INFLUENCE OF RELATIVE TRAFFIC DISTRIBUTION IN NODES WITH BLOCKING: AN ANALYTICAL MODEL

Herwig Bruneel
Willem Mélange
Bart Steyaert
Dieter Claeys
Joris Walraevens*
SMACS Research Group
Department of Telecommunications and Information Processing
Ghent University
Sint-Pietersnieuwstraat 41
B 9000 Ghent, Belgium
E-mail:{hb,wmelange,bs,dclaeys,jw}@telin.ugent.be
* The fifth author is a Postdoctoral Fellow with the Fund
for Scientific Research, Flanders (F.W.O.-Vlaanderen), Belgium.

KEYWORDS
Blocking, Relative Traffic Distribution, Queueing Model.

ABSTRACT

In nodes where the arriving packets are stored in one common buffer, packets with a given destination may have to wait for the transmission of packets with other destinations, even when the corresponding output channel is free. Although this so-called blocking effect has attracted considerable attention in literature, the influence of the relative distribution of the traffic according to destination has been largely overlooked. We therefore develop and analyze an appropriate discrete-time queueing model for a node whereby all arriving packets are accommodated in one common buffer and with two output channels that lead to distinct destinations. We study the stability of and the number of packets in the node. We then compare these results with those obtained for an analogous node with individual buffers for the distinct output channels. We demonstrate that the relative distribution of the traffic according to destination can have a major impact on the blocking effect and hence on the overall performance of the node.

INTRODUCTION

In nodes in telecommunication networks, information packets with a given destination A may have to wait for the transmission of packets destined to node B that arrived earlier, even when the link to destination A is free, if the arriving packets are accommodated in one common buffer. The underlying reason is that in general the first-come-first-served (FCFS) policy is adopted. As a result, when the two eldest packets in the node are heading for the same destination, one of those packets is transmitted, whereas the other becomes the head of the buffer and blocks the access of other packets requiring transmission over distinct output channels, simply because of the FCFS transmission policy.

Although the blocking effect has attracted considerable attention in literature (Beekhuizen and Resing 2009; Laevens 1999; Liew 1994; Mandelbaum and Reiman 1998; Stolyar 2004; Van Dijk and Van der Sluis 2008; Van Woensel and Vandaele 2006; Van Woensel and Vandaele 2007), we believe that the influence of the relative distribution of the traffic according to destination has been largely overlooked. For this reason, we develop and analyze an analytical model for a node whereby all packets are accommodated in one common buffer and with two output channels that lead to distinct destinations. It is a discrete-time queueing model with general independent arrivals, two (uncorrelated) packet classes and two class-specific servers. The servers hereby correspond with output channels and distinct packet classes with packets requiring transmission over different output channels.

The remainder of the paper is organized as follows: to start with, the investigated model is described in detail. Then, the analysis of the model is carried out. We study the condition under which the node is able to transmit all packets within a finite time, which is referred to as stability condition throughout the paper. We also analyze the system content, i.e., the number of packets in the node at random slot boundaries, those under transmission included. Next, analogous results in case of individual buffers are briefly summarized. These results then allow us to investigate the impact of blocking and the relative distribution of the traffic on the overall performance. Finally, some conclusions are drawn and directions for future research are given.

ANALYTICAL MODEL

As model, we consider the discrete-time queueing system depicted in Fig.1, with infinite queue capacity (i.e., all packets can be stored in the queue), two servers, named A and B, and two types (classes) of packets, named 1 and 2. The system corresponds with the node, servers with output channels, the queue with the buffer and packet classes with destinations. Each of the two servers is dedicated to a given class of packets, i.e., server A can only serve (transmit) packets of type (with destination) 1 and server B can only serve packets of type 2. Service times (transmission times) of all packets are deterministically equal to 1 slot each. Packets are served in their order of arrival, regardless of the class they belong to.

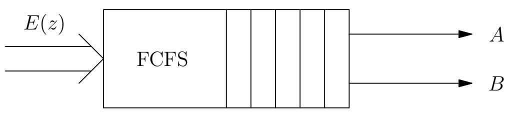

Figure 1: Illustration of the Queueing Model

The arrival process of new packets in the system is characterized in two steps. First, we model the total (aggregated) arrival stream of new packets by means of a sequence of independent and identically distributed (i.i.d.) discrete random variables with common probability mass function (pmf) $e(n)$ and common probability generating function (pgf) $E(z)$ respectively. More specifically,

$$e(n) \triangleq \Pr\left[n \text{ arrivals in one slot}\right] \quad , \quad n \geq 0 \ ,$$
$$E(z) \triangleq \sum_{n=0}^{\infty} e(n)\, z^n \ .$$

The (total) mean number of arrivals per slot, in the sequel referred to as the (total) mean arrival rate, is given by

$$\lambda \triangleq E'(1) \ .$$

Next, we describe the relative distribution of the arrival stream amongst the packet classes. We assume that an arriving packet belongs to the first class with probability σ, and to the second class with probability $\overline{\sigma} \triangleq 1-\sigma$, independently from packet to packet. The mean per-class arrival rates, λ_1 for class 1 and λ_2 for class 2, are then given by

$$\lambda_1 = \sigma\lambda \quad ; \quad \lambda_2 = \overline{\sigma}\lambda \ .$$

It can be seen that the two-server system described here includes the blocking effect: whenever the two "eldest" packets in the system, i.e., the two packets at the front of the queue, are of the same type, only one of them can then be served (by its own dedicated server) and the other "blocks" the access to the second server for packets of the opposite type further in the queue.

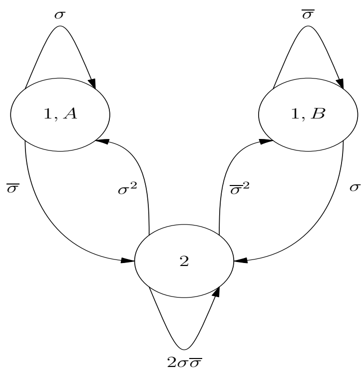

Figure 2: State Transition Diagram of the Markov Chain of the Server State in a Nearly Saturated System

ANALYSIS OF THE MODEL

Stability Condition

Consider a (nearly) saturated system, i.e., assume that at each slot mark at least two packets are present (during a long period of time). The state of the two servers can then be characterized by one of the following three descriptions:

- $(1, A)$: server A is active, server B is idle
- $(1, B)$: server B is active, server A is idle
- (2): both servers are active

It is easily seen that the server states during consecutive slots form a Markov chain with state transition diagram as depicted in Fig. 2. Note that no direct transitions are possible between the states $(1, A)$ and $(1, B)$. Indeed, $(1, A)$ entails that the second eldest packet is also of type 1, otherwise server B would be processing a packet of type 2. As a result, the eldest packet in the next slot is of type 1, so that server A is active during the next slot, which implies that state $(1, B)$ cannot be reached from state $(1, A)$. A similar reasoning holds in the opposite direction. The Markov chain has the following equilibrium distribution:

$$p_{1,A} = \frac{\sigma^3}{1-2\sigma\overline{\sigma}} \ ,$$
$$p_{1,B} = \frac{\overline{\sigma}^3}{1-2\sigma\overline{\sigma}} \ ,$$
$$p_2 = \frac{\sigma\overline{\sigma}}{1-2\sigma\overline{\sigma}} \ ,$$

where $p_{1,A}$, $p_{1,B}$ and p_2 represent the equilibrium probabilities of finding the servers in state $(1, A)$, $(1, B)$ or (2) respectively, during a random slot. The mean number of packets that can be processed during a random slot is thus given by

$$p_{1,A} + p_{1,B} + 2p_2 = \frac{1-\sigma\overline{\sigma}}{1-2\sigma\overline{\sigma}} \ .$$

As a result, the stability condition of the system reads

$$\lambda < \frac{1-\sigma\overline{\sigma}}{1-2\sigma\overline{\sigma}} \ . \tag{1}$$

System Content

This section is subdivided in three parts. First, we explain how the operation of the system during a slot can be described concisely, which is called system state description. Then we deduce the pgf of the system content at random slot boundaries and finally, we extract the average value from this pgf.

System State Description

As opposed to the section about the stability condition, we now consider a non-saturated system, assuming that (1) is satisfied. Whenever the system was empty or contained exactly one packet (two cases) at the previous slot mark, then at the current slot boundary the system contains only those packets that have arrived during the previous slot. As the type of a packet is independent of the types of previous packets, the type of the packet served (if any) during the previous slot has no influence on the types of the packets present at the current slot mark and consequently has no influence on the number of packets that can be served during the current slot. This implies that, in these two cases, the system state does not require any class-related information. If, on the other hand, multiple packets were present at the previous slot mark, then the types of the two eldest packets at the previous slot boundary may have an influence on the type of the eldest packet at the current slot boundary. Indeed, when the two eldest packets in the previous slot were of the same type, the eldest packet in the current slot was the second eldest packet in the previous slot and is thus of the same type. This, in turn, implies that the system state needs to include (at least) information on the type of the eldest packet, in these cases. On account of these observations, the evolution of the system from slot to slot can be described by a Markov chain with state space

$$\{(0),(1),(i,n), i=\{1,2\}, n\geq 2\} \ ,$$

where (0) represents an empty system, (1) denotes a system containing one packet and (i,n) characterizes a system with n packets, the eldest packet being of type i.

Pgf of the System Content

Let u_k represent the system content (including the packets in service, if any) at slot mark k. The pgf of the system content at a random slot boundary in steady state is denoted by $U(z)$. Next, we indicate the type of the eldest packet in slot k, when at least two packets are present, by t_k. The steady-state probabilities corresponding to the states (0), (1), (i,n) are designated by p_0, p_1 and $p(i,n)$, respectively, i.e.,

$$p_n \triangleq \lim_{k\to\infty} \Pr[u_k = n] \ , \qquad n=\{0,1\} \ ,$$
$$p(i,n) \triangleq \lim_{k\to\infty} \Pr[t_k = i, u_k = n] \ , \qquad i=\{1,2\}, n\geq 2 \ .$$

We then have that

$$U(z) = p_0 + p_1 z + Q_1(z) + Q_2(z) \ ,$$

where

$$Q_i(z) \triangleq \sum_{n=2}^{\infty} p(i,n) z^n \ , \qquad i=\{1,2\} \ .$$

We now calculate $Q_1(z)$ and $Q_2(z)$. Therefore, we start from the balance equation for state $(1,n)$, $n\geq 2$:

$$\begin{aligned} p(1,n) =\ & p_0\sigma e(n) + p_1\sigma e(n) + \sum_{m=2}^{n+2}[\sigma e(n-m+1) \\ & + \overline{\sigma}\sigma e(n-m+2)]p(1,m) \\ & + \sum_{m=2}^{n+2}\sigma^2 e(n-m+2)p(2,m) \ , \end{aligned} \tag{2}$$

where $e(-1)=0$. Multiplying both sides of (2) by z^n, then summing over n from 2 to infinity and taking into account the definition of $Q_1(z)$ yields

$$\begin{aligned} Q_1(z) =\ & \sigma(p_0+p_1)\,[E(z)-e(0)-e(1)z] \\ & + \sigma\sum_{m=2}^{\infty} p(1,m)\sum_{n=y}^{\infty} e(n-m+1)z^n \\ & + \overline{\sigma}\sigma\sum_{m=2}^{\infty} p(1,m)\sum_{n=y}^{\infty} e(n-m+2)z^n \\ & + \sigma^2\sum_{m=2}^{\infty} p(2,m)\sum_{n=y}^{\infty} e(n-m+2)z^n, \\ y =\ & \max(2,m-2) \ . \end{aligned}$$

This equation can be further transformed into

$$\begin{aligned} z^2Q_1(z) =\ & \sigma[p_0+p_1+\overline{\sigma}p(1,2)+\sigma p(2,2)]z^2 \\ & \cdot[E(z)-e(0)-e(1)z] + [E(z)-e(0)] \\ & \cdot\sigma[p(1,2)+\overline{\sigma}p(1,3)+\sigma p(2,3)]z^3 \\ & + \sigma[z\sum_{m=3}^{\infty}p(1,m)z^m + \overline{\sigma}\sum_{m=4}^{\infty}p(1,m)z^m \\ & + \sigma\sum_{m=4}^{\infty}p(2,m)z^m]E(z) \ . \end{aligned}$$

Relying on the definitions of $Q_1(z)$ and $Q_2(z)$ produces

$$\begin{aligned}&[z^2-\sigma\overline{\sigma}E(z)-\sigma zE(z)]Q_1(z)-\sigma^2E(z)Q_2(z)\\ =&\ \sigma(p_0+p_1)z^2\left[E(z)-e(0)-e(1)z\right]-\sigma p(1,2)e(0)z^3\\ &-\sigma z^2[\overline{\sigma}p(1,2)+\sigma p(2,2)][e(0)+e(1)z]\\ &-\sigma z^3[\overline{\sigma}p(1,3)+\sigma p(2,3)]e(0)\ . \end{aligned} \tag{3}$$

Before proceeding, we note that the balance equation for state 1 reads

$$\begin{aligned}p_1=&\ p_0e(1)+p_1e(1)+p(1,2)\sigma e(0)+p(2,2)\sigma e(1)\\ &+p(1,2)\overline{\sigma}e(1)+p(2,2)\overline{\sigma}e(0)\\ &+p(1,3)\overline{\sigma}e(0)+p(2,3)\sigma e(0)\ ,\end{aligned}$$

or equivalently

$$\begin{aligned}&[\overline{\sigma}p(1,3)+\sigma p(2,3)]e(0)\\ =&-p_0e(1)+p_1(1-e(1))-p(1,2)[\sigma e(0)+\overline{\sigma}e(1)]\\ &-p(2,2)[\sigma e(1)+\overline{\sigma}e(0)]\ .\end{aligned}$$

Invoking this relationship in (3) yields the following linear relation between $Q_1(z)$ and $Q_2(z)$:

$$\begin{aligned}&[z^2-\sigma\overline{\sigma}E(z)-\sigma zE(z)]Q_1(z)-\sigma^2E(z)Q_2(z)\\ =&\ \sigma p_0z^2[E(z)-e(0)]+\sigma p_1z^2[E(z)-e(0)-z]\\ &-\sigma p(2,2)e(0)z^2[\sigma-\overline{\sigma}z]\\ &-\sigma\overline{\sigma}p(1,2)e(0)z^2(z+1)\ .\end{aligned} \tag{4}$$

Note that this equation has been deduced by starting from the balance equation for state $(1,n)$. Completely analogously, a second relation between $Q_1(z)$ and $Q_2(z)$ can be obtained by starting from the balance equation for state $(2,n)$. This eventually produces:

$$\begin{aligned}&\left[z^2-\sigma\overline{\sigma}E(z)-\overline{\sigma}zE(z)\right]Q_2(z)-\overline{\sigma}^2E(z)Q_1(z)\\ =&\ \overline{\sigma}p_0z^2\left[E(z)-e(0)\right]+\overline{\sigma}p_1z^2\left[E(z)-e(0)-z\right]\\ &-\overline{\sigma}p(1,2)e(0)z^2[\overline{\sigma}-\sigma z]\\ &-\sigma\overline{\sigma}p(2,2)e(0)z^2(z+1).\end{aligned} \tag{5}$$

The unknown partial generating functions $Q_1(z)$ and $Q_2(z)$ can now be found by solving the set of linear (algebraic) equations (4) and (5), which yields

$$\begin{aligned}Q_1(z)=&\sigma z^2\Bigg[\Bigg\{-(1-e(0))z^2+[\{1+\overline{\sigma}(1-e(0))\}\\ &\cdot E(z)-1]z+E(z)[1-\sigma e(0)-\overline{\sigma}E(z)]\Bigg\}p_0\\ &+\Bigg\{-(1-e(0))z^2+\{1+\overline{\sigma}(1-e(0))\}zE(z)\\ &-E(z)\left[\sigma e(0)+\overline{\sigma}E(z)\right]\Bigg\}p_1\\ &+e(0)\Bigg\{z^2-\{\overline{\sigma}z+\sigma\}E(z)\Bigg\}p(2,2)\Bigg]\\ &/\left[z^3-z^2E(z)+\sigma\overline{\sigma}zE(z)(E(z)-2)\right.\\ &\left.+\sigma\overline{\sigma}E(z)^2\right]\ ,\end{aligned} \tag{6}$$

$$\begin{aligned}Q_2(z)=&\overline{\sigma}z^2\Bigg[\Bigg\{-(1-e(0))z^2+[\{1+\sigma(1-e(0))\}\\ &\cdot E(z)-1]z+E(z)[1-\overline{\sigma}e(0)-\sigma E(z)]\Bigg\}p_0\\ &+\Bigg\{-(1-e(0))z^2+\{1+\sigma(1-e(0))\}zE(z)\\ &-E(z)\left[\overline{\sigma}e(0)+\sigma E(z)\right]\Bigg\}p_1\\ &+e(0)\Bigg\{z^2-[\sigma z+\overline{\sigma}]E(z)\Bigg\}p(1,2)\Bigg]\\ &/\left[z^3-z^2E(z)+\sigma\overline{\sigma}zE(z)(E(z)-2)\right.\\ &\left.+\sigma\overline{\sigma}E(z)^2\right]\ .\end{aligned} \tag{7}$$

Furthermore, we still have the balance equation for state 0:

$$p_0=p_0e(0)+p_1e(0)+p(1,2)\overline{\sigma}e(0)+p(2,2)\sigma e(0)\ ,$$

which is equivalent with

$$p(1,2)=\frac{[1-e(0)]p_0-p_1e(0)-p(2,2)\sigma e(0)}{\overline{\sigma}e(0)}\ . \tag{8}$$

On account of (6)-(8), $U(z)$ is equal to

$$\begin{aligned}U(z)=&\ p_0+p_1z+Q_1(z)+Q_2(z)\\ =&\ (z-1)E(z)\Bigg[\Bigg\{(2\sigma-1)\sigma z^2e(0)+\overline{\sigma}\\ &\cdot[(2\sigma+1)z^2-2\sigma(E(z)-1)z-\sigma E(z)]\Bigg\}p_0\\ &+\Bigg\{(2\sigma-1)\sigma z^2e(0)+\overline{\sigma}\sigma z(2z-E(z))\Bigg\}p_1\\ &+(2\sigma-1)\sigma z^2e(0)p(2,2)\Bigg]\\ &/\left[z^3-z^2E(z)+\sigma\overline{\sigma}zE(z)(E(z)-2)\right.\\ &\left.+\sigma\overline{\sigma}E(z)^2\right]\ .\end{aligned} \tag{9}$$

This equation still contains three unknown probabilities p_0, p_1 and $p(2,2)$. It is, however, possible to express $p(2,2)$ as a function of p_0 and p_1 by invoking the “rate-in-rate-out” principle. The “rate-in-rate-out” principle expresses that in steady state the average number of packets leaving the system in a slot equals the mean arrival rate λ, which leads to:

$$\lambda=\ p_1+(2-\sigma)Q_1(1)+(1+\sigma)Q_2(1)\ . \tag{10}$$

Invoking (6) and (7) (for $z=1$) in (10) enables us to express $p(2,2)$ in terms of p_0 and p_1:

$$\begin{aligned}p(2,2)=&\Bigg[\big\{\overline{\sigma}(1+\sigma)-e(0)\sigma(1-2\sigma)\big\}p_0\\ &+\sigma\big\{\overline{\sigma}-e(0)(1-2\sigma)\big\}p_1+\sigma\overline{\sigma}\\ &+\lambda\{1-2\sigma\overline{\sigma}\}-1\Bigg]\Bigg/\Bigg[(1-2\sigma)\sigma e(0)\Bigg]\ .\end{aligned}$$

It can be shown that the same result can be obtained from the normalizing equation of the pgf $U(z)$, i.e., from the condition $U(1) = 1$. Using this relation in (9) finally results into

$$\begin{aligned} U(z) =& (z-1)E(z)\Big[p_1\sigma\overline{\sigma}z\{z-E(z)\} \\ &+ p_0\sigma\overline{\sigma}\{z^2+2z(1-E(z))-E(z)\} \\ &- z^2\left\{\sigma\overline{\sigma}-1+\lambda[1-2\sigma\overline{\sigma}]\right\}\Big] \\ &/\Big[z^3 - z^2E(z) + \sigma\overline{\sigma}E(z)(E(z)-2)z \\ &+ \sigma\overline{\sigma}E(z)^2\Big] \ . \end{aligned} \quad (11)$$

Note that the denominator of (11) can be rewritten as

$$(z-1)[z-\sigma E(z)][z-\overline{\sigma}E(z)] \\ +[z-E(z)][z-2\sigma\overline{\sigma}E(z)] \ ,$$

and that, by means of Rouché's theorem (Bruneel and Kim 1993; Takagi 1993), it is not difficult to prove that this function has exactly three zeroes inside the closed complex unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$, one of which equals 1. Let z_1 and z_2 indicate the two other zeroes. Then the two remaining unknowns p_0 and p_1 in (11) can be calculated by solving the set of linear equations

$$U_N(z_i) = 0 \ , \quad i = 1, 2 \ ,$$

where $U_N(z)$ is the numerator of $U(z)$ in equation (11).

Mean System Content

Once the two probabilities p_0 and p_1 have been computed, several performance measures of the system can be extracted from (11). For instance, the total mean system content can be deduced by taking the first derivative of (11) at $z = 1$:

$$\begin{aligned} \mathrm{E}[u] =& \Big[2\sigma\overline{\sigma}[(1-p_0)(3\lambda-2)+p_1(1-\lambda)]+2\lambda(1-\lambda) \\ &+[\sigma^2+\overline{\sigma}^2]E''(1)\Big]/\Big[2[1-\lambda-\sigma\overline{\sigma}(1-2\lambda)]\Big] \ . \end{aligned} \quad (12)$$

INDIVIDUAL BUFFERS

In this section, we summarize results for an analogous model as described above, with the only exception that packets cannot be blocked by packets of other types. This thus corresponds, as illustrated in Fig. 3, to a system with two individual single-server queues with single-slot service times, one with mean arrival rate $\sigma\lambda$ and the other with mean arrival rate $\overline{\sigma}\lambda$.

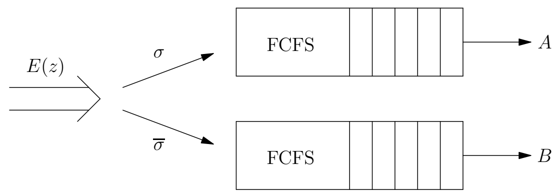

Figure 3: Illustration of the System with Individual Buffers

Stability Condition

As the stability conditions for the first and second individual queue read

$$\sigma\lambda < 1 \ ,$$

and

$$\overline{\sigma}\lambda < 1 \ ,$$

respectively (Bruneel and Kim 1993; Takagi 1993), the stability condition for the entire system is given by

$$\lambda < \min\left\{\frac{1}{\sigma}, \frac{1}{\overline{\sigma}}\right\} \ . \quad (13)$$

This should be compared with the inequality (1) in case of one common queue. It is straightforward to prove that

$$\frac{1-\sigma\overline{\sigma}}{1-2\sigma\overline{\sigma}} \leq \min\left(\frac{1}{\sigma}, \frac{1}{\overline{\sigma}}\right) \ , \quad (14)$$

for $0 < \sigma < 1$. Hence, in general, the stability condition is more stringent in case of blocking, meaning that the maximum tolerable arrival rate is smaller. When $\sigma = 0$ or $\sigma = 1$, both systems are equivalent with a single-server queue fed by an arrival process with mean arrival rate λ and stability condition $\lambda < 1$. It is worth noting that the inequality (14) also implies that the stability condition (1) in case of blocking not only guarantees global stability (for the total system content), but also individual stability for each type of packets.

Mean System Content

The pgf of the system content at random slot boundaries in one individual queue is given by the well-known formula for a discrete-time single-server system with service times of one slot (Bruneel and Kim 1993; Takagi 1993):

$$U_i(z) = \frac{\left(1-E_i'(1)\right)(z-1)E_i(z)}{z-E_i(z)} \ , \quad i = 1, 2 \ ,$$

where $U_i(z)$ is the pgf of the system content in the class-i queue. The corresponding mean system content of type i is given by

$$E[u_i] = U_i'(1) = E_i'(1) + \frac{E_i''(1)}{2\left[1-E_i'(1)\right]} \ .$$

Here the function $E_i(z)$ $(i = 1, 2)$ denotes the pgf of the number of type-i arrivals per slot. In view of the lack of interclass correlation in the arrival process, $E_1(z)$ and $E_2(z)$ are given by

$$E_1(z) = E(\overline{\sigma} + \sigma z) \ ,$$
$$E_2(z) = E(\sigma + \overline{\sigma} z) \ ,$$

which leads to the following explicit expressions for $E[u_1]$ and $E[u_2]$:

$$E[u_1] = \sigma\lambda + \frac{\sigma^2 E''(1)}{2(1-\sigma\lambda)} \ ,$$
$$E[u_2] = \overline{\sigma}\lambda + \frac{\overline{\sigma}^2 E''(1)}{2(1-\overline{\sigma}\lambda)} \ .$$

The system contents in buffers 1 and 2 are, in general, not independent. As a result, the pgf of the total system content is (in general) not just the product of the individual pgf's. However, the mean value of the total system content ($\mathrm{E}[u]$) is always given by the sum of the mean values of the system contents in the individual queues, leading to

$$\mathrm{E}[u] = \lambda + E^{''}(1)\left(\frac{\sigma^2}{2[1-\sigma\lambda]} + \frac{\overline{\sigma}^2}{2[1-\overline{\sigma}\lambda]}\right) \ . \qquad (15)$$

INFLUENCE OF BLOCKING AND RELATIVE TRAFFIC DISTRIBUTION

In this section, we apply our results to investigate the influence of blocking (common queue) and the relative traffic distribution on the behavior of the node. We have therefore first depicted the maximum tolerable arrival rate λ, as derived from equations (1) and (13), versus the traffic-distribution parameter σ in Fig. 4. On the other hand, the total mean system content $\mathrm{E}[u]$, as defined by equations (12) and (15), is shown versus λ and versus σ in Figs. 5 and 6 respectively. Curves are plotted for the common queue (i.e., blocking) as well as for individual queues (i.e., without blocking). In Figs. 5 and 6, we have assumed that the number of arrivals per slot has a geometric distribution with mean λ, i.e.,

$$E(z) = \frac{1}{1+\lambda-\lambda z} \ .$$

We observe from Fig. 5 that blocking only has a minor impact on the mean system content when $\lambda < 1$. When $\lambda < 1$, fewer packets compete for transmission, so that an arriving packet more probably enters a sparsely populated node, and thus experiences nearly no blocking of packets of the other type. On the other hand, when $\lambda > 1$, an arriving packet is more likely to arrive in a densely populated node and is thus more likely to be hindered by packets of the other type. This leads to a larger mean system content and a smaller range of tolerable combinations of λ and σ as compared to two individual queues. Note, for instance, that when $\sigma = 0.5$, the maximum tolerable arrival rate is 1.5 in case of one common queue instead of 2 in case of individual queues (Fig. 4). Indeed, the probability that the second eldest packet is of the same type as the eldest equals 0.5, in which case only one output channel transmits, whereas with probability 0.5 the two eldest packets are of distinct types so that both output channels are active. Hence, the mean number of active output channels per slot in a nearly saturated node with one common queue and $\sigma = 0.5$ equals $1{\cdot}0.5{+}2{\cdot}0.5 = 1.5$. The figures further exhibit that the node, regardless of whether a common queue is adopted or not, performs best when $\sigma = 0.5$, in terms of smaller $\mathrm{E}[u]$ and larger maximum tolerable arrival rate. The reason is that both output channels transmit an equal fraction of the packets, i.e., work is spread fairly amongst the output channels. Figs. 4 and 5 further show that, regardless of the policy, the node performs worst when $\sigma = 0$ (or, $\sigma = 1$). In these cases, all packets have to be transmitted by the same output channel, whereas the other output channel is superfluous. As a result, the node (system) degrades to a node with one output channel (single-server system), with stability condition $\lambda < 1$, which is reflected clearly in Figs. 4 and 5.

Note that we have only shown values of $\sigma \leq 0.5$ in Fig. 5. The reason is that $\sigma = \alpha$ and $\sigma = 1-\alpha$ $(0 \leq \alpha \leq 1)$ lead to the same results, even in case of a common queue. The key observation to understand the latter is that for the operation of the node the exact types of the two eldest packets are irrelevant: only equality or non equality of these two types is of importance. Whether the two eldest packets are both of type 1 or both of type 2, the node only transmits one packet anyway. In fact, a node with $\sigma = 1-\alpha$ can be conceived as a node with $\sigma = \alpha$ whereby the names of the types 1 and 2 have been "swapped". There thus exists a kind of symmetry in the packet types around the value $\sigma = 0.5$. This symmetry can be observed

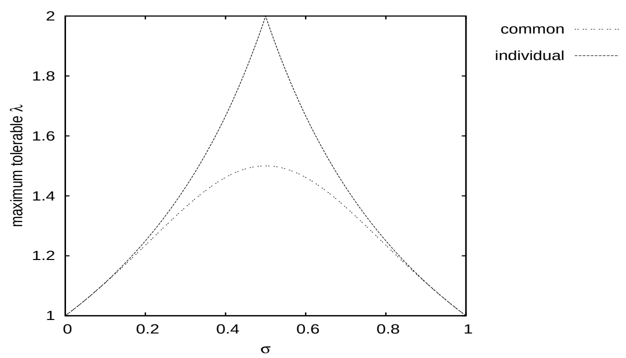

Figure 4: Maximum Tolerable Arrival Rate λ versus the Traffic-Distribution Parameter σ

clearly in Figs. 4 and 6: the curves are symmetric around their best case $\sigma = 0.5$ and the more σ differs from 0.5, the worse the node behaves. Indeed, the more σ differs from 0.5, the more the node becomes similar to a node

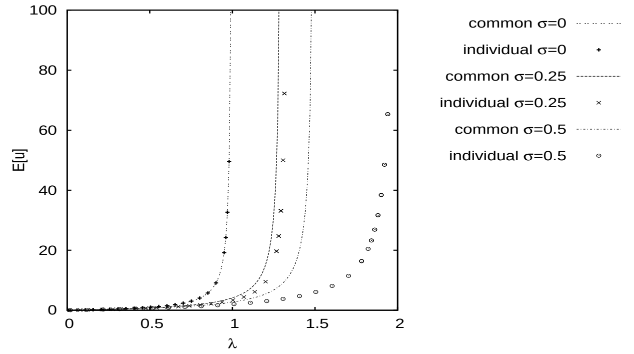

Figure 5: Mean System Content $\mathrm{E}[u]$ versus the Mean Arrival Rate λ, for Various Values of the Traffic-Distribution Parameter σ

with single output channel. As σ deviates more from 0.5, the stability condition is eventually violated for $\lambda = 1.4$ in Fig. 6, which explains the vertical asymptotes in that case.

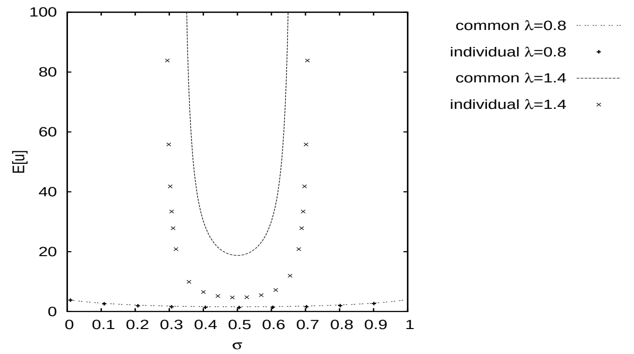

Figure 6: Mean System Content $\mathrm{E}[u]$ versus the Traffic-Distribution Parameter σ, for Various Values of the Mean Arrival Rate λ

CONCLUSIONS AND FUTURE RESEARCH

In this paper, we have developed a discrete-time queueing model for nodes with blocking, caused by accomodating packets that require transmission over different output channels in one common buffer. In the model, two types (classes - corresponding to destinations) of packets, both to be served (transmitted) by their own dedicated server (output channel), are accommodated in one common queue (buffer). We have assumed that the total aggregated arrival stream of new packets during consecutive slots forms an independent and identically distributed process. In addition, a packet belongs to the first class with probability σ and to the second class with probability $\overline{\sigma}$, independently from packet to packet. We have deduced the stability condition and have calculated the pgf of the number of packets in the node, whereafter we have compared these results with those whereby individual buffers are provided for the two output channels (i.e., no blocking).

We have demonstrated that when the mean total arrival rate (λ) is smaller than 1, a common buffer only has a minor impact, whereas the opposite holds when $\lambda > 1$: more packets are waiting in the node and the range of tolerable combinations of arrival rate and relative-traffic parameter σ is narrower. We have also shown that the performance is symmetric around $\sigma = 0.5$. When $\sigma = 0.5$, the work is spread fairly amongst both output channels. The more σ differs from 0.5, the more the node degrades to a node with one output channel (single-server queue), and a node with $\sigma = \alpha$ and a node with $\sigma = 1-\alpha$ ($0 \leq \alpha \leq 1$) lead to the same results, because a node with $\sigma = 1-\alpha$ can be conceived as a node with $\sigma = \alpha$ whereby the names of the packet types have been "swapped".

There are a number of possible extensions to this work. First, the independence assumption of the types of consecutive packets in the arrival stream could be relaxed. The simplest possible extension in this respect would probably be to assume that the types of consecutive packets form a first-order Markov chain. This comes down to assuming that the probability that the next packet belongs to class 1 or 2 depends on the type of the previous packet. In fact, a very specific special case of this kind of model was considered in our earlier paper (Bruneel et al. 2012), where we introduced a "cluster parameter" in the description of the arrival process, which denotes the probability that the next packet has the *same type* as the previous packet. In (Bruneel et al. 2012), however, we assumed that the cluster parameter did not depend on the type of the previous packet, which basically comes down to assuming equal loads for both packet classes. This could be relaxed to two class-dependent cluster parameters, i.e., arbitrary transition probabilities for the Markov chain mentioned above. However, it is to be expected that the analysis of this more general case would be considerably more complicated than the analyses in [3] and in the current paper.

Of course, even more general assumptions than first-order Markov could be envisaged, such as alternating periods (of random length) in which only packets of type 1 or 2 respectively, arrive in the system, and so on.

Another restriction of the current work is the assumption that all service times are deterministically equal to 1 slot. Although this assumption greatly simplifies the analysis of the model, it does imply that packets can never "overtake" each other while being served. If the service times, however, were random (and, hence variable), the latter phenomenon could occur and possibly affect the blocking in the system. We plan to tackle such generalizations of the model in future research.

REFERENCES

Beekhuizen, P. and J. Resing. 2009. "Performance Analysis of Small Non-Uniform Packet Switches". *Performance Evaluation*, No.66, 640–659.

Bruneel, H. and B.G. Kim. 1993. *Discrete-Time Models for Communication Systems Including ATM*. Kluwer Academic, Boston, USA.

Bruneel, H.; Mélange, W.; Steyaert, B.; Claeys, D. and J. Walraevens. 2012. "Impact of Blocking when Customers of Different Classes are Accommodated in One Common Queue". In *Proceedings of the 1st International Conference on Operations Research and Enterprise Systems (ICORES)* (Villamoura, Portugal, Feb.4–6). ICORES'12, 31–38.

Laevens, K. 1999. "A Processor-Sharing Model for Input-Buffered ATM-Switches in a Correlated Traffic Environment". *Microprocessors and Microsystems*, No.22, 589–596.

Liew, S. 1994. "Performance of Various Input-Buffered and Output-Buffered ATM Switch Design Principles under Bursty Traffic: Simulation Study". *IEEE Transactions on Communications*, No.42, 1371–1379.

Mandelbaum, A. and M.I. Reiman. 1998. "On Pooling in Queueing Networks". *Management Science*, No.44, 971–981.

Stolyar, A.L. 2004. "MaxWeight Scheduling in a Generalized Switch: State Space Collapse and Workload Minimization in Heavy Traffic". *Annals of Applied Probability*, No.14, 1–53.

Takagi, H. 1993. *Queueing Analysis - Vol. 3: Discrete-Time Systems*. North Holland.

Van Dijk, N.M. and E. Van der Sluis. 2008. "To Pool or Not To Pool in Call Centers". *Production and Operations Management*, No.17, 1–10.

Van Woensel, T. and N. Vandaele. 2006. "Empirical Validation of a Queueing Approach to Uninterrupted Traffic Flows". *4OR, A Quarterly Journal of Operations Research*, No.4, 59–72.

Van Woensel, T. and N. Vandaele. 2007. "Modeling Traffic Flows with Queueing Models: A Review". *Asia-Pacific Journal of Operational Research*, No.24, 435–461.

BIOGRAPHY

HERWIG BRUNEEL was born in Zottegem, Belgium, in 1954. He received the masters degree in Electrical Engineering, the masters 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 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 400 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, 1993). From October 2001 to September 2003, he served as the Academic Director for Research Affairs at Ghent University. Since 2009, he holds a career-long Methusalem grant from the Flemish Government at Ghent University, specifically on Stochastic Modeling and Analysis of Communication Systems.

WILLEM MÉLANGE was born in Kortrijk, Belgium, in 1987. He received the degree of Master in Computer Sciences in 2010 from Ghent University. In 2010, he joined the SMACS Research Group, Department of Telecommunications and Information Processing, at the same university. The main topic of his research is the analysis of continuous-time queueing models with global FCFS service discipline and class clustering.

BART STEYAERT was born in Roeselare, Belgium, in 1964. He received the degrees of Licentiate in Physics and Licentiate in Computer Science in 1987 and 1989, respectively, and a Phd in Engineering Sciences in 2008, all from the University of Ghent, Belgium. Since January 1990, he has been working as a researcher at the SMACS Research Group, Department for Telecommunications and Information Processing (TELIN), at the same university, in the context of various research projects. His main research interests include discrete-time queueing models, traffic control, and stochastic modelling of high-speed communications networks.

DIETER CLAEYS obtained the masters degree in Computer Science and the Ph.D. degree in Engineering in 2006 and 2011 respectively, both from Ghent University, Belgium. In 2007, he received the ORBEL award, which is accredited annually to the best Belgian master thesis in the field of operations research. Since 2006, he has been working at the SMACS Research Group within the Department of Telecommunications and Information Processing at Ghent University. His main research interests include the analysis of discrete-time queueing models and its applications to operations research and telecommunications systems.

JORIS WALRAEVENS 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 of Telecommunications and Information Processing, at the same university. In 2007, he performed a research visit at the EURANDOM Research Institute (Eindhoven, The Netherlands). From 2008 onwards, he is a part-time professor at Ghent University. His main research interests include discrete-time queueing models and performance analysis of heterogeneous telecommunication networks.