Analysis of a discrete-time preemptive resume priority buffer

Joris Walraevens*, Bart Steyaert and Herwig Bruneel

SMACS Research Group
Department of Telecommunications and Information Processing (IR07)
Ghent University - UGent
Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium.
Phone: +32-9-2648902, Fax: +32-9-2644295
E-mail: {jw,bs,hb}@telin.UGent.be

Abstract

In this paper, we analyze a discrete-time preemptive resume priority queue. We consider two classes of customers which have to be served, where customers of one class have preemptive resume priority over customers of the other. Both classes contain customers with generally distributed service times. We show that the use of probability generating functions is beneficial for analyzing the system contents and customer delays of both classes. It is shown (theoretically as well as by some practical procedures) how moments and approximate tail probabilities of system contents and customer delays are calculated. The influence of the priority scheduling discipline and the service time distributions on the performance measures is shown by some numerical examples.

Key words: Queueing, preemptive resume priority, general service times

1 Introduction

In this paper, we present the analysis of a discrete-time preemptive resume (PR) priority queue. Time is divided into slots and the initiation of service is synchronized with respect to slot boundaries. Customers of two classes (class-1 and class-2) arrive in a single-server queueing system and the customers of class-1 are scheduled for service with *priority* over class-2 customers. So, when the server becomes available, a class-1 customer is served next (if any).

*Corresponding author

If no class-1 customers are present, a class-2 customer starts service (if any). The scheduling type is *preemptive* which means that newly arriving class-1 customers interrupt an on-going service of a class-2 customer. Furthermore, an interrupted class-2 customer can *resume* its service upon returning in the server (after all class-1 customers have left the system).

PR priority scheduling can be applied in many areas, such as in multitasking operating systems (see e.g. [14] and references therein), call centers (see e.g. [8] and references) and telecommunications (see [4]). In multitasking operating systems, tasks that need real-time computing have PR priority over other, less urgent tasks. Examples of such systems are command and control systems, flight control systems and process control systems. In call centers, PR priority is e.g. given to answering the telephone over responding to E-mails. It is quite clear that the priority is of the PR type since the telephone has to be answered at the moment it rings and responding to the E-mail can be resumed afterwards. Finally, in telecommunications, real-time applications (such as telephony, multimedia applications, ...) have transmission priority over data-applications (ftp-sessions, E-mail,...) in packet-based networks (e.g., IP (Internet Protocol) based networks). On the link-layer, non-preemptive priority is given to real-time packets. PR priority models however arise when the link layer implements link level fragmentation. E.g. when IP over ATM (Asynchronous Transfer Mode) is applied, an IP packet consist of a number of ATM cells and the ATM cells are scheduled in a non-preemptive priority fashion (on the lower link layer). On the higher IP layer, IP packets are scheduled in a PR priority fashion, or in other words, arriving high-priority IP-packets interrupt the transmission of the ATM cells of a low-priority IP packet. Since one wants to analyze the QoS (Quality of Service) on the higher layer, the PR priority model is an accurate model in this case.

In the literature, there have been a number of contributions with respect to priority scheduling. An overview of some basic priority queueing models can be found in [7], [13] and [16] and the references therein. In [6], [9], [12], [15], [19] and [22] discrete-time priority queues with deterministic service times equal to one slot are studied. Furthermore, preemptive resume priority queues have been analyzed in [10], [17], [18], [23] and [24]. Machihara [10] analyzes waiting times when high-priority arrivals are distributed according to a Markovian arrival process. In [17], the delay of the low-priority customers is analyzed using probability generating functions. Takine and Hasegawa [18] study the waiting times of customers arriving to a queue according to independent Markovian arrival processes. Finally in [23] and [24], we have analyzed the system contents and customer delay when the service times of high-priority customers are geometrically and generally distributed respectively and the service times of low-priority customers are geometrically distributed.

In this paper, we analyze the system contents and customer delays of high-priority and low-priority customers in a discrete-time single-server buffer for a preemptive resume priority scheme and per-slot i.i.d. (independent and identically distributed) numbers of arrivals. The service times of the customers are assumed to be generally distributed. These distributions

are class-dependent, i.e., the service times of the high-priority customers can have a different (common) distribution from the service times of the low-priority customers.

The contribution of this paper concerns the model that is considered, the solution technique that is used, as well as the results that are generated. First, as far as the model is concerned, the arrival processes of the different types of customers are not mutually independent. Note that this arrival model was already used in [24] as well. This correlation between the high- and low-priority arrival process is however ignored in most other papers on priority queues. As a result of this correlation in the arrival process, the different classes can not be analyzed separately (i.e., as a model with server interruptions for low-priority customers - see [5]), which complicates the analysis. Furthermore, the service times of both classes are generally distributed. Many specific distributions of customers' service times of both classes are thus incorporated in this model. In our previous papers, at least one of the two classes was assumed to consist of customers with geometrically distributed service times. In these cases, we made use of the memoryless property of the geometric distribution, or more precisely, the property that the probability of a customer being served needing another slot of service is independent of the amount of slots that it is already being served at that time instant. As a consequence, it was not necessary to keep track of the amount of slots that a service was already taking place. This property is even more significant for the low-priority service times since the analysis does not have to keep track of the number of slots the customer is already being served when interrupted by high-priority customers. A *general* distribution for the service times of both classes (and especially for the low-priority service times) obviously highly complicates the analysis. Since most variables in practice are not geometrically distributed, this extension to generally distributed service times is a necessary one to obtain accurate results. In the paper, we will show some examples where the results are highly influenced by the distribution of the class-2 service times. Note further that the distributions of the service times can be different for the two classes. We can conclude that the model is thus quite general and many applications and or arrival and service patterns fit into the model. Analyzing this rather complex model by using probability generating functions is a key result of this paper. In particular we (had to) calculate a 4-dimensional pgf to be able to analyze the system contents and customer delays. Finally, determining the (tail) distributions of the system contents and customer delays is also a main contribution of the paper. Note that this paper is based on the results presented in [21] where this tail behavior was not yet analyzed. It turns out that the tail behavior of the low-priority variables may be non-geometric.

The remainder of this paper is structured as follows. In the following section, we present the mathematical model. In sections 3 and 4, we will then analyze the steady-state system contents and customer delays of both classes. In section 5, we show how to theoretically and practically calculate moments and (approximate) tail probabilities of these stochastic variables. Some numerical examples are treated in section 6. Finally, some conclusions are formulated in section 7.

2 Mathematical model

We consider a discrete-time single-server queueing system with infinite buffer space. Time is assumed to be slotted. There are two types of customers arriving to the system, namely customers of class-1 and customers of class-2. The numbers of per-slot arrivals are i.i.d. The numbers of class-j arrivals during slot k are denoted by $a_{j,k}$ $(j = 1, 2)$. Their joint probability generating function (pgf) is defined as

$$A(z_1, z_2) \triangleq \mathrm{E}[z_1^{a_{1,k}} z_2^{a_{2,k}}].$$

Note that the number of arrivals of both classes can be correlated during one slot. The marginal pgf of the number of arrivals of class-j is denoted by $A_j(z)$ $(j = 1, 2)$ and they are given by $A(z, 1)$ and $A(1, z)$ respectively. The total number of arrivals during slot k is denoted by $a_{T,k} \triangleq a_{1,k} + a_{2,k}$ and its pgf is given by $A_T(z) = A(z, z)$. We will furthermore denote the mean arrival rate of class-j customers during a slot by $\lambda_j \triangleq \mathrm{E}[a_{j,k}] = A_j'(1)$ $(j = 1, 2)$.

The service times of the class-j customers, i.e., the number of slots a class-j customer is effectively being served, are i.i.d. and generally distributed and their pgf is denoted by $S_j(z)$ $(j = 1, 2)$. The mean service time of a class-j customer is denoted by $\mu_j = S_j'(1)$ $(j = 1, 2)$.

The class-1 customers are assumed to have preemptive resume priority over the class-2 customers and within one class the scheduling is FCFS (First-Come-First-Served). Finally, the load offered by class-j customers is given by $\rho_j \triangleq \lambda_j \mu_j$. The total load is then given by $\rho_T \triangleq \rho_1 + \rho_2$. We assume a stable system, i.e., $\rho_T < 1$.

3 System contents

In this section, we analyze the system contents. We denote the system contents of class-1 customers and class-2 customers at the beginning of slot k by $u_{1,k}$ and $u_{2,k}$ respectively. Their joint pgf is defined as

$$U_k(z_1, z_2) \triangleq \mathrm{E}\left[z_1^{u_{1,k}} z_2^{u_{2,k}}\right].$$

Since the service times are generally distributed, the set $\{(u_{1,k}, u_{2,k}), k \geq 1\}$ does not form a Markov chain. Therefore, we introduce two new stochastic variables $r_{j,k}$ $(j = 1, 2)$ as follows: $r_{1,k}$ indicates the remaining number of slots needed to serve the class-1 customer in service at the beginning of slot k, if $u_{1,k} > 0$, and $r_{1,k} = 0$ if $u_{1,k} = 0$; $r_{2,k}$ indicates the remaining number of slots service time of the "oldest" class-2 customer - i.e., the class-2 customer longest present in the system - at the beginning of slot k, if $u_{2,k} > 0$, and $r_{2,k} = 0$ if $u_{2,k} = 0$. With this definition, $\{(r_{1,k}, u_{1,k}, r_{2,k}, u_{2,k}), k \geq 1\}$ is easily seen to constitute a Markovian state description of the system at the beginning of slot k. If $s_{j,k}^*$ $(j = 1, 2)$ indicates the service time of the next class-j customer to receive service at the beginning of slot k, the following system equations

can be established:

1. If $r_{1,k} = 0$ (and hence $u_{1,k} = 0$):
   (a) If $r_{2,k} = 0$ (and hence $u_{2,k} = 0$):

$$u_{j,k+1} = a_{j,k} \quad ; \quad r_{j,k+1} = \begin{cases} 0 & \text{if } a_{j,k} = 0 \\ s^*_{j,k} & \text{if } a_{j,k} > 0 \end{cases},$$

   with $j = 1, 2$. The only customers present in the system at the beginning of slot $k+1$ are the customers that arrived during the previous slot. If there have been new arrivals of class-j customers during slot k, the remaining number of slots needed to serve the first class-j customer is that customer's full service time.
   (b) If $r_{2,k} = 1$:

$$\begin{array}{l} u_{1,k+1} = a_{1,k} \\ r_{1,k+1} = \begin{cases} 0 & \text{if } a_{1,k} = 0 \\ s^*_{1,k} & \text{if } a_{1,k} > 0 \end{cases} \end{array} \quad ; \quad \begin{array}{l} u_{2,k+1} = u_{2,k} - 1 + a_{2,k}; \\ r_{2,k+1} = \begin{cases} 0 & \text{if } u_{2,k} - 1 + a_{2,k} = 0 \\ s^*_{2,k} & \text{if } u_{2,k} - 1 + a_{2,k} > 0 \end{cases} \end{array},$$

   i.e., the class-2 customer in service at the beginning of slot k leaves the system at the end of slot k.
   (c) If $r_{2,k} > 1$:

$$\begin{array}{l} u_{1,k+1} = a_{1,k} \\ r_{1,k+1} = \begin{cases} 0 & \text{if } a_{1,k} = 0 \\ s^*_{1,k} & \text{if } a_{1,k} > 0 \end{cases} \end{array} \quad ; \quad \begin{array}{l} u_{2,k+1} = u_{2,k} + a_{2,k}; \\ r_{2,k+1} = r_{2,k} - 1, \end{array}$$

   i.e., the class-2 customer in service at the beginning of slot k remains in the system (not necessarily in the server - it only remains in the server if there are no new class-1 arrivals, due to the priority scheduling). Its remaining service time is decreased by one.

2. If $r_{1,k} = 1$:
   (a) If $r_{2,k} = 0$ (and hence $u_{2,k} = 0$):

$$\begin{array}{l} u_{1,k+1} = u_{1,k} - 1 + a_{1,k} \\ r_{1,k+1} = \begin{cases} 0 & \text{if } u_{1,k} - 1 + a_{1,k} = 0 \\ s^*_{1,k} & \text{if } u_{1,k} - 1 + a_{1,k} > 0 \end{cases} \end{array} \quad ; \quad \begin{array}{l} u_{2,k+1} = a_{2,k}; \\ r_{2,k+1} = \begin{cases} 0 & \text{if } a_{2,k} = 0 \\ s^*_{2,k} & \text{if } a_{2,k} > 0 \end{cases} \end{array},$$

   i.e., the class-1 customer in service at the beginning of slot k, leaves the system at the end of slot k. There were no class-2 customers in the system at the beginning of slot k.

(b) If $r_{2,k} > 0$:

$$u_{1,k+1} = u_{1,k} - 1 + a_{1,k} \quad ; \quad u_{2,k+1} = u_{2,k} + a_{2,k};$$
$$r_{1,k+1} = \begin{cases} 0 & \text{if } u_{1,k} - 1 + a_{1,k} = 0 \\ s^*_{1,k} & \text{if } u_{1,k} - 1 + a_{1,k} > 0 \end{cases} \quad ; \quad r_{2,k+1} = r_{2,k},$$

i.e., the class-1 customer in service at the beginning of slot k, leaves the system at the end of slot k. The remaining service of the oldest class-2 customer stays the same.

3. If $r_{1,k} > 1$:

(a) If $r_{2,k} = 0$ (and hence $u_{2,k} = 0$):

$$u_{1,k+1} = u_{1,k} + a_{1,k} \quad ; \quad u_{2,k+1} = a_{2,k};$$
$$r_{1,k+1} = r_{1,k} - 1 \quad ; \quad r_{2,k+1} = \begin{cases} 0 & \text{if } a_{2,k} = 0 \\ s^*_{2,k} & \text{if } a_{2,k} > 0 \end{cases},$$

i.e., the class-1 customer in service at the beginning of slot k stays in the server at the beginning of slot $k+1$. Its remaining service time is decreased by one. There were no class-2 customers in the system at the beginning of slot k.

(b) If $r_{2,k} > 0$:

$$u_{1,k+1} = u_{1,k} + a_{1,k} \quad ; \quad u_{2,k+1} = u_{2,k} + a_{2,k};$$
$$r_{1,k+1} = r_{1,k} - 1 \quad ; \quad r_{2,k+1} = r_{2,k}.$$

The difference with the previous case is that there was at least one class-2 customer in the system at the beginning of slot k.

Now, let us define $P_k(x_1, z_1, x_2, z_2)$ as the joint pgf of the state vector $(r_{1,k}, u_{1,k}, r_{2,k}, u_{2,k})$:

$$P_k(x_1, z_1, x_2, z_2) \triangleq E[x_1^{r_{1,k}} z_1^{u_{1,k}} x_2^{r_{2,k}} z_2^{u_{2,k}}].$$

Taking the z-transform of the system equations, we find a relation between $P_k(x_1, z_1, x_2, z_2)$ and $P_{k+1}(x_1, z_1, x_2, z_2)$:

$$\begin{aligned}
&P_{k+1}(x_1, z_1, x_2, z_2) \\
&= [A(0,0) + (A(0,z_2) - A(0,0))S_2(x_2) + (A(z_1,0) - A(0,0))S_1(x_1) \\
&+ (A(z_1,z_2) - A(z_1,0) - A(0,z_2) + A(0,0))S_1(x_1)S_2(x_2)]P_k(0,0,0,0) + A(0,0)R_{2,k}(0) \\
&+ [A(0,z_2)R_{2,k}(z_2) - A(0,0)R_{2,k}(0)]\, S_2(x_2) + (A(z_1,0) - A(0,0))R_{2,k}(0)S_1(x_1) \\
&+ [(A(z_1,z_2) - A(0,z_2))R_{2,k}(z_2) - (A(z_1,0) - A(0,0))R_{2,k}(0)]\, S_1(x_1)S_2(x_2) \\
&+ \frac{A(0,z_2) + (A(z_1,z_2) - A(0,z_2)S_1(x_1)}{x_2} [P_k(0,x_2,0,z_2) - x_2 z_2 R_{2,k}(z_2) - P_k(0,0,0,0)]
\end{aligned}$$

$$
\begin{aligned}
&+\left[A(0,0)+(A(0,z_2)-A(0,0))S_2(x_2)\right]R_{1,k}(0,0,0) \qquad (1)\\
&+[A(z_1,0)R_{1,k}(z_1,0,0)-A(0,0)R_{1,k}(0,0,0)]S_1(x_1)\\
&+[(A(z_1,z_2)-A(z_1,0))R_{1,k}(z_1,0,0)-(A(0,z_2)-A(0,0))R_{1,k}(0,0,0)]S_1(x_1)S_2(x_2)\\
&+A(0,z_2)[R_{1,k}(0,x_2,z_2)-R_{1,k}(0,0,0)]\\
&+[A(z_1,z_2)(R_{1,k}(z_1,x_2,z_2)-R_{1,k}(z_1,0,0))+A(0,z_2)(R_{1,k}(0,x_2,z_2)-R_{1,k}(0,0,0))]S_1(x_1)\\
&+\frac{A(z_1,0)+(A(z_1,z_2)-A(z_1,0)S_2(x_2)}{x_1}\left[P_k(x_1,z_1,0,0)-x_1z_1R_{1,k}(z_1,0,0)-P_k(0,0,0,0)\right]\\
&+\frac{A(z_1,z_2)}{x_1}[(P_k(x_1,z_1,x_2,z_2)-P_k(x_1,z_1,0,0))-x_1z_1(R_{1,k}(z_1,x_2,z_2)-R_{1,k}(z_1,0,0))\\
&-(P_k(0,0,x_2,z_2)-P_k(0,0,0,0))],
\end{aligned}
$$

where the functions $R_{1,k}(z_1,x_2,z_2)$ and $R_{2,k}(z_2)$ are defined as

$$
\begin{aligned}
R_{1,k}(z_1,x_2,z_2) &\triangleq \mathrm{E}\left[z_1^{u_{1,k}-1}x_2^{r_{2,k}}z_2^{u_{2,k}}1\{r_{1,k}=1\}\right];\\
R_{2,k}(z_2) &\triangleq \mathrm{E}\left[z_2^{u_{2,k}-1}1\{r_{1,k}=u_{1,k}=0,r_{2,k}=1\}\right],
\end{aligned}
$$

with $1\{X\}$ the indicator function of X. Note that $R_{1,k}(z_1,x_2,z_2)$ is the partial pgf of the class-1 *queue* content (i.e. not counting the customer in service), the residual class-2 service time and the class-2 queue content at the beginning of slot k given that a class-1 customer is being served during that slot and given that this customer leaves the system at the end of that slot. Analogously, $R_{2,k}(z_2)$ is the partial pgf of the class-2 *queue* content at the beginning of slot k given that a class-2 customer is being served during slot k and this customer leaves the system at the end of that slot.

We assume that the system is stable (implying that the equilibrium condition $\rho_T<1$ is met) and as a result $P_k(x_1,z_1,x_2,z_2)$ and $P_{k+1}(x_1,z_1,x_2,z_2)$ converge both to a common steady-state value $P(x_1,z_1,x_2,z_2)=\lim_{k\to\infty}P_k(x_1,z_1,x_2,z_2)$. By taking the $k\to\infty$ limit in (1) we obtain (taking into account the statistical independence of the random variables $(r_{1,k},u_{1,k},r_{2,k},u_{2,k})$, $(a_{1,k},a_{2,k})$, s_1^* and s_2^* respectively):

$$
\begin{aligned}
&[x_1-A(z_1,z_2)]P(x_1,z_1,x_2,z_2)\\
&\qquad=\Bigg[x_1A(0,0)(1-S_1(x_1))(1-S_2(x_2))+\frac{x_1}{x_2}A(0,z_2)(1-S_1(x_1))(x_2S_2(x_2)-1)\\
&\qquad\quad+A(z_1,0)(x_1S_1(x_1)-1)(1-S_2(x_2))\\
&\qquad\quad+\frac{1}{x_2}A(z_1,z_2)(x_1x_2S_1(x_1)S_2(x_2)-x_1S_1(x_1)-x_2S_2(x_2)+x_2)\Bigg]P(0,0,0,0)\\
&\qquad\quad+x_1[A(0,0)(1-S_1(x_1))+A(z_1,0)S_1(x_1)](1-S_2(x_2))R_2(0)\\
&\qquad\quad+x_1(A(0,z_2)-A(0,0))(1-S_1(x_1))(S_2(x_2)-1)R_1(0,0,0)
\end{aligned}
$$

$$
\begin{aligned}
&+ (A(z_1,z_2) - A(z_1,0))(S_2(x_2)-1)P(x_1,z_1,0,0) \qquad (2)\\
&+ x_1(A(z_1,z_2) - A(z_1,0))(z_1 - S_1(x_1))(1-S_2(x_2))R_1(z_1,0,0)\\
&+ \frac{1}{x_2}[x_1A(0,z_2)(1-S_1(x_1)) + A(z_1,z_2)(x_1S_1(x_1)-x_2)]P(0,0,x_2,z_2)\\
&+ x_1[A(0,z_2)(1-S_1(x_1)) + A(z_1,z_2)S_1(x_1)](S_2(x_2)-z_2)R_2(z_2)\\
&+ x_1A(0,z_2)(1-S_1(x_1))R_1(0,x_2,z_2) + x_1A(z_1,z_2)(S_1(x_1)-z_1)R_1(z_1,x_2,z_2),
\end{aligned}
$$

with $R_1(z_1,x_2,z_2) \triangleq \lim_{k\to\infty} R_{1,k}(z_1,x_2,z_2)$ and $R_2(z_2) \triangleq \lim_{k\to\infty} R_{2,k}(z_2)$. It now remains for us to determine the unknown boundary functions $P(x_1,z_1,0,0)$, $P(0,0,x_2,z_2)$, $R_2(z_2)$, $R_1(z_1,0,0)$, $R_1(0,x_2,z_2)$, $R_1(z_1,x_2,z_2)$ and the unknown constants $P(0,0,0,0)$, $R_2(0)$ and $R_1(0,0,0)$. This can be done in a few steps which are summarized as follows:

1. (a) Firstly, due to the fact that $r_{j,k} = 0$ $(j = 1,2)$ if and only if $u_{j,k} = 0$, the joint pgfs that were defined above must satisfy $R_1(z,x_2,0) = R_1(z,0,0)$, $P(x_1,z,x_2,0) = P(x_1,z,0,0)$ and $P(x_1,0,x_2,z) = P(0,0,x_2,z)$ for all x_j and z, $|x_j| \le 1$ and $|z| \le 1$. Invoking these properties in (2) eventually leads to the following formulas for $P(x_1,z_1,0,0)$ and $P(0,0,x_2,z_2)$:

$$
\begin{aligned}
P(x_1,z_1,0,0) &= \frac{1}{x_1 - A(z_1,0)}\Big\{[x_1(1-S_1(x_1)) + A(z_1,0)(x_1S_1(x_1)-1)]P(0,0,0,0)\\
&\quad + x_1A(z_1,0)S_1(x_1)R_2(0) + x_1A(z_1,0)(S_1(x_1)-z_1)R_1(z_1,0,0)\Big\}; \quad (3)\\
P(0,0,x_2,z_2) &= \frac{1}{x_2 - A(0,z_2)}\Big\{[x_2(1-S_2(x_2)) + A(0,z_2)(x_2S_2(x_2)-1)]P(0,0,0,0)\\
&\quad + x_2A(0,z_2)(S_2(x_2)-1)R_1(0,0,0) + x_2A(0,z_2)(S_2(x_2)-z_2)R_2(z_2)\\
&\quad + x_2A(0,z_2)R_1(0,x_2,z_2)\Big\}. \quad (4)
\end{aligned}
$$

   (b) The identities in (a) are in particular valid for $z = 0$, leading to the additional identities $R_1(0,x_2,0) = R_1(0,0,0)$ and $P(x_1,0,x_2,0) = P(0,0,0,0)$. These identities constitute the following relation between the three constants $R_1(0,0,0)$, $R_2(0)$ and $P(0,0,0,0)$:

$$
P(0,0,0,0) = A(0,0)\left[P(0,0,0,0) + R_2(0) + R_1(0,0,0)\right]. \qquad (5)
$$

2. The remaining derivations heavily rely on the observation that the respective (joint) pgfs are bounded inside the complex unit disk, and consist of the following steps:

   (a) The function $P(x_1,z_1,x_2,z_2)$ is bounded for $x_1 = A(z_1,z_2)$, $|x_2| \le 1$ and $|z_j| \le 1$ $(j = 1,2)$, since $|A(z_1,z_2)| \le 1$ for all such x_2 and z_j. This implies that the left-hand side of (2) vanishes when substituting x_1 by $A(z_1,z_2)$. This leads to the following equation for $R_1(z_1,x_2,z_2)$ (by also substituting $P(x_1,z_1,0,0)$ and $P(0,0,x_2,z_2)$ by

their expressions obtained in step 1):

$$\begin{aligned}
&A(z_1,z_2)(x_2-A(0,z_2))(z_1-S_1(A(z_1,z_2)))[(S_2(x_2)-1)R_1(z_1,0,0)+R_1(z_1,x_2,z_2)]\\
&\quad=(A(z_1,z_2)-A(0,z_2))S_1(A(z_1,z_2))S_2(x_2)(x_2-1)P(0,0,0,0)\\
&\qquad+A(0,z_2)(A(z_1,z_2)-x_2)S_1(A(z_1,z_2))(S_2(x_2)-1)R_1(0,0,0)\\
&\qquad+x_2(A(z_1,z_2)-A(0,z_2))S_1(A(z_1,z_2))(S_2(x_2)-z_2)R_2(z_2)\\
&\qquad+A(0,z_2)(A(z_1,z_2)-x_2)S_1(A(z_1,z_2))R_1(0,x_2,z_2).
\end{aligned}\tag{6}$$

(b) It can be proved by means of Rouché's theorem that $z_1 = S_1(A(z_1,z_2))$ has exactly one solution for z_1 with $|z_1| \leq 1$ for all z_2 with $|z_2| \leq 1$. This solution is denoted by $Y(z_2)$. The above implies that if we insert $z_1 = Y(z_2)$ in equation (6), where $|z_2| \leq 1$, the left-hand side of this equation vanishes, yielding

$$\begin{aligned}
&A(0,z_2)(x_2-A(Y(z_2),z_2))[(S_2(x_2)-1)R_1(0,0,0)+R_1(0,x_2,z_2)] \qquad (7)\\
=\;&(A(Y(z_2),z_2)-A(0,z_2))[S_2(x_2)(x_2-1)P(0,0,0,0)+x_2(S_2(x_2)-z_2)R_2(z_2)].
\end{aligned}$$

(c) Finally, substituting x_2 by $A(Y(z_2),z_2)$ in (7) provides the following relation for $R_2(z_2)$:

$$R_2(z_2) = P(0,0,0,0)\frac{S_2(A(Y(z_2),z_2))(A(Y(z_2),z_2)-1)}{A(Y(z_2),z_2)(z_2-S_2(A(Y(z_2),z_2)))}. \tag{8}$$

By using the expressions obtained in this step, the unknown functions and constants are further reduced to $P(x_1,z_1,0,0)$, $R_1(z_1,0,0)$, $R_1(0,0,0)$, $R_2(0)$ and $P(0,0,0,0)$. We further already have relation (5) between the three constants.

3. These still unknown functions and constants can be obtained by repeating steps 2(a) and 2(b) in the special case when $z_2 = 0$.

(a) Substituting x_1 by $A(z_1,0)$ in expression (3) leads to

$$A(z_1,0)R_1(z_1,0,0) = \frac{S_1(A(z_1,0))}{z_1-S_1(A(z_1,0))}[(A(z_1,0)-1)P(0,0,0,0)+A(z_1,0)R_2(0)]. \tag{9}$$

(b) Substituting z_1 by $Y(0)$ in (9) yields

$$R_2(0) = P(0,0,0,0)\frac{1-A(Y(0),0)}{A(Y(0),0)}. \tag{10}$$

4. An almost fully determined version for $P(x_1,z_1,x_2,z_2)$ can then be derived by substituting all functions and constants found throughout this procedure in expression (2). The only still unknown is $P(0,0,0,0)$. This constant can be found by using the normalisation condition $P(1,1,1,1) = 1$. Doing so, we obtain the expected result for the probability of an empty system, i.e., $P(0,0,0,0) = 1-\rho_T$.

We finally find the following expression for $P(x_1, z_1, x_2, z_2)$:

$$\begin{aligned}
&P(x_1,z_1, x_2, z_2)\\
&\quad=(1-\rho_T)\left[1+\frac{x_1z_1(A(z_1,0)-A(Y(0),0))(S_1(x_1)-S_1(A(z_1,0)))(1-S_2(x_2))}{A(Y(0),0)(x_1-A(z_1,0))(z_1-S_1(A(z_1,0)))}\right.\\
&\quad+x_1z_1\frac{(A(z_1,z_2)-A(Y(z_2),z_2))(S_1(x_1)-S_1(A(z_1,z_2)))}{(x_1-A(z_1,z_2))(z_1-S_1(A(z_1,z_2)))(z_2-S_2(A(Y(z_2),z_2)))}\\
&\quad\left\{\frac{S_2(A(Y(z_2),z_2))(z_2-S_2(x_2))}{A(Y(z_2),z_2)}-z_2\frac{(1-x_2)(S_2(x_2)-S_2(A(Y(z_2),z_2)))}{x_2-A(Y(z_2),z_2)}\right\}\\
&\quad\left.-x_2z_2\frac{(1-A(Y(z_2),z_2))(S_2(x_2)-S_2(A(Y(z_2),z_2)))}{(x_2-A(Y(z_2),z_2))(z_2-S_2(A(Y(z_2),z_2)))}\right].
\end{aligned}\tag{11}$$

From this pgf, several joint and marginal pgfs can be calculated. First, we calculate the joint pgf of the system contents of class-1 customers and the remaining service time of the class-1 customer in service at the beginning of an arbitrary slot in steady-state:

$$\begin{aligned}
P_1(x,z) &\triangleq \lim_{k\to\infty}\mathrm{E}\left[x^{r_{1,k}}z^{u_{1,k}}\right]=P(x,z,1,1)\\
&= (1-\rho_1)\left[1-xz\frac{(1-A_1(z))(S_1(x)-S_1(A_1(z)))}{(x-A_1(z))(z-S_1(A_1(z)))}\right].
\end{aligned}$$

This joint pgf is independent of the amount of class-2 customers, due to the preemptive priority scheduling discipline. For class-1 customers it is as if they are the only customers in the system. This pgf was also already calculated in [1], wherein a single-class GI-G-1 buffer is analyzed. Secondly, we also calculate the joint pgf of the system contents of class-2 customers and the remaining service time of the oldest class-2 customer in the system at the beginning of a randomly chosen slot (note that this oldest class-2 customer is not necessarily in service) from equation (11), yielding

$$\begin{aligned}
P_2(x,z) &\triangleq \lim_{k\to\infty}\mathrm{E}\left[x^{r_{2,k}}z^{u_{2,k}}\right]=P(1,1,x,z)\\
&= (1-\rho_T)\left[\frac{A_2(0)(1-A(Y(0),0))-(A_2(0)-A(Y(0),0))S_2(x)}{A(Y(0),0)(1-A_2(0))}\right.\\
&\quad+\frac{S_2(A(Y(z),z))(A_2(z)-A(Y(z),z))(z-S_2(x))}{A(Y(z),z)(1-A_2(z))(z-S_2(A(Y(z),z)))}\\
&\quad+z\frac{(x-1)(A_2(z)-A(Y(z),z))(S_2(x)-S_2(A(Y(z),z)))}{(1-A_2(z))(x-A(Y(z),z))(z-S_2(A(Y(z),z)))}\\
&\quad\left.+xz\frac{(A(Y(z),z)-1)(S_2(x)-S_2(A(Y(z),z)))}{(x-A(Y(z),z))(z-S_2(A(Y(z),z)))}\right].
\end{aligned}$$

Thirdly, and most importantly, we can calculate the joint pgf of the system contents of class-1 and class-2 customers from equation (11). It is given by:

$$\begin{aligned} U(z_1,z_2) &\triangleq \lim_{k\to\infty} \mathrm{E}\left[z_1^{u_{1,k}} z_2^{u_{2,k}}\right] = P(1,z_1,1,z_2) \\ &= (1-\rho_T)\frac{S_2(A(Y(z_2),z_2))(z_2-1)}{z_2 - S_2(A(Y(z_2),z_2))} \\ &\quad \times \left[1 + z_1 \frac{(A(z_1,z_2)-A(Y(z_2),z_2))(S_1(A(z_1,z_2))-1)}{A(Y(z_2),z_2)(A(z_1,z_2)-1)(z_1 - S_1(A(z_1,z_2)))}\right]. \end{aligned} \tag{12}$$

If we assume $S_2(z) = \dfrac{(1-\beta_2)z}{1-\beta_2 z}$, i.e., the special case of geometrical service times for the low-priority class, we obtain the same equation as found in [24], as expected. From the two-dimensional pgf $U(z_1,z_2)$, we can easily derive an expression for the pgf of the total system contents at the beginning of an arbitrary slot - denoted by $U_T(z)$ - yielding

$$\begin{aligned} U_T(z) &\triangleq \lim_{k\to\infty} \mathrm{E}\left[z^{u_{T,k}}\right] = U(z,z) \\ &= (1-\rho_T)\frac{S_2(A(Y(z),z))(z-1)}{z - S_2(A(Y(z),z))} \\ &\quad \times \left[1 + z \frac{(A_T(z)-A(Y(z),z))(S_1(A_T(z))-1)}{A(Y(z),z)(A_T(z)-1)(z - S_1(A_T(z)))}\right]. \end{aligned}$$

We can also derive the expressions for the pgf of the system contents of class-1 customers and class-2 customers at the beginning of an arbitrary slot from expression (12), yielding

$$\begin{aligned} U_1(z) &\triangleq \lim_{k\to\infty} \mathrm{E}\left[z^{u_{1,k}}\right] = U(z,1) \\ &= (1-\rho_1)\frac{S_1(A_1(z))(z-1)}{z - S_1(A_1(z))}; \end{aligned} \tag{13}$$

$$\begin{aligned} U_2(z) &\triangleq \lim_{k\to\infty} \mathrm{E}\left[z^{u_{2,k}}\right] = U(1,z) \\ &= (1-\rho_T)\frac{A_2(z)}{A(Y(z),z)}\frac{1-A(Y(z),z)}{1-A_2(z)}\frac{S_2(A(Y(z),z)(z-1)}{z - S_2(A(Y(z),z))}. \end{aligned} \tag{14}$$

4 Delay

The customer delay is defined as the total amount of time a customer spends in the system, or more precisely, the number of slots between the end of the customer's arrival slot and the end of its departure slot. We can analyze the customer delay of class-1 customers as if they are the only customers in the system. This is e.g. done in [2] and the pgf of the customer delay of class-1 customers is given by

$$D_1(z) = \frac{1-\rho_1}{\lambda_1}\frac{S_1(z)(z-1)}{z - A_1(S_1(z))}\frac{1-A_1(S_1(z))}{1-S_1(z)}. \tag{15}$$

Because of the priority discipline, the analysis of the class-2 delay is more involved. We tag a class-2 customer that enters the buffer during slot k. Let us refer to the customers in the system at the end of slot k, but that have to be served before the tagged customer as the "primary customers". So, basically, the tagged class-2 customer can enter the server, when all primary customers and all class-1 customers that arrived after slot k are transmitted. In order to analyze the delay of the tagged class-2 customer, the number of class-1 customers and class-2 customers that are served between the arrival slot of the tagged class-2 customer and its departure slot is important, not the precise order in which they are served. Therefore, in order to facilitate the analysis, we will consider an equivalent virtual system with an altered service discipline. We assume that from slot $k+1$ on, the order of service for class-1 customers (those in the queue at the end of slot k and newly arriving ones) is LCFS (Last-Come-First-Served) instead of FCFS in the equivalent system (the service order of class-2 customers remains FCFS). So, a primary customer can enter the server, when the system becomes free (for the first time) of class-1 customers that arrived during and after the service time of the primary customer that preceded it according to the new service discipline. Let $v_{1,m}^{(i)}$ denote the length of the time period during which the server is occupied by the m-th class-1 customer that arrives during slot i and its class-1 "successors", i.e., the time period starting at the beginning of the service of that customer and terminating when the system becomes free (for the first time) of class-1 customers which arrived during and after its service time. Analogously, let $v_{2,m}^{(i)}$ denote the length of the time period during which the server is occupied by the m-th class-2 customer that arrives during slot i and its class-1 "successors". The $v_{j,m}^{(i)}$'s ($j = 1, 2$) are called sub-busy periods, caused by the m-th class-j customer that arrived during slot i. The service time of the tagged class-2 customer is denoted by s_2^*. We further denote the delay of the tagged class-2 customer by d_2.

When the tagged class-2 customer arrives, the system is in one of the following states:

1. $r_{1,k} = 0$ (and hence $u_{1,k} = 0$):
   (a) $r_{2,k} = 0$ (and hence $u_{2,k} = 0$):

$$d_2 = \sum_{j=1}^{2} \sum_{m=1}^{f_{j,k}} v_{j,m}^{(k)} + s_2^* + \sum_{i=1}^{s_2^*-1} \sum_{m=1}^{a_{1,l_i}} v_{1,m}^{(l_i)}, \tag{16}$$

with $f_{j,k}$ defined as the number of class-j customers arriving during slot k, but that have to be served before the tagged customer. Slots l_i are defined as the slots during which the tagged customer receives service ($i = 1, \ldots, s_2^*$). $f_{1,k}$ class-1 primary customers and $f_{2,k}$ class-2 primary customers that arrived during slot k and their class-1 successors have to be served before the tagged class-2 customer. During the service time of the tagged class-2 customer, new class-1 customers may arrive and interrupt the tagged customer's service. The last two terms take this part of the delay into

account.

(b) $r_{2,k} > 0$:

$$d_2 = (r_{2,k}-1) + \sum_{i=1}^{r_{2,k}-1} \sum_{m=1}^{a_{1,k+i}} v_{1,m}^{(n_i)} + \sum_{j=1}^{2} \sum_{m=1}^{f_{j,k}} v_{j,m}^{(k)} + \sum_{m=1}^{u_{2,k}-1} \tilde{v}_{2,m} + s_2^* + \sum_{i=1}^{s_2^*-1} \sum_{m=1}^{a_{1,l_i}} v_{1,m}^{(l_i)}, \tag{17}$$

with the n_i-th slots ($i = 1, \ldots, r_{2,k}-1$) the slots that the oldest class-2 customer receives service and the $\tilde{v}_{2,m}$'s are defined as the sub-busy periods, caused by the m-th class-2 customer already in the queue at the beginning of slot k. The residual service time of the customer in service during slot k contributes in the first term, the sub-busy periods of the class-1 customers arriving during the residual service time contribute in the second term, the sub-busy periods of the class-1 and class-2 customers arriving during slot k, but that have to be served before the tagged class-2 customer contribute in the third term, the sub-busy periods of the class-2 customers already in the queue at the beginning of slot k contribute in the fourth term and finally the service time of the tagged class-2 customer itself and the sub-busy periods of the class-1 customers arriving during this service time (except for its last slot) contribute in the last two terms.

2. $r_{1,k} > 0$:

(a) $r_{2,k} = 0$ (and hence $u_{2,k} = 0$):

$$d_2 = (r_{1,k}-1) + \sum_{i=1}^{r_{1,k}-1} \sum_{m=1}^{a_{1,k+i}} v_{1,m}^{(k+i)} + \sum_{j=1}^{2} \sum_{m=1}^{f_{j,k}} v_{j,m}^{(k)} + \sum_{m=1}^{u_{1,k}-1} \tilde{v}_{1,m} + s_2^* + \sum_{i=1}^{s_2^*-1} \sum_{m=1}^{a_{1,l_i}} v_{1,m}^{(l_i)}, \tag{18}$$

with $\tilde{v}_{1,m}$ the sub-busy period, caused by the m-th class-1 customer already in the queue at the beginning of slot k. The expression is almost the same as in the previous case, with the difference that in this case a class-1 customer was being served during slot k.

(b) $r_{2,k} > 0$:

$$d_2 = (r_{1,k}-1) + \sum_{i=1}^{r_{1,k}-1} \sum_{m=1}^{a_{1,k+i}} v_{1,m}^{(k+i)} + \sum_{j=1}^{2} \sum_{m=1}^{f_{j,k}} v_{j,m}^{(k)} + \sum_{m=1}^{u_{1,k}-1} \tilde{v}_{1,m} \tag{19}$$

$$+r_{2,k}+\sum_{i=1}^{r_{2,k}}\sum_{m=1}^{a_{1,n_i}} v_{1,m}^{(n_i)}+\sum_{m=1}^{u_{2,l}-1}\tilde{v}_{2,m}+s_2^*+\sum_{i=1}^{s_2^*}\sum_{m=1}^{a_{1,l_i}} v_{1,m}^{(l_i)}.$$

This case is a combination of the former two cases.

Due to the initial assumptions and since the lengths of different sub-busy periods only depend on the number of class-1 customer arrivals during different slots and the service times of the corresponding primary customers, the sub-busy periods associated with the primary customers of class-1 and class-2 form a set of i.i.d. random variables and their pgfs will be presented by $V_1(z)$ and $V_2(z)$ respectively. Notice that $f_{1,k}$ and $f_{2,k}$ are correlated; in section 2 it was explained that $a_{1,k}$ and $a_{2,k}$ may be correlated as well. Once again, applying a z-transform technique to equations (16)-(19) and taking into account the previous remarks, we can derive an expression for $D_2(z)$:

$$\begin{aligned}
&D_2(z)\triangleq \mathrm{E}[z^{d_2}]=\mathrm{E}\left[z^{d_2}1\{r_{1,k}=r_{2,k}=0\}\right]+\mathrm{E}\left[z^{d_2}1\{r_{1,k}=0,r_{2,k}>0\}\right]\\
&+\mathrm{E}\left[z^{d_2}1\{r_{1,k}>0,r_{2,k}=0\}\right]+\mathrm{E}\left[z^{d_2}1\{r_{1,k}>0,r_{2,k}>0\}\right]\\
&=F(V_1(z),V_2(z))\frac{S_2(zA_1(V_1(z)))}{A_1(V_1(z))}\Bigg\{P(0,0,0,0)\\
&+\frac{P(0,0,zA_1(V_1(z)),V_2(z))-P(0,0,0,0)}{zA_1(V_1(z))V_2(z)}+\frac{P(zA_1(V_1(z)),V_1(z),0,0)-P(0,0,0,0)}{zA_1(V_1(z))V_1(z)}\\
&+[P(zA_1(V_1(z)),V_1(z),zA_1(V_1(z)),V_2(z))-P(0,0,zA_1(V_1(z)),V_2(z))\\
&-P(zA_1(V_1(z)),V_1(z),0,0)+P(0,0,0,0)]\frac{1}{zA_1(V_1(z))V_1(z)V_2(z)}\Bigg\},
\end{aligned}\tag{20}$$

with $F(z_1,z_2)\triangleq E[z_1^{f_{1,k}}z_2^{f_{2,k}}]$ and $P(x_1,z_1,x_2,z_2)$ as defined in the previous section. The random variables $f_{1,k}$ and $f_{2,k}$ can be shown to have the following joint pgf (extension of a technique used in e.g. [2]):

$$F(z_1,z_2)=\frac{A(z_1,z_2)-A_1(z_1)}{\lambda_2(z_2-1)}.\tag{21}$$

Finally, we have to find expressions for $V_1(z)$ and $V_2(z)$. These pgfs satisfy the following relations:

$$V_j(z)=S_j(zA_1(V_1(z))),\tag{22}$$

with $j=1,2$. This can be understood as follows: when the m-th class-j customer that arrived during slot i enters service, $v_{j,m}^{(i)}$ consists of two parts: the service time of that customer itself, and the service times of the class-1 customers that arrive during its service time and of their class-1 successors. This leads to equation (22). Equation (20) together with equations (21) and

(11) leads to a fully determined version for $D_2(z)$:

$$D_2(z) = \frac{1-\rho_T}{\lambda_2} \frac{V_2(z)}{A_1(V_1(z))} \frac{1 - zA_1(V_1(z))}{1 - V_2(z)} \frac{A(V_1(z), V_2(z)) - A_1(V_1(z))}{zA_1(V_1(z)) - A(V_1(z), V_2(z))}. \tag{23}$$

As for the system contents, if we assume $S_2(z) = \dfrac{(1-\beta_2)z}{1-\beta_2 z}$, i.e., the special case of geometrical service times for the low-priority customers, we obtain the same equation for the pgf of the class-2 delay as found in [24].

5 Performance measures

In this section, we will show how to calculate moments and tail probabilities of the customer delays, both theoretically (subsections 5.1 and 5.2) and practically (subsection 5.3). The calculations of the moments and tail probabilities of the system contents are similar and therefore omitted here.

5.1 Calculation of moments

The functions $V_1(z)$ and $V_2(z)$ can only be explicitly found in case of some simple arrival processes. Their derivatives for $z = 1$, necessary to calculate the moments of the customer delay, on the contrary, can be calculated in closed-form since $V_j(1) = 1$. For example, the first derivatives of $V_j(z)$ for $z = 1$ are given by

$$V_j'(1) = \frac{\mu_j}{1-\rho_1},$$

with $j = 1, 2$. Let us define λ_{ij} and μ_{jj} as

$$\lambda_{ij} \triangleq \left.\frac{\partial^2 A(z_1, z_2)}{\partial z_i \partial z_j}\right|_{z_1=z_2=1} \quad ; \quad \mu_{jj} \triangleq \left.\frac{d^2 S_j(z)}{dz^2}\right|_{z=1},$$

with $i, j = 1, 2$. Now we can calculate the mean customer delay of both classes by taking the first derivatives of the respective pgfs for $z = 1$. We find

$$\mathrm{E}[d_1] = \mu_1 + \frac{\lambda_{11}\mu_1 + \lambda_1^2 \mu_{11}}{2\lambda_1(1-\rho_1)}, \tag{24}$$

for the mean customer delay of a class-1 customer and

$$\mathrm{E}[d_2] = \mu_2 + \frac{\rho_1(\mu_2 - 1)}{1-\rho_1} + \frac{\lambda_{22}\mu_2}{2\lambda_2(1-\rho_T)} + \frac{\lambda_2 \mu_{22}}{2(1-\rho_T)(1-\rho_1)} + \frac{\lambda_{12}\mu_1}{\lambda_2(1-\rho_T)} \tag{25}$$

$$+\frac{\lambda_{11}\mu_1^2+\lambda_1\mu_{11}}{2(1-\rho_T)(1-\rho_1)},$$

for the mean customer delay of class-2.

In a similar way, expressions for the variance (and higher moments) of the customer delays can be calculated by taking the appropriate derivatives of the respective pgfs as well (expressions are omitted because they are too elaborate; we will show some figures of variances in the next section though).

5.2 Tail probabilities

Not only the moments of the customer delays are important, but also, and especially, the (tail) distributions of these quantities. From the pgfs of the customer delay of class-1 and class-2 customers derived in section 4, approximations of the probability mass functions can be derived using Darboux's theorem (see Appendix). In order to determine the asymptotic behavior of the distribution of a random variable, the dominant singularity of the steady-state pgf of this random variable is important. It is commonly known that the dominant singularity of the pgf of a random variable lies on the positive real axis and is larger than or equal to 1. Note that the calculations in the remainder only apply in case of 'traditional' (pgfs of) arrival and service processes. More precisely, we assume that the pgfs and their derivatives diverge on their radii of convergence. This is however not a very restrictive assumption since it is fulfilled for most processes that occur in practice.

The dominant singularity of $D_1(z)$ (expression (15)) is a zero with multiplicity 1 of $z - A_1(S_1(z))$, denoted by $\hat{z}_H$. So, in the neighborhood of its dominant pole $\hat{z}_H$, we can approximate $D_1(z)$ by

$$D_1(z) \approx \frac{K_1}{\hat{z}_H - z}. \tag{26}$$

K_1 can be found by substituting $z = \hat{z_H}$ in (26) and using expression (15) for $D_1(z)$:

$$\begin{aligned} K_1 &= \lim_{z\to\hat{z}_H} D_1(z)(\hat{z}_H - z) \\ &= \frac{1-\rho_1}{\lambda_1}\frac{S_1(\hat{z}_H)(\hat{z}_H-1)^2}{(S_1(\hat{z}_H)-1)(A_1'(S_1(\hat{z}_H))S_1'(\hat{z}_H)-1)}, \end{aligned} \tag{27}$$

where we have used de l'Hôpital's rule. Using Darboux's theorem on (26) (see Appendix) we find the well-known geometric tail behavior for the high-priority delay:

$$\text{Prob}[d_1 = n] \approx K_1 \hat{z}_H^{-n-1},$$

for large enough n. Substituting (27) in this expression yields

$$\text{Prob}[d_1 = n] \approx \frac{1-\rho_1}{\lambda_1} \frac{S_1(\hat{z}_H)(\hat{z}_H - 1)^2}{\hat{z}_H(S_1(\hat{z}_H) - 1)(A_1'(S_1(\hat{z}_H))S_1'(\hat{z}_H) - 1)} \hat{z}_H^{-n}. \tag{28}$$

The tail behavior of the delay of class-2 customers is a bit more involved since the nature of the dominant singularity of $D_2(z)$ may differ. This is due to the occurrence of the function $V_1(z)$ in (23), which is only implicitly defined. First we take a closer look at that function $V_1(z)$ on the (positive) real axis. The first derivative of $V_1(z)$ is given by

$$V_1'(z) = \frac{S_1'(zA_1(V_1(z)))A_1(V_1(z))}{1 - zS_1'(zA_1(V_1(z)))A_1'(V_1(z))}, \tag{29}$$

Consequently, $V_1(z)$ has a singularity, denoted as $\hat{z}_B$, where the denominator of $V_1'(z)$ becomes 0, i.e., $\hat{z}_B S_1'(\hat{z}_B A_1(V_1(\hat{z}_B)))A_1'(V_1(\hat{z}_B)) = 1$. Note that $V_1(\hat{z}_B)$ is finite. A singularity of this type is called a branch point. In the neighborhood of $\hat{z}_B$, $V_1(z)$ is approximately given by (see [3])

$$V_1(z) \approx V_1(\hat{z}_B) - K_V\sqrt{\hat{z}_B - z}. \tag{30}$$

K_V can be found from expression (30) as follows:

$$\begin{aligned} K_V^2 &= \lim_{z\to\hat{z}_B} \frac{\left(V_1(\hat{z}_B) - V_1(z)\right)^2}{\hat{z}_B - z} \\ &= \lim_{z\to\hat{z}_B} \left[2\left(V_1(\hat{z}_B) - V_1(z)\right)V_1'(z)\right], \end{aligned}$$

where we have used de 'l Hôpital's rule. Using expression (29) for $V_1'(z)$, we obtain

$$K_V^2 = 2S_1'(\hat{z}_B A_1(V_1(\hat{z}_B)))A_1(V_1(\hat{z}_B)) \lim_{z\to z_B} \frac{V_1(\hat{z}_B) - V_1(z)}{1 - zS_1'(zA_1(V_1(z)))A_1'(V_1(z))}.$$

Applying de 'l Hôpital's rule once more and using the fact that $V_1'(z) \to \infty$ for $z \to \hat{z}_B$ ultimately leads to

$$K_V = \sqrt{\frac{2A_1(V_1(\hat{z}_B))}{\hat{z}_B[\hat{z}_B^2(A_1'(V_1(\hat{z}_B)))^3 S_1''(\hat{z}_B A_1(V_1(\hat{z}_B))) + A_1''(V_1(\hat{z}_B))]}}. \tag{31}$$

Since $V_1(z)$ appears in expression (23) of $D_2(z)$, $\hat{z}_B$ is also a branch point of $D_2(z)$. A second singularity of $D_2(z)$ is given by the dominant zero $\hat{z}_L$ of $zA_1(V_1(z)) - A(V_1(z), V_2(z))$ on the real axis. The tail behavior of the class-2 packet delay is thus characterized by $\hat{z}_L$ or $\hat{z}_B$, depending on which one is the dominant (i.e., smallest) singularity. Three situations may thus occur, namely when $\hat{z}_L$ is solely dominant, $\hat{z}_B$ is solely dominant, and $\hat{z}_L = \hat{z}_B$. We will first study

the (approximate) behavior of $D_2(z)$ in the neighborhood of its dominant singularity for the three cases separately. Afterwards, we will use Darboux's theorem to find expressions for the tail probabilities of the class-2 delay. In the first case, the single pole $\hat{z}_L$ is dominant and thus

$$D_2(z) \approx \frac{K_2^{(1)}}{\hat{z}_L - z},$$

for $z \to \hat{z}_L$. $K_2^{(1)}$ can be calculated by substituting expression (23) in the previous expression for $z = \hat{z}_L$ (in a similar way as in the calculation of (27)). This yields

$$K_2^{(1)} = \frac{(1-\rho_T)V_2(\hat{z}_L)(\hat{z}_L A_1(V_1(\hat{z}_L)) - 1)(\hat{z}_L - 1)}{\lambda_2(V_2(\hat{z}_L) - 1)Q_2(\hat{z}_L)}, \tag{32}$$

with

$$Q_2(z) = \frac{dA(V_1(z), V_2(z))}{dz} - A_1(V_1(z)) - zA_1'(V_1(z))V_1'(z).$$

In the second case, when the branch point $\hat{z}_B$ is solely dominant, $D_2(z)$ inherits the behavior of $V_1(z)$ in the neighborhood of $\hat{z}_B$, or:

$$D_2(z) \approx D_2(\hat{z}_B) - K_2^{(3)} (\hat{z}_B - z)^{1/2}.$$

$K_2^{(3)}$ is found as follows:

$$\begin{aligned} K_2^{(3)} =& K_V \lim_{z \to \hat{z}_B} \frac{D_2(\hat{z}_B) - D_2(z)}{V_1(\hat{z}_B) - V_1(z)} \\ =& K_V \lim_{z \to \hat{z}_B} \frac{D_2'(z)}{V_1'(z)}. \end{aligned}$$

Taking the first derivative of expression (23), substituting $D_2'(z)$ for this result in the former expression and taking into account that $V_1'(z) \to \infty$ for $z \to \hat{z}_B$, we find

$$K_2^{(3)} = \frac{(1-\rho_T)K_V Q_3(\hat{z}_B)}{\lambda_2 A_1(V_1(\hat{z}_B))^2 (V_2(\hat{z}_B) - 1)^2 (\hat{z}_B A_1(V_1(\hat{z}_B)) - A(V_1(\hat{z}_B), V_2(\hat{z}_B)))^2}, \tag{33}$$

with

$$\begin{aligned} Q_3(z) = & \big\{ A_1(V_1(z)) \left(A^{(1)}(V_1(z), V_2(z)) + A^{(2)}(V_1(z), V_2(z)) S_2'(zA_1(V_1(z)))zA_1'(V_1(z)) \right) \\ & - A(V_1(z), V_2(z))A_1'(V_1(z)) \big\} (z-1)(zA_1(V_1(z)) - 1)(V_2(z) - 1)A_1(V_1(z))V_2(z) \\ & + (A(V_1(z), V_2(z)) - A_1(V_1(z)))(A(V_1(z), V_2(z)) - zA_1(V_1(z))) \{ A_1(V_1(z)) \\ & (zA_1(V_1(z)) - 1)S_2'(zA_1(V_1(z)))zA_1'(V_1(z)) - A_1'(V_1(z))(V_2(z) - 1)V_2(z) \}. \end{aligned}$$

Finally in the third case, when $\hat{z}_L = \hat{z}_B$, it can be proved that $D_2(z)$ behaves as

$$D_2(z) \approx \frac{K_2^{(2)}}{(\hat{z}_B - z)^{1/2}}$$

in the neighborhood of $\hat{z}_B$. $K_2^{(2)}$ is found as follows:

$$\begin{aligned} K_2^{(2)} &= \lim_{z \to \hat{z}_B} D_2(z)(\hat{z}_B - z)^{1/2} \\ &= \frac{1}{K_V} \lim_{z \to \hat{z}_B} \frac{V_1(\hat{z_B}) - V_1(z)}{1/D_2(z)} \\ &= \frac{1}{K_V} \lim_{z \to \hat{z}_B} \frac{V_1'(z)(D_2(z))^2}{D_2'(z)}, \end{aligned}$$

after using de l'Hôpital's rule once more. Using expression (23) in this expression and taking into account that $V_1'(z) \to \infty$ for $z \to \hat{z}_B$ and that $A(V_1(\hat{z}_B), V_2(\hat{z}_B)) = \hat{z}_B A_1(V_1(\hat{z}_B))$ - since $\hat{z_B} = \hat{z}_L$ in this case, we find

$$K_2^{(2)} = \frac{(1-\rho_T)(\hat{z}_B - 1)V_2(\hat{z}_B)(\hat{z}_B A_1(V_1(\hat{z}_B)) - 1)}{\lambda_2 K_V (V_2(\hat{z}_B) - 1) Q_4(\hat{z}_B)}, \tag{34}$$

with

$$Q_4(z) = A^{(1)}(V_1(z), V_2(z)) + \left(A^{(2)}(V_1(z), V_2(z)) S_2'(zA_1(V_1(z))) - 1\right) zA_1'(V_1(z)). \tag{35}$$

Summarizing, $D_2(z)$ can be approximated in the neighborhood of its dominant singularity by:

$$D_2(z) \approx \begin{cases} \dfrac{K_2^{(1)}}{\hat{z}_L - z} & \text{if } \hat{z}_L \text{ dominant} \\ \dfrac{K_2^{(2)}}{\sqrt{\hat{z}_B - z}} & \text{if } \hat{z}_L = \hat{z}_B \text{ dominant} \\ D_2(\hat{z}_B) - K_2^{(3)} \sqrt{\hat{z}_B - z} & \text{if } \hat{z}_B \text{ dominant,} \end{cases}$$

with the constants $K_2^{(i)}$ $(i = 1, 2, 3)$ given by (32), (34) and (33) respectively. By applying Darboux's theorem (see Appendix) on these expressions, the asymptotic behavior of the class-2 customer delay probabilities is given by

$$\text{Prob}[d_2 = n] \approx \begin{cases} K_2^{(1)} \hat{z}_L^{-n-1} & \text{if } \hat{z}_L \text{ dominant} \\ \dfrac{K_2^{(2)} n^{-1/2} \hat{z}_B^{-n}}{\sqrt{\hat{z}_B \pi}} & \text{if } \hat{z}_L = \hat{z}_B \text{ dominant} \\ \dfrac{K_2^{(3)}}{2} \sqrt{\dfrac{\hat{z}_B}{\pi}} n^{-3/2} \hat{z}_B^{-n} & \text{if } \hat{z}_B \text{ dominant.} \end{cases} \tag{36}$$

5.3 Calculations in practice

We now conclude this section about the performance measures of this PR priority queue by going into some of the more practice-oriented aspects. We will more precisely summarize how the performance measures of the class-2 packet delay are calculated (a similar reasoning is possible for the system contents).

Firstly, one needs to 'obtain' all input pgfs, most notably $A(z_1, z_2)$, $S_1(z)$ and $S_2(z)$. These can either be given or calculated from measurements. In the latter case, this is done by calculating the z-transform of the (measured) probability mass functions.

To calculate the n-th (central) moment of the class-2 packet delay, the following procedure can be used:

- **Step 1:** Calculation of the required derivatives of $V_1(z)$ and $V_2(z)$ and their evaluation for $z = 1$
  - Define $V_1^{(j)}$ and $V_2^{(j)}$ as $\frac{d^j}{dz^j}V_1(1)$ and $\frac{d^j}{dz^j}V_2(1)$ respectively, for $j = 0, \ldots, n+1$.
  - Start with $V_1^{(0)} = 1$ and $V_2^{(0)} = 1$.
  - Take subsequent derivatives of both sides of $V_1(z) = S_1(zA_1(V_1(z)))$ and evaluate in 1. This iteratively leads to explicit expressions for $V_1^{(j)}$, $j = 1, \ldots, n+1$.
  - Take subsequent derivatives of both sides of $V_2(z) = S_2(zA_1(V_1(z)))$ and evaluate in 1. Substituting the results for $V_1^{(j)}$ yields explicit expressions for $V_2^{(j)}$, $j = 1, \ldots, n+1$.
- **Step 2:** Calculation of the required derivatives of $D_2(z)$ and their evaluation for $z = 1$
  - Define $D_2^{(0)}(z) = D_2(z)$.
  - For $j = 1, \ldots, n$ do
    - Take the first derivative of $D_2^{(j-1)}(z)$ and denote it by $D_2^{(j)}(z)$.
    - Denote numerator and denominator of $D_2^{(j)}(z)$ by $T_j(z)$ and $N_j(z)$ respectively. Both have a zero in $z = 1$ of multiplicity $2(j+1)$.
    - Calculate $D_2^{(j)}(1) = \frac{d^j}{dz^j}D_2(1)$ as $\frac{\frac{d^{2(j+1)}}{dz^{2(j+1)}}T_j(1)}{\frac{d^{2(j+1)}}{dz^{2(j+1)}}N_j(1)}$. This yields an expression with unknowns $V_1^{(i)}$ and $V_2^{(i)}$, $i = 1, \ldots, j+1$, which are already calculated in step 1. Substituting these thus yields $\frac{d^j}{dz^j}D_2(1)$.
- **Step 3:** Calculation of the moments
  - The first n factorial moments $\mathrm{E}[\prod_{i=0}^{j-1}(d_2 - i)]$, $j = 1, \ldots, n$ are given by $\frac{d^j}{dz^j}D_2(1)$, $j = 1, \ldots, n$ respectively, as calculated in step 2.
  - The n-th moment ($\mathrm{E}[d_2^n]$) and the n-th central moment ($\mathrm{E}[(d_2 - \mathrm{E}[d_2])^n]$) can be expressed in terms of these factorial moments and can thus be calculated (if desired).

For instance the second central moment is the variance and can be calculated using the procedure with $n = 2$ - yielding the first two factorial moments $\mathrm{E}[d_2]$ and $\mathrm{E}[d_2(d_2-1)]$ - and the relation $\mathrm{Var}[d_2] = \mathrm{E}[d_2(d_2-1)] + \mathrm{E}[d_2] - (\mathrm{E}[d_2])^2$ between the variance and these first two factorial moments.

Finally, the following procedure to calculate the tail probabilities for certain 'input' distributions and/or parameter sets, that avoids the explicit calculation of $V_1(z)$ and $V_2(z)$ (which is in general not possible), looks as follows:

- **Step 1:** Calculation of $\hat{z}_L$ and $V_1(\hat{z}_L)$
  - Find the dominant numerical solution x^* of $x - A(S_1(x), S_2(x)) = 0$ in the range $]1, \infty[$. The found x^* is $\hat{z}_L A_1(V_1(\hat{z}_L))$, if $\hat{z}_L$ exists.
  - Calculate $V_1(\hat{z}_L)$ as $S_1(x^*)$ and $\hat{z}_L$ as $\dfrac{x^*}{A_1(S_1(x^*))}$.
- **Step 2:** Calculation of $\hat{z}_B$ and $V_1(\hat{z}_B)$
  - Set $z_{min} = 1$ and $z_{max} = 2$.
  - If $z_{max} S_1'(z_{max})\lambda_1 > 1$, go to the next line. Else, solve $z_{max} S_1'(z_{max} A_1(V)) A_1'(V) = 1$, $V > 1$ numerically for V. If $S_1(z_{max} A_1(V)) < V$, increase z_{min} and z_{max} by one and repeat this line. Else go to the next line ($\hat{z}_B$ then lies in between z_{min} and z_{max}).
  - Define $z_{new} = (z_{min} + z_{max})/2$. If $z_{new} S_1'(z_{new})\lambda_1 > 1$, assign z_{new} to z_{max}. Solve $z_{new} S_1'(z_{new} A_1(V)) A_1'(V) = 1$, $V > 1$ numerically in V. If $S_1(z_{new} A_1(V)) < V$, assign z_{new} to z_{min}. Else, assign z_{new} to z_{max}. Repeat this step until the required precision is reached (e.g., until $z_{max} - z_{min} < 10^{-14}$).
  - Calculate $\hat{z}_B$ as z_{new} and $V_1(\hat{z}_B)$ as V.
- **Step 3:** Determination of the dominant singularity
  - If $V_1(\hat{z}_L) < V_1(\hat{z}_B)$, $\hat{z}_L$ is dominant.
  - If $V_1(\hat{z}_L) = V_1(\hat{z}_B)$, $\hat{z}_L = \hat{z}_B$ is dominant.
  - If $V_1(\hat{z}_L) > V_1(\hat{z}_B)$, $\hat{z}_B$ is dominant.
- **Step 4:** Calculation of the tail probabilities
  - If $\hat{z}_L$ is dominant, use the first formula of (36) with $K_2^{(1)}$ given by (32).
  - If $\hat{z}_L = \hat{z}_B$ is dominant, use the second formula of (36) with $K_2^{(2)}$ given by (34).
  - If $\hat{z}_B$ is dominant, use the third formula of (36) with $K_2^{(3)}$ given by (33).

We finally make some remarks concerning this last procedure. Firstly, we note that the calculation of the pole in step 1 is only equal to $\hat{z}_L$ if the latter one exists. This is only the case when the calculated $V_1(\hat{z}_L)$ in step 1 is smaller than or equal to $V_1(\hat{z}_B)$ (see step 3). For more details, we refer to a similar analyzed problem in [11]. Secondly, the procedure in step 2 is largely based on an algorithm in [20]. Finally, the required technique to numerically solve an

equation can be rather simple: a simple bisection algorithm and/or the use of a mathematical software program generally suffices.

6 Numerical example

In this section, we present some numerical examples. We will focus on the impact of the distribution of the (class-2) service times on the performance measures. For the impact of other parameters (e.g. the influence of the load and of the mean service times) on the performance measures, we refer to [23, 24].

We assume the customers of the two classes to be arriving according to a two-dimensional binomial process. Its two-dimensional pgf is given by:

$$A(z_1, z_2) = (1 - \frac{\lambda_1}{N}(1 - z_1) - \frac{\lambda_2}{N}(1 - z_2))^N. \tag{37}$$

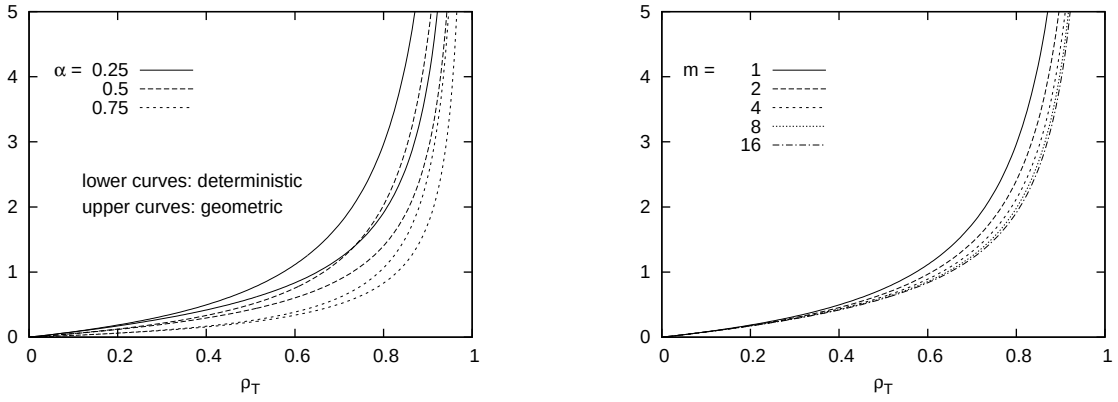
The arrival rate of class-j customers is thus given by λ_j $(j = 1, 2)$. In the remainder of this section, we assume that $N = 16$. We furthermore denote the fraction of the high-priority load in the total load by α, i.e., $\alpha = \rho_1/\rho_T$.

Figure 1 shows the mean system contents of class-2 as a function of the total load. The service times of class-1 are deterministically equal to 2 slots. The mean class-2 service time equals 16 slots. In Figure 1a., the mean system contents are compared for geometrically distributed and deterministic class-2 service times, for different values of α. In Figure 1b., the class-2 service times are assumed to be negative binomially distributed with parameters m and p (with $m/p = \mu_2 = 16$), i.e.,

$$S_2(z) = \left(\frac{pz}{1 - (1 - p)z}\right)^m.$$

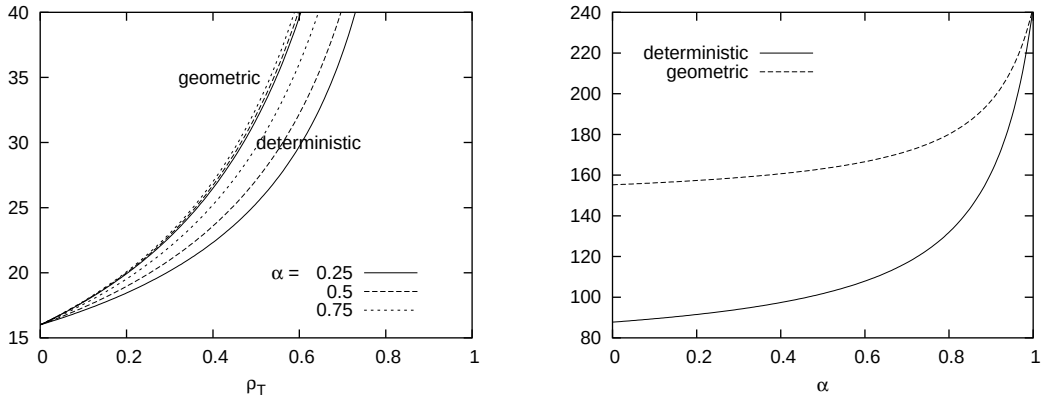
α equals 0.25 in this figure. By increasing m while keeping m/p constant, the variance of the class-2 service times is decreased while keeping their mean value constant. It may be noted that $m = 1$ corresponds with the geometric distribution, while $m = 16$ corresponds with deterministic service times. It is seen from these figures that a higher variance of the class-2 service times leads to higher mean system contents. It is further seen that the influence of the variance of the class-2 service times is bigger for smaller α and higher ρ_T.

In Figure 2, we illustrate the influence of the distribution of the class-2 service times on the mean class-2 customer delay. We assume deterministic class-1 service times of 2 slots and the mean value of the class-2 service times equals 16 slots. In Figure 2a., the mean class-2 customer delay is shown as a function of the total load for $\alpha = 0.25, 0.5$ and 0.75. Figure 2b. depicts the mean class-2 delay versus α for $\rho_T = 0.9$. In both figures, we compare the results for deterministically and geometrically distributed class-2 service times. It becomes apparent



a. geometric versus deterministic distribution b. several negative binomial distributions

Figure 1: Mean class-2 system contents versus the total load for different distribution of the class-2 service times



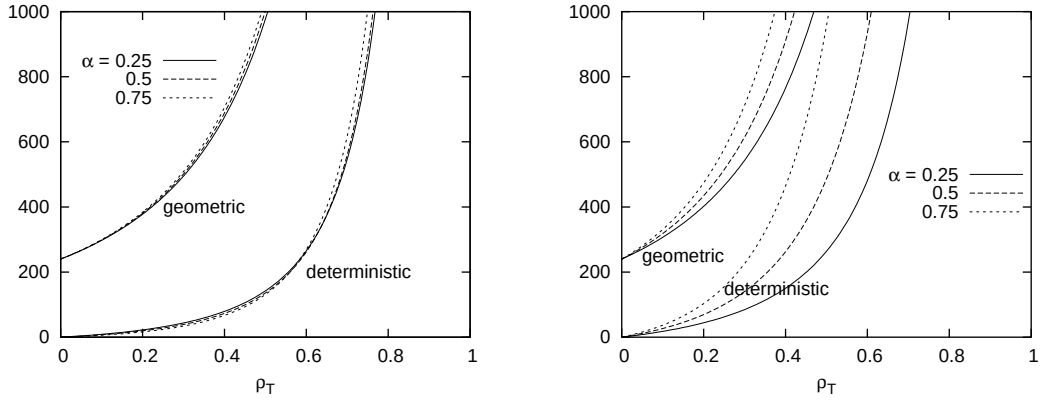
a. versus ρ_T b. versus α

Figure 2: Influence of class-2 service time distributions on the mean class-2 customer delay

from these figures that the mean class-2 customer delay also depends highly on the distribution of the class-2 service times, especially for low α. Modeling the service times as geometrically distributed stochastic variables can lead to considerable errors in the estimation of the mean class-2 customer delay.

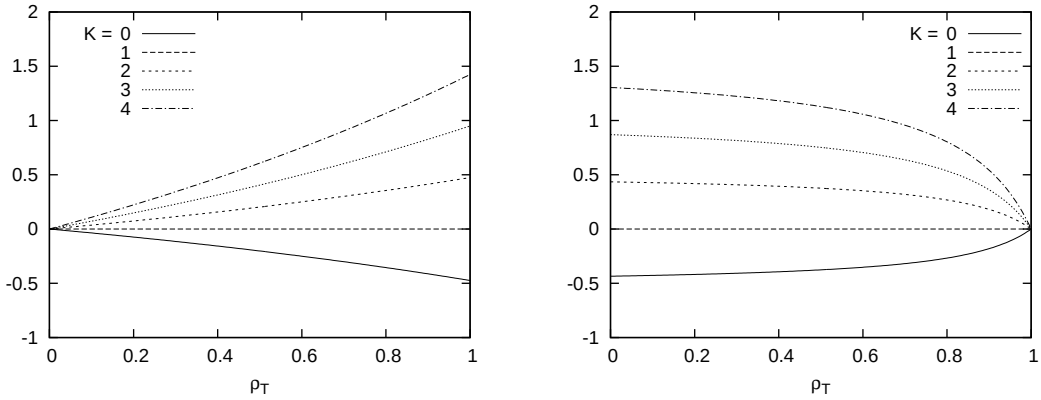
Figure 3 depicts the variance of the class-2 delay as a function of the total load, for $\alpha = 0.25$, 0.5 and 0.75 and for geometrically distributed and deterministic class-2 service times (with mean 16 slots). In Figure 3a., the class-1 service times are assumed to be equal to 2 while these are chosen equal to 16 in Figure 3b. Both figures show that the distribution of class-2 service times has a big impact on the variance of the class-2 customer delay. It is also seen by comparing both figures that the fraction of high-priority packets in the overall mix plays almost no role for small class-1 service times, while this impact is significant for larger class-1 service times.

To conclude the figures of the moments, we show in Figure 4 the relative deviation of the mean class-2 delay when the variance of the class-2 service times is varied, versus the total load (Figure 4a.) and versus the fraction of class-1 load (Figure 4b.). For all curves the class-1



a. $\mu_1 = 2$ b. $\mu_1 = 16$

Figure 3: Variance of the class-2 customer delay versus the total load



a. versus ρ_T b. versus α

Figure 4: Relative deviation of class-2 delay with respect to the geometric distribution

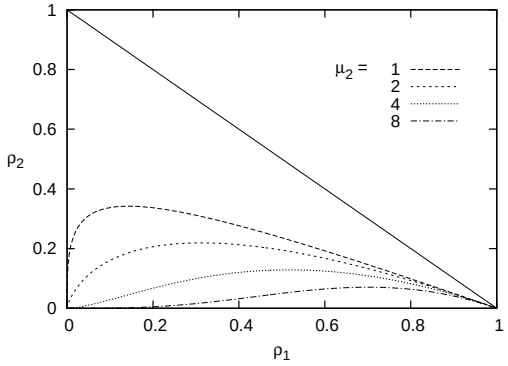
service times are deterministically equal to 2 and the mean class-2 service time equals 16. α is furthermore equal to 0.25 in 4a., while $\rho_T = 0.9$ in Figure 4b. The variance of the class-2 service times is assumed to be equal to $K(\mu_2^2 - \mu_2)$. For several values of K we have plotted the relative deviation of the mean class-2 customer delay, defined as

$$\frac{\mathrm{E}[d_2]_K - \mathrm{E}[d_2]_{K=1}}{\mathrm{E}[d_2]_{K=1}}.$$

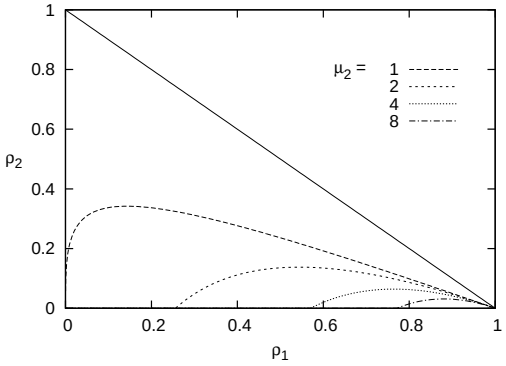
Note that the case $K = 1$ corresponds with the geometric distribution. The case $K = 0$ corresponds with the deterministic case while $K > 1$ corresponds with distributions that have a larger variance than the geometric distribution. Note that a variance with $K > 1$ can be easily constructed by using a mix of geometric distributions. We once again see from this figure that the variance of the class-2 service times has a big impact on the mean class-2 delay, especially for a large load and/or many class-2 packets in the traffic mix. E.g. for a doubled *variance* of the class-2 service times ($K = 2$ vs. $K = 1$), the relative deviation of the *mean* class-2 delay can be up to a 1/2.

In the next figures, we illustrate the tail behavior of the customer delay. We have shown in section 5, that the tail probabilities of the class-2 customer delay can have 3 types of behavior, depending on which singularity of $D_2(z)$ is dominant. In case of the arrival process considered in this section, Figure 5 shows for which combination of class-1 and class-2 loads the transition type behavior occurs for the customer delay when $\mu_1 = 2$ and for several values of μ_2, i.e., for which combination of loads the regular pole and the branch point coincide. We have shown this tail behavior for different combinations of deterministic and geometrically distributed class-1 and class-2 service times (Figure 5a.. . . 5d.). In the region above each of the curves, the tail behavior is geometric for the respective ρ_1 and ρ_2 values, while below the curves the tail behavior is typically non-geometric. Note that in the area above the boundary defined by $\rho_1 + \rho_2 = 1$ in the figures, the total load is larger than 1, and as a result, the system becomes unstable. As can be seen from the figures, the higher the mean service time of class-2 customers, the smaller the region where the tail behavior is non-geometric. By comparing the 4 figures (and from other extensive examples not mentioned here), we see that the transition between geometric and non-geometric tails highly depends on the service time distribution of the high- and low-priority customers. From this figure it can e.g. be concluded that the region where the tail behavior is non-geometric increases when the class-1 service times are changed from deterministic to geometrically distributed, while a reverse influence of the class-2 service times is observed.

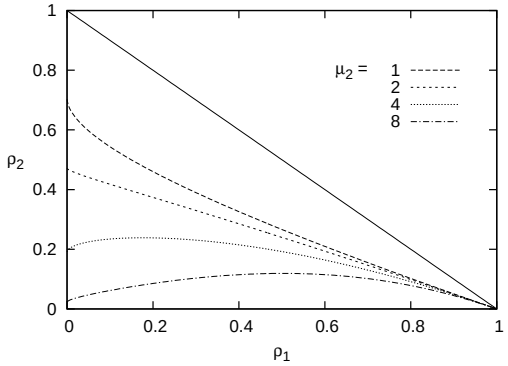
Figure 6a. shows the tail behavior of the customer delay of class-1 and class-2 customers for deterministic service times ($\mu_1 = \mu_2 = 2$), if $\rho_1 = 0.4$ and $\rho_2 = 0.1$ (non-geometric behavior), approximately 0.21 (transition type behavior) and 0.4 (geometric behavior) respectively. Tail behavior of customer delay of class-1 customers is shown as comparison material and is of course the same for the three cases, since the arrival process of class-1 customers is identical in all cases, and class-2 customers are 'invisible' for the high-priority class-1 customers due to the preemptive service discipline. We have also compared our approximations with simulation results (marks in the figures). The figures show that the approximations for the tail probabilities of the delay of both classes is very good. Finally Figure 6b. depicts the influence of the distribution of the service times of both classes on the tail probabilities. The mean service times of both classes equal 2. The loads of class-1 and class-2 are assumed 0.1 and approximately 0.12. This latter load is chosen such that the transition type tail behavior is observed in the case that both classes have geometrically distributed service times. It is seen that the distribution of the service times plays a non-negligible role, or more precisely, if service times of at least one of the two classes is non-deterministic, the tail probability of the class-2 delay is considerably higher, leading to a worse performance in terms of delay (and also system content, which is not shown here).



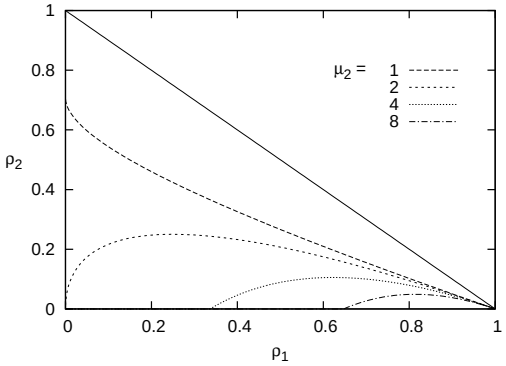
a. deterministic class-1 service times
deterministic class-2 service times



b. deterministic class-1 service times
geometric class-2 service times

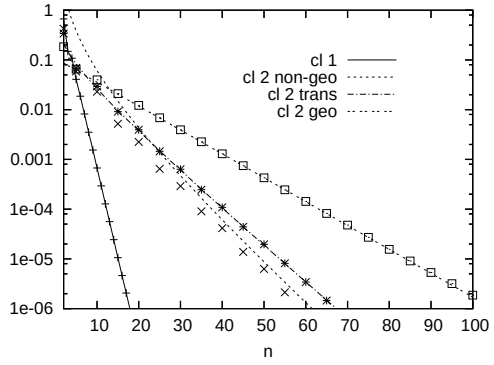


c. geometric class-1 service times
deterministic class-2 service times

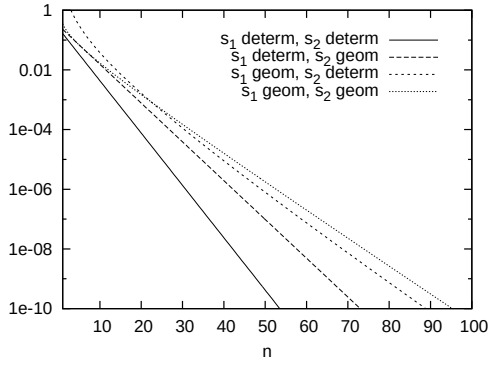


d. geometric class-1 service times
geometric class-2 service times

Figure 5: Regions for the different tail behaviors in the (ρ_1, ρ_2)-plane



a. different types of tail behavior



b.different service time distributions

Figure 6: Tail probabilities of the class-2 customer delay

7 Conclusion

In this paper, we have analyzed a discrete-time queue with a preemptive resume priority scheduling, two priority classes and generally distributed service times. We have first constructed a 4-dimensional Markov-chain which led to the calculation of a 4-dimensional pgf. We have derived the joint pgf of the system contents of both classes and the pgfs of the customer delays

of both classes from this 4-dimensional pgf. These pgfs are not explicitly found, but we have proved that the moments and tail distributions of the respective stochastic variables can be calculated explicitly in terms of the system parameters. Procedures to practically calculate these performance measures are further proposed. We have shown the impact of the priority scheduling and the influence of the distributions of the service times on the performance characteristics by some numerical examples.

Acknowledgment

This paper is an extended version of our paper presented at the Networking 2002 conference (Pisa, May 19-24, 2002) and published in LNCS 2345. The first author is a Postdoctoral Fellow with the Fund for Scientific Research, Flanders (F.W.O.-Vlaanderen), Belgium. The authors wish to thank the anonymous referees for their valuable comments.

Appendix

Theorem 1 (Darboux's theorem) *Suppose* $X(z) = \sum_{n=0}^{\infty} x(n)z^n$ *with positive real coefficients* $x(n)$ *is analytic near 0 and has only algebraic singularities* α_k *on its circle of convergence* $|z| = R$*, in other words, in a neighborhood of* α_k *we have*

$$X(z) \sim (1 - \frac{z}{\alpha_k})^{-\omega_k} G_k(z), \tag{38}$$

where $\omega_k \neq 0, -1, -2, \ldots$ *and* $G_k(z)$ *denotes a nonzero analytic function near* α_k*. Let* $\omega = max_k Re(\omega_k)$ *denote the maximum of the real parts of the* ω_k*. Then we have*

$$x(n) = \sum_j \frac{G_j(\alpha_j)}{\Gamma(\omega_j)} n^{\omega_j - 1} \alpha_j^{-n} + o(n^{\omega - 1} R^{-n}), \tag{39}$$

with $\omega = Re(\omega_j)$ *and* $\Gamma(\omega)$ *the Gamma-function of* ω *(with* $\Gamma(n) = (n-1)!$ *for* n *discrete).*

References

[1] H. Bruneel. Performance of discrete-time queueing systems. *Computers and Operations Research*, 20(3):303–320, 1993.

[2] H. Bruneel and B. Kim. *Discrete-time models for communication systems including ATM.* Kluwer Academic Publisher, Boston, 1993.

[3] M. Drmota. Systems of functional equations. *Random Structures & Algorithms*, 10(1-2):103–124, 1997.

[4] M. Fidler and R. Persaud. M/G/1 priority scheduling with discrete pre-emption points: on the impacts of fragmentation on IP QoS. *Computer Communications*, 27(12):1183–1196, 2004.

[5] D. Fiems, B. Steyaert, and H. Bruneel. Analysis of a discrete-time gi-g-1 queueing model subjected to bursty interruptions. *Computers and Operations Research*, 30(1):139–153, 2002.

[6] A. Khamisy and M. Sidi. Discrete-time priority queues with two-state markov modulated arrivals. *Stochastic Models*, 8(2):337–357, 1992.

[7] L. Kleinrock. *Queueing systems volume II: Computer applications*. John Wiley & Sons, New York, 1976.

[8] G. Koole and A. Mandelbaum. Queueing models of call centers: an introduction. *Annals of Operations Research*, 113(1-4):41–59, 2002.

[9] K. Laevens and H. Bruneel. Discrete-time multiserver queues with priorities. *Performance Evaluation*, 33(4):249–275, 1998.

[10] F. Machihara. A bridge between preemptive and nonpreemptive queueing models. *Performance Evaluation*, 23(2):93–106, 1995.

[11] T. Maertens, J. Walraevens, and H. Bruneel. Priority queueing systems: from probability generating functions to tail probabilities. *Queueing Systems*, to appear.

[12] M. Mehmet Ali and X. Song. A performance analysis of a discrete-time priority queueing system with correlated arrivals. *Performance Evaluation*, 57(3):307–339, 2004.

[13] R. Miller. Priority queues. *Annals of Mathematical Statistics*, 31:86–103, 1960.

[14] K. Ramamritham and J. Stankovic. Scheduling algorithms and operating systems support for real-time systems. *Proceedings of the IEEE*, 82(1):55–67, 1994.

[15] M. Sidi and A. Segall. Structured priority queueing systems with applications to packet-radio networks. *Performance Evaluation*, 3(4):265–275, 1983.

[16] H. Takagi. *Queueing analysis: a foundation of performance evaluation volume 1: vacation and priority systems, part 1*. North-Holland, 1991.

[17] Y. Takahashi and O. Hashida. Delay analysis of discrete-time priority queue with structured inputs. *Queueing Systems*, 8(2):149–164, 1991.

[18] T. Takine and T. Hasegawa. The workload in the MAP/G/1 queue with state-dependent services: its application to a queue with preemptive resume priority. *Communications in Statistics - Stochastic Models*, 10(1):183–204, 1994.

[19] T. Takine, B. Sengupta, and T. Hasegawa. An analysis of a discrete-time queue for broadband ISDN with priorities among traffic classes. *IEEE Transactions on Communications*, 42(2-4):1837–1845, 1994.

[20] J. Van Velthoven, B. Van Houdt, and C. Blondia. The impact of buffer finiteness on the loss rate in a priority queueing system. *Lecture Notes in Computer Science*, 4054:211–225, 2006.

[21] J. Walraevens, B. Steyaert, and H. Bruneel. *Performance analysis of a GI-G-1 preemptive resume priority buffer*, pages 745–756. LNCS 2345 (Proceedings of the Networking 2002 Conference, Pisa). Springer Verlag, 2002.

[22] J. Walraevens, B. Steyaert, and H. Bruneel. Performance analysis of a single-server ATM queue with a priority scheduling. *Computers & Operations Research*, 30(12):1807–1829, 2003.

[23] J. Walraevens, B. Steyaert, and H. Bruneel. Performance analysis of a GI-Geo-1 buffer with a preemptive resume priority scheduling discipline. *European Journal of Operational Research*, 157(1):130–151, 2004.

[24] J. Walraevens, B. Steyaert, and H. Bruneel. A packet switch with a priority scheduling discipline: Performance analysis. *Telecommunication Systems*, 28(1):53–77, 2005.