal{N}_n - u\mathcal{A}_n) = c^{-1} \mathcal{L}(a'c, a', b')$$

where $a' := \lambda e^{-t}$; $b' := s + \lambda + u + c - 1$.

In particular,

$$\Omega_c^*(s - u, t, u) = \frac{\lambda}{c} \frac{I_c(s, \lambda)}{I_{c-1}(s, \lambda)} \frac{\sum_{k=0}^{c-1} \binom{c-1}{k} \frac{\lambda^k}{k!} I_{2k}(s + c - 1 - k, \lambda)}{\sum_{k=0}^c \binom{c}{k} \frac{\lambda^k}{k!} I_{2k}(s + c - k, \lambda)}.$$

Proof.

$$\begin{aligned} \Omega_n^*(s, t, u) &= \mathbb{E} \exp(-s\mathcal{D}_n - t\mathcal{N}_n - u\mathcal{A}_n) \\ &= \mathbb{E} \exp(-(s + u)\mathcal{D}_n - t\mathcal{N}_n - u(\mathcal{A}_n - \mathcal{D}_n)) \\ &= \mathbb{E} \exp\left(-(s + u)T_{n-1}^{(0)} - t - \sum_{i=1}^{G_{n-1}-1} \left[(s + u)T_{n-1}^{(i)} + (s + u)\mathcal{D}_{n-1} + t\mathcal{N}_{n-1} + u\mathcal{A}_{n-1}\right]\right) \end{aligned}$$

$$\begin{aligned}
&= T_{n-1}^*(s+u)p_n e^{-t} [1 - (1-p_n)T_{n-1}^*(s+u)\Omega_{n-1}^*(s+u, t, u)]^{-1} \\
&= \frac{\lambda e^{-t}}{n-1+s+u+\lambda - (n-1)\Omega_{n-1}^*(s+u, t, u)}.
\end{aligned} \tag{29}$$

Let $x_n = (n+c)\Omega_{n+c}^*(s-nu, t, u)$. Then (29) falls in the framework of Lemma 2 with $a = a'c$; $b = a'$; $c = b'$. \square

6.3. Moments of the C-intercongestion period quantities

The derivations of the moments and joint expectations of the intercongestion-period quantities are analogous to the derivation of the congestion-period quantities in Sections 3 and 4, although there is a large difference in obtaining the starting conditions. As the system can never have less than 0 customers, all quantities corresponding to level 0 are 0 themselves, e.g., $\mathbb{E}\mathcal{D}_0^n = 0$, $\mathbb{E}\mathcal{N}_0^n = 0$, $\mathbb{E}\mathcal{A}_0^n = 0$, $\mathbb{E}\mathcal{D}_0\mathcal{N}_0 = 0$, $\mathbb{E}\mathcal{D}_0\mathcal{A}_0 = 0$, $\mathbb{E}\mathcal{N}_0\mathcal{A}_0 = 0$.

6.3.1. Moments of the duration of an C-intercongestion period

$$\mathbb{E}\mathcal{D}_C = \frac{1}{\lambda} \frac{(C-1)!}{\rho^{C-1}} \sum_{j=0}^{C-1} \frac{\rho^j}{j!}, \tag{30}$$

$$\begin{aligned}
\mathbb{E}\mathcal{D}_C^2 &= \frac{(C-1)!}{\rho^{C-1}} \frac{2}{\lambda^2} + 2 \frac{(C-1)!}{\rho^C} \sum_{j=1}^{C-1} \frac{\rho^j}{(j-1)!} \frac{1}{v_j} (\mathbb{E}\mathcal{D}_{j+1} + \mathbb{E}\mathcal{D}_j) \\
&\quad + 2 \frac{(C-1)!}{\rho^C} \sum_{j=1}^{C-1} \frac{\rho^j}{(j-1)!} (\mathbb{E}\mathcal{D}_{j+1}\mathbb{E}\mathcal{D}_j) + \frac{2}{\lambda} \frac{(C-1)!}{\rho^{C-1}} \sum_{j=1}^{C-1} \frac{\rho^j}{j!} \frac{1}{v_j}.
\end{aligned} \tag{31}$$

6.3.2. Moments of the number of arrivals during a C-intercongestion period

$$\begin{aligned}
\mathbb{E}\mathcal{N}_C &= \frac{(C-1)!}{\rho^C} \sum_{j=1}^{C-1} \frac{\rho^j}{(j-1)!} + 1, \\
\mathbb{E}\mathcal{N}_C^2 &= \frac{(C-1)!}{\rho^{C-1}} + 2 \frac{(C-1)!}{\rho^C} \sum_{j=1}^{C-1} \frac{\rho^j}{(j-1)!} \mathbb{E}\mathcal{N}_{j+1}\mathbb{E}\mathcal{N}_j + \frac{(C-1)!}{\rho^{C-1}} \sum_{j=1}^{C-1} \frac{\rho^j}{j!} \frac{1}{v_j}.
\end{aligned}$$

6.3.3. Moments of the area swept under C during a C-intercongestion period

$$\begin{aligned}
\mathbb{E}\mathcal{A}_C &= \frac{1}{\lambda} \frac{(C-1)!}{\rho^{C-1}} + \frac{(C-1)!}{\rho^{C-1}} \sum_{j=1}^{C-1} \frac{\rho^j}{(j-1)!} \mathbb{E}\mathcal{D}_j + \frac{1}{\lambda} \frac{(C-1)!}{\rho^{C-1}} \sum_{j=1}^{C-1} \frac{\rho^j}{j!}, \\
\mathbb{E}\mathcal{A}_C^2 &= \frac{C-1}{\rho^{C-1}} \frac{2}{\lambda^2} + \frac{(C-1)!}{\rho^C} \sum_{j=1}^{C-1} \frac{\rho^j}{(j-1)!} (\mathbb{E}\mathcal{D}_j^2 + 2\mathbb{E}[\mathcal{D}_{j-1}\mathcal{A}_{j-1}] + 2\mathbb{E}\mathcal{A}_{j-1}\mathbb{E}\mathcal{A}_j + 2\mathbb{E}\mathcal{D}_{j-1}\mathbb{E}\mathcal{A}_j) \\
&\quad + \frac{(C-1)!}{\rho^C} \sum_{j=1}^{C-1} \frac{\rho^j}{(j-1)!} \frac{1}{v_j} (\mathbb{E}\mathcal{A}_{j-1} + \mathbb{E}\mathcal{D}_{j-1} + \mathbb{E}\mathcal{A}_j) + \frac{2}{\lambda} \frac{(C-1)!}{\rho^{C-1}} \sum_{j=1}^{C-1} \frac{\rho^j}{j!}.
\end{aligned}$$

6.3.4. Joint expectations of a C-intercongestion period

The joint expectations can be obtained in a similar fashion; they can be found in Section 6.4 of [10].

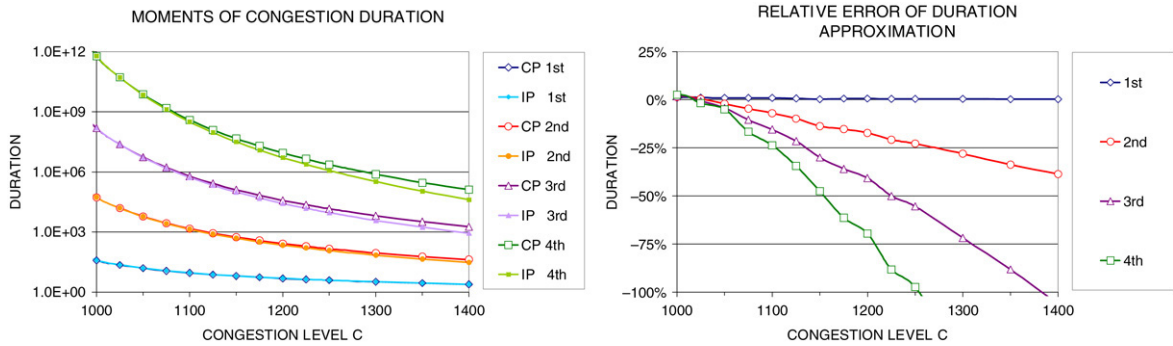


Fig. 1. Approximation of the simulated C-congestion period (CP) duration by an analytically derived $(2\rho - C)$ -intercongestion period (IP). Left: duration moments. Right: Relative error between simulated C-congestion period and derived $(2\rho - C)$ -intercongestion period.

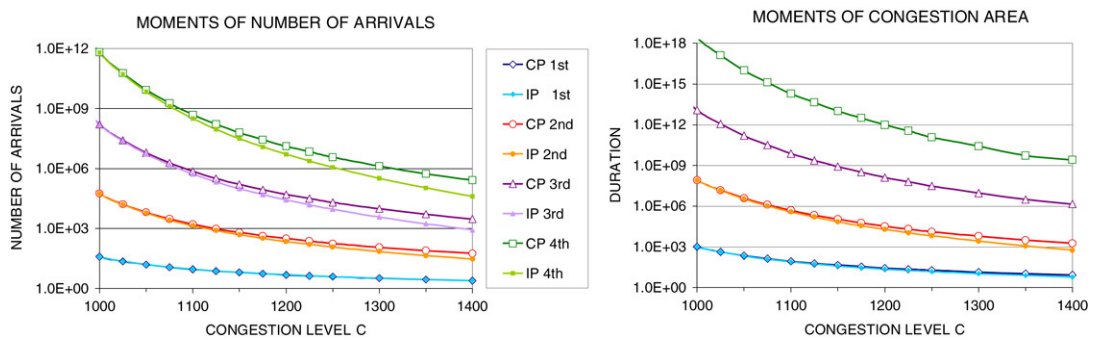


Fig. 2. Approximation of a simulated C-congestion period (CP) by an analytically derived $(2\rho - C)$ -intercongestion period (IP). Left: Number of arrivals. Right: Area swept above C.

7. Intercongestion period as an approximation of a congestion period

From a numerical perspective a drawback of the congestion period recursions is that the starting condition corresponds to a busy period; for high loads the system will hardly ever be empty, and hence the busy-period quantities will tend to grow large, thus resulting in numerical instability. The intercongestion period recursions do not have this problem; as remarked before, all moments of the quantities of level 0 are 0, and consequently the recursions are numerically stable.

Congestion and intercongestion periods are similar in the sense that a C-congestion period is the duration that the system is *above* level C and a C-intercongestion period is the duration that the system is *below* level C. For C close to ρ , the birth rate (λ) and the death rate ($\rho\mu$) are (almost) identical, hence $P_\rho \approx 1 - P_\rho$ in distribution. As $T_n \approx T_{\rho-(n-\rho)}$ in distribution for n close to ρ , it follows that a ρ -congestion and a ρ -intercongestion period exhibit similar stochastic behavior; a ρ -congestion period can be approximated by a ρ -intercongestion period. More generally, C-congestion periods can be approximated by $(\rho - (C - \rho))$ -intercongestion periods by the observation $P_n \approx 1 - P_{\rho-(n-\rho)}$ for n close to ρ . In particular, this approximation is expected to work well for C close to the average load ρ .

Figs. 1 and 2 present numerical results of the proposed approximation for arrival rate $\lambda = 1$ and mean service time $\mu^{-1} = 1000$, so the average load ρ is 1000. The moments of the congestion period quantities are obtained by simulations; the recursions are numerically unstable as the busy periods are very large due to the high average load. The intercongestion period quantities are obtained analytically by the recursive relations presented in Section 6. The left graph of Fig. 1 shows the first four moments of C-congestion period duration and an approximation of the duration; the approximation is the duration of a $(2\rho - C)$ -intercongestion period. The right graph presents the relative error of the approximation. We see that for values of C in the neighborhood of ρ the approximation is close to the simulated results. Especially for ‘lower’ moments the approximation is accurate; as could be expected, for higher moments the relative error becomes larger. In the range of $C = (\rho, \dots, 1100)$ the error of the second moment is less than 7%; here it is important to notice that the system will hardly ever have more than 1100 customers (probability is in the order of

0.001). Fig. 2 presents the results for the number of arrivals and the area for the same scenario. The results for these quantities are also accurate, so the intercongestion period seems to be a very good approximation for the congestion period, in particular for C close to average load ρ .

Another approximation was proposed earlier by Guillemin and Simonian [3]. They argue that a C -congestion period converges (after a specific scaling) to an $M/M/1$ busy period for large C . They propose to use the death-rate $C\mu$ of the $M/M/\infty$ queue as the death-rate for the $M/M/1$ queue, which results in an accurate approximation for C large compared to ρ . For C close to ρ the approximation is not so good; the behavior of the $M/M/\infty$ -congestion period differs significantly from the $M/M/1$ -busy period. However, as concluded earlier, the approximation of a congestion period by an intercongestion period is very accurate for C close to the average load ρ . We remark that the regime in which C is close to ρ is from a practical point of view perhaps the most relevant regime: networks are usually dimensioned such that C is exceeded only a small fraction of time. Hence, our main conclusion is that our approximation (for C close to ρ) nicely complements the one proposed by Guillemin and Simonian.

8. Concluding remarks

This paper studied the quantities duration, number of arrivals, and area for C -congestion periods of an $M/M/\infty$ -queue. We presented a derivation using recursive relations thus obtaining all moments and ‘joint expectations’ of the above quantities. The starting conditions of the recursions correspond to the busy period (a 0-congestion period); it is noted that the derivation of the higher moments and the joint expectations of these busy-period quantities were far from trivial, and followed from tedious calculations.

Furthermore, this paper introduced C -intercongestion periods, which are the intervals in which the system is *below* level C . Analogously to C -congestion periods, recursive relations are presented for the moments and joint expectations of the quantities. These are also solved in terms of a starting condition, but in contrast with C -congestion periods, the starting conditions of C -intercongestion period quantities are easily obtained: all moments and joint expectations of 0-intercongestion period quantities are 0. For the C -intercongestion period we also derived the Laplace transforms of the duration and the so-called intercongestion triple.

Finally, it was shown that an intercongestion period can be used in an approximation of a congestion period, in particular for C close to the average load ρ . This approximation is especially useful as the calculations of the intercongestion period are numerically more stable than those of the congestion periods. The proposed approximation complements other approximations proposed in literature, as these tend to be less accurate for C close to the average load ρ .

Acknowledgements

This work was (partly) carried out within the EQUANET project, an ‘ICT-doorbraakproject’ which is supported by the Dutch Ministry of Economic Affairs via its agency Senter.

References

- [1] J.L. van den Berg, M.R.H. Mandjes, R. van de Meent, A. Pras, F. Roijers, P.H.A. Venemans, QoS aware bandwidth provisioning for IP network links, *Computer Networks* 50 (5) (2006) 631–647.
- [2] F. Guillemin, D. Pinchon, On a random variable associated with excursions in an $M/M/\infty$ system, *Queueing Systems* 32 (1999) 305–318.
- [3] F. Guillemin, A. Simonian, Transient characteristics of an $M/M/\infty$ system, *Advances in Applied Probability* 27 (1995) 862–888.
- [4] J. Keilson, *Markov Chain Models — Rarity and Exponentiality*, Springer-Verlag, 1979.
- [5] C. Knessl, Y.P. Yang, Asymptotic expansions for the congestion period for the $M/M/\infty$ -queue, *Queueing Systems* 39 (2001) 213–256.
- [6] W. Liu, D.-H. Shi, Busy period in $G1^X/G/\infty$, *Journal of Applied Probability* 33 (1996) 815–829.
- [7] J. Preater, $M/M/\infty$ transience revisited, *Journal of Applied Probability* 34 (1997) 1061–1067.
- [8] J. Preater, On the severity of $M/M/\infty$ congested periods, *Journal of Applied Probability* 39 (2002) 228–230.
- [9] P. Robert, *Stochastic Networks and Queues*, Springer, 2003.
- [10] F. Roijers, F. Mandjes, J.L. van den Berg, Analysis of congestion periods of an $M/M/\infty$ -queue. CWI-Report. Available at: <http://ftp.cwi.nl/CWIreports/PNA/PNA-E0606.pdf>, 2006.
- [11] D.N. Shanbhag, On infinite server queues with batch arrivals, *Journal of Applied Probability* 3 (1966) 274–279.
- [12] W. Stadje, Busy period of the queueing system $M/G/\infty$, *Journal of Applied Probability* 22 (1985) 697–704.
- [13] L. Takács, *Introduction to the Theory of Queues*, Oxford University Press, New York, 1962.



Frank Roijers received his M.Sc. degree in Business Mathematics and Informatics from the Vrije Universiteit Amsterdam (VU) in 1998. In 2000 he joined TNO Information and Communication Technology (TNO ICT), former KPN Research. His research interests include performance analysis and traffic engineering of communication networks (IP, Wireless LAN, ad-hoc networks) and queueing theory. Since 2004 he has been working towards a Ph.D. in cooperation with the Centre for Mathematics and Computer Science (CWI) and from 2006 he has been part-time affiliated with the University of Amsterdam, Amsterdam.



Michel Mandjes received M.Sc. (in both mathematics and econometrics) and Ph.D. degrees from the Vrije Universiteit (VU), Amsterdam, the Netherlands. After having worked as a member of the technical staff at KPN Research (Leidschendam, the Netherlands) and Bell Laboratories/Lucent Technologies (Murray Hill NJ, USA), as a part-time full professor at the University of Twente, and as department head at the Centre for Mathematics and Computer Science (CWI), Amsterdam, he currently holds a full professorship at the University of Amsterdam, the Netherlands. His research interests include performance analysis of communication networks, queueing theory, Gaussian traffic models, traffic management and control, and pricing in multi-service networks.



Hans van den Berg received the M.Sc. and Ph.D. degree in mathematics (stochastic operations research) from the University of Utrecht, the Netherlands, in 1986 and 1990 respectively. From 1986, he worked at the Centre for Mathematics and Computer Science (CWI), Amsterdam. In 1990, he joined KPN Research (now TNO Information and Communication Technology, since January 2003). He is particularly working on performance analysis, traffic management and QoS provisioning for wired and wireless multi-service communication networks (ATM, IP, UMTS, WLAN, ad-hoc networks). Currently, he is a Senior Research Member and leader of the QoS group within TNO ICT, Delft, the Netherlands. Since July 2003, Hans van den Berg has had a part-time position as full professor within the faculty of Electrical Engineering, Mathematics and Computer Science at the University of Twente.