The asymptotic behavior of queueing systems: Large deviations theory and dominant pole approximation

Piet Van Mieghem

Alcatel-Telecom Research Division, Francis Wellesplein 1, B-2018 Antwerp, Belgium

Received 31 August 1995; revised 18 July 1996

This paper presents the exact asymptotics of the steady state behavior of a broad class of single-node queueing systems. First we show that the asymptotic probability functions derived using large deviations theory are consistent (in a certain sense) with the result using dominant pole approximations. Then we present an exact asymptotic formula for the cumulative probability function of the queue occupancy and relate it to the "cell loss ratio", an important performance measure for service systems such as ATM networks. The analysis relies on a new generalization of the Taylor coefficients of a complex function which we call "characteristic coefficients". Finally we apply our framework to obtain new results for the M/D/1 system and for a more intricate multiclass M/D/n system.

Keywords: Queueing systems, asymptotics, dominant pole approximation, large deviations theory, ATM.

1. Introduction

The asynchronous transfer mode (ATM), the current driving force in telecommunication, has dramatically increased the importance of the asymptotic behavior of queues. In ATM, services are offered with a certain guaranteed quality of service (QOS). An important QOS performance measure in ATM is the *cell loss ratio* (besides, e.g., the mean cell delay, the cell delay variation and the end-to-end delay [14]), where a "cell" denotes the unit of flow similar to a "job" or "customer" in queueing theory. The cell loss ratio, which will be precisely defined in section 7, is closely related to the buffer overflow probability which is the steady state probability that the limited buffer is fully occupied. Before allowing a new connection to send ATM cells into the network, the connection admission control (CAC) management checks whether the QOS of all existing connections and the new one can be guaranteed. To provide ATM services with such stringent QOS as a cell loss ratio on the order of 10^{-10} , the CAC of the network must know the asymptotic behavior of the queue lengths in the network. This article focuses on two methods, *the large deviations theory* and *the dominant pole approximation*, to explore the asymptotic behavior of queues employing as a mathematical device the concept of *characteristic coefficients* introduced in the next section.

Large deviations theory (see, e.g., Dembo and Zeitouni [8] and Weiss [27] and references therein) is the natural vehicle to study the asymptotic behavior of stochastic phenomena. Large deviations theory furnishes the theoretical justification for the important concept of equivalent bandwidth (see, e.g., [15, 11, 10, 13, 16] and the recent journal JSAC, Vol. 13, No. 6, 1995). Equivalent bandwidth is the effective bandwidth needed for a connection to obey its QOS requirements and is widely used to formulate practical CAC-rules in ATM. However, most of the large deviations results assume the limit of some quantity (e.g., the buffer size K) at infinity. In practice, therefore, it is highly desirable to know when this asymptotic regime applies or what the error is if large deviations theory is applied for finite values of the quantity under consideration.

The dominant pole approximation is derived from the partial fraction expansion of the probability generating function (pgf) G(s) of a random variable of interest \mathcal{G} and is broadly used in discrete-time analyses [5]. The relation between both methods will be discussed in section 3. Section 4 presents an asymptotic expansion for the overflow probability and questiones the use of the well-known rate-function I(s) when $\log G(s)$ is highly asymmetrical around its minimum. The remainder of the article is devoted to the dominant pole approximation. A formal expansion for the dominant pole in queues with a particular form for the pgf is proposed and applied to the M/D/1 and the convolved M/D/1 systems. The last section uses these results in computations of the cell loss ratio in ATM networks.

2. Characteristic coefficients of a complex function

The problem of finding the Taylor coefficients of G(f(z)), where G(z) is analytic, in terms of the Taylor coefficients a_k of f(z) has inspired the introduction of characteristic coefficients. Although Taylor coefficients belong to the basics of the theory of complex functions, we only found expressions for the first few Taylor coefficients of G(f(z)) in terms of the Taylor coefficients of f(z) listed in Abramowitz and Stegun [1, section 3.6]. At first glance, the lack of regularity in the patterns suggests that only very little can be gained that may lead to new results. However, while concentrating on this problem, we noticed that, under certain constraints, Taylor series of simple functions of a function f(z) could be elegantly written in terms of re-appearing quantities, that we called, therefore, *characteristic* coefficients. To our knowledge, this concept of characteristic coefficients has not been used before.

In this section, we briefly introduce characteristic coefficients and refer to the appendix for used expansions in this article. However, we must refer to [20] for the complete mathematical framework because this article merely illustrates its applications in queueing theory.

Let f(z) be an analytic function of z in a disk around z_0 with power series $f(z) = \sum_{k=0}^{\infty} a_k(z_0) (z - z_0)^k$. Our general result, from which most of the subsequent formulae are derived, is

$$G(f(z)) = G(a_0(z_0)) + \sum_{m=1}^{\infty} \left(\sum_{k=1}^{m} \frac{1}{k!} \frac{d^k G(p)}{dp^k} \Big|_{p=a_0(z_0)} s[k,m] \right) (z-z_0)^m, \quad (1)$$

where the characteristic coefficients s[k,m] are defined as

$$s[k,m] = \sum_{\sum_{i=1}^{k} j_i = m} \prod_{i=1}^{k} a_{j_i}(z_0) \quad (j_i > 0),$$
(2)

where the sum being over all possible ways one can write m as a sum of k integers consists of $\binom{m-1}{k-1}$ terms. The characteristic coefficients satisfy the recursive equation

$$s[1,m] = a_m(z_0),$$

$$s[k,m] = \sum_{j=k}^m s[1,m-j+1] \ s[k-1,j-1] \quad (k>1)$$

$$= \sum_{j=1}^{m-k+1} s[1,j] \ s[k-1,m-j] \quad (k>1).$$
(3)

In particular, this recursive set enables to compute the presented expansion formally up to an arbitrary order provided the coefficients $a_k(z_0)$ are known. As the technique of characteristic coefficients is formal, aspects of convergence are not investigated.

3. The dominant pole approximation and large deviations

In this section, we relate asymptotic results from the theory of generating functions [5] to those of the theory of large deviations. In the next section, we present our asymptotic expansion in continuous-time and compare it with established large deviations results.

Concentrating on the buffer overflow probability in discrete-time, we will demonstrate the following proposition.

Proposition 1. The buffer overflow probability derived from large deviations theory is asymptotically equivalent to its dominant pole approximation.

The first approach using the generating function G(z) of \mathcal{G} is an immediate consequence of Lemma 1.

Lemma 1. If G(z) is meromorphic with residues r_k at the (simple) poles p_k ordered as $0 < |p_0| \leq |p_1| \leq |p_2| \leq \cdots$ and if $G(z) = o(z^{N+1})$ as $z \to \infty$, then the following holds:

$$G(z) = \sum_{k=0}^{N} g(k) z^{k} + \sum_{k=0}^{\infty} \frac{r_{k}}{p_{k}^{N+1}(z-p_{k})} z^{N+1}$$
(4)

$$=\sum_{k=0}^{N} g(k)z^{k} + \sum_{k=0}^{\infty} r_{k} \left(\frac{1}{z - p_{k}} + \sum_{m=0}^{N} \frac{z^{m}}{p_{k}^{m+1}} \right).$$
(5)

The normalization condition G(1) = 1 implies that

$$\mathbf{P}[\mathcal{G} > k] = 1 - \sum_{k=0}^{N} g(k) = \sum_{k=0}^{\infty} \frac{r_k}{p_k^{N+1}(1-p_k)}.$$
(6)

The lemma follows from [26, section 3.21]. Rewriting (5) gives

$$G(z) = \sum_{k=0}^{N} g(k) z^{k} - \sum_{j=N+1}^{\infty} \left(\sum_{k=0}^{\infty} \frac{r_{k}}{p_{k}^{j+1}} \right) z^{j}$$
(7)

and hence,

$$g(j) = \sum_{k=0}^{\infty} \frac{r_k}{p_k^{j+1}} \quad (j > N).$$
(8)

The cumulative density function for j > N follows from (8) as

$$\sum_{k=j+1}^{\infty} g(k) = \sum_{k=0}^{\infty} \frac{r_k}{p_k^{j+1} (p_k - 1)} \quad (j > N).$$
(9)

This Lemma 1 means that, if the plot g(j) versus j exhibits a kink at j = N, then $G(z) = O(z^N)$ as $z \to \infty$. Alternatively, the asymptotic regime does not start earlier than $j \ge N$. In terms of the queue occupancy in ATM, the initial g(j)-regime for j < N reflects the cell scale, while the asymptotic regime $j \ge N$ refers to the burst scale. For large K, only the pole with smallest modulus, p_0 , will dominate. Hence,

$$\mathbf{P}[\mathcal{G} > K] \approx \frac{r_0}{p_0^{K+1} \left(p_0 - 1\right)}.$$
(10)

This approximation we call the *dominant pole* approximation.

The second approach is a large deviations approximation in discrete-time. We have

$$-\log \mathbf{P}[\mathcal{G} > K] = -\log \sum_{j=K+1}^{\infty} g(j)$$

$$\geq -\log \sum_{j=K+1}^{\infty} x^{j-K-1} g(j) \quad (x \in \mathbb{R} \text{ and } x \geq 1)$$

$$\geq -\log \left(x^{-K-1} \sum_{j=0}^{\infty} x^{j} g(j) \right)$$

$$= (K+1) \log x - \log G(x). \tag{11}$$

This inequality holds for all real $x \ge 1$. To get the tightest bound, we determine the maximizer x_{\max} of (11), thus $I(K) = \sup_{x\ge 1} [(K+1) \log x - \log G(x)]$. There exists such a supremum on account of the convexity of I(K) because G(x) and $\log G(x)$ are convex [3] for $x \ge 1$. Assuming that the maximum, say x_{\max} exists, then it is the solution of

$$x_{\max} = (K+1)\frac{G(x_{\max})}{G'(x_{\max})}$$

and the large deviations estimate becomes

$$P[\mathcal{G} > K] \leq e^{-[(K+1)\log x_{\max} - \log G(x_{\max}))]} = G(x_{\max}) x_{\max}^{-(K+1)}$$
(12)

Comparing (12) and (10), we observe that, for large K, we have that $x_{\text{max}} = p_0$ because

$$\lim_{K \to \infty} \frac{-\log \mathbb{P}[\mathcal{G} > K]}{K} = p_0 = x_{\max}$$

This illustrates Proposition 1.

4. An asymptotic expansion for $P[\mathcal{G} > K]$

Consider a continuous random variable \mathcal{G} with probability density function $g(x) = (d/dx)P[\mathcal{G} < x]$ and probability generating function G(s) where

$$G(s) = \int_0^\infty e^{sx} g(x) \,\mathrm{d}x. \tag{13}$$

We assume that all moments of \mathcal{G} exists or, equivalently, that G(s) converges for $0 < \operatorname{Re}(s) < c^*$. Since (13) is a Laplace transform, by inversion we have

$$g(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} G(s) e^{-sx} ds \quad (c^* > c > 0).$$
(14)

Formally for all $K \ge 0$ holds that

$$P[\mathcal{G} > K] = \int_{K}^{\infty} g(x) dx$$

$$= \frac{1}{2\pi i} \int_{K}^{\infty} \int_{c-i\infty}^{c+i\infty} G(s) e^{-sx} ds dx$$

$$= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} G(s) ds \int_{K}^{\infty} e^{-sx} dx$$

$$= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} G(s) \frac{e^{-Ks}}{s} ds$$

$$= \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{-J(s)} ds = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-J(c+it)} dt, \qquad (15)$$

where

$$J(s) = Ks - \log \frac{G(s)}{s}.$$
(16)

The order of integration can be reversed because the Laplace transform exists. Let s_{μ} be an extremum of J(s) with $J'(s_{\mu}) = 0$ and $\operatorname{Re}(J''(s_{\mu})) < 0$. In case s_{μ} is real, it is clearly a maximum. From large deviations theory, it is known that there exists a supremum. We further assume that the maximum at s_{μ} also exists. Hence, we can expand $J(s) = \sum_{k=0}^{\infty} a_k(s_{\mu}) (s - s_{\mu})^k$. The convergence of this series requires that $\lim_{k\to\infty} a_k(s_{\mu}) = 0$.

At this point, we invoke the concept of characteristic coefficients s[k, m] and use (80) for all $K \ge 0$,

$$P[\mathcal{G} > K] = \frac{e^{-J(s_{\mu})}}{2\sqrt{\pi |a_2(s_{\mu})|}} \left(1 + \sum_{m=1}^{\infty} \frac{R_m}{|a_2(s_{\mu})|^m}\right),\tag{17}$$

where

$$R_m = \sum_{k=1}^{2m} (-1)^k \frac{\Gamma(k+m+\frac{1}{2})}{\sqrt{\pi} \, k!} \, \frac{s^{**}[k,2m](s_\mu)}{a_2^k(s_\mu)}.$$

The first three terms of this expansion are explicitly computed in the appendix (81). Our formula gives the complete asymptotic series and is therefore more complete than the previous work of Daniels [7] on the saddle point method in statistics. Formula (17) exhibits the importance of the maximum s_{μ} that obeys

$$K - \frac{G'(s_{\mu})}{G(s_{\mu})} + \frac{1}{s_{\mu}} = 0$$
(18)

and shows that s_{μ} is a function of K. In particular, from

$$\frac{\partial s_{\mu}}{\partial K} = \frac{1}{\frac{1}{\frac{1}{s_{\mu}^2} + \frac{G_j''(s_{\mu})}{G_j(s_{\mu})} - \left(\frac{G_j'(s_{\mu})}{G_j(s_{\mu})}\right)^2}} = -\frac{1}{2a_2(s_{\mu})} > 0$$
(19)

it is seen that s_{μ} is strictly increasing in K.

We will immediately contrast our result with classical large deviations theory. Elwalid et al. [9] have estimated $P[\mathcal{G} \ge K]$ for $K \to \infty$ as

$$\mathbf{P}[\mathcal{G} \ge K] = \frac{\mathrm{e}^{-I(s^*)}}{s^* \,\sigma(s^*) \,\sqrt{2\pi}} \big[1 + \mathrm{o}(1) \big], \tag{20}$$

where I(s) is the rate function

$$I(s) = Ks - \log G(s) \tag{21}$$

in large deviations theory and where $I(s^*) = \sup_{s \ge 0} I(s)$ and $\sigma^2(s) = \partial^2 \log I(s)/\partial^2 s$ or, explicitly, $\sigma^2(s) = G''_j(s)/G_j(s) - (G'_j(s)/G_j(s))^2$. Ignoring the limit assumption $K \to \infty$, the formula of Elwalid et al. (20) is very close to the first term in our exact asymptotic expansion (17) and basically differs from ours in that they use, to determine the maximum, I(s) instead of J(s) and thus obtain s^* instead of s_{μ} . Hence, large deviations theory neglects the small difference of $\log s$ between J(s) and I(s).

We now consider the situation in the limit when $K \to \infty$. Since s_{μ} is strictly increasing in K, we observe from (18) that for sufficiently large K, $s_{\mu} \to s^*$ and simultaneously that $J(s_{\mu}) \to I(s^*)$. Furthermore, from the convexity of $\log G(s)$ for all s > 0 and $\lim_{K\to\infty} s_{\mu} = \infty$, it follows that for $K \to \infty$ that $|a_k(s_{\mu})| \to \infty$ for k < 3. Thus, even if all $a_k(s_{\mu})$ grow to infinity at nearly the same rate (which is certainly not the case for sufficiently large k on the convergence of the power series of J(s) around s_{μ} as $\lim_{k\to\infty} a_k(s_{\mu}) = 0$ for all s_{μ}), the *m*-sum in (17) vanishes in the limit $K \to \infty$, because in each *m*-term $a_2(s_{\mu})$ appears in the numerator to a higher power than whatever other coefficient in the denominator (see, e.g., (81) although this can be proved in general using properties of s[k, m]). Hence, our formula (17) tends to the large deviations result (20) in the limit $K \to \infty$.

However, from a practical point of view, it is of interest to know how fast the asymptotic 'large deviations regime' is attained. Unfortunately, all derivatives

$$a_{k}(s_{\mu}) = -\left[\frac{1}{k!} \frac{\mathrm{d}^{k-1}}{\mathrm{d}s^{k-1}} \left(\frac{G'(s)}{G(s)}\right) - \frac{(-1)^{k}}{k \, s^{k}}\right]\Big|_{s=s_{\mu}}$$

for $k \ge 2$ are function of K via s_{μ} . From (81) (with x = 1), it can be seen that rapid convergence requires that $a_2(s_{\mu}) \gg a_k(s_{\mu})$ for k > 2. If all derivatives are about the same order of magnitude, i.e., if the plot $\log G(s)$ versus s is highly asymmetrical around s_{μ} even though $\log G(s)$ is convex for all s > 0, convergence is expected to be slow and the asymptotic expansion is of limited practical use, and therefore, so is the large deviations result (12) or (20).

In conclusion, we have presented an exact asymptotic series for the overflow probability from which limitations of the large deviations result (12) are deduced. Specifically, we propose to use J(s) defined in (16) instead of the rate function I(s) in (21). Furthermore, we suggest to calculate some terms of our expansion (17). In case these terms (of the *m*-summation in (17)) are small compared to unity, the large deviations result (20) is accurate. Otherwise, as follows from the general theory of asymptotic series [12], the last omitted term in (17) is approximately a measure for the accuracy of the truncated series.

5. The exact dominant pole ζ

For many queueing systems, the probability generating function (pgf) Q(z) of a queue-related random variable Q (such as the number of cells or the waiting time of cells in the queue) is of the Pollaczek–Khintchine form

$$Q(z) = \frac{z - 1}{z - U(z)} F(z) \equiv \sum_{k=0}^{\infty} q(k) z^k,$$
(22)

where $U(z) = \sum_{k=0}^{\infty} u(k) z^k$ is a pgf of the random variable \mathcal{U} related to the input process (e.g., the number of arrivals per unit time) while F(z) is an analytic function in the unit disk $|z| \leq 1$ with F(1) = 1 - U'(1). The probability density function (pdf) of the random variable \mathcal{Q} is denoted as $q(k) = P[\mathcal{Q} = k]$. For example, in a GI/D/1 system, the steady state buffer occupancy pgf equals

$$Q_{\rm GI/D/1}(z) = \left(1 - U'(1)\right) \frac{z - 1}{z - U(z)},\tag{23}$$

where U(z) is the pgf of the general independent arrival process. This section is limited to the characteristic coefficient approach of (22)-forms.

The asymptotic behavior of (22)-forms depends on the dominant zero ζ of z - U(z) exceeding unity as $q(K) = O(1/\zeta^K)$ for $K \to \infty$ as explained in section 3. It is well known [5] that this zero ζ is real and larger than one. Moreover, it has the smallest modulus of all the zeros of z - U(z) apart from the trivial zero at z = 1. In addition, ζ also equals the radius of convergence for the power series in (22) and it is a single zero if all moments of Q exist [4, appendix].

Theorem 1. Let the discrete-time pgfs Q(z) and U(z) of the random variables Q and U be related as Q(z) = ((z-1)/(z-U(z)))F(z) where F(z) is analytic at least inside (and on) the disk $|z| \leq \zeta$. Then, the probability density function of Q for large K equals

$$q(K) \approx \frac{(\zeta - 1)F(\zeta)}{(U'(\zeta) - 1)\zeta^{K+1}},$$
(24)

where

$$\zeta = 1 + \frac{1 - \lambda}{u_2(1)} + \sum_{n=2}^{\infty} g_n (1 - \lambda)^n$$
(25)

with

$$g_n = \frac{(-1)^n}{u_2^n(1)} \left[\sum_{k=1}^{n-1} \frac{\binom{n+k-1}{k-1}}{k (u_2(1))^k} s^*[k, n-1] \right]$$
(26)

and $E[\mathcal{U}] = \lambda = u_1(1)$, $u_k(1) = U^{(k)}(1)/k!$ and $s^*[k, m]$ are the characteristic coefficients defined in appendix A.

Proof. Formula (24) follows from Lemma 1.

Since z = 1 is a trivial zero of z - U(z) in (22), but not a pole of Q(z), we will determine the zero of (z - U(z))/(z - 1). The series expansion of U(z) around z = 1,

$$U(z) = \sum_{k=0}^{\infty} u_k(1)(z-1)^k$$

exists with $u_0(1) = 1$ and is related to the Taylor series around z = 0 by

$$u_k(1) = \frac{U^{(k)}(1)}{k!} = \sum_{j=k}^{\infty} {j \choose k} u_j(0).$$
(27)

These derivatives $u_k(1)$ are real and positive since $u_j(0) = u(j)$ are probabilities. Hence, we have

$$\frac{z - U(z)}{z - 1} = \frac{1}{z - 1} \left(z - 1 - \sum_{k=1}^{\infty} u_k(1)(z - 1)^k \right)$$
$$= 1 - \sum_{k=1}^{\infty} u_k(1)(z - 1)^{k-1}$$
$$= 1 - u_1(1) - \sum_{k=1}^{\infty} u_{k+1}(1)(z - 1)^k.$$
 (28)

The closest zero ζ to z = 1 follows from (77) for $z_0 = 1$ and

$$a_0(1) = 1 - u_1(1) = 1 - \lambda, \tag{29}$$

$$a_k(1) = -u_{k+1}(1). (30)$$

This proves relation (25).

Explicitly summing the first five terms in (25) gives

$$\begin{aligned} \zeta &\approx 1 + \frac{1-\lambda}{u_2(1)} - \frac{u_3(1)}{u_2(1)} \left(\frac{1-\lambda}{u_2(1)}\right)^2 + \left[2\left(\frac{u_3(1)}{u_2(1)}\right)^2 - \frac{u_4(1)}{u_2(1)}\right] \left(\frac{1-\lambda}{u_2(1)}\right)^3 \\ &+ \left[-5\left(\frac{u_3(1)}{u_2(1)}\right)^3 + 5\frac{u_4(1)u_3(1)}{u_2^2(1)} - \frac{u_5(1)}{u_2(1)}\right] \left(\frac{1-\lambda}{u_2(1)}\right)^4 \\ &+ \left[14\left(\frac{u_3(1)}{u_2(1)}\right)^4 - 21\frac{u_4(1)}{u_2(1)} \left(\frac{u_3(1)}{u_2(1)}\right)^2 + 3\left(\frac{u_4(1)}{u_2(1)}\right)^2 \\ &+ 6\frac{u_5(1)u_3(1)}{u_2(1)} - \frac{u_6(1)}{u_2(1)}\right] \left(\frac{1-\lambda}{u_2(1)}\right)^5. \end{aligned}$$
(31)

Expressions (25) and (31) illustrate that $\zeta \to 1$ in the heavy traffic limit $\lambda \to 1$. Also, when λ is close to unity, only a few terms in (25) are needed to achieve sufficiently high accuracy. For intermediate or small values of λ , the series (25) is very likely alternating in which case we advise Euler's method¹ of summation in order to accelerate convergence. In some cases, it may be instructive to use moments $M_n = E[\mathcal{U}^n] = \sum_{k=0}^{\infty} k^n u(k)$ instead of the derivatives $u_k(1)$ both connected via

$$U^{(n)}(1) = \sum_{m=1}^{n} S_n^{(m)} M_m,$$
(33)

$$M_n = \sum_{m=1}^n \mathcal{S}_n^{(m)} U^{(m)}(1), \tag{34}$$

where $S_n^{(m)}$ and $S_n^{(m)}$ are Stirling numbers of the first and second kind respectively [1]. How fast (in terms of K) q(K) approaches the asymptotic (24) mainly depends

on the distance of other poles of Q(z) to ζ . This is in general a difficult problem.

The interest of (25) for CAC in ATM networks lies in the fact that the overflow probability $P[Q > K] \approx F(\zeta)/(U'(\zeta) - 1)\zeta^{K+1}$ is merely needed for relatively high loads λ . In this case, only a few moments of the arriving traffic need to be known to estimate the overflow probability P[Q > K] or the cell loss ratio [24].

6. Application to the M/D/1 queue

6.1. The dominant pole ζ of the M/D/1 queue

The pgf of the buffer occupancy in a M/D/1 queue is given by $Q(z) = (1 - \lambda)(z-1)/(z - e^{\lambda(z-1)})$, and, hence, $U(z) = e^{\lambda(z-1)}$ and $F(z) = 1 - \lambda$ in (22). The expansion² of (z - U(z))/(z - 1) is readily found as

$$\frac{z - U(z)}{z - 1} = 1 - \lambda - \sum_{k=1}^{\infty} \frac{\lambda^{k+1}}{(k+1)!} (z - 1)^k$$
(36)

¹ The Euler summation method

$$\sum_{n=0}^{\infty} a_n x^{n+1} = \sum_{n=0}^{\infty} \left(\sum_{k=0}^n \binom{n}{k} q^{n-k} a_k \right) \left[\frac{x}{1+qx} \right]^{n+1}$$
(32)

is particularly useful for alternating series [22]. The influence of the complex number q in (32) has been discussed by Hardy [12].

² We demonstrate that it is necessary to consider the zero of $(z - e^{\lambda(z-1)})/(z-1)$ rather than that of $z - e^{\lambda(z-1)}$ because of the trivial zero appearing at z = 1. Alternatively, this implies that a Lagrange series expansion around the origin is inadequate, since it will converge to this trivial zero. Indeed, we have that $z - e^{\lambda(z-1)} = 0$ is equivalent to $ze^{-\lambda z} = e^{-\lambda}$. Let $f(z) = ze^{-\lambda z}$ and $w = e^{-\lambda}$, then

and, hence, $u_k(1) = \lambda^k/k!$ for k > 1. The coefficients (26), $g_n = (-1)^{n+1}(b_n/\lambda^{n+1})$, are calculated exactly in appendix B up to order 40. Since the series (25) for the M/D/1 queue is clearly alternating,

$$\zeta = 1 - \frac{1}{\lambda} \left(\sum_{n=1}^{\infty} (-1)^n b_n \left(\frac{1}{\lambda} - 1 \right)^n \right),\tag{37}$$

we invoke Euler's summability method (32) to obtain

$$\zeta = 1 + \frac{2}{\lambda} \left(\sum_{n=0}^{\infty} c_n (1-\lambda)^{n+1} \right).$$
(38)

The coefficients

$$c_n = \sum_{k=0}^n \binom{n}{k} (-1)^{1+k} b_{(1+k)}$$
(39)

are again computed in appendix B. The resulting dominant zero, accurate to order 11, is

$$\zeta = 1 + \frac{2}{\lambda} \bigg[(1-\lambda) + \frac{(1-\lambda)^2}{3} + \frac{2(1-\lambda)^3}{9} + \frac{22(1-\lambda)^4}{135} + \frac{52(1-\lambda)^5}{405} + \frac{20(1-\lambda)^6}{189} + \frac{3824(1-\lambda)^7}{42525} + \frac{1424(1-\lambda)^8}{18225} + \frac{15856(1-\lambda)^9}{229635} + \frac{11714672(1-\lambda)^{10}}{189448875} + \frac{44536288(1-\lambda)^{11}}{795685275} + o((1-\lambda)^{11}) \bigg].$$
(40)

The numerical data enclosed in appendix B clearly underline the power of the Euler summability: the slowly converging alternating series (37) is transformed into a much more rapidly converging series (38) with positive terms ($\lambda < 1$). The positive sign of all terms in the transformed series (40) indicates that the numerical result computed by truncating the series at a finite number of terms, always forms a lower bound for the real zero. In the heavy traffic limit $\lambda \rightarrow 1$, the dominant zero is approximately equal to $\zeta \approx 1 + 2(1 - \lambda)/\lambda$ and the resulting tail asymptotic for the buffer occupancy pdf

clearly we obtain $z = f^{-1}(w)$. Using Lagrange series, for the sought zero we find

$$z = e^{-\lambda} \sum_{n=0}^{\infty} \frac{(1+n)^{n-1}}{n!} \left(\lambda e^{-\lambda}\right)^n.$$
 (35)

However, since $e^{bz} = 1 + b \sum_{n=1}^{\infty} \frac{(b+an)^{n-1}}{n!} (z e^{-az})^n$ (see [19, II]), we readily observe that in (35) z = 1 for all $0 \le \lambda < 1$.

is $q(K) \sim e^{-K \log \zeta} \sim e^{-2K((1-\lambda)/\lambda)}$ corresponding to the classical result (see, e.g., [23]).

The interest of the summation (37) lies in the observation that the coefficients b_n are all close to 1.0 for n > 1. Hence

$$\zeta \approx 1 - \frac{1}{\lambda} \left(1 - \frac{1}{\lambda} \right) \left[2 + \sum_{n=1}^{\infty} \left(1 - \frac{1}{\lambda} \right)^n \right]$$
$$= 1 - \frac{1}{\lambda} \left(1 - \frac{1}{\lambda} \right) \left[2 + \frac{1}{1 - \left(1 - \frac{1}{\lambda} \right)} - 1 \right]$$
$$= \frac{1}{\lambda^2}.$$
 (41)

Expression (41) is a particularly simple approximation for the zero ζ of $e^{\lambda(z-1)} - z$. The numerical data show that $\zeta = 1/\lambda^2$ is within 1% for $\lambda > 0.84$. In some cases of practical interest (see section 7.2 below) the load λ lies in this regime. Moreover, $\zeta = 1/\lambda^2$ is always larger than the exact zero.

6.2. The convolved M/D/1 queue

In telecommunication, several connections to a same direction i are multiplexed onto one physical link towards i. Often, the N input links to a switching fabric share one large buffer, called a *shared buffer* (sha), which is served by N deterministic servers S_i , one for each direction i, as depicted in Fig. 1. But, since the ith server only processes cells of link i in FIFO order, the present shared buffer system is not work conservative [18] because it violates the basic property that no servers should be *idle as long as the buffer is not empty*. Especially when considering output queueing, we may assume that the aggregate flow on link i is Poissonean with parameter λ_i and independent of flow j because the aggregate usually exist of many, none-of-them dominant single connections towards a same direction. Then, the buffer occupation pdf of a infinitely long shared buffer is the N-fold convolution of the separate M/D/1 queues corresponding to a direction i.

This kind of queueing system has not received much attention in the literature. Moreover, the dominant pole approximation really offers a computational advantage as explained below. Finally, the buffer pdf of a convolved M/D/1 system is an example of a pdf with a non-geometric tail.

The exact expression for the buffer occupation pdf $q_N[m]$ with pgf

$$Q_{\rm sha}(z) = \prod_{i=1}^{N} \frac{(1-\lambda_i) (z-1)}{z - e^{\lambda_i (z-1)}}$$
(42)



Fig. 1. A sketch of a shared buffer. The dashed open boxes represent the 'logical' output queues for the different N links. In practice, cells belonging to direction i are stored in positions known to the corresponding server S_i and served in FIFO order to obey the sequence integrity as required in ATM.

is in general difficult to compute. Fortunately, in the special case where all $\lambda_i = \lambda$ are the same, an exact formula for $q_N[m]$ can be derived as follows:

$$Q_{\rm sha}(z) = \frac{(1-\lambda)^N (1-z)^N}{(e^{\lambda(z-1)} - z)^N} \equiv \sum_{k=0}^{\infty} q_N[k] z^k$$

$$= \frac{(1-\lambda)^N (1-z)^N e^{\lambda N(1-z)}}{(1-z)^N e^{\lambda(1-z)} N}.$$
(43)

This symmetrical traffic case is especially important for CAC in ATM [21]. Introducing the expansion $(1-x)^{-m-1} = \sum_{k=m}^{\infty} {k \choose m} x^{k-m}$ for |z| < 1 [1] gives

$$\begin{aligned} Q_{\rm sha}(z) &= (1-\lambda)^N \ (1-z)^N \ e^{\lambda N(1-z)} \sum_{k=N-1}^\infty \binom{k}{N-1} z^{k-N+1} \ e^{(k-N+1)\lambda(1-z)} \\ &= (1-\lambda)^N \ (1-z)^N \sum_{k=0}^\infty \binom{k+N-1}{N-1} z^k \ e^{(k+N)\lambda} \ e^{-(k+N)\lambda z} \\ &= (1-\lambda)^N \ (1-z)^N \sum_{k=0}^\infty \binom{k+N-1}{N-1} z^k \ e^{(k+N)\lambda} \ \sum_{n=0}^\infty \frac{(-(k+N)\lambda)^n}{n!} z^n \\ &= (1-\lambda)^N \ \sum_{m=0}^N \binom{N}{m} (-1)^m \ z^m \ \sum_{s=0}^\infty \beta_s \ z^s \\ &= (1-\lambda)^N \ \sum_{l=0}^\infty \left[\sum_{m=0}^{\min(N,l)} \binom{N}{m} (-1)^m \ \beta_{l-m} \right] z^l, \end{aligned}$$



Fig. 2. The exact buffer pdf (45) and the exact dominant pole approximation (49) for several servers N and $\lambda = 0.8$. Observe the numerical instability of the exact buffer pdf (dotted line) given by (45) for m > 20.

where

$$\beta_s = \sum_{n=0}^{s} \binom{s-n+N-1}{N-1} e^{(s-n+N)\lambda} \frac{(-(s-n+N)\lambda)^n}{n!} .$$
(44)

Finally the buffer occupation pdf for this special case reads

$$q_{N}[0] = (1 - \lambda)^{N} e^{N\lambda},$$

$$q_{N}[l] = (1 - \lambda)^{N} \sum_{m=0}^{\min(N,l)} {\binom{N}{m}} (-1)^{m} \beta_{l-m}.$$
(45)

Although an exact result, formula (45) is highly unfavourable for numerical computations as the buffer position increases (see Fig. 2, dotted line). In case N = 1, a transform to a Lagrange series (see (69) below) is possible. However, this Lagrange series converges very slowly for values of λ close to unity.

We now present the dominant pole approximation. We have for the general case

$$q_N[m] = -\frac{\prod_{i=1}^N (1-\lambda_i)}{2\pi \mathbf{i}} \int_L \frac{(z-1)^N}{z^{m+1}} \frac{\mathrm{d}z}{\prod_{i=1}^N (z-\mathrm{e}^{\lambda_i(z-1)})}, \quad m > 0, \qquad (46)$$

where the contour encloses the whole complex plain except for the origin. Approximating this integral by the residue at the dominant poles ζ_i corresponding to λ_i yields

$$q_N[m] \approx -\prod_{i=1}^N (1-\lambda_i) \sum_{i=1}^N \frac{(\zeta_i - 1)^N}{\zeta_i^{m+1} (1-\zeta_i \lambda_i)} \prod_{j=1; j \neq i}^N \frac{1}{\zeta_i - e^{\lambda_j(\zeta_i - 1)}}.$$
 (47)

When all $\lambda_i = \lambda$, expression (46) reduces to

$$q_N[m] = -\frac{(1-\lambda)^N}{2\pi i} \int_L \frac{(z-1)^N}{z^{m+1}} \frac{dz}{(z-e^{\lambda(z-1)})^N}.$$
 (48)

However, the dominant pole approximation requires the computation of a Nthorder residue at the dominant pole ζ ,

$$q_{N}[m] \approx -\frac{(1-\lambda)^{N}}{(N-1)!} \lim_{z \to \zeta} \frac{\mathrm{d}^{N-1}}{\mathrm{d}z^{N-1}} \left[\frac{(z-1)^{N} (z-\zeta)^{N}}{z^{m+1} (z-e^{\lambda(z-1)})^{N}} \right]$$
$$\approx -\frac{(1-\lambda)^{N}}{(N-1)!} \sum_{k=0}^{N-1} \binom{N-1}{k} \frac{\mathrm{d}^{N-1-k}}{\mathrm{d}z^{N-1-k}} \left[\frac{(z-1)^{N}}{z^{m+1}} \right]_{z=\zeta} h_{k}, \qquad (49)$$

where we used the Leibniz rule with

$$h_j = \lim_{z \to \zeta} \frac{\mathrm{d}^j}{\mathrm{d}z^j} \left[\frac{(z - \zeta)^N}{(z - \mathrm{e}^{\lambda(z-1)})^N} \right].$$
(50)

Again invoking characteristic coefficients [20] where $s^*[k, m]$ is defined in appendix A, we can compute h_j exactly as

$$h_j = j! \sum_{k=1}^{j} (-1)^k \binom{n+k-1}{k} s^*[k,m] a_1^{-n-k}(\zeta),$$
(51)

where $a_k(\zeta) = -\zeta \lambda^k / k!$ (for k > 1) and $a_1(\zeta) = 1 - \lambda \zeta$ are the Tailor coefficients of $z - e^{\lambda(z-1)}$ around $z = \zeta$.

The comparison between the exact buffer occupancy pdf (45) and the dominant pole approximation (49) is plotted in Fig. 2, for $\lambda = 0.8$ and $N = 1, \ldots, 5$ and illustrates the limited use of the exact result (45). The dominant pole approximation is remarkably good except for the very small values of m around zero. For a single queue N = 1, the approximation

$$q(m) \approx \frac{1-\lambda}{\lambda\zeta - 1} \frac{\zeta - 1}{\zeta^{m+1}}$$

is surprisingly accurate for $m \ge 2$ and $\lambda \ge 0.5$. For $\lambda = 0.5$ we found that at least 3 significant digits are correct if $m \ge 2$. The reason for the success of the dominant pole



Fig. 3. The exact dominant pole approximation (49) for a large number of servers (and links) N and $\lambda = 0.8$.

approximation in a M/D/1 system is due to the relatively large separation between the dominant and its next neighbour pole together with decreasing residus.

Figure 3 illustrates the behavior of the shared buffer occupancy pdf as function of the buffer position m and the number of links N. Observe that the maximum of $q_N[m]$ is reached around the mean buffer occupancy in a shared buffer (sha) with infinite length $Q'_{sha}(1) = N\lambda^2/2(1-\lambda)$ which equals N times the mean buffer occupancy in the corresponding single buffer. Further the plot also explains the efficiency gain of a shared buffer over a single buffer. Indeed, it is much less probable to have an empty shared buffer in contrast to a single buffer that is empty most of the time.

7. The cell loss ratio

Due to its importance in ATM, the QOS performance measure, the cell loss ratio, deserves some attention. The cell loss ratio clr is defined as the ratio of the long-time average number of lost cells because of buffer overflow to the long-time average number of cells that arrive in steady state. There are typically two different views to describe the cell loss ratio: a conservation-based and a combinatorial one. The conservation law simply states that cells entering the system also must leave it. The average number of entering cells are all those offered per time slot minus the ones that have been rejected, thus $(1 - clr)\lambda$. On the other side, the average number

of cells that leave the system are related to the server activity as $(1 - q(0))\mu$, where μ is the service rate. Hence, we have

$$(1 - clr)\lambda = (1 - q(0))\mu.$$
(52)

In the combinatorial view, only the arrival process is viewed from a position in the buffer and the number of ways in which cells are lost are counted, leading to

$$clr = \frac{1}{U'(1)} \sum_{i=0}^{\infty} i \sum_{j=0}^{K} q(K-j)u(j+i)$$
(53)

with $U'(1) = \lambda$. Although equation (52) is simple, its practical use is limited since the quantities involved are to be known with extreme high accuracy if clr is of the order of 10^{-10} . Therefore, we confine ourselves to the combinatorial result and present an other form for (53) as

$$clr U'(1) = \frac{\mathrm{d}S(z)}{\mathrm{d}z}\Big|_{z=1},$$
(54)

where

$$S(z) = \sum_{i=0}^{\infty} z^{i} \sum_{j=0}^{K} q(K-j)u(j+i)$$

$$= \sum_{i=0}^{\infty} z^{i} \sum_{j=0}^{K} q(K-j)z^{-j}u(j+i)z^{j}$$

$$= \sum_{j=0}^{K} q(K-j)z^{-j} \left(\sum_{i=0}^{\infty} u(j+i)z^{j+i} \right)$$

$$= \sum_{j=0}^{K} q(K-j)z^{-j} \left(U(z) - \sum_{i=0}^{j-1} u(i)z^{i} \right)$$

$$= U(z) \ z^{-K} \sum_{j=0}^{K} q(K-j)z^{K-j} - \sum_{j=0}^{K} q(K-j)z^{-j} \sum_{i=0}^{j-1} u(i)z^{i}$$

$$= z^{-K}U(z)Q(z) - z^{-K} \sum_{j=0}^{K} q(j)z^{j} \sum_{i=0}^{K-j-1} u(i)z^{i},$$
(55)

where we have introduced the generating function for the buffer occupancy:

$$Q(z) = \sum_{i=0}^{K} q(i)z^{i},$$
(57)

since q(i) = 0 for i > K. In order to express the cell loss ratio entirely in terms of the generating functions U(z) and Q(z), we need the following result:

$$\sum_{i=0}^{n} y[i]z^{i} = \sum_{i=0}^{n} z^{i} \left(\frac{1}{2\pi i} \int_{C} \frac{Y(\omega)}{\omega^{i+1}} d\omega \right)$$
$$= \frac{1}{2\pi i} \int_{C} \frac{Y(\omega)}{\omega - z} \left[1 - \left(\frac{z}{\omega} \right)^{n+1} \right] d\omega$$
$$= Y(z) - \frac{1}{2\pi i} \int_{C} \frac{Y(\omega)}{\omega - z} \left(\frac{z}{\omega} \right)^{n+1} d\omega,$$
(58)

where C is a contour enclosing the origin AND the point z and lying within the convergence region of Y(z). Combining (56) and (58), we rewrite S(z) as

$$S(z) = z^{-K}U(z)Q(z) - z^{-K}Q(z)U(z) + \frac{1}{2\pi i} \int_{C} \frac{U(\omega)Q(\omega)}{(\omega - z)\omega^{K}} d\omega$$
$$= \frac{1}{2\pi i} \int_{C} \frac{U(\omega)Q(\omega)}{(\omega - z)\omega^{K}} d\omega.$$
(59)

Finally, our expression for the cell loss ratio in an GI/G/1/K system, reads

$$clr = \frac{1}{2\pi i U'(1)} \int_C \frac{U(\omega)Q(\omega)}{(\omega-1)^2 \omega^K} \,\mathrm{d}\omega,\tag{60}$$

where the contour encloses both the origin and the point z = 1 and lies in the convergence region of U(z). Usually, U(z) is known while Q(z) proves to be much more complicated to obtain. The product Q(z)U(z) is known as the pgf of the system content.

7.1. Properties and applications

Lemma 2. If Q(z) and U(z) are meromorphic functions and if

$$\lim_{z \to \infty} \left| \frac{U(z) \ Q(z)}{(z-1)^2 \ z^{K-1}} \right| = 0,$$

the contour C in (60) can be closed over $|\omega| > 1$ to get

$$clr = -\frac{1}{U'(1)} \sum_{p} \underset{\omega \to p}{\operatorname{Res}} \frac{U(\omega)Q(\omega)}{\omega^{K} (\omega - 1)^{2}},$$
(61)

where p are the poles of U(z)Q(z) outside the unit circle.

Hence, if the conditions of Lemma 2 are met, a non-trivial evaluation of the cell loss ratio can be obtained. As frequently done, one may approximate the buffer pgf of the finite system by that of the infinite system. In case the buffer pgf of the finite system is known, then Q(z) is a polynomial of degree at most K so that the only pole of $Q(z)/z^{K}$ is zero and $\lim_{z\to\infty} Q(z)/z^{K} = q(K) \leq 1$. In this case the requirements of Lemma 2 simplify to $\lim_{z\to\infty} |z U(z)/(z-1)^2| = 0$. Executing (60) then leads to

$$clr = -\frac{1}{U'(1)} \sum_{p} \frac{Q(p)}{p^{K} (p-1)^{2}} \mathop{\rm Res}_{\omega \to p} U(\omega),$$
 (62)

where only the poles p of the arrival process U(z) play a role. As a simple application of (62), if the number of arrivals $u(k) = (1 - \alpha)\alpha^k$ with $0 \le \alpha \le 1$ has a geometric distribution with generating function [2], $U_{\text{geo}}(z) = (1 - \alpha)/(1 - \alpha z)$, then the above requirements are met and we obtain from (62)

$$clr_{\text{geo}} = \alpha^K Q\left(\frac{1}{\alpha}\right).$$
 (63)

An important class excluded from (61) consists of integral functions U(z) that possess a Taylor series expansion converging for all complex variables z. The generating function of a Poisson process with parameter λ , $U_{\text{Poisson}}(z) = e^{\lambda (z-1)}$, is an important representative of that class. For a Poissonian arrival process, (60) reads

$$clr_{\text{Poisson}} = \frac{e^{-\lambda}}{2\pi i \lambda} \int_C \frac{e^{\lambda \omega} Q(\omega)}{(\omega - 1)^2 \, \omega^K} \, \mathrm{d}\omega.$$
(64)

Deforming the contour to enclose the negative half ω -plane (Re(ω) < c), we obtain

$$clr_{\text{Poisson}} = \frac{e^{-\lambda}}{2\pi i \lambda} \int_{c-i\infty}^{c+i\infty} \frac{e^{\lambda\omega}Q(\omega)}{(\omega-1)^2 \ \omega^K} \, \mathrm{d}\omega, \tag{65}$$

where the real number c exceeds unity. This expression is recognized as an inverse Laplace transform,

$$clr_{\text{Poisson}} = \frac{e^{-\lambda}}{\lambda} L_{\lambda}^{-1} \left[\frac{Q(\omega)}{(\omega-1)^2 \ \omega^K} \right].$$
(66)

Since the argument of L_{λ}^{-1} is a rational function, an exact evaluation of (66) is possible leading, however, again to (53). Hence, the combinatorial view does not offer much insight immediately suggesting to consider a conservation-based approach. Indeed, it is well known that, owing to the PASTA property [28, 29], an exact expression [4, 25, 6] in continuous-time for the cell loss ratio in a M/G/1/K system can be derived with as result

$$clr_{M/G/1/K;cont} = (1-\rho) \frac{P[Q > K-1]}{1-\rho P[Q > K-1]},$$
 (67)

where, as usual, the traffic intensity $\rho = \lambda/\mu$ and P[Q > K - 1] is the overflow probability in the corresponding infinite system, M/G/1. Transforming (67) to discrete-time using

$$clr_{\rm disc} = rac{clr_{\rm cont}}{
ho(1 - clr_{\rm cont})}$$

yields

$$clr_{M/G/1/K;disc} = \frac{1-\rho}{\rho} \frac{P[Q > K-1]}{1-P[Q > K-1]}.$$
 (68)

7.2. The cell loss ratio in a discrete-time M/D/1/K system

Although for a M/D/1 the exact expression of the overflow probability [25] is known as

$$\mathbf{P}[\mathcal{Q} > K - 1] = (1 - \lambda) \lambda^K \sum_{n=1}^{\infty} \frac{n^{n+K}}{(n+K)!} (\lambda \mathrm{e}^{-\lambda})^n, \tag{69}$$

this form is severely handicaped by the slow convergence for high traffic intensities $\rho = \lambda$ (since $\mu = 1$) so that fast executable expressions are desirable.

From (68) it follows that the cell loss ratio in a M/D/1/K system is reasonably well approximated for sufficiently large K as

$$clr(\lambda) \approx \frac{1-\lambda}{\lambda} \mathbf{P}[\mathcal{Q} > K-1].$$
 (70)

Invoking the exact dominant pole approximation

$$\mathbb{P}[\mathcal{Q} > K - 1] = \frac{1 - \lambda}{\lambda \zeta - 1} \zeta^{-K}$$

gives

$$clr(\lambda) \approx \frac{(1-\lambda)^2}{\lambda(\lambda z - 1)} \zeta^{-K},$$
(71)

where ζ is the non-trivial solution of $z = e^{\lambda(z-1)}$ with |z| > 1 as computed in section 6.1. For sufficiently high loads $\lambda > 0.8$ we use the approximation (41), $\zeta \approx \lambda^{-2}$, to obtain the new result

$$clr(\lambda) \approx (1-\lambda) \ \lambda^{2K}.$$
 (72)

As the estimate for the root is always larger than the real root, our approximate relation (72) is always a lower bound for (71).

Formula (72) is of a particularly remarkable form since it frequently appears in probability theory [17]. First of all $(1-\rho) \rho^k$ represents the geometric distribution with parameter ρ and can be regarded as the probability distribution of a stochastic variable \mathcal{Z} that counts the number of successes – the occurrence of each, independent success

has the probability ρ – prior to the first failure. In an M/M/1 system, $p_k = (1 - \rho) \rho^k$ is the steady state probability that this systems contains k members (or equivalent is in state k). And finally, the exact expression for the cell loss ratio in an M/M/1/K system [17] reads

$$clr_{M/M/1/K}(\rho) = \frac{(1-\rho) \ \rho^K}{1-\rho^{K+1}}.$$
 (73)

When using (68) instead of (70), we find the more precise result with $\lambda = \rho$

$$clr_{M/D/1/K}(\rho) = (1-\rho) \frac{\rho^{2K}}{1-\rho^{2K+1}}.$$
 (74)

Comparing both (73) and (74), we observe that, for sufficiently high ρ , the *M*-server (in continuous-time) needs *approximately twice* as much place to guarantee a same cell loss ratio as in the corresponding *D*-server system (in discrete-time). As well-known for continuous-time [17, pp. 191], the mean waiting time in a M/M/1 system is *exactly twice* of that in a M/D/1.

Further, our simple formula (72) is particularly useful to engineer buffers or to dimension simple queueing networks. The buffer length K in case of Poisson traffic given the cell loss ratio requirement clr and the traffic intensity $\lambda > 0.8$ is immediate from (72) as

$$K = \left[\frac{\log c l r - \log(1 - \lambda)}{2 \log \lambda}\right] + 1,$$
(75)

where [x] denotes the integral part of x. In addition, denoting $clr = f_K(\lambda)$, then the inverse function is $\lambda = f_K^{-1}(clr)$. For (72) we find the relatively simple $f_K^{-1}(w) = \lim_{n\to\infty} f_n(w)$ with $f_1(w) = w^{1/(2K)}$ and

$$f_n(w) = \frac{w^{1/(2K)}}{[1 - f_{n-1}(w)]^{1/(2K)}}.$$
(76)

This continued fraction (76) converges rapidly.

8. Summary

Characteristic coefficients have been employed to generate asymptotic expansions of the steady state behavior of queues in both the large deviations settings as in the generating function approach.

The dominant pole approximation, a natural asymptotic method arising from the generating function approach, has been applied to M/D/1-like systems. A series expansion for the dominant pole in a M/D/1 queue has been given explicitly and the probability density function of the N-fold convolved M/D/1 system is calculated. The important quality of service parameter in ATM, the cell loss ratio, has been studied and an elegant, approximate result for the M/D/1 queue is presented.

Acknowledgement

We are grateful to Jan Van Mieghem for his useful comments.

Appendix

A. Useful results

The zero $\zeta(z_0)$ of f(z) closest to z_0 is given in terms of the coefficients $a_k(z_0)$ of the power series of f(z) around z_0 as

$$\zeta(z_0) = f^{-1}(0) = z_0 - \frac{a_0(z_0)}{a_1(z_0)} + \sum_{n=2}^{\infty} \left[\sum_{k=1}^{n-1} \frac{(-1)^k \binom{n+k-1}{k-1}}{k (a_1(z_0))^k} s^*[k, n-1] \right] \left(-\frac{a_0(z_0)}{a_1(z_0)} \right)^n, \tag{77}$$

where $s^*[k, n] = s[k, n]|_{\forall m: a_m(z_0) \to a_{m+1}(z_0)}$. Formula (77) is another representation for Lagrange's expression [19, II, pp. 88]

$$f^{-1}(0) = z_0 + \sum_{n=1}^{\infty} \frac{1}{n!} \left[\frac{\mathrm{d}^{n-1}}{\mathrm{d}z^{n-1}} \left(\frac{z - z_0}{f(z) - f(z_0)} \right)^n \right] \Big|_{z=z_0} (-1)^n f^n(z_0).$$
(78)

For most functions, Lagrange's formula (78) rapidly becomes unfeasible to calculate, while our formula (77), due to its recursive character, can be formally computed to an arbitrary order provided the Taylor coefficients $a_k(z_0)$ are known. For example, the series (37) could only be computed with conventional execution of (78) and with current personal computers to order 11, while with characteristic coefficients (77), no such limitations were encountered. Therefore, we found it instructive to give the exact coefficients up to order 40 in section B.

Explicitly summing the first five terms ($n \leq 5$) in (77) yields

$$\begin{aligned} \zeta(z_0) &= z_0 - \frac{a_0(z_0)}{a_1(z_0)} - \frac{a_2(z_0)}{a_1(z_0)} \left(\frac{a_0(z_0)}{a_1(z_0)}\right)^2 + \left[-2\left(\frac{a_2(z_0)}{a_1(z_0)}\right)^2 + \frac{a_3(z_0)}{a_1(z_0)} \right] \left(\frac{a_0(z_0)}{a_1(z_0)}\right)^3 \\ &+ \left[-5\left(\frac{a_2(z_0)}{a_1(z_0)}\right)^3 + 5\frac{a_3(z_0)}{a_1(z_0)}\frac{a_2(z_0)}{a_1(z_0)} - \frac{a_4(z_0)}{a_1(z_0)} \right] \left(\frac{a_0(z_0)}{a_1(z_0)}\right)^4 \\ &+ \left[-14\left(\frac{a_2(z_0)}{a_1(z_0)}\right)^4 + 21\frac{a_3(z_0)}{a_1(z_0)}\left(\frac{a_2(z_0)}{a_1(z_0)}\right)^2 - 3\left(\frac{a_3(z_0)}{a_1(z_0)}\right)^2 \right] \\ &- 6\frac{a_4(z_0)}{a_1(z_0)}\frac{a_2(z_0)}{a_1(z_0)} + \frac{a_5(z_0)}{a_1(z_0)} \right] \left(\frac{a_0(z_0)}{a_1(z_0)}\right)^5 + O\left(\frac{a_0(z_0)}{a_1(z_0)}\right)^6. \end{aligned}$$
(79)

Let f(z) be an analytic function of z in a disk around the extremum z_{μ} with expansion

$$f(z) = \sum_{k=0}^{\infty} a_k(z_\mu) \left(z - z_\mu\right)^k,$$

where with $a_1(z_{\mu}) = 0$. The asymptotic expansion for

$$R = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{x f(s)} ds$$

reads

$$R = \frac{e^{x a_0(z_\mu)}}{2\sqrt{x\pi a_2(z_\mu)}} \times \left(1 + \sum_{m=1}^{\infty} \left[\sum_{k=1}^{2m} (-1)^k \frac{\Gamma(k+m+\frac{1}{2})}{\sqrt{\pi} k!} \frac{s^{**}[k,2m]}{a_2^k(z_\mu)}\right] \frac{(-1)^m}{(a_2(z_\mu) x)^m}\right), \quad (80)$$

where $s^{**}[k, n] = s[k, n]|_{\forall m: a_m(z_\mu) \to a_{m+2}(z_\mu)}$ and $\operatorname{Re}(x a_2(z_\mu)) < 0$. This expansion is only exact if f(z) is an integral function. In the other cases (as is most usual for asymptotic expansions [12]), the infinite series diverges. Hence, the practical applicability is limited to the first few terms in the *m*-sum that are decreasing in magnitude.

Although compact in form, when written explicitly, (80) rapidly becomes impressive. For only three terms in the *m*-summation, we find

$$R = \frac{e^{x a_0(z_{\mu})}}{2\sqrt{x \pi a_2(z_{\mu})}} \left\{ 1 - \left[\frac{15a_3^2(z_{\mu})}{16a_2^3(z_{\mu})} - \frac{3a_4(z_{\mu})}{4a_2^2(z_{\mu})} \right] \frac{1}{x} + \left[\frac{3465a_3^4(z_{\mu})}{512a_2^6(z_{\mu})} - \frac{945a_3^2(z_{\mu}) a_4(z_{\mu})}{64a_2^5(z_{\mu})} + \frac{105(a_4^2(z_{\mu}) + 2a_3(z_{\mu}) a_5(z_{\mu}))}{32a_2^4(z_{\mu})} - \frac{15a_6(z_{\mu})}{8a_2^3(z_{\mu})} \right] \frac{1}{x^2} - \left[\frac{765765a_3^6(z_{\mu})}{8192a_2^9(z_{\mu})} - \frac{675675a_3^4(z_{\mu}) a_4(z_{\mu})}{2048a_2^8(z_{\mu})} + \frac{45045a_3^2(z_{\mu})[3a_4^2(z_{\mu}) + 2a_3(z_{\mu})a_5(z_{\mu})]}{512a_2^7(z_{\mu})} - \frac{3465[a_4^3(z_{\mu}) + 6a_3(z_{\mu}) a_4(z_{\mu}) a_5(z_{\mu}) + 3a_3^2(z_{\mu}) a_6(z_{\mu})]}{128a_2^6(z_{\mu})} - \frac{945[a_2^2(z_{\mu}) + 2a_4(z_{\mu}) a_6(z_{\mu}) + 2a_3(z_{\mu}) a_7(z_{\mu})]}{64a_2^5(z_{\mu})} - \frac{105a_8(z_{\mu})}{16a_2^4(z_{\mu})} \right] \frac{1}{x^3} + o\left(\frac{1}{x^3}\right) \right\}.$$

)

B. Numerical coefficients of the dominant pole expansion for the M/D/1 queue

Since the expansion for ζ is an exact result, not obtainable with conventional techniques as explained in the previous appendix, we list the first fourty coefficients in (25) $g_n = (-1)^{n+1} \frac{b_n}{\lambda^{n+1}}$:

$$b_{1} = 2, \qquad b_{2}$$

$$b_{3} = \frac{10}{9} = 1.11111, \qquad b_{4}$$

$$b_{5} = \frac{386}{405} = 0.953086, \qquad b_{6}$$

$$b_{7} = \frac{38698}{42525} = 0.910006, \qquad b_{8}$$

$$b_{9} = \frac{1040686}{1148175} = 0.906383, \qquad b_{10}$$

$$b_{11} = \frac{732086318}{795685275} = 0.92007, \qquad b_{12}$$

$$b_{13} = \frac{438366017438}{465475885875} = 0.941759, \qquad b_{14}$$

$$b_{15} = \frac{321365923826}{332482775625} = 0.966564, \qquad b_{16}$$

$$b_{17} = \frac{12357217339126666}{12463116844303125} = 0.991503, \qquad b_{18} = \frac{25157310094776004}{25072858592656875} = 1.00337, \qquad b_{19} = \frac{356767997597062240718}{351646841762012671875} = 1.01456, \qquad b_{20} = \frac{11381426452523809832}{11104637108274084375} = 1.02493, \qquad b_{21} = \frac{4582801328220180237002}{4430750206201359665625} = 1.03432, \qquad b_{22} = \frac{227680409399602632710572}{218372688734209869234375} = 1.04262, \qquad b_{23} = \frac{517202743969010750316878}{492692099375531357859375} = 1.04975, \$$

$$b_{2} = \frac{4}{3} = 1.33333,$$

$$b_{4} = \frac{136}{135} = 1.00741,$$

$$b_{6} = \frac{524}{567} = 0.924162,$$

$$b_{8} = \frac{16496}{18225} = 0.90513,$$

$$b_{10} = \frac{172739156}{189448875} = 0.911798,$$

$$b_{12} = \frac{13121928056}{14105329875} = 0.930282,$$

$$b_{14} = \frac{102475021588}{107417512125} = 0.953988,$$

$$b_{16} = \frac{369799304676448}{377670207403125} = 0.979159,$$

$$\begin{split} b_{24} &= \frac{2736149442418265808257104}{2591988870627795404390625} = 1.05562, \\ b_{25} &= \frac{4740230682068871716934972106}{4471180801832947072573828125} = 1.06017, \\ b_{26} &= \frac{570547150460378858384921252}{536541696219953648708859375} = 1.06338, \\ b_{27} &= \frac{29675590935592413532223620618}{27858895765266824067575390625} = 1.06521, \\ b_{28} &= \frac{258546091475864119095437139689808}{2426147655509791907572938043359375} = 1.06567, \\ b_{28} &= \frac{22142177787795638219432824420246}{207955513329410734934823260859375} = 1.06476, \\ b_{30} &= \frac{124001539976502789031080018848217028}{2460468256419300381525812775390625} = 1.06251, \\ b_{31} &= \frac{913942879519589340576250740113307218374}{863041374756220726321383385474013671875} = 1.05898, \\ b_{32} &= \frac{12578379301403075966500674215988023744}{11931447577274940916885484591806640625} = 1.05422, \\ b_{33} &= \frac{542842766602892099722217105771465167274}{517824824853732435792830031284408203125} = 1.04831, \\ b_{34} &= \frac{2515578482169070458311758207863571382948}{241568972403775717215355409101904296875} = 1.04135, \\ b_{35} &= \frac{395033002495493164785568330959976069381726}{393144993202427842093318701198359349091796875} = 1.02469, \\ b_{37} &= \frac{78660797482558424580284131747519877025764452082318}{77479461939200296965786487928736146463000732421875} = 1.01525, \\ b_{38} &= \frac{631506936948695573831692392667417028650004084836}{6282118535610834889117823335071478256438511694658} = 0.994837, \\ b_{39} &= \frac{365113126798745178721423333007187865438511694658}{3670079776057324878534441650453964114052978515625} = 0.994837, \\ b_{40} &= \frac{6493310939847396596343462122216690160482900896929136}{659767418205344067239428078125} = 0.984182. \\ b_{40} &= \frac{6493310939847396596343462122216690160482900896929136}{6597674182053440672394280164393147241118600830078125} = 0.984182. \\ b_{40} &= \frac{6493310939847396596343462122216690160482900896929136}{6597674182053440672394280164393147241118600830078125} = 0.984182. \\ b_{40} &= \frac{6493310939847396596343466122216690160482900896929136}{65976741820534406723942801643931472411186008300$$

The exact coefficients c_k of the considerably faster converging Euler series (38) are

$$c_{0} = 1, \qquad c_{1} = \frac{1}{3} = 0.3$$

$$c_{2} = \frac{2}{9} = 0.222222, \qquad c_{3} = \frac{22}{135} = 0$$

$$c_{4} = \frac{52}{405} = 0.128395, \qquad c_{5} = \frac{20}{189} = 0$$

$$c_{6} = \frac{3824}{42525} = 0.0899236, \qquad c_{7} = \frac{1424}{18225} = 0$$

$$c_{8} = \frac{15856}{229635} = 0.0690487, \qquad c_{9} = \frac{117146}{1894483}$$

$$c_{10} = \frac{44536288}{795685275} = 0.0559722, \qquad c_{11} = \frac{72097}{141053}$$

$$c_{12} = \frac{446696704}{9499507875} = 0.0470231, \qquad c_{13} = \frac{46761}{107417}$$

$$c_{14} = \frac{5773668352}{1242492618125} = 0.0405191, \qquad c_{15} = \frac{157414}{415437}$$

$$c_{16} = \frac{443454518548736}{12463116844303125} = 0.0355813,$$

$$c_{17} = \frac{4203929353540352}{50235263108858953125} = 0.0317067,$$

$$c_{19} = \frac{238484326615184896}{7931883648767203125} = 0.0300665,$$

$$c_{20} = \frac{2302874708091894784}{80559094658206539375} = 0.0285862,$$

$$c_{21} = \frac{5949200657650573179904}{218372688734209869234375} = 0.0272433,$$

$$c_{22} = \frac{141017453840496297955328}{5419613093130844936453125} = 0.0249005,$$

$$c_{24} = \frac{1172952349518476898910208}{491338549651972205777734375} = 0.0238726,$$

$$c_{1} = \frac{1}{3} = 0.333333,$$

$$c_{3} = \frac{22}{135} = 0.162963,$$

$$c_{5} = \frac{20}{189} = 0.10582,$$

$$c_{7} = \frac{1424}{18225} = 0.0781344,$$

$$c_{9} = \frac{11714672}{189448875} = 0.0618355,$$

$$c_{11} = \frac{720976352}{14105329875} = 0.0511138,$$

$$c_{13} = \frac{4676141056}{107417512125} = 0.0435324,$$

$$c_{15} = \frac{157416967271936}{4154372281434375} = 0.0378919,$$



References

- [1] M. Abramowitz and I.A. Stegun, Handbook of Mathematical Functions (Dover, New York, 1968).
- [2] Arnold O. Allen, *Probability, Statistics, and Queueing Theory*, Computer Science and Applied Mathematics (Academic Press, Orlando, 1978).
- [3] P. Billingsley, Probability and Measure (Wiley, New York, 1986).
- [4] C. Bisdikian, J.S. Lew and A.N. Tantawi, On the tail approximation of the blocking probability of single server queues with finite buffer capacity, in: *Queueing Networks with Finite Capacity*, *Proc. 2nd Int. Conf.* (May 28–29, 1992) pp. 267–280.
- [5] H. Bruneel and B.G. Kim, *Discrete-Time Models for Communication System Including ATM* (Kluwer Academic Publishers, Boston, 1993).
- [6] B. Steyaert and H. Bruneel, Analytic derivation of the cell loss probability in finite multiserver buffers, from infinite buffer results, *Proceedings of the Second Workshop on Performance Modelling and Evaluation of ATM Networks, Bradford, UK* (June, 1994) 18.1–11.
- [7] H.E. Daniels, Saddlepoint approximations in statistics, *The Annals of Mathematical Statistics* 25 (1954) 631–650.
- [8] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Appications* (Jones & Bartlett, Boston, 1992).
- [9] A. Elwalid, D. Heyman, T.V. Lakshman, D. Mitra and A. Weiss, Fundamental bounds and approximations for ATM multiplexers with applications to video teleconferencing, *IEEE Journal on Selected Areas in Communications* 13(6) (August, 1995) 1004–1016.
- [10] R.J. Gibbens and P.J. Hunt, Effective bandwidths for the multi-type UAS channel, Queueing Systems 9 (1991) 17–28.
- [11] R. Guerin, H. Ahmadi and M. Naghshineh, Equivalent capacity and its application to bandwidth allocation in high-speed networks, *IEEE Journal on Selected Areas in Communications* 9(7) (September, 1991) 968–981.
- [12] G.H. Hardy, Divergent Series (Oxford Univ. Press, London, 1948).
- [13] J.Y. Hui, Resource allocation for broadband networks, IEEE Journal on Selected Areas in Communications 6(9) (December, 1988) 1598–1608.
- [14] ITU-T, Traffic control and congestion control in B-ISDN, *Recommendation I.371* (November, 1995).
- [15] F.P. Kelly, Effective bandwidths at multi-class queues, Queueing Systems 9 (1991) 5–16.
- [16] G. Kesidis, J. Walrand and C.-S. Chang, Effective bandwidths for multiclass markov fluids and other ATM sources, *IEEE/ACM Trans. Networking* 1(4) (August, 1993) 424–428.
- [17] L. Kleinrock, Queueing Systems, Volume 1: Theory (Wiley, New York, 1975).
- [18] Arthur Y-M. Lin and John A. Silvester, Priority queueing strategies and buffer allocation protocols for traffic control at an ATM integrated broadband switching system, *IEEE Journal on Selected Areas in Communications* 9(9) (1991) 1524–1536.
- [19] A.I. Markushevich, *Theory of Functions of a Complex Variable*, Vols I–III (Chelsea, New York, 1985).
- [20] P. Van Mieghem, The characteristic coefficients of a complex function (unpublished, 1995).
- [21] P. Van Mieghem, J. David and G.H. Petit, Performance of cell loss priority management schemes in shared buffers with poisson arrivals, *IEEE Journal on Selected Areas in Communications* (submitted, 1996).
- [22] W.H. Press, S.A. Teukolsky, W.T. Vetterling and B.P. Flannery, *Numerical Recipes in C* (Cambridge Univ. Press, New York, second edition, 1992).
- [23] J.W. Roberts, Performance Evaluation and Design of Multiservice Networks, Volume COST 224 of Information Technologies and Sciences (Commission of the European Communities, Luxembourg, October, 1991).

- [24] K. Sohraby, On the theory of general on-off sources with applications in high-speed networks, *IEEE Infocom* (1993) (4a.3) 1-10.
- [25] R. Syski, Introduction to Congestion Theory in Telephone Systems, Studies in Telecommunication, Vol. 4 (North-Holland, Amsterdam, second edition, 1986).
- [26] E.C. Titchmarsh, The Theory of Functions (Oxford Univ. Press, Amen House, London EC4, 1964).
- [27] A. Weiss, An introduction to large deviations for communication networks, *IEEE Journal on Selected Areas in Communications* 13(6) (1995) 938–952.
- [28] R.W. Wolff, Poisson arrivals see time averages, Operations Research 30(2) (1982) 223-231.
- [29] R.W. Wolff, Stochastic Modeling and the Theory of Queues (Prentice-Hall International Editions, New York, 1989).