

Available online at www.sciencedirect.com





Performance Evaluation 65 (2008) 641-652

www.elsevier.com/locate/peva

Time-dependent performance analysis of a discrete-time priority queue

Joris Walraevens*, Dieter Fiems, Herwig Bruneel

SMACS Research Group, Department of Telecommunications and Information Processing (IR07), Ghent University - UGent, Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium

> Received 8 June 2006; received in revised form 21 August 2007; accepted 13 February 2008 Available online 16 February 2008

Abstract

We present the transient analysis of the system content in a two-class discrete-time $M^X/D/1$ priority queue. In particular, we derive an expression for the generating function of the transient system contents of both classes at the beginning of slots. Performance measures are calculated from this generating function. To illustrate our approach we conclude with some examples. © 2008 Elsevier B.V. All rights reserved.

Keywords: Transient analysis; Discrete-time priority queue

1. Introduction

In recent years, there has been much research devoted to the incorporation of multimedia applications in packetbased networks (e.g. IP networks). Different types of traffic need different QoS standards. For real-time interactive applications, it is important that mean delay and delay-jitter are bounded, while for non real-time applications, the loss ratio and throughput are the restrictive quantities. In order to guarantee acceptable delay boundaries to delaysensitive traffic (such as voice/video), several scheduling schemes – for switches, routers, etc – have been proposed and analyzed, each with their own specific algorithmic and computational complexity.

Amongst these scheduling disciplines are some well-known strategies like weighted round-robin (WRR) [12], weighted fair queueing (WFQ) or generalized processor sharing (GPS) [16], earliest deadline first (EDF) [11], probabilistic priority (PP) [19], Head-Of-Line (HOL) priority scheduling (see e.g. [23]) and HOL priority scheduling with priority jumps (HOL-PJ) [14]. All these scheduling disciplines give some kind of transmission priority to delay-sensitive traffic over delay-insensitive traffic. The most drastic one in this respect is the strict priority scheduling discipline. With this scheduling, packets are classified in a number of priority classes, based on their delay constraints. As long as packets of the highest priority class are present in the queueing system, this type of traffic is served. Packets of lower priority can only be transmitted when no higher priority traffic is present in the system. As already mentioned, this is the most drastic way to meet the QoS constraints of delay-sensitive traffic (and thus the scheduling with the most disadvantageous consequences to the delay-insensitive traffic), but also the easiest one to implement.

* Corresponding author. Tel.: +32 9 264 89 02; fax: +32 9 264 42 95.

E-mail addresses: jw@telin.UGent.be (J. Walraevens), df@telin.UGent.be (D. Fiems), hb@telin.UGent.be (H. Bruneel).

^{0166-5316/\$ -} see front matter © 2008 Elsevier B.V. All rights reserved. doi:10.1016/j.peva.2008.02.002

In this paper, we focus on the transient behavior of a discrete-time $M^X/D/1$ queue with a HOL priority scheduling discipline and two priority classes. The numbers of per-slot arrivals are series of independent and identically distributed (i.i.d.) random variables, the service times are deterministically equal to one slot and there is one server to transmit the packets. One type of arriving traffic (type 1) has priority over another type of traffic (type 2). Thus, a packet of type-1 traffic is transmitted at the beginning of a slot, if at least one is present. Only when there are no packets of type 1, a packet of type 2 can be scheduled for transmission.

The steady-state behavior of HOL priority queueing models is well understood, see e.g. [18] for an overview of some basic models. In the recent literature, emphasis is put on more complex priority queueing models. Examples are multi-server priority queueing models [6,9], priority queueing models with correlation in the arrival process [3,13,15, 20,22,24], priority queueing models with acceptance control [1,4] and priority models with server vacations [8].

Results on the *transient* behavior of priority queueing systems however are scarce. One of the few exceptions is [10], wherein the time-dependent and steady-state behavior of a continuous-time non-preemptive priority queue is analyzed. Transient performance measures are useful for a whole number of reasons, including (a) the determination of the time necessary to reach the steady state, (b) the analysis of systems with time-dependent or quasi-stationary arrival rates and (c) the characterization of the buffer size in overload situations.

Transient analyses of queueing models are notoriously hard. For particular (non-priority) queueing systems, one may find explicit expressions for the time-dependent probability generating function of the queue content. This is e.g. the case for the M/M/1 queueing system [17] and the M/E - r/1 queueing system [5]. Inversion of these transform functions is possible but involve Bessel functions. Various authors have obtained expressions for the *z*-transform of the sequence of the probability generating functions of the queue content at consecutive epochs in time. Bruneel [2] performs a transient analysis of the discrete-time $M^X/Geo/1$ queue. The *z*-transform of the probability generating functions and the *z*-transform of the probability that the queue is empty at the consecutive slot boundaries. An ad-hoc method is provided to find the latter probabilities. A similar approach is followed in [7], wherein a discrete-time queue with a correlated train arrival process is studied.

Here, we perform a transient analysis by using a generating functions approach. From the obtained generating functions, expressions are calculated for some interesting performance measures such as the time-dependent probability of having an empty buffer and the time-dependent mean system contents of both priority classes. The queueing model is identical to the model in [21]. In [21], we analyzed (a.o.) the steady-state system contents of both classes, while in this paper, we focus on the transient system contents. We put hereby great emphasis on *how to obtain the exact transient performance measures* (such as the moments) from the generating functions, which is a non-trivial problem. We therefore extend the transient analysis in buffers with a FIFO queueing discipline described in [2] in order to have an algorithm to obtain these transient performance measures.

The outline is as follows. In the following section, we introduce the mathematical model. In Sections 3 and 4, we investigate the time-dependent system contents. Section 3 relates the generating function of the time-dependent system contents to some unknown boundary functions, which are investigated in Section 4. We illustrate our approach by means of a few examples in Section 5. Finally, some conclusions are drawn in Section 6.

2. Mathematical model

We consider a single-server queueing system with infinite buffer space. Time is assumed to be slotted and one server transmits packets at the rate of one packet per slot. Two types of traffic arrive at the system, namely packets of type 1 and packets of type 2. We denote the number of arrivals of type j during slot k by $a_j^{(k)}$ (j = 1, 2). The numbers of arrivals are assumed to be i.i.d. from slot-to-slot and are characterized by their joint probability mass function $a(m, n) = \text{Prob}[a_1^{(k)} = m, a_2^{(k)} = n]$ and their joint probability generating function (pgf) $A(z_1, z_2)$, defined and given by

$$A(z_1, z_2) \triangleq \mathbb{E}\left[z_1^{a_1^{(k)}} z_2^{a_2^{(k)}}\right] = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a(m, n) z_1^m z_2^n.$$

Notice that the numbers of arrivals from both types within a slot can be correlated. Furthermore, we denote the total number of arriving packets during slot k by $a_T^{(k)} \triangleq a_1^{(k)} + a_2^{(k)}$ and let $A_T(z) \triangleq E\left[z^{a_T^{(k)}}\right] = A(z, z)$ denote the

corresponding pgf. The pgfs of the number of arrivals of type 1 and type 2 are defined as $A_1(z) \triangleq \mathbb{E}\left[z^{a_1^{(k)}}\right] = A(z, 1)$ and $A_2(z) \triangleq \mathbb{E}\left[z^{a_2^{(k)}}\right] = A(1, z)$ respectively. Finally, the arrival rate of type-*j* packets (j = 1, 2) is denoted and given by $\lambda_j = A'_j(1)$, while the total arrival rate is denoted and given by $\lambda_T = A'_T(1)$.

3. Analysis

We denote the type-*j* system content at the beginning of slot k + 1 by $u_j^{(k)}$ (j = 1, 2). Furthermore, the joint pgf of $u_1^{(k)}$ and $u_2^{(k)}$ is denoted by $U^{(k)}(z_1, z_2)$, i.e., $U^{(k)}(z_1, z_2) \triangleq \mathbb{E}\left[z_1^{u_1^{(k)}} z_2^{u_2^{(k)}}\right]$. The system contents are characterized by the following set of system equations:

$$u_1^{(k)} = \left[u_1^{(k-1)} - 1\right]^+ + a_1^{(k)}$$
(1)
$$\int \left[u_1^{(k-1)} - 1\right]^+ + a_1^{(k)} \quad \text{if } u_1^{(k-1)} = 0$$

$$u_{2}^{(k)} = \begin{cases} u_{2}^{(k-1)} + u_{2}^{(k)} & \text{if } u_{1}^{(k-1)} > 0, \end{cases}$$
(2)

where $[.]^+$ denotes the maximum of the argument and 0. The first equation can be understood by noticing that transmission of type-1 packets is not influenced by type-2 packets. That is, one type-1 packet is transmitted in a slot when there are packets of that type present at the beginning of that slot. A type-2 packet on the other hand can only be transmitted, if there are no type-1 packets waiting in the queue at the end of the previous slot (because of the priority scheduling). This leads to the second equation. The system equations then yield the following relation between $U^{(k)}(z_1, z_2)$ and $U^{(k-1)}(z_1, z_2)$ by means of some standard z-transform manipulations:

$$U^{(k)}(z_1, z_2) = A(z_1, z_2) \left\{ \frac{z_2 - 1}{z_2} U^{(k-1)}(0, 0) + \frac{z_1 - z_2}{z_1 z_2} U^{(k-1)}(0, z_2) + \frac{1}{z_1} U^{(k-1)}(z_1, z_2) \right\}.$$
(3)

Define $U(x, z_1, z_2)$ as the z-transform of the sequence $U^{(k)}(z_1, z_2)$ with respect to the time parameter k, i.e., $U(x, z_1, z_2) \triangleq \sum_{k=0}^{\infty} U^{(k)}(z_1, z_2) x^k$. Multiplying both sides of Eq. (3) with x^k and summing over all $k = 1, ..., \infty$ then leads to the following expression for $U(x, z_1, z_2)$:

$$U(x, z_1, z_2) = \frac{z_1 U^{(0)}(z_1, z_2) + \frac{xA(z_1, z_2)}{z_2} [z_1(z_2 - 1)U(x, 0, 0) + (z_1 - z_2)U(x, 0, z_2)]}{z_1 - xA(z_1, z_2)}.$$
(4)

Under the assumption that $U^{(0)}(z_1, z_2)$ - the joint pgf of the system contents of both types at the beginning of the first slot - is known, there are still two unknown functions in the right-hand side of Eq. (4), namely $U(x, 0, z_2)$ and U(x, 0, 0). U(x, 0, 0) is the z-transform of the sequence $U^{(k)}(0, 0)$, i.e., U(x, 0, 0) describes the evolution in time of the probability of an empty buffer. $U(x, 0, z_2)$ is the z-transform of the sequence $U^{(k)}(0, z_2)$ and is thus the evolution in time of the partial pgf of the class-2 system content given that there are no class-1 packets present. Once we have found expressions for these generating functions, the transient behavior of the type-1 and type-2 system contents can be found by expanding Eq. (4) about x = 0. More specifically, we show in the following subsections how to calculate means (and higher moments) of the total system content and of the system contents of type 1 and type 2 from expression (4). Determination of the unknown boundary functions (U(x, 0, 0) and $U(x, 0, z_2)$) and/or the corresponding sequences ($\{U^{(k)}(0, 0), k \ge 0\}$ and $\{U^{(k)}(0, z_2), k \ge 0\}$) is the subject of Section 4.

3.1. Total system content

By substituting z_1 and z_2 by z in Eq. (4), we find:

$$U_T(x,z) = \frac{zU_T^{(0)}(z) + xA_T(z)(z-1)U_T(x,0)}{z - xA_T(z)},$$
(5)

with $U_T(x, z) \triangleq U(x, z, z)$ and $U_T^{(k)}(z) = U^{(k)}(z, z)$. So, concerning the transient total system content, the only unknown is $U_T(x, 0)$, or equivalently, the only unknowns are the transient probabilities of an empty system $U_T^{(k)}(0), k \ge 0$.

 $U_T^{(k)}(0), k \ge 0.$ The moment generating property of pgfs then leads to expressions of various (time-dependent) moments of the total system content. For example, if $\mathbb{E}\left[u_T^{(k)}\right]$ denotes the mean total system content at the beginning of slot k + 1, or,

 $\mathbb{E}\left[u_T^{(k)}\right] \triangleq \left. \frac{dU_T^{(k)}(z)}{dz} \right|_{z=1}, \text{ then the transform function of the sequence } \left\{\mathbb{E}\left[u_T^{(k)}\right], k \ge 0\right\} \text{ can be found from } U_T(x, z) \text{ as:}$

$$\sum_{k=0}^{\infty} \mathbf{E} \left[u_T^{(k)} \right] x^k = \left. \frac{\partial U_T(x, z)}{\partial z} \right|_{z=1}.$$
(6)

Taking the first derivative in z of both sides of Eq. (5), substituting z by 1 and using Eq. (6), we obtain:

$$\sum_{k=0}^{\infty} \mathbf{E} \left[u_T^{(k)} \right] x^k = \frac{\mathbf{E} \left[u_T^{(0)} \right] + x U_T(x,0)}{1-x} - \frac{(1-\lambda_T)x}{(1-x)^2}.$$
(7)

Since the coefficients in x^k have to be equal in both sides of Eq. (7), this equation is equivalent with

$$E\left[u_{T}^{(k)}\right] = E\left[u_{T}^{(k-1)}\right] + \lambda_{T} - 1 + U_{T}^{(k-1)}(0),$$
(8)

for all $k \ge 1$. Eq. (8) can thus be used to recursively calculate the mean total system content $E\left[u_T^{(k)}\right]$, once the sequence $\left\{U_T^{(k)}(0), k \ge 0\right\}$ is known. Note that this last equation can also be directly retrieved from the system equations (1)–(2). If one wants to also calculate higher moments though, the calculated pgfs come in handy.

As already mentioned, we show how to calculate the unknown sequence $\left\{U_T^{(k)}(0), k \ge 0\right\}$ (or its transform function) in the next section.

3.2. Type-1 system content

The substitution of z_2 by 1 (and z_1 by z for ease of notation) in Eq. (4) yields

$$U_1(x,z) = \frac{zU_1^{(0)}(z) + xA_1(z)(z-1)U_1(x,0)}{z - xA_1(z)},$$
(9)

with $U_1(x, z) \triangleq U(x, z, 1)$ and $U_1^{(k)}(z) = U^{(k)}(z, 1)$. So, this equation is very similar to Eq. (5). Therefore, we get a similar recursive expression (like expression (8)) for the mean type-1 system content:

$$\mathbf{E}\left[u_{1}^{(k)}\right] = \mathbf{E}\left[u_{1}^{(k-1)}\right] + \lambda_{1} - 1 + U_{1}^{(k-1)}(0), \tag{10}$$

for all $k \ge 1$. Eq. (10) can thus be used to recursively calculate the mean type-1 system content $E\left[u_1^{(k)}\right]$, once the sequence $\left\{U_1^{(k)}(0), k \ge 0\right\}$ is known.

3.3. Type-2 system content

By substituting z_1 by 1 (and z_2 by z) in Eq. (4), we find

$$U_2(x,z) = \frac{zU_2^{(0)}(z) + xA_2(z)(1-z)[U(x,0,z) - U_T(x,0)]}{z(1-xA_2(z))},$$
(11)

with $U_2(x, z) \triangleq U(x, 1, z)$ and $U_2^{(k)}(z) = U^{(k)}(1, z)$. So, in this case the remaining two unknowns are U(x, 0, z) and $U_T(x, 0)$, or equivalently, the sequences $\{U^{(k)}(0, z), k \ge 0\}$ and $\{U_T^{(k)}(0), k \ge 0\}$.

Again, the moment generating property of pgfs leads to an expression of the mean type-2 system content $E\left[u_{2}^{(k)}\right]$ (and higher moments) during the consecutive slots, i.e.,

$$\sum_{k=0}^{\infty} \mathbb{E}\left[u_2^{(k)}\right] x^k = \left. \frac{\partial U_2(x,z)}{\partial z} \right|_{z=1}.$$
(12)

Taking the first derivative in z of both sides of Eq. (11), and substituting z by 1, we obtain:

$$\sum_{k=0}^{\infty} \mathbb{E}\left[u_{2}^{(k)}\right] x^{k} = \frac{\mathbb{E}\left[u_{2}^{(0)}\right] + x[U_{T}(x,0) - U_{1}(x,0)]}{1-x} + \frac{\lambda_{2}x}{(1-x)^{2}}.$$
(13)

Since the coefficients in x^k have to be equal in both sides of Eq. (13), we find the following recursive equation for the mean type-2 system content,

$$\mathbf{E}\left[u_{2}^{(k)}\right] = \mathbf{E}\left[u_{2}^{(k-1)}\right] + \lambda_{2} + U_{T}^{(k-1)}(0) - U_{1}^{(k-1)}(0), \tag{14}$$

for all $k \ge 1$. From Eq. (14), the mean type-2 system content $\mathbb{E}\left[u_2^{(k)}\right]$ can be recursively calculated once the sequences $\left\{U_T^{(k)}(0), k \ge 0\right\}$ and $\left\{U_1^{(k)}(0), k \ge 0\right\}$ are known.

Notice that Eqs. (8), (10) and (14) fulfil the relation $E\left[u_T^{(k)}\right] = E\left[u_1^{(k)}\right] + E\left[u_2^{(k)}\right]$, as expected. Furthermore, higher moments and cross moments (such as the covariance) of the type 1 and type 2 system content can also be recursively calculated by computing higher-order derivatives of $U(x, z_1, z_2)$ (i.e., of expression (4)). Notice that – although for the *mean* type-2 system contents it suffices to calculate $U_T(x, 0)$ and $U_1(x, 0) - U(x, 0, z)$ will have to be determined if one wants to calculate *higher* moments of the type-2 system contents and/or *cross* moments of the system contents of both types. As an example, we show the expression of the variance of $u_2^{(k)}$:

$$\operatorname{Var}\left[u_{2}^{(k)}\right] = \operatorname{Var}\left[u_{2}^{(k-1)}\right] + \lambda_{2}(1-\lambda_{2}) + \operatorname{Var}\left[a_{2}^{(k-1)}\right] + 2\lambda_{2}\operatorname{E}\left[u_{2}^{(k)}\right] - \left(\operatorname{E}\left[u_{2}^{(k)}\right] - \operatorname{E}\left[u_{2}^{(k-1)}\right]\right) - \left(\left(\operatorname{E}\left[u_{2}^{(k)}\right]\right)^{2} - \left(\operatorname{E}\left[u_{2}^{(k-1)}\right]\right)^{2}\right) - 2\operatorname{E}\left[u_{2}^{(k-1)}|u_{1}^{(k-1)} = 0\right]\operatorname{Prob}\left[u_{1}^{(k-1)} = 0\right].$$
(15)

Since

$$E\left[u_{2}^{(k-1)}|u_{1}^{(k-1)}=0\right]\operatorname{Prob}\left[u_{1}^{(k-1)}=0\right] = \left.\frac{\partial U^{(k-1)}(0,z_{2})}{\partial z_{2}}\right|_{z_{2}=1}$$
(16)

the variance of the transient class-2 system content can be recursively calculated from expression (15) when $U(x, 0, z_2)$ or the sequence $\{U^{(k)}(0, z_2), k \ge 0\}$ are known.

4. Determination of the unknowns

In this section we show how to calculate the boundary functions U(x, 0, 0) and $U(x, 0, z_2)$, or alternatively, the sequences $\{U_T^{(k)}(0), k \ge 0\}$ and $\{U^{(k)}(0, z_2), k \ge 0\}$ respectively. We apply two different methods. The first method is only useful for specific arrival processes, while the second one is more generally applicable.

4.1. Calculation of the boundary functions

Applying Rouché's theorem, it can be shown that the denominator of the right-hand side of Eq. (4) has one zero in the unit circle for z_1 for given values of x and z_2 (|x| < 1, $|z_2| < 1$), namely $X(x, z_2) \triangleq xA(X(x, z_2), z_2)$. One easily shows that $U(x, z_1, z_2)$ is analytic for all x and z_j with |x| < 1 and $|z_j| < 1$ (j = 1, 2). Therefore $X(x, z_2)$

must also be a zero of the numerator of the right-hand side of (4), i.e.,

$$X(x, z_2)U^{(0)}(X(x, z_2), z_2) + \frac{xA(X(x, z_2), z_2)}{z_2}[X(x, z_2)(z_2 - 1)U(x, 0, 0) + (X(x, z_2) - z_2)U(x, 0, z_2)] = 0.$$
(17)

This can be transformed into

$$U(x, 0, z_2) = \frac{z_2 U^{(0)}(X(x, z_2), z_2) + X(x, z_2)(z_2 - 1)U(x, 0, 0)}{z_2 - X(x, z_2)},$$
(18)

if $A(0, z_2) \neq 0$. When $A(0, z_2) = 0$, $X(x, z_2) = 0$ and (17) is always valid. We discuss this special case at the end of this subsection. Applying Rouché's theorem once more, it can be proved that the denominator of the right-hand side of Eq. (18) has one zero in the unit circle for z_2 for given value of x (|x| < 1), namely $Y(x) \triangleq X(x, Y(x))$. Since $U(x, 0, z_2)$ is finite in the unit circle, Y(x) must also be a zero of the numerator of the right-hand side of (18), i.e.,

$$Y(x)U^{(0)}(X(x, Y(x)), Y(x)) + X(x, Y(x))(Y(x) - 1)U(x, 0, 0) = 0.$$

This yields

$$U(x,0,0) = \frac{U^{(0)}(Y(x),Y(x))}{1-Y(x)}.$$
(19)

From (4), (18) and (19), it is obvious that the two boundary functions $U(x, 0, z_2)$ and U(x, 0, 0) and thus also $U(x, z_1, z_2)$ itself are fully determined, once the functions $X(x, z_2)$ and Y(x) are explicitly known. Since these functions are only implicitly defined, this proves to be a hard task in general. However for specific choices of $A(z_1, z_2)$, this is rather straightforward.

Finally, we need to invert the obtained expression of $U(x, z_1, z_2)$ to obtain the sequence $\{U^{(k)}(z_1, z_2), k \ge 0\}$. We discuss this further in Section 5, when we touch upon some examples.

Special case: If $A(0, z_2^*) = 0$ for a particular z_2^* ($|z_2^*| < 1$), $X(x, z_2^*) = 0$ irrespective of x. Taking the limit of expression (18) for $z_2 \rightarrow z_2^*$, we find

$$U(x, 0, z_2^*) = U^{(0)}(0, z_2^*).$$
⁽²⁰⁾

So in this case $U(x, z_1, z_2^*)$ is always explicitly found by this first method and we thus exclude the case $A(0, z_2) = 0$ in the second method.

4.2. Calculation of the transient sequences

In this section we show how to calculate the sequences directly, since the determination of $z_1 = X(x, z_2)$ as a solution of

$$z_1 - xA(z_1, z_2) = 0, (21)$$

for given values of x and z_2 , may be hard for general $A(z_1, z_2)$. To do that we observe that solving Eq. (21) for the unknown x, for given values z_1 and z_2 is straightforward. This solution is given by

$$x = f(z_1, z_2) \triangleq \frac{z_1}{A(z_1, z_2)}.$$
(22)

It is clear from Eq. (22), that if $|z_1| < 1$ is chosen sufficiently small – say $|z_1| < \varepsilon$ – and $|z_2| < 1$, $|x| = |f(z_1, z_2)|$ is also smaller than unity as $|f(z_1, z_2)| \approx |z_1|/|A(0, z_2)|$ for small $|z_1|$ and since we have assumed $|A(0, z_2)| > 0$ (see also the special case in the previous subsection).

In the remainder of this subsection, we first show in detail how to calculate the $U_T^{(k)}(0)$ from (5). $U_T(x, z)$ must be analytic for (x, z) = (f(z, z), z) and $|z| < \varepsilon$, since |f(z, z)| < 1 for $|z| < \varepsilon$. Since the denominator of (5) is zero for (x, z) = (f(z, z), z), the numerator must vanish too, or,

$$zU_T^{(0)}(z) + f(z,z)A_T(z)(z-1)U_T(f(z,z),0) = 0.$$

646

Since $f(z, z) = z/A_T(z)$, this leads to

$$U_T(f(z,z),0) = \frac{U_T^{(0)}(z)}{1-z},$$
(23)

for $|z| < \varepsilon$. We now show that the former expression is sufficient to derive the probabilities $U_T^{(k)}(0)$. $U_T^{(k)}(0)$ is by definition the coefficient of x^k in the expansion of $U_T(x, 0)$ about x = 0. In other words, using complex analysis, $U_T^{(k)}(0)$ can be obtained as the residue of the function $x^{-k-1}U_T(x, 0)$ in x = 0. This residue can be calculated as follows:

$$U_T^{(k)}(0) = \frac{1}{2\pi i} \oint_{C_0} U_T(x,0) x^{-k-1} dx,$$
(24)

with $i = \sqrt{-1}$ and C_0 a small contour which surrounds x = 0, but no singularities of $U_T(x, 0)$. Changing the variable x by the variable z with x = f(z, z) (with $f(z_1, z_2)$ as defined in (22)), this expression transforms into

$$U_T^{(k)}(0) = \frac{1}{2\pi i} \oint_{C_1} U_T(f(z, z), 0) f(z, z)^{-k-1} \frac{df(z, z)}{dz} dz,$$
(25)

with C_1 the transformation of C_0 in the z-plane, which circles the point z = 0 exactly once. We may choose C_0 such that $|z| < \varepsilon$ on the contour C_1 . Therefore, we can use expression (23), leading to

$$U_T^{(k)}(0) = \frac{1}{2\pi i} \oint_{C_1} \frac{U_T^{(0)}(z) A_T(z)^{k-1} [A_T(z) - zA_T'(z)]}{1 - z} z^{-k-1} dz.$$
(26)

Since this expression only contains known functions and quantities, it leads to the determination of $U_T^{(k)}(0)$. Performing the contour integration in expression (26) – by means of any method – yields the $U_T^{(k)}(0)$'s. A practical interpretation of (26) is the following: $U_T^{(k)}(0)$ is the coefficient of z^k in the expansion about z = 0 of

$$P_T^{(k)}(z) = \frac{U_T^{(0)}(z)A_T(z)^{k-1}[A_T(z) - zA_T'(z)]}{1 - z}.$$
(27)

The calculation of this coefficient, using any method – depending on the particular form of $U_T^{(0)}(z)$ and $A_T(z)$ – thus gives $U_T^{(k)}(0)$, for this specific k. In Section 5, we will discuss through an example how this can be done.

A similar technique can also be used to calculate the sequence $\{U^{(k)}(0, z_2), k \ge 0\}$. $U(x, 0, z_2)$ must be analytic for $(x, z_2) = (f(z_1, z_2), z_2), |z_2| < 1$ and $|z_1| < \varepsilon$, since $|f(z_1, z_2)| < 1$ for $|z_1| < \varepsilon$ and for any z_2 within the unit disk. Since the denominator of (4) is zero for $(x, z_2) = (f(z_1, z_2), z_2)$, the numerator must vanish too:

$$z_1 U^{(0)}(z_1, z_2) + \frac{f(z_1, z_2)A(z_1, z_2)}{z_2} [z_1(z_2 - 1)U_T(f(z_1, z_2), 0) + (z_1 - z_2)U(f(z_1, z_2), 0, z_2)] = 0.$$

Using (22), we find:

$$U(f(z_1, z_2), 0, z_2) = \frac{z_2 U^{(0)}(z_1, z_2) + z_1(z_2 - 1) U_T(f(z_1, z_2), 0)}{z_2 - z_1},$$
(28)

for $|z_1| < \varepsilon$ and $|z_2| < 1$. $U^{(k)}(0, z_2)$ is by definition the coefficient of x^k in the expansion of $U(x, 0, z_2)$ about x = 0. We thus have

$$U^{(k)}(0, z_2) = \frac{1}{2\pi i} \oint_{C'_0} U(x, 0, z_2) x^{-k-1} dx,$$
(29)

with C'_0 a small contour which surrounds x = 0, but no singularities of $U(x, 0, z_2)$. Changing the variable x by the variable z_1 with $x = f(z_1, z_2)$ this expression transforms into

$$U^{(k)}(0, z_2) = \frac{1}{2\pi i} \oint_{C_1'} U(f(z_1, z_2), 0, z_2) f(z_1, z_2)^{-k-1} \frac{\partial f(z_1, z_2)}{\partial z_1} dz_1,$$
(30)

with C'_1 the transformation of C'_0 in the z_1 -plane, which circles the point $z_1 = 0$ exactly once. Choosing C'_0 such that $|z| < \varepsilon$ on contour C'_1 , we can use expression (28), leading to

$$U^{(k)}(0, z_2) = \frac{1}{2\pi i} \oint_{C_1} \frac{z_2 U^{(0)}(z_1, z_2) + z_1(z_2 - 1) U_T(z_1 / A(z_1, z_2), 0)}{z_2 - z_1} \frac{A(z_1, z_2) - z_1 \frac{\partial A(z_1, z_2)}{\partial z_1}}{A(z_1, z_2)^{-k+1}} \times z_1^{-k-1} dz_1.$$
(31)

Since this expression only contains known functions and quantities, it leads to the determination of $U^{(k)}(0, z_2)$. $U^{(k)}(0, z_2)$ is for instance the coefficient of z_1^k in the expansion about $z_1 = 0$ of

$$P^{(k)}(z_1, z_2) = \frac{z_2 U^{(0)}(z_1, z_2) + z_1(z_2 - 1) U_T(z_1 / A(z_1, z_2), 0)}{z_2 - z_1} \frac{A(z_1, z_2) - z_1 \frac{\partial A(z_1, z_2)}{\partial z_1}}{A(z_1, z_2)^{-k+1}}.$$
(32)

Recall that obtaining the unknown function $U_T(z, 0)$ – which still appears in this expression – was the topic of the first part of this section. Note further that by substituting z_2 by 1 in expression (31) the sequence $\{U_1^{(k)}(0), k \ge 0\}$ is calculated.

5. Examples

In this section, we demonstrate our methods by two examples. The joint pgf of the numbers of arrivals of class 1 and class 2 in a random slot is given by

$$A(z_1, z_2) = \left(1 - \frac{\lambda_1}{N}(1 - z_1) - \frac{\lambda_2}{N}(1 - z_2)\right)^N.$$
(33)

This is in fact the arrival process in an output queue of a non-blocking output-queueing switch with N inlets and N outlets, a Bernoulli process with arrival rate λ_T and independent and uniform routing. An arriving packet is assumed to be of class j with probability λ_j/λ_T (j = 1, 2).

We demonstrate the two methods discussed in Section 4 in detail for particular values of N: the method in Section 4.1 for N = 2 and the one in Section 4.2 for $N = \infty$.

5.1. The case
$$N = 2$$

In this case, explicit expressions for $X(x, z_2)$ and Y(x) – as defined in Section 4.1 – can be found:

$$X(x, z_2) = \frac{2 - \lambda_1 (2 - \lambda_1) x + \lambda_1 \lambda_2 x (1 - z_2) - 2\sqrt{1 - \lambda_1 (2 - \lambda_1) x + \lambda_1 \lambda_2 x (1 - z_2)}}{\lambda_1^2 x},$$
(34)

and

$$Y(x) = \frac{2 - \lambda_T (2 - \lambda_T) x - 2\sqrt{1 - \lambda_T (2 - \lambda_T) x}}{\lambda_T^2 x},$$
(35)

respectively. One then obtains an explicit expression for $U(x, z_1, z_2)$ by substituting expressions (18)–(19) and (34)–(35) in expression (4). To calculate the required performance measures from this expression such as the probability of the system contents being zero, the mean system contents, higher moments of the system contents and/or the cross moments of the system contents, two steps have to be taken:

- an expansion about x = 0 and
- appropriate derivatives for z_1 and z_2 in 0 or 1.

The order in which these steps are executed is not fixed, albeit that it plays an important role. When the expansion about x = 0 is done first, the pgfs $U^{(k)}(z_1, z_2)$ are found and these contain all information about the distribution of the transient system contents of both classes. This means that all moments can be calculated from these pgfs by taking the derivatives in the second step. As a consequence, only the second step has to be applied multiple times when

more than one moment is required. The expansion about x = 0 however requires a recursive computation (as will be demonstrated later). If this expansion is done before the derivatives in z_1 and z_2 are taken, this recursive computation is performed on functions in two variables, which results in a computational burden when many recursions are necessary. If on the other hand the derivatives for z_1 and z_2 in 0 or 1 are taken first, the expansion about x = 0 leads to a recursive computation on numbers, which is computational more efficient. In that case both steps have to be repeated though when more than one moment is to be calculated.

As mentioned, the expansion about x = 0 is computationally the hardest step, because it leads to a recursive procedure in the time index. For the specific arrival process discussed in this subsection, a number of square-root functions appear in the final expression of $U(x, z_1, z_2)$ (see expressions (34) and (35)), which have to be expanded about x = 0. As an example, we show how to (recursively) calculate the coefficients $H_j(z_2)$ of the square root in expression (34), i.e., of the function

$$H(x, z_2) \triangleq \sqrt{1 - \lambda_1 (2 - \lambda_1) x + \lambda_1 \lambda_2 x (1 - z_2)}.$$
(36)

Taking the logarithmic derivative of this expression with respect to x, we obtain

$$\frac{\frac{\partial H(x,z_2)}{\partial x}}{H(x,z_2)} = \frac{-\lambda_1(2-\lambda_1)+\lambda_1\lambda_2(1-z_2)}{2(1-\lambda_1(2-\lambda_1)x+\lambda_1\lambda_2x(1-z_2))}.$$

Replacing $H(x, z_2)$ and $\partial H(x, z_2)/\partial x$ by their respective series expansions $\sum_{j=0}^{\infty} H_j(z_2) x^j$ and $\sum_{j=1}^{\infty} j H_j(z_2) x^{j-1}$, we find

$$2(1 - \lambda_1(2 - \lambda_1)x + \lambda_1\lambda_2x(1 - z_2))\sum_{j=1}^{\infty} jH_j(z_2)x^{j-1} = (-\lambda_1(2 - \lambda_1) + \lambda_1\lambda_2(1 - z_2))\sum_{j=0}^{\infty} H_j(z_2)x^j.$$

By identification of the coefficients of x^j in both sides of this equation, the following recursive calculation of $H_j(z_2)$ is found:

$$H_1(z_2) = \frac{-\lambda_1(2-\lambda_1) + \lambda_1\lambda_2(1-z_2)}{2}H_0(z_2)$$

and

$$H_j(z_2) = \frac{2j-3}{2j} (\lambda_1(2-\lambda_1) - \lambda_1\lambda_2(1-z_2))H_{j-1}(z_2), \quad j \ge 2.$$

Since $H_0(z_2) = H(0, z_2) = 1$, all $H_i(z_2)$ can be calculated recursively.

To demonstrate typical transient behavior, we show a couple of figures in the remainder of this subsection. In Fig. 1, the transient mean contents of the total, class-1 and class-2 system are shown for $\lambda_1 = \lambda_2 = 0.4$, starting from an empty system (a) or from a system with a FIFO scheduling discipline which is in the steady state (b). The latter figure thus gives the transient mean system contents when the scheduling discipline is switched from a FIFO scheduling discipline to a priority discipline. As can be seen, the mean values go to their respective steady-state values for $k \to \infty$ (for the steady-state values, see [21]). Notice that the steady-state total system content is identical for the FIFO and the priority scheduling disciplines. Therefore, the mean total system contents remains constant in time in Fig. 1b.

In Fig. 2(a), we depict the variances of the transient total, class-1 and class-2 system contents starting from an empty system. In Fig. 2(b), the correlation coefficient of the transient class-1 and class-2 system contents is shown, also when the system is empty at the beginning. The correlation coefficient increases for increasing k, but stays negative. This can be explained as follows: the buffer is empty in the beginning and since the numbers of per-slot arrivals of both classes are heavily negatively correlated, the system contents of both classes will also be heavily negatively correlated in the beginning. The further in time though, the more packets of both types that are buffered. As a result the correlation coefficient increases.

5.2. The case $N = \infty$

In this case, Eq. (33) becomes

$$A(z_1, z_2) = e^{\lambda_1(z_1-1)} e^{\lambda_2(z_2-1)}$$

(37)



Fig. 1. Transient mean total, class-1 and class-2 system contents.



Fig. 2. Second-order moments.

i.e., the numbers of per-slot arrivals of class 1 and class 2 are two mutually independent Poisson distributions. Expression (21) is a transcendental equation in z_1 in this case and therefore we use the method proposed in Section 4.2 in this example. Obtaining the required performance measures – i.e., the transient probabilities of the system, the class-1 buffer or the class-2 buffer being empty, the mean transient system contents of both classes, etc – requires mostly the calculation of (27) and/or (32), expanding this expression about z = 0 ($z_1 = 0$ respectively) and calculating the coefficient of z^k (z_1^k respectively) in this expansion, and this for all k. Note that a similar discussion on that in the previous example is possible about the order in which the expansions about 0 and the derivatives for z_1 and z_2 in 0 or 1 (to obtain the required moment) are taken.

We explain here in detail how the sequence $\{U_T^{(k)}(0), k \ge 0\}$ can be calculated from expression (27). The expansion of (32) is similar albeit more complex. Eq. (27) can be written as

$$P_T^{(k)}(z) = P_T^{(0)}(z)Q_k(z)$$
(38)

in this case, with

$$P_T^{(0)}(z) = \frac{1 - \lambda_T z}{1 - z} U_T^{(0)}(z),$$

$$Q_k(z) = e^{k\lambda_T(z-1)}$$

and $\lambda_T = \lambda_1 + \lambda_2$. Let $p_T^{(0)}(n)$, $u_T^{(0)}(n)$ and $q_k(n)$ denote the coefficients in the series expansion about z = 0 of $P_T^{(k)}(z)$, $U_T^{(0)}(z)$ and $Q_k(z)$ respectively. Since $U_T^{(k)}(0)$ is the coefficient of z^k in the series expansion of $P_T^{(k)}(z)$ about z = 0 and since the product of generating functions translates to a convolution of the corresponding coefficients, it follows from (38) that

$$U_T^{(k)}(0) = \sum_{n=0}^k p_T^{(0)}(n)q_k(k-n).$$
(39)

It thus remains for us to calculate $p_T^{(0)}(n)$ and $q_k(n)$ for $0 \le n \le k$. It is easily seen that the $p_T^{(0)}(n)$'s are given by

$$p_T^{(0)}(n) = u_T^{(0)}(n) + (1 - \lambda_T) \sum_{l=0}^{n-1} u_T^{(0)}(l)$$
(40)

and that the $q_k(n)$'s are given by

$$q_k(n) = e^{-k\lambda_T} \frac{(k\lambda_T)^n}{n!}, \quad n \ge 0.$$
(41)

Substituting (40) and (41) in (39) yields the sequence $\left\{ U_T^{(k)}(0), k \ge 0 \right\}$.

As already mentioned, (32) can be expanded about $z_1 = 0$ in a similar way. Note that the $U_T^{(k)}(0)$'s show up in these calculations since $U_T(x, 0)$ appears in (32).

6. Conclusions

In this paper, we studied the transient behavior of a priority queueing system. Packets of two types arrive in the system and packets of type 1 have priority over packets of type 2. Using generating functions, we analyzed the transient generating functions of the system contents of both classes at the beginning of slots. Furthermore, we showed how to calculate the moments of the transient system contents of both types and of the total system contents. We illustrated our analytic approach by means of a couple of examples.

Acknowledgments

The first two authors are Postdoctoral Fellows with the Fund for Scientific Research, Flanders (F.W.O.-Vlaanderen), Belgium. The authors wish to thank the referee for insightful comments that improved the presentation of the paper.

References

- I. Awan, Analysis of multiple-threshold queues for congestion control of heterogeneous traffic streams, Simulation Modelling Practice and Theory 14 (6) (2006) 712–724.
- [2] H. Bruneel, Exact derivation of transient behavior for buffers with random output interruptions, Computers Networks and ISDN Systems 22 (1991) 277–285.
- [3] S. Drekic, W. Grassmann, An eigenvalue approach to analyzing a finite source priority queueing model, Annals of Operations Research 112 (1–4) (2002) 139–152.
- [4] D. Fiems, J. Walraevens, H. Bruneel, Performance of a partially shared priority buffer with correlated arrivals, in: Proceedings of the 20th International Teletraffic Congress, ITC20, in: LNCS, vol. 4516, Ottawa, 2007, pp. 582–593.
- [5] J. Griffiths, G. Leonenko, J. Williams, The transient solution to M/E-k/1 queue, Operations Research Letters 34 (3) (2006) 349-354.
- [6] M. Harchol-Balter, T. Osogami, A. Scheller-Wolf, A. Wierman, Multi-server queueing systems with multiple priority classes, Queueing Systems 51 (3–4) (2005) 331–360.
- [7] F. Kamoun, Performance analysis of a discrete-time queuing system with a correlated train arrival process, Performance Evaluation 63 (4–5) (2006) 315–340.
- [8] T. Katayama, K. Kobayashi, Analysis of a nonpreemptive priority queue with exponential timer and server vacations, Performance Evaluation 64 (6) (2007) 495–506.
- [9] K. Laevens, H. Bruneel, Discrete-time multiserver queues with priorities, Performance Evaluation 33 (4) (1998) 249-275.

- [10] C. Langaris, A. Katsaros, Time-dependent analysis of a queue with batch arrivals and n levels of nonpreemptive priority, Queueing Systems 19 (3) (1995) 269–288.
- [11] J. Liebeherr, D. Wrege, Priority queue schedulers with approximate sorting in output-buffered switches, IEEE Journal on Selected Areas in Communications 17 (6) (1999) 1127–1144.
- [12] K. Liu, D. Petr, V. Frost, H. Zhu, C. Braun, W. Edwards, Design and analysis of a bandwidth management framework for ATM-based broadband ISDN, IEEE Communications Magazine (1997) 138–145.
- [13] Y. Liu, W. Gong, On fluid queueing systems with strict priority, IEEE Transactions on Automatic Control 48 (12) (2003) 2079–2088.
- [14] T. Maertens, J. Walraevens, H. Bruneel, On priority queues with priority jumps, Performance Evaluation 63 (12) (2006) 1235–1252.
- [15] M. Mehmet Ali, X. Song, A performance analysis of a discrete-time priority queueing system with correlated arrivals, Performance Evaluation 57 (3) (2004) 307–339.
- [16] A. Parekh, R. Gallager, A generalized processor sharing approach to flow control in integrated services networks: The multiple node case, IEEE/ACM Transactions on Networking 2 (2) (1994) 137–150.
- [17] P. Parthasarathy, A transient solution to an M/M/1 queue a simple approach, Advances in Applied Probability 19 (4) (1987) 997–998.
- [18] H. Takagi, Queueing Analysis: A Foundation of Performance Evaluation, Volume 1: Vacation and Priority Systems, Part 1, North-Holland, 1991.
- [19] C. Tham, Q. Yao, Y. Jiang, Achieving differentiated services through multi-class probabilistic priority scheduling, Computer Networks 40 (4) (2002) 577–593.
- [20] B. Van Houdt, C. Blondia, Analyzing priority queues with 3 classes using tree-like processes, Queueing Systems 54 (2) (2006) 99–109.
- [21] J. Walraevens, B. Steyaert, H. Bruneel, Performance analysis of a single-server ATM queue with a priority scheduling, Computers & Operations Research 30 (12) (2003) 1807–1829.
- [22] J. Walraevens, S. Wittevrongel, H. Bruneel, A discrete-time priority queue with train arrivals, Stochastic Models 23 (3) (2007) 489-512.
- [23] X. Xiao, L. Ni, Internet QoS: A big picture, IEEE Network 13 (2) (1999) 8-18.
- [24] J. Zhao, B. Li, X. Cao, I. Ahmad, A matrix-analytic solution for the DBMAP/PH/1 priority queue, Queueing Systems 53 (3) (2006) 127–145.



Joris Walraevens was born in Zottegem, Belgium, in 1974. He received the M.S. degree in Electrical Engineering and the Ph.D. degree in Engineering in 1997 and 2004 respectively, all from Ghent University, Belgium. In September 1997, he joined the SMACS Research Group, Department for Telecommunications and Information Processing, at the same university. His main research interests include discrete-time queueing models and performance analysis of heterogeneous telecommunication networks. His personal webpage can be found at http://telin.UGent.be/~jw.



Dieter Fiems was born in Ghent, Belgium, in 1973. He received an engineering degree at KAHO-St-Lieven in 1997, the post-graduate degree in Computer Science at Ghent University in 1998 and the Ph.D. degree in engineering at Ghent University in 2004. Since 1998, he is a researcher at the Department of Telecommunications and Information Processing of Ghent University, as a Member of the SMACS Research Group. His research interests include a.o. discrete-time queueing models and stochastic modeling of telecommunication networks.



Herwig Bruneel was born in Zottegem, Belgium, in 1954. He received the master's degree in Electrical Engineering, the master's degree in Computer Science, and the Ph.D. degree in Computer Science in 1978, 1979 and 1984 respectively, all from Ghent University, Belgium. He is full Professor in the Faculty of Engineering and head of the Department of Telecommunications and Information Processing at the same university. He also leads the SMACS Research Group within this department. His main personal research interests include stochastic modeling and analysis of communication systems, discrete- time queueing theory, and the study of ARQ protocols. He has published more than 300 papers on these subjects and is coauthor of the book H. Bruneel and B. G. Kim, "Discrete-Time Models for Communication Systems Including ATM" (Kluwer Academic Publishers, Boston, 1993). From October 2001 to September 2003, he served as the Academic Director for Research Affairs at Ghent University.