Priority Queueing with Finite Buffer Size and Randomized Push-Out Mechanism ⋆

Konstantin E. Avrachenkov *

INRIA Sophia Antipolis, 2004, Route des Lucioles, B.P. 93, 06902, France

Nikita O. Vilchevsky

St.Petersburg State Polytechnic University, 29, Polytechnicheskaya Str., St.Petersburg, 195251, Russia

Georgy L. Shevlyakov

Dept. of Information and Communications, Gwangju Institute of Science and Technology (GIST), 1 Oryong-dong, Buk-gu, Gwangju, 500-712, South Korea

Abstract

The non-preemptive priority queueing with a finite buffer is considered. We introduce a randomized push-out buffer management mechanism which allows to control very efficiently the loss probability of priority packets. The packet loss probabilities for priority and non-priority traffic are calculated using the generating function approach. In the particular case of the standard non-randomized push-out scheme we obtain explicit analytic expressions. The theoretical results are illustrated by numerical examples. The randomized push-out scheme is compared with the threshold based push-out scheme. It turns out that the former is much easier to tune than the latter. The proposed scheme can be applied to the Differentiated Services of the Internet.

Key words: non-preemptive priority queueing, finite buffer, randomized push-out

⋆ This work is supported by a Research Grant from INRIA Liapunov Institute.

* Corresponding author.

Email addresses: K.Avrachenkov@sophia.inria.fr (Konstantin E. Avrachenkov), vilch@kit.amd.stu.neva.ru (Nikita O. Vilchevsky), shev@gist.ac.kr (Georgy L. Shevlyakov).

URL: http://www-sop.inria.fr/maestro/personnel/K.Avrachenkov/me.html (Konstantin E. Avrachenkov).

1 Introduction

Consider the non-preemptive priority queueing system with two classes of packets. Class 1 packets have priority over class 2 packets. The packets of class 1 (2) arrive into the buffer according to the Poisson process with rate λ_1 (λ_2, resp.). The service time has the exponential distribution with the same rate μ for each class. The service times are independent of the arrival processes. The buffer has a finite size N and it is shared by both types of traffic. If the buffer is full, a new coming packet of class 1 can push out of the buffer a packet of class 2 with the probability α. Note that if $\alpha = 1$ we retrieve the standard non-randomized push-out mechanism.

The infinite buffer priority queueing was studied thoroughly in the past [11,15,19]. The case of finite buffer priority queueing received considerably less attention. Kapadia et al [12,13] analyzed the M/M/C/K type finite buffer non-preemptive priority queueing with non-randomized push-out mechanism. Bondi [3] analyzed the M/M/1/K type preemptive and non-preemptive priority queueing with the following buffer management schemes: complete partitioning, complete sharing and sharing with minimum allocation. Wagner and Krieger [20] analyzed the M/M/1/K type non-preemptive priority queueing with the complete sharing buffer management scheme and with the class-dependent service rates. Recently Sharma and Virtamo [18] studied a finite buffer priority queueing with complete sharing and complete partitioning buffer management schemes. The novelty of the model in [18] is that the buffer size limits the amount of work and not the number of packets. In [5] Cheng and Akyildiz considered the priority queueing with general service time distributions and a general service discipline function. They analyzed the push-out with threshold as the buffer management scheme. Another push-out scheme with threshold, which makes better utilization of the buffer space, was proposed in [8].

Most of the above works use recursive relations to solve steady state Kolmogorov equations. We use the generating function based approach, which only requires the solution of a linear system of N equations in contrast to approximately $N^2/2$ Kolmogorov equations. Furthermore, the derived system has quasi-triangular form and is solved by efficient recursive formulae. The computational complexity of the recursive formulae is $O(N^2)$ which is significantly less than the computational complexity of the Folding Algorithm $O(N^3 \log_2(N))$ [21] and of the Linear Level Reduction, Block-Gaussian Elimination Algorithms $O(N^4)$ [7,17] for the general level-dependent QBD processes. The logarithmic reduction algorithm [4,16] can also be applied to the present problem. However, one iteration of the logarithmic reduction algorithm has the computational complexity $O(N^3)$. Furthermore, our analytical approach is particularly suited for the exact arithmetics implementation. The latter allows to eliminate numerical instability even in the case of large values of N.

To our best knowledge, the randomized push-out mechanism is analyzed for the first time. In particular, we show that with the randomized push-out it is easy to control the loss probability of priority packets in a very large range. Furthermore, in the particular case of non-randomized push-out we obtain explicit analytic expressions for the loss probabilities that are simpler than the recurrent expressions in Kapadia et al [12,13]. Finally we present some numerical examples and compare the randomized push-out scheme with the threshold based push-out scheme [8]. It turns out that the proposed scheme is easier to tune than the threshold based scheme.

Priority queueing discipline has a number of important applications in telecommunications and computer networks, e.g., Differentiated Services (DiffServ) architecture for the Internet [2,14]. In the context of the DiffServ Expedited Forwarding [10], the proposed scheme can be employed to guarantee the required Quality of Service level for the priority traffic and at the same time to prevent the non-priority traffic from starvation. The major part of traffic in the Assured Forwarding Differentiated Service [9] is carried by TCP, the protocol which adjusts its sending rate based on packet losses. Thus, the randomized push-out priority queueing which provides an easy control of the packet loss probabilities leads to simple and efficient differentiation between AF traffic classes. Since there can be more than two AF traffic classes, it is interesting to consider a generalization of the present two class model to a multi-class model where an incoming packet can push out a packet from a lower priority class with a given probability that depends on the class indices. This is a compelling topic for future research. The derivation of asymptotics for the packet loss probabilities for large values of N is another interesting future research direction.

The present work is a significantly extended version of [1].

2 Main results

Denote by $p(i,n)$ the stationary probability of the event that there are n packets in the queue including i packets of class 1. Let p_0 be the stationary probability of the event that there are no packets in the system. These probabilities satisfy the following stationary Kolmogorov equations:

$$(\lambda_1+\lambda_2)p_0 = \mu p(0,0);$$

- $n=0$

$$(\lambda_1+\lambda_2+\mu)p(0,0) = \mu p(1,1) + \mu p(0,1) + (\lambda_1+\lambda_2)p_0;$$

- $0<n<N$

$$
\begin{aligned}
(\lambda_1+\lambda_2+\mu)p(0,n) &= \mu p(1,n+1) + \mu p(0,n+1) + \lambda_2 p(0,n-1),\\
(\lambda_1+\lambda_2+\mu)p(i,n) &= \mu p(i+1,n+1) + \lambda_1 p(i-1,n-1) + \lambda_2 p(i,n-1),\\
(\lambda_1+\lambda_2+\mu)p(n,n) &= \mu p(n+1,n+1) + \lambda_1 p(n-1,n-1);
\end{aligned}
$$

- $n = N$

$$
\begin{aligned}
(\alpha\lambda_1+\mu)p(0,N) &= \lambda_2 p(0,N-1),\\
(\alpha\lambda_1+\mu)p(i,N) &= \lambda_1 p(i-1,N-1) + \lambda_2 p(i,N-1) + \alpha\lambda_1 p(i-1,N),\\
\mu p(N,N) &= \lambda_1 p(N-1,N-1) + \alpha\lambda_1 p(N-1,N).
\end{aligned}
$$

Next we introduce the generating function for $p(i,n)$ by index i

$$F_n(x) = \sum_{i=0}^{n} p(i,n)x^i.$$

Using the above given Kolmogorov equations, we obtain relations for the generating functions $F_n(x), n = 0, 1, ..., N$:

- $n = 0$

$$(\lambda_1+\lambda_2+\mu)F_0(x) = \frac{\mu}{x}\left[F_1(x) - p(0,1)\right] + \mu p(0,1) + (\lambda_1+\lambda_2)p_0,$$

- $0 < n < N$

$$(\lambda_1+\lambda_2+\mu)F_n(x) = \frac{\mu}{x}\left[F_{n+1}(x) - p(0,n+1)\right] + \mu p(0,n+1) + (\lambda_1 x+\lambda_2)F_{n-1}(x).$$

In particular, we get the following boundary condition

- $n = N$

$$
\begin{aligned}
(\alpha\lambda_1+\mu)F_N(x) - \alpha\lambda_1 p(N,N)x^N &= \\
(\lambda_1 x+\lambda_2)F_{N-1}(x) + \alpha\lambda_1 x F_N(x) &- \alpha\lambda_1 x^{N+1} p(N,N).
\end{aligned}
\tag{1}
$$

Now introduce the generating function for $F_n(x)$ by index n

$$\Phi(x,y) = \sum_{n=0}^{N-1} F_n(x)y^n.$$

The generating function $\Phi(x,y)$ satisfies equation (2) given in Lemma 1 below.

Lemma 1 *The generating function $\Phi(x,y)$ satisfies the following equation*

$$[(\rho+1)xy - xy^2(\rho_1 x+\rho_2)-1]\Phi(x,y) = -y^{N+1}x(\rho_1 x+\rho_2)F_{N-1}(x) + y^N F_N(x) \quad (2)$$

$$+y(x-1)A(y) + (xy-1)\rho p_0,$$

where $\rho_i = \lambda_i/\mu$, $\rho = \rho_1+\rho_2$ and $A(y) = \sum_{n=0}^{N-1} p(0,n+1)y^n$.

In the next theorem we determine the generating function $\Phi(x,y)$.

Theorem 2 *The generating function $\Phi(x,y)$ is given by*

$$\begin{aligned}\Phi(x,y) = & \frac{[1-xy+\alpha\rho_1 xy(x-1)]y^N V_{N-1}(x) + y(x-1)A(y)}{(\rho+1)xy - xy^2(\rho_1 x+\rho_2)-1} \\ & + \frac{[1-xy]x^N y^N p(N,N) + \rho[xy-1]p_0}{(\rho+1)xy - xy^2(\rho_1 x+\rho_2)-1},\end{aligned}$$

where

$$\begin{aligned}V_{N-1}(x) = & \sum_{k=0}^{N-1} x^k p(k,N), \\ A(y) = & -\alpha\rho y^{N-1}p(0,N) \\ & + \sum_{k=1}^{N-1}[\rho_2 y^{N-k}\frac{U_{k-1}(t)}{\rho_1^{(k+1)/2}} - \alpha\rho y^{N-k-1}\frac{U_k(t)}{\rho_1^{k/2}} + \alpha y^{N-k-1}\frac{U_{k-1}(t)}{\rho_1^{(k-1)/2}}]p(k,N) \\ & + \rho_2\frac{U_{N-1}(t)}{\rho_1^{(N+1)/2}}p(N,N)\end{aligned}$$

with $t = (\rho+1-\rho_2 y)/(2\rho_1^{1/2})$, and where probabilities $p(k,N), k=0,...,N$ can be obtained as a solution to the following system of linear equations

- $s=0$

$$\alpha\rho_1 C^1_{N-1}(t_0)p(N-1,N) + \left[\rho C^1_{N-1}(t_0) - {\rho_1}^{1/2}C^1_N(t_0)\right]p(N,N) + \rho{\rho_1}^{(N+1)/2}p_0 = 0,$$

- $0<s<N$

$$\begin{aligned}& \sum_{k=0}^{s-1}\left[\rho\frac{C^{s-k}_{N-s-1}(t_0){\rho_1}^{k+1}}{(-\rho_2)^{k+1}} - {\rho_1}^{3/2}(1+\alpha\rho)\frac{C^{s-k}_{N-s}(t_0){\rho_1}^k}{(-\rho_2)^{k+1}}\right. \\ & \left. +\rho_1\alpha\frac{C^{s-k+1}_{N-s-1}(t_0){\rho_1}^k}{(-\rho_2)^k}\right]p(N-1-k,N) + \alpha{\rho_1}^{s+1}\frac{C^1_{N-s-1}(t_0)}{(-\rho_2)^s}p(N-1-s,N) \\ & + \left[\rho C^{s+1}_{N-s-1}(t_0) - {\rho_1}^{1/2}C^{s+1}_{N-s}(t_0)\right]p(N,N) = 0,\end{aligned}$$

- $s = N$

$$-{\rho_1}^{3/2}(1+\alpha\rho)\sum_{k=0}^{N-1}\frac{C_0^{N-k}(t_0){\rho_1}^k}{(-\rho_2)^{k+1}}p(N-1-k,N) - {\rho_1}^{1/2}C_0^{N+1}(t_0)p(N,N) = 0$$

with $U_n(x)$ and $C_n^{\nu}(x)$ denoting the Chebyshev polynomials of the second kind and the Gegenbauer polynomials [6]*, respectively, and*

$$p_0 = (1-\rho)/(1-\rho^{N+2}), \qquad t_0 = (\rho+1)/(2\rho_1^{1/2}).$$

PROOF: given in Appendix.

Once we know the value of $p(N,N)$, we can calculate the loss probabilities of class 1 and class 2 packets.

Theorem 3 *The loss probabilities of class 1 and class 2 packets are given by the following formulae*

$$P_{loss}^{(1)} = p(N,N) + (1-\alpha)[P_N - p(N,N)], \tag{3}$$

$$P_{loss}^{(2)} = P_N + \alpha\frac{\rho_1}{\rho_2}[P_N - p(N,N)], \tag{4}$$

where

$$P_N = \frac{1-\rho}{1-\rho^{N+2}}\rho^{N+1}.$$

PROOF: A priority packet can be lost either when the whole buffer is filled only with priority packets or when there are some packets of class 2 but with probability $1-\alpha$ the push-out mechanism is not enabled. The probability of the first event is $p(N,N)$ and the probability of the second event is $\sum_{k=0}^{N-1} p(k,N) = P_N - p(N,N)$. Thus, we obtain formula (3).

The stream of lost packets of class 2 consists of the stream of packets with rate $\lambda_2 P_N$ lost when the buffer is full and the stream of packets with rate $\alpha\lambda_1(P_n - p(N,N))$ pushed out by packets of class 1. Since the system is ergodic, we obtain formula (4).

□

Note that if $\alpha = 0$ (no push-out), the loss probabilities for two classes coincide and are equal to P_N. We also would like to note that due to the fact that the service time distribution is the same for the two classes, the expressions for p_0, $F_N(1)$ and $\Phi(1,1)$ could be obtained immediately by elementary considerations.

In the particular case of the non-randomized push-out mechanism, that is, when $\alpha = 1$, we are able to calculate the loss probabilities explicitly.

Theorem 4 *The loss probabilities of class 1 and class 2 packets in the case of non-randomized push-out mechanism are given by*

$$P_{loss}^{(1)} = \rho\rho_1^N \frac{(1-\rho_1)(1-\rho^{N+1})}{(1-\rho_1^{N+1})(1-\rho^{N+2})}, \tag{5}$$

$$P_{loss}^{(2)} = P_N + \frac{\rho_1}{\rho_2}[P_N - P_{loss}^{(1)}]. \tag{6}$$

PROOF: In the case of non-randomized push-out mechanism ($\alpha = 1$), the equation for the generating function (2) takes the form

$$[(\rho+1)xy - xy^2(\rho_1 x + \rho_2) - 1]\Phi(x,y) = y^N[1 - xy + \rho_1 x(x-1)y]F_N(x) \tag{7}$$

$$+y(x-1)A(y) + \rho_1(1-x)x^{N+1}y^{N+1}p(N,N) + (xy-1)\rho p_0.$$

Setting $x = 1$ in (7), and then reducing it by the term $(y-1)$, we get

$$(1-\rho y)\Phi(1,y) = \rho p_0 - y^N F_N(1).$$

Then in the above equation we take subsequently $y = 1$ and $y = 1/\rho$ to obtain

$$(1-\rho)\Phi(1,1) = \rho p_0 - F_N(1) \tag{8}$$

and

$$0 = \rho p_0 - \frac{1}{\rho^N} F_N(1). \tag{9}$$

Solving equations (8) and (9) together with the normalization condition

$$\Phi(1,1) + p_0 + F_N(1) = 1,$$

we obtain the following expressions for p_0, $F_N(1)$ and $\Phi(1,1)$:

$$p_0 = \frac{1-\rho}{1-\rho^{N+2}}, \qquad F_N(1) = \frac{1-\rho}{1-\rho^{N+2}}\rho^{N+1}, \qquad \Phi(1,1) = \frac{1-\rho^{N+1}}{1-\rho^{N+2}}\rho.$$

Next we take $y = 1$ in equation (7) and then reduce it by the term $(x-1)$

$$(1-\rho_1 x)\Phi(x,1) = -(1-\rho_1 x)F_N(x) + A(1) - \rho_1 x^{N+1} p(N,N) + \rho p_0.$$

We now set subsequently $x = 1$ and $x = 1/\rho_1$ in the above equation. This results in the following two equations:

$$(1-\rho_1)\Phi(1,1) = -(1-\rho_1)F_N(1) + A(1) - \rho_1 p(N,N) + \rho p_0, \tag{10}$$

$$0 = A(1) - \frac{1}{\rho_1^N} p(N,N) + \rho p_0. \tag{11}$$

Solving equations (10) and (11), we obtain

$$p(N,N) = \frac{(1-\rho_1)(1-\rho^{N+1})}{(1-\rho_1^{N+1})(1-\rho^{N+2})} \rho \rho_1^N.$$

The loss probability of class 1 packets $P_{loss}^{(1)}$ is given by $p(N,N)$. Then, we note that the stream of lost packets of class 2 consists of the stream of packets with rate $\lambda_2 F_N(1)$ lost when the buffer is full and the stream of packets with rate $\lambda_1(F_N(1) - p(N,N))$ pushed out by packets of class 1. Hence, using the ergodicity property of the system, we obtain formula (6) for $P_{loss}^{(2)}$.

□

3 Numerical Examples and Conclusions

In order to calculate the coefficients of the linear system for $p(i,N), i = 0, ..., N$ in Theorem 2, we need to compute the Gegenbauer polynomials. We suggest to use the recursive formulae

$$(n+1)C_{n+1}^k(t) = 2(n+k)tC_n^k(t) - (n+2k-1)C_{n-1}^k(t),$$

with $C_0^k(t) = 1$ and $C_1^k = 2kt$ [6, v.2, p.175]. Next we note that the system of linear equations in Theorem 2 can be written in the following form

$$\begin{bmatrix} \underline{a}^T & 1 \\ A & \underline{b} \end{bmatrix} \begin{bmatrix} \underline{p} \\ p(N,N) \end{bmatrix} = -c \begin{bmatrix} 0 \\ \underline{e} \end{bmatrix}, \tag{12}$$

where $\underline{p} = [p(0,N), ..., p(N-1,N)]^T \in \mathrm{IR}^{\mathrm{N}\times 1}$, $\underline{e} = [0, ..., 0, 1]^T \in \mathrm{IR}^{\mathrm{N}\times 1}$, $c = \rho \rho_1^{(N+1)/2} p_0$, $\underline{a}^T \in \mathrm{IR}^{1\times \mathrm{N}}$ with $a_i = (1+\alpha\rho)(-\rho_1/\rho_2)^{N-i+1}$, and $\underline{b} \in \mathrm{IR}^{\mathrm{N}\times 1}$ with $b_i = \rho C_{i-1}^{N-i+1}(t_0) - \rho_1^{1/2} C_i^{N-i+1}(t_0)$. The matrix A is triangular with the entries

$$a_{ij} = \begin{cases} [\rho C_{i-1}^{j-i}(t_0) - \rho_1^{1/2}(1+\alpha\rho)C_i^{j-i}(t_0) - \alpha\rho_2 C_{i-1}^{j-i+1}(t_0)](-\rho_1/\rho_2)^{N-j+1}, & \text{if } j > i, \\ \alpha\rho_1 C_{i-1}^1(t_0)(-\rho_1/\rho_2)^{N-i}, & \text{if } j = i, \\ 0, & \text{if } j < i. \end{cases}$$

The solution of (12) can be written as follows:

$$p(N,N) = c \frac{\underline{a}^T A^{-1} \underline{e}}{1 - \underline{a}^T A^{-1} \underline{b}},$$

$$\underline{p} = -p(N,N)A^{-1}\underline{b} - cA^{-1}\underline{e}.$$

Let us introduce a vector $\underline{v}$ such that $A^T\underline{v} = \underline{a}$. Then,

$$p(N,N) = c\frac{v_N}{1-\underline{v}^T\underline{b}}. \tag{13}$$

Since A has a triangular structure, the elements of the vector v are easily calculated by the recursive formulae

$$v_j = \frac{1}{\alpha\rho_2 C^1_{j-1}(t_0)}\times \tag{14}$$

$$\left[\sum_{i=1}^{j-1}\left(\rho C^{j-i}_{i-1}(t_0) - \rho_1^{1/2}(1+\alpha\rho)C^{j-i}_{i}(t_0) - \alpha\rho_2 C^{j-i+1}_{i-1}(t_0)\right)v_i - 1 - \alpha\rho\right],$$

for $j = 1, ..., N$, with $v_1 = -(1+\alpha\rho)/(\alpha\rho_2)$.

We would like to note that the computational complexity of the proposed scheme is $O(N^2)$ which is significantly less than the computational complexity of the Folding Algorithm $O(N^3\log_2(N))$ [21] and of the Linear Level Reduction, Block-Gaussian Elimination Algorithms $O(N^4)$ [7,17] for the general level-dependent QBD processes. The logarithmic reduction algorithm [4,16] is also applicable to the present problem. However, one iteration of the logarithmic reduction algorithm has the computational complexity $O(N^3)$. Furthermore, our analytical approach is particularly suited for the exact arithmetics implementation. The latter allows to eliminate numerical instability even in the case of large values of N.

Let us now consider a numerical example with the following values for the system parameters: $\rho_1 = 0.2$, $\rho_2 = 0.9$ and $N = 30$. This is a typical scenario when the intensity of arrival of higher priority packets is smaller than the intensity of arrival of lower priority packets. Once the value of $p(N,N)$ is computed by formulae (14) and (13), we can calculate packet loss probabilities by the formulae given in Theorem 3. In Figure 1 we plot the packet loss probabilities for two classes as a function of parameter α. In the particular cases, $\alpha = 0$ and $\alpha = 1$, we can calculate the loss probabilities using the explicit analytic formulae (the formula for P_N and the formulae in Theorem 4). As one can see, the numerical solutions for $\alpha = 0$ and $\alpha = 1$ coincide with the explicit analytical solutions.

There are at least two important conclusions that we can draw from Figure 1. First, by changing parameter α we tune the loss probability of the priority packets in a very large range, that is, in our particular example, from the order 10^{-22} to 10^{-1}. At the same time, we note that with the increase of α the loss of non-priority packets does not deteriorate as quickly as the acceptance of priority packets improves. Namely, the loss probability of the non-priority packets only changes by 22%. Second, in the considered scenario we note that the dependence of the packet loss probabilities for both classes on the parameter α is very close to linear. In fact, for this particular example the relative error between

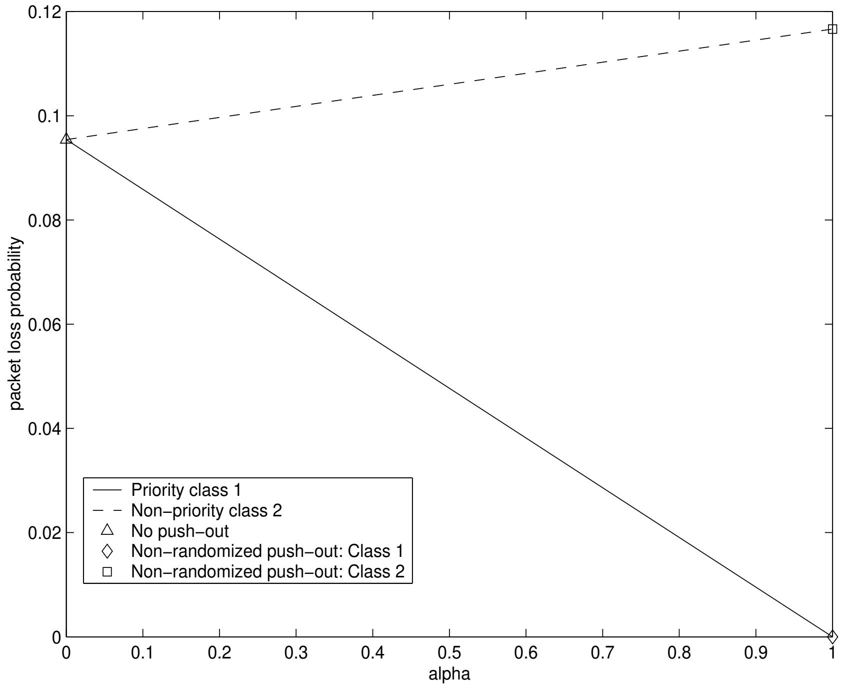

Fig. 1. Numerical example with $\rho_1 = 0.2$, $\rho_2 = 0.9$ and $N = 30$.

the calculated values and the linear approximation is of the order 10^{-7}. Of course, the dependence of the packet loss probabilities on α is not close to linear in all cases. This dependence is significantly non-linear when the high rate of the priority traffic leads to starvation of the low priority traffic (see Figure 2). We note that in the intermediate case when ρ_1 and ρ_2 are close to each other (e.g., see the plots in Figure 3 for $\rho_1 = 0.45$ and $\rho_2 = 0.5$) the dependence of loss probabilities on α is still close to linear. As we vary the buffer size N, the shapes of curves in Fugures 1-3 remain qualitatively same.

Thus, in the case of no starvation of the non-priority traffic the randomized push-out mechanism can easily be applied for the engineering of the priority queueing systems. Namely, one calculates the packet loss probabilities by the analytic formulae for the boundary points $\alpha = 0$ and $\alpha = 1$ and then one uses the linear approximation for $0 < \alpha < 1$.

Finally, we would like to compare the randomized push-out scheme with the threshold based push-out scheme proposed in [8]. In the push-out scheme proposed in [8] the priority and non-priority traffic also share a common buffer. Furthermore, when the buffer is full, an arriving priority packet can push out a non-priority packet if the number of non-prioiry packets in the buffer is above a given threshold. In Figure 4 we plot the packet loss probabilities of the priority and non-priority traffic for different values of the threshold. To compare with the randomized push-out scheme, we take the same values of the parameters: $\rho_1 = 0.2$, $\rho_2 = 0.9$ and $N = 30$. One can see that the threshold based scheme is too sensitive for the threshold values close to 30. One may also prefer the randomized push-out scheme over the threshold based push-out scheme because it allows continuous tuning of the loss probabilities, whereas in the threshold based scheme the packet loss probabilities take their values from a discrete set.

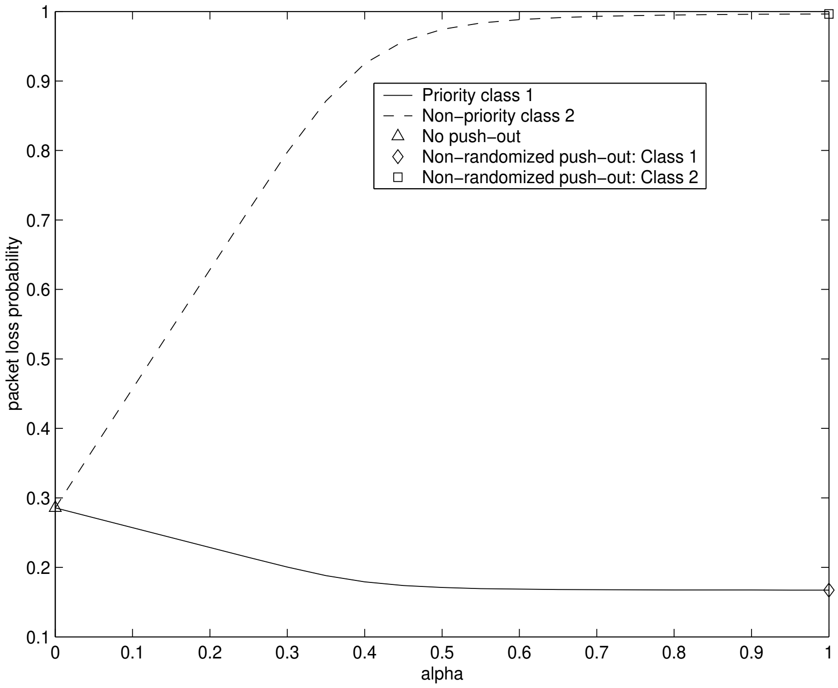

Fig. 2. Numerical example with $\rho_1 = 1.2$, $\rho_2 = 0.2$ and $N = 30$.

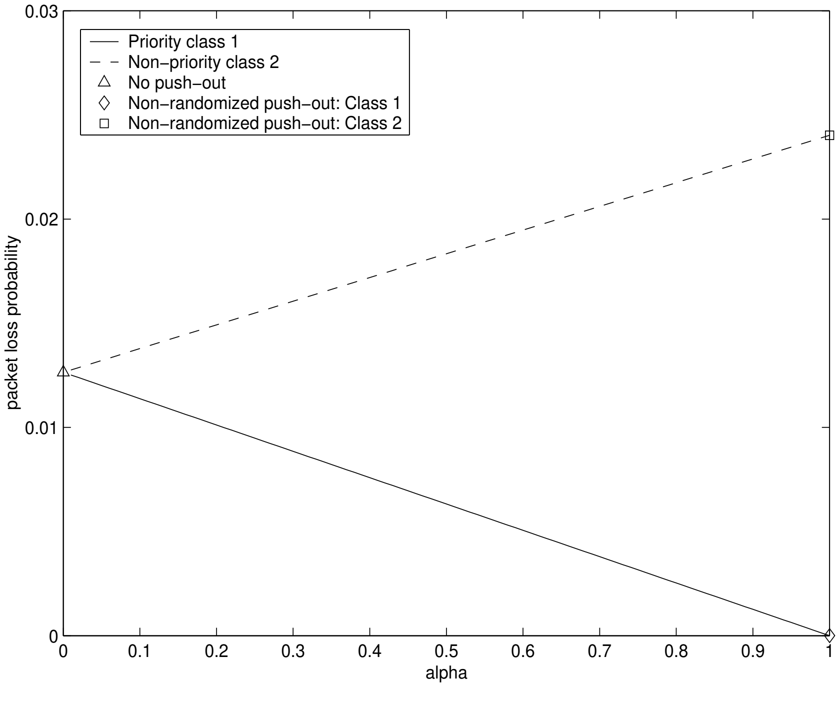

Fig. 3. Numerical example with $\rho_1 = 0.45$, $\rho_2 = 0.5$ and $N = 30$.

Acknowledgments

The authors would like to thank N. Hegde for providing the data for Figure 4 and A. Verbenko for numerical computations.

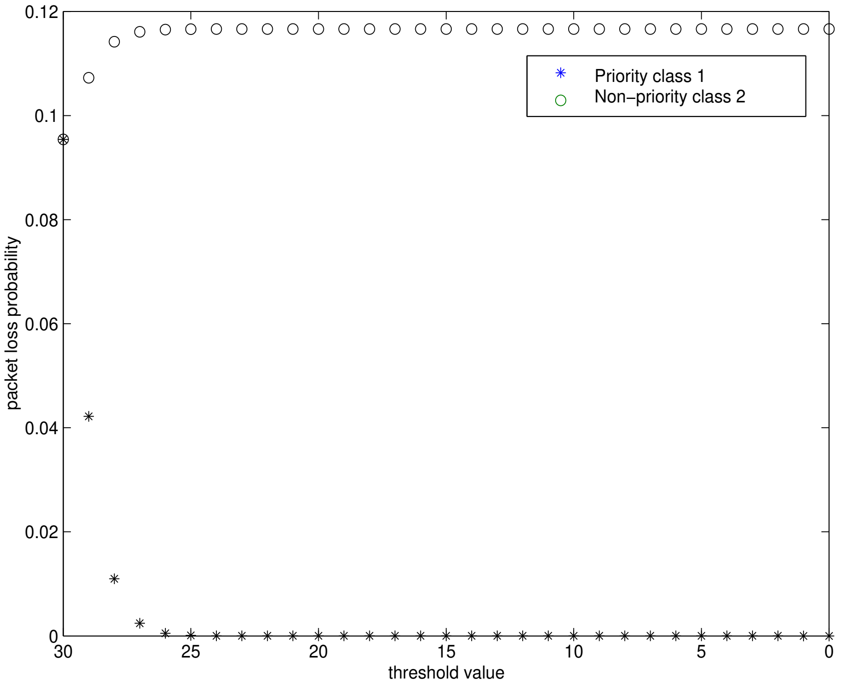

Fig. 4. Threshold based scheme with $\rho_1 = 0.2$, $\rho_2 = 0.9$ and $N = 30$.

Appendix. Proof of Theorem 2

By substituting boundary condition (1) into equation (2) for the generating function $\Phi(x, y)$, we get

$$[(\rho+1)xy - xy^2(\rho_1 x + \rho_2) - 1]\Phi(x, y) = [1 - xy + \alpha\rho_1 xy(x-1)]y^N V_{N-1}(x)$$

$$+[1 - xy]x^N y^N p(N, N) + y(x-1)A(y) + \rho[xy-1]p_0, \tag{15}$$

where $V_{N-1}(x) = \sum_{i=0}^{N-1} x^i p(i, N)$, and hence the expression for $\Phi(x, y)$.

Next, we set $z := xy$ and rewrite equation (15) as follows:

$$[(\rho_1 + \rho_2 + 1)z - \rho_1 z^2 - \rho_2 yz - 1]\Phi\left(\frac{z}{y}, y\right) = [(1-z)y + \rho_1\alpha(z-y)z]y^{N-1}V_{N-1}\left(\frac{z}{y}\right)$$

$$+(z-y)A(y) + (1-z)z^N p(N, N) + \rho(z-1)p_0.$$

Let us now consider the analyticity condition for the generating function $\Phi(z/y, y)$. Namely, the following two conditions have to be satisfied simultaneously

$$(\rho_1 + \rho_2 + 1)z - \rho_1 z^2 - \rho_2 yz - 1 = 0,$$

$$[(1-z)y + \rho_1\alpha(z-y)z]y^{N-1}V_{N-1}\left(\frac{z}{y}\right) + (z-y)A(y) + (1-z)z^N P(N, N) + \rho(z-1)p_0 = 0.$$

The first condition can be rewritten as

$$\rho_2(y-z)z = (1-z)(\rho z - 1),$$

which gives

$$y - z = \frac{(1-z)(\rho z-1)}{\rho_2 z}.$$

Substitute the above expression for $y-z$ into the first two terms of the second analyticity condition and then reduce it by $1-z$, to get

$$\left(y - \frac{\rho_1}{\rho_2}\alpha(\rho z - 1)\right) y^{N-1} V_{N-1}\left(\frac{z}{y}\right) - \frac{\rho z - 1}{\rho_2 z} A(y) + z^N p(N,N) - (\rho_1+\rho_2)p_0 = 0. \quad (16)$$

Next we denote by a and b the roots of the following quadratic equation with respect to the variable z

$$(\rho_1+\rho_2+1)z - \rho_1 z^2 - \rho_2 yz - 1 = 0.$$

Substitute subsequently the roots a and b into (16), which allows us to eliminate $A(y)$

$$\frac{\rho b-1}{b}\left(y - \frac{\rho_1}{\rho_2}\alpha(\rho a-1)\right) y^{N-1}V_{N-1}\left(\frac{a}{y}\right) - \frac{\rho a-1}{a}\left(y - \frac{\rho_1}{\rho_2}\alpha(\rho b-1)\right) y^{N-1}V_{N-1}\left(\frac{b}{y}\right)$$

$$+\left(\frac{\rho b-1}{b}a^N - \frac{\rho a-1}{a}b^N\right)p(N,N) - \rho\left(\frac{\rho b-1}{b} - \frac{\rho a-1}{a}\right)p_0 = 0.$$

Taking into account the properties of the roots of the quadratic equation

$$ab = 1/\rho_1, \quad (\rho a-1)(\rho b-1) = \frac{\rho_2}{\rho_1}(\rho y - 1),$$

we have

$$((\rho-\rho_1 a)y - q(\rho y-1)\rho_1 a)y^{N-1}V_{N-1}\left(\frac{a}{y}\right) - ((\rho-\rho_1 b)y - q(\rho y-1)\rho_1 b)y^{N-1}V_{N-1}\left(\frac{a}{y}\right)$$

$$+(\rho(a^N-b^N) - \rho_1(a^{N+1}-b^{N+1}))p(N,N) + \rho\rho_1(a-b)p_0 = 0,$$

or, equivalently,

$$\rho y^N\left(V_{N-1}\left(\frac{a}{y}\right) - V_{N-1}\left(\frac{b}{y}\right)\right) - \rho_1(y+q(\rho y-1))\left(aV_{N-1}\left(\frac{a}{y}\right) - bV_{N-1}\left(\frac{b}{y}\right)\right)y^{N-1}$$

$$+(\rho(a^N-b^N) - \rho_1(a^{N+1}-b^{N+1}))p(N,N) + \rho\rho_1(a-b)p_0 = 0.$$

Substituting into the above equation $V_{N-1}(x)$ in the form $V_{N-1}(x) = \sum_{i=0}^{N-1} v_i x^i$, we get

$$\rho y \sum_{i=1}^{N-1} v_i(a^i - b^i)y^{N-1-i} - \rho_1(y+q(\rho y-1))\sum_{i=0}^{N-1} v_i(a^{i+1}-b^{i+1})y^{N-1-i}$$

$$+(\rho(a^N-b^N)-\rho_1(a^{N+1}-b^{N+1}))p(N,N)+\rho\rho_1(a-b)p_0=0. \tag{17}$$

By denoting $\cos\varphi=(\rho+1-\rho_2 y)/(2{\rho_1}^{1/2})$, the roots a and b can be written in the form

$$a=\frac{\exp(i\varphi)}{{\rho_1}^{1/2}},\quad b=\frac{\exp(-i\varphi)}{{\rho_1}^{1/2}}.$$

Then equation (17) can be rewritten as

$$\rho y\sum_{i=1}^{N-1}v_iU_{i-1}(t)\frac{y^{N-1-i}}{{\rho_1}^{i/2}}-\rho_1(y+q(\rho y-1))\sum_{i=0}^{N-1}v_iU_i(t)\frac{y^{N-1-i}}{{\rho_1}^{(i+1)/2}}$$

$$+\left(\rho U_{N-1}(t)\frac{1}{{\rho_1}^{(N)/2}}-\rho_1U_N(t)\frac{1}{{\rho_1}^{(N+1)/2}}\right)p(N,N)+\rho{\rho_1}^{1/2}p_0=0, \tag{18}$$

where $t:=\cos\varphi=(\rho+1-\rho_2 y)/(2{\rho_1}^{1/2})$ and $U_s(t)$ are the Chebyshev polynomials of the second kind [6]

$$U_s(\cos\varphi)=\frac{\sin(s+1)\varphi}{\sin\varphi}.$$

The Taylor series for the function $U_s(t)$ with respect to y, being actually a polynomial in this case, has the following form

$$U_s(t(y))=\sum_{s=0}^{s}\frac{U_s^{(i)}(t_0)}{i!}(-1)^i\frac{\rho_2^i y^i}{2^i{\rho_1}^{i/2}}$$

with $t_0=(\rho+1)(2{\rho_1}^{1/2})$. By changing the order of summation in the expressions

$$\sum_{i=1}^{N-1}v_iU_{i-1}(t)\frac{y^{N-1-i}}{{\rho_1}^{i/2}}=\sum_{l=0}^{N-2}y^l\sum_{k=0}^{l}v_{N-1-k}\frac{U_{N-k-2}^{(l-k)}(t_0)(-\rho_2)^{l-k}}{(l-k)!2^{l-k}{\rho_1}^{(N-1-2k+l)/2}},$$

$$\sum_{i=0}^{N-1}v_iU_i(t)\frac{y^{N-1-i}}{{\rho_1}^{(i+1)/2}}=\sum_{l=0}^{N-1}y^l\sum_{k=0}^{l}v_{N-1-k}\frac{U_{N-k-1}^{(l-k)}(t_0)(-\rho_2)^{l-k}}{(l-k)!2^{l-k}{\rho_1}^{(N-2k+l)/2}},$$

we rewrite equation (18) as follows:

$$\rho\sum_{s=1}^{N-1}y^s\sum_{k=0}^{s-1}v_{N-1-k}\frac{U_{N-k-2}^{(s-k-1)}(t_0)(-\rho_2)^{s-k-1}}{(s-k-1)!2^{s-k-1}{\rho_1}^{(N-2-2k+s)/2}}$$

$$-\rho_1(1+\alpha\rho)\sum_{s=1}^{N}y^s\sum_{k=0}^{s-1}v_{N-1-k}\frac{U_{N-k-1}^{(s-k-1)}(t_0)(-\rho_2)^{s-k-1}}{(s-k-1)!2^{s-k-1}{\rho_1}^{(N-2k+s-1)/2}}$$

$$+\rho_1\alpha\sum_{s=0}^{N-1}y^s\sum_{k=0}^{s}v_{N-1-k}\frac{U_{N-k-1}^{(s-k)}(t_0)(-\rho_2)^{s-k}}{(s-k)!2^{s-k}{\rho_1}^{(N-2k+s)/2}}$$

$$+\left(\rho\sum_{s=0}^{N-1} y^s \frac{U_{N-1}^{(s)}(t_0)(-\rho_2)^s}{(s)!2^s{\rho_1}^{(N+s)/2}} - \rho_1 \sum_{s=0}^{N} y^s \frac{U_N^{(s)}(t_0)(-\rho_2)^s}{(s)!2^s{\rho_1}^{(N+s+1)/2}}\right) p(N,N) + \rho{\rho_1}^{1/2}p_0 = 0.$$

Next we use the relation between the derivatives of the Chebyshev polynomials and Gegenbauer polynomials [6, v.2, p.186]

$$U_n^{(m)}(x) = 2^m m! C_{n-m}^{m+1}(x)$$

to get

$$\rho\sum_{s=1}^{N-1} y^s \sum_{k=0}^{s-1} v_{N-1-k} \frac{C_{N-s-1}^{s-k}(t_0)(-\rho_2)^{s-k-1}}{{\rho_1}^{(N-2-2k+s)/2}}$$
$$-\rho_1(1+\alpha\rho)\sum_{s=1}^{N} y^s \sum_{k=0}^{s-1} v_{N-1-k} \frac{C_{N-s}^{s-k}(t_0)(-\rho_2)^{s-k-1}}{{\rho_1}^{(N-2k+s-1)/2}}$$
$$+\rho_1\alpha\sum_{s=0}^{N-1} y^s \sum_{k=0}^{s} v_{N-1-k} \frac{C_{N-s-1}^{s-k+1}(t_0)(-\rho_2)^{s-k}}{{\rho_1}^{(N-2k+s)/2}}$$
$$+\left(\rho\sum_{s=0}^{N-1} y^s \frac{C_{N-s-1}^{s+1}(t_0)(-\rho_2)^s}{{\rho_1}^{(N+s)/2}} - \rho_1 \sum_{s=0}^{N} y^s \frac{C_{N-s}^{s+1}(t_0)(-\rho_2)^s}{{\rho_1}^{(N+s+1)/2}}\right) p(N,N) + \rho{\rho_1}^{1/2}p_0 = 0.$$

Collecting the terms with the same power of y, we obtain the required system of equations:

- $s = 0$

$$\alpha\rho_1 C_{N-1}^1(t_0)v_{N-1} + \left[\rho C_{N-1}^1(t_0) - {\rho_1}^{1/2}C_N^1(t_0)\right] p(N,N) + \rho{\rho_1}^{(N+1)/2}p_0 = 0,$$

- $0 < s < N$

$$\rho\sum_{k=0}^{s-1} \frac{C_{N-s-1}^{s-k}(t_0){\rho_1}^{k+1}}{(-\rho_2)^{k+1}} v_{N-1-k} - {\rho_1}^{3/2}(1+\alpha\rho)\sum_{k=0}^{s-1} \frac{C_{N-s}^{s-k}(t_0){\rho_1}^{k}}{(-\rho_2)^{k+1}} v_{N-1-k}$$
$$+\alpha\rho_1\sum_{k=0}^{s} \frac{C_{N-s-1}^{s-k+1}(t_0){\rho_1}^{k}}{(-\rho_2)^{k}} v_{N-1-k}$$
$$+\left[\rho C_{N-s-1}^{s+1}(t_0) - {\rho_1}^{1/2}C_{N-s}^{s+1}(t_0)\right] p(N,N) = 0,$$

- $s = N$

$$-{\rho_1}^{3/2}(1+\alpha\rho)\sum_{k=0}^{N-1} \frac{C_0^{N-k}(t_0){\rho_1}^{k}}{(-\rho_2)^{k+1}} v_{N-1-k} - {\rho_1}^{1/2}C_0^{N+1}(t_0)p(N,N) = 0,$$

or, equivalently,

- $s = 0$

$$\alpha\rho_1 C_{N-1}^1(t_0)v_{N-1} + \left[\rho C_{N-1}^1(t_0) - {\rho_1}^{1/2}C_N^1(t_0)\right] p(N,N) + \rho{\rho_1}^{(N+1)/2}p_0 = 0,$$

- $0 < s < N$

$$\sum_{k=0}^{s-1}\left[\rho\frac{C^{s-k}_{N-s-1}(t_0){\rho_1}^{k+1}}{(-\rho_2)^{k+1}} - {\rho_1}^{3/2}(1+\alpha\rho)\frac{C^{s-k}_{N-s}(t_0){\rho_1}^{k}}{(-\rho_2)^{k+1}}\right.$$
$$\left.+\rho_1\alpha\frac{C^{s-k+1}_{N-s-1}(t_0){\rho_1}^{k}}{(-\rho_2)^{k}}\right]v_{N-1-k} + \alpha\frac{C^{1}_{N-s-1}(t_0){\rho_1}^{s+1}}{(-\rho_2)^{s}}v_{N-1-s}$$
$$+\left[\rho C^{s+1}_{N-s-1}(t_0) - {\rho_1}^{1/2}C^{s+1}_{N-s}(t_0)\right]p(N,N) = 0,$$

- $s = N$

$$-{\rho_1}^{3/2}(1+\alpha\rho)\sum_{k=0}^{N-1}\frac{C_0^{N-k}(t_0){\rho_1}^{k}}{(-\rho_2)^{k+1}}v_{N-1-k} - {\rho_1}^{1/2}C_0^{N+1}(t_0)p(N,N) = 0.$$

Finally, to obtain an expression for $A(y)$ in terms of $p(k,N), k = 0, ..., N$ and Chebyshev polynomials, we again substitute subsequently the roots a and b into (16) and subtract one equation from another

$$y^N\sum_{k=0}^{N-1}\frac{a^k - b^k}{y^k}p(k,N) - \frac{\rho_1}{\rho_2}\alpha\rho y^{N-1}\sum_{k=0}^{N-1}\frac{a^{k+1} - b^{k+1}}{y^k}p(k,N)$$

$$+\frac{\rho_1}{\rho_2}\alpha y^{N-1}\sum_{k=0}^{N-1}\frac{a^k - b^k}{y^k}p(k,N) + (a^N - b^N)p(N,N) - \frac{\rho_1}{\rho_2}A(y)(a-b) = 0.$$

As above, taking into account that

$$\frac{a^k - b^k}{a-b} = \frac{U_{k-1}(t)}{\rho_1^{(k-1)/2}},$$

we can express $A(y)$ in terms of $p(k,N), k = 0, ..., N$ and the Chebyshev polynomials of the second type.

□

References

[1] K.E. Avrachenkov, N.O. Vilchevsky, and G.L. Shevlyakov, Priority queueing with finite buffer size and randomized push-out mechanism, Extended Abstract, *ACM SIGMETRICS'03*, 2003, San Diego, 2 pages.

[2] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, An architecture for differentiated services, *RFC* 2475, December 1998.

[3] A.B. Bondi, An analysis of finite capacity queues with priority scheduling and common or reserved waiting areas, *Comput. & Ops. Res.*, 16(3), 1989, pp.217-233.

[4] L. Bright and P. Taylor, Calculating the equilibrium distribution in level dependent quasi-birth-and-death processes, *Stochastic Models*, 11(3), 1995, pp.497-525.

[5] X. Cheng and F. Akyildiz, A finite buffer two class queue with different scheduling and push-out schemes, *IEEE INFOCOM'92*, 1992, pp. 231-241.

[6] A. Erdelyi and H. Bateman, *Higher transcendental functions*, 2nd ed., v.1,2, Robert E. Krieger Publishing Company, 1985.

[7] D.P. Gaver, P.A. Jacobs, and G. Latouche, Finite birth-and-death models in randomly changing environments, *Adv. Appl. Prob.*, 16, 1984, pp.715-731.

[8] N. Hegde and K.E. Avrachenkov, Service differentiation and guarantees for TCP-based elastic traffic, in *From QoS Provisioning to QoS Charging. QofIS/ICQT 2002*, Burkhard Stiller, Michael Smirnow, Martin Karsten, and Peter Reichl, Eds., October 2002, v. 2511 in Lecture Notes in Computer Science, pp. 159–168.

[9] J. Heinanen, F.Baker, W. Weiss, and J. Wroclawski, Assured forwarding PHB group, *RFC* 2597, June 1999.

[10] V. Jacobson, K. Nichols, and K. Poduri, (1999), An expedited forwarding PHB group, *RFC* 2598, June 1999.

[11] N.K. Jaiswal, *Priority queues*, Academic Press, 1968.

[12] A.S. Kapadia, M.F. Kazmi, and A.C. Mitchell, Analysis of a finite capacity non-preemptive priority queue, *Comput. & Ops. Res.*, 11(3), 1984, pp.337-343.

[13] A.S. Kapadia, Y.K. Chiang, and M.F. Kazmi, Finite capacity priority queues with potential health applications, *Comput. & Ops. Res.*, 12(4), 1985, pp.411-420.

[14] K. Kilkki, (1999) *Differentiated services for the Internet*, Macmillan Technical Publishing, 1999.

[15] L. Kleinrock, *Queueing systems, v.2: Computer applications*, Wiley, 1976.

[16] G. Latouche and V. Ramaswami, *J. Appl. Prob.*, 30, 1993, pp.650-674.

[17] G. Latouche and V. Ramaswami, *Introduction to matrix analytic methods in stochastic modeling*, ASA-SIAM Series on Statistics and Applied Probability, 1999.

[18] V. Sharma and J.T. Virtamo, A finite buffer queue with priorities, *Performance Evaluation*, 47, 2002, pp.1-22.

[19] H. Takagi, *Queueing analysis, Vacation and Priority Systems, Part 1.* North Holland, 1991.

[20] D. Wagner and U.R. Krieger, Analysis of a finite buffer with non-preemptive priority scheduling, *Comm.Stat.-Stochastic Models*, 15(2), 1999, pp.345-365.

[21] J. Ye and S.Q. Li, Folding algorithm: A computational method for finite QBD processes with level dependent transitions, *IEEE Trans. Commun.*, 42, 1994, pp.625-639.