Industrial Engineering
& Management Systems
Vol 12, No 3, September 2013, pp.274-280
ISSN 1598-7248 | EISSN 2234-6473 |

http://dx.doi.org/10.7232/iems.2013.12.3.274

A Space Merging Approach to the Analysis of the Performance of Queueing Models with Finite Buffers and Priority Jumps

Youngjin Oh, Chesoong Kim*
Department of Industrial Engineering, Sangji University, Wonju, Korea

Agassi Melikov
Department of Information Technologies, National Academy of Aviation, Baku, Azerbaijan

(Received: March 28, 2013 / Revised: July 5, 2013 / Accepted: September 13, 2013)

ABSTRACT

This paper proposes a space merging approach to studying the queuing models with finite buffers and jump priorities. Upon the arrival of a call with low priority, one call of such kind is assumed to be transferred to the end of the queue of high priority calls. The transfer probabilities depend on the state of the queue of the heterogeneous calls. We developed the algorithms to calculate the quality of service metrics of such queuing models, and the results of the numerical experiments are shown.

Keywords: Space Merging, Queueing, Performance, Priority Jumps Style

* Corresponding Author, E-mail: dowoo@sangji.ac.kr

1. INTRODUCTION

In modern packet-switching networks, heterogeneous traffic (voice, video, computer data, etc.) imposes various demands with respect to the quality of service (QoS) metrics. As a result, whereas the real-time traffic is sensitive to possible delays, the non-real-time traffic demands the loss to be as small as possible with respect to its packets. An effective way to satisfy the conflicting requirements of heterogeneous calls (packets) is to use priority schemes. Various types of priority schemes in queuing models have already been thoroughly investigated. Detailed information on the subject can be found in the book of Kleinrock (1976). Therefore, the known results are skipped here. Instead, we provide only the necessary facts that are connected with priority schemes.

Let us briefly consider the works in which similar problems are investigated. First, note that during the last decade, new types of priority schemes, which are called multiple priorities, have been intensively investigated (Demoor *et al.*, 2011; Kim *et al.*, 2007; Lee and Choi, 2001; Melikov *et al.*, 2006). Unlike the classical priority scheme, multiple priorities pertaining to real-time calls have high time priorities and low space priorities, and non-real-time calls have low time priorities and high space priorities. Whereas space priorities are used for the solution of the conflict situations that are connected with the occupation of the buffer store upon the arrival of calls, time priorities define the call selection order from the buffer to transfer to the channel.

On the basis of variability, classical priority schemes can be divided into the following classes: static, dynamic on-time and dynamic on-state. Static priorities (preemptive or non-preemptive) do not change during the operation of the system. Thus, in the case of non-preemptive static priority scheme, for example, traffic has a fixed priority level, and at the moment of channel clearing, the call is removed from the head of the queue that has the highest priority level among all of the nonempty queues. These priorities are called "HOL-priorities" (where "HOL" is an abbreviation for "Head-Of-Line"). In the dynamic on-time priority scheme, the priority level of a call of each type changes (i.e., it increases or decreases), depending on its waiting time in the queue (Kleinrock, 1964). Various functions that define the law of changes in these priorities can be used. In the dynamic on-state priority scheme (situational priorities), the priority level of a call of each type is defined, depending on the state of the system, where the state of the system is usually set by means of a vector whose components specify the number of heterogeneous calls in the system (Ponomarenko *et al.*, 2010).

In recent years, new types of HOL-priorities were studied (Lim and Kobza, 1990; Maertens *et al.*, 2006a, 2006b, 2007, 2008; Walraevens *et al.*, 2003). The authors of the work (Lim and Kobza, 1990) denoted the investigated priority scheme as "Head-Of-Line with Priority Jumps" (HOL-PJ). Briefly, the model consists of a single channel system with infinite separate queues and N types of traffic, $N > 1$ (Lim and Kobza, 1990). Calls of type i are assumed to have high (non-preemptive) priority level over calls of type $i+1,\ i=1, 2, \cdots, N-1$. For the traffic i, the parameters $D_i,\ 0 < D_1 < D_2 < \cdots < D_N \le \infty$ are defined. If the waiting time of a call i (at the head of queue i) reaches the value $D_i - D_{i-1}$, the call jumps to the end of queue $i-1,\ i=2, \cdots, N.$ This procedure continues until a call of any type reaches the queue with the highest priority level. An exact analysis of the state of the queues and the distribution of waiting times in the given model are difficult problems. Therefore, the formulas for the calculation of an average waiting time of the heterogeneous calls are described (Lim and Kobza, 1990). Note that the HOL-PJ formulas described in the paper (Lim and Kobza, 1990) are inconvenient to implement, as they require the use of additional hardware to monitor the waiting time of the heterogeneous calls.

Various types of HOL-PJ for queuing models in discrete time (i.e., systems in which time is divided into time slots) are investigated (Maertens *et al.*, 2006a, 2006b, 2007, 2008; Walraevens *et al.*, 2003). In these models, two types of calls, namely, calls of a high priority (H-calls) and calls of a low priority (L-calls) were offered. For the buffering of calls of each type, there are infinite queues. In the Head-Of-Line Merge-By-Probability (HOL-MBP) scheme at the end of each slot, all L-calls jump to the tail of an H-calls queue according to a probability of $\beta,\ 0 < \beta < 1$. In other words, at the end of each slot, the queues of the H-calls and the L-calls have a probability β of becoming merged.

A scheme to modify HOL-MBP was studied in paper (Maertens *et al.*, 2007). The scheme is called the Head-Of-Line Jump-Or-Serve (HOL-JOS) scheme, and only the L-call at the head of the L-queue can jump to the H-queue, unlike the previous scheme. This possible jump at the beginning of each slot depends on the content of the H-queue at the beginning of the slot, i.e., when the H-queue is non-empty, the call jumps; otherwise, the L-call is immediately transmitted to the channel. In the scheme HOL-JIA^1 (Head-Of-Line Jump-If-Arrival; Maertens *et al.*, 2006b), unlike the scheme HOL-JOS, the possible jump at the end of a slot does not depend exclusively on the contents of the H-queue at the beginning of the slot. Instead, it also depends on the number of types of L-calls that arrive in that slot. Specifically, during a slot in which a call of the H-queue is transmitted, the call at the head of the L-queue jumps to the H-queue if and only if L-calls arrive during that slot. In this scheme, arriving L-calls are not allowed to jump immediately upon arrival. Finally, in the paper (Maertens *et al.*, 2008), the HOL-JIA^2 scheme was proposed. The unique difference between scheme HOL-JIA^1 and scheme HOL-JIA^2 is that in the latter scheme, arriving L-calls are allowed to jump to the H-queue immediately upon arrival. In papers (Maertens *et al.*, 2006a, 2006b, 2007, 2008; Walraevens *et al.*, 2003), the authors have developed formulas for the generating functions of the lengths of the queues of calls of both types and their waiting times in the appropriate queues.

A principal objective of HOL-PJ is the elimination of too long a waiting time in a queue of L-calls in systems with HOL-priorities. This problem is especially acute in systems where the loading of H-calls greatly exceeds the loadings of L-calls. Thus, it is obvious that the introduction of HOL-PJ allows for the solution of the specified problem at the expense of increasing the waiting time of H-calls. Hence, upon the introduction of HOL-PJ, it is necessary to consider the admissible bounds for the increase in the waiting times of H-calls.

Note that the abovementioned papers (Lim and Kobza, 1990; Maertens *et al.*, 2006a, 2006b, 2007, 2008; Walraevens *et al.*, 2003) are devoted to the research of models with infinite queues. These models cannot be accepted as adequate models of realistic telecommunication systems because, as a rule, real telecommunication systems have finite buffers for the temporary storage of heterogeneous calls. In other words, for the widespread use of priorities of type HOL-PJ, it is necessary to define their efficiency in real systems.

This paper proposes a new class of HOL-PJ in systems with finite queues. These priorities allow a jump from an L-queue to an H-queue only upon the arrivals of L-calls and the probability of a jump depends on the number of L-calls in the queue. The introduction of restrictions on the size of the buffers for heterogeneous calls leads, by necessity, to the definition of a new QoS metric: call loss probability (CLP). Another distinction of this paper, compared to papers (Lim and Kobza, 1990; Maertens *et al.*, 2006a, 2006b, 2007, 2008; Walraevens *et al.*, 2003), is that in this paper, the state space merging (SSM) approach (Ponomarenko *et al*., 2010) is used for the system analysis. By using the proposed approach, simple computational procedures are developed to determine all QoS metrics of the investigated models. Note that the alternative approach to the solution of the given problem consists of using the results of paper (Bruneel *et al*., 1994).

This paper is organized as follows. In Section 2, a mathematical model of the investigated system with HOL-PJ is formulated. An approximate method to calculate the QoS metrics of the given model is proposed in Section 3. The results of numerical experiments are discribed in Section 4, considering the problem of selecting the appropriate values of the buffer size for H-calls to satisfy the given constraints to calls transfer delay. Concluding remarks are given in Section 5.

2. MATHEMATICAL MODEL

We consider a continuous time system with two (separate) queues of finite capacity and one (common) transmission channel. Two types of Poisson traffic arrive at the system: whereas the first type of traffic represents the traffic of calls in real-time (H-calls), the second type of traffic is the traffic of calls in non-real-time (L-calls). The intensity of traffic i is equal to λ_i, i = 1, 2 and the channel occupancy time is a random variable that has an exponential distribution, with the parameter μ being used for calls of both types.

In the system, there are separate buffers for waiting in the queue of heterogeneous calls, where the size of the buffer for calls of type i is equal to R_i, $0 < R_i < \infty$, i = 1, 2. The limitation of separate buffers means that if, upon arrival, a call of any type finds its corresponding buffer completely filled, it is lost, irrespective of the condition of the other buffers. H-calls have a high non-preemptive priority level over L-calls, i.e., as long as there are H-calls in the system, this traffic has transmission priority over L-calls, irrespective of both the number of L-calls in the queue and their waiting time in the queue. In each type of traffic, the discipline first-comes first-served is used.

It is clear that the HOL-priority scheme provides good performance for H-calls, and that, at the same time, the performance of L-calls can be severely degraded, especially in the event that the system is highly loaded with H-calls. Thus, for the purpose of increasing the chances of L-calls to be served, a comprehensible time HOL-PJ scheme is introduced. However, note that these priority schemes lead to insignificant worsening of the QoS metrics of H-calls. The basic questions in introducing HOL-PJ are the definition of the moment of the jumps from the L-queue to the H-queue and the number of L-calls transferred to the H-queue. In this paper, the HOL-PJ scheme is defined as follows. First, H-calls are always accepted with probability of 1 if, at the moment of their arrival, there is at least one empty place in the H-buffer; otherwise, H-calls are lost with a probability of 1.

If, upon the arrival of an L-call, the number of calls of the given type in the buffer equals k, $k < R_2$, and if there is an empty place in the H-queue, then with a probability of $\alpha(k)$, one L-call immediately jumps to the tail of the H-queue (to be precise, we will assume that an L-call standing at the head of an L-queue jumps to the H-buffer); with the complementary probability of $1\text{-}\alpha(k)$, the arriving L-call joins the queue if there is an empty place in the queue. In the case of a successful jump, the L-call joins the H-queue, and further, the L-call is served as an H-call in accordance with the HOL-priorities. However, if upon the arrival of an L-call, there is no empty place in the H-queue, then with a probability of 1 the arriving L-call joins the L-queue if there is an empty place in the L-queue; otherwise, with a probability of 1, this L-call is lost.

Let us note some important special schemes introduced above regarding the jump priorities.

1) *The uniform scheme*: In this scheme, the probabilities $\alpha(k)$ are constants and do not depend on the number of L-calls in the buffer, i.e., $\alpha(k) = \alpha$ for any $k = 0, 1, \cdots, R_2$. In the special case of $\alpha(k) = 0$, we obtain the classical HOL-priorities.
2) *The threshold scheme*: In this scheme, the threshold parameters L_i, $i = 1, \cdots, r$ are introduced, and the probabilities $\alpha(k)$ are defined as follows: $\alpha(k) = \alpha_i$, if $L_{i\text{-}1} \le k < L_i$, $i = 0, 1, 2, \cdots, r$. Here, it is assumed that L_0: = 0, L_r: = R_2. These probabilities α_i, $i = 1, \cdots, r$, can be defined in various ways.

3. METHODS OF CALCULATING QoS METRICS

The basic QoS metrics of the investigated system are the call loss probability (CLP_i), the average numbers of calls of each type in the buffers (Q_i) and the average call transfer delay (CTD_i), i = 1, 2.

The state of the buffers at any moment in time can be described by means of a two-dimensional vector $\boldsymbol{n} = (n_1, n_2)$, where n_i denotes the number of calls of type i in the buffer, i = 1, 2. Thus, the operational space of the given system is described by a two-dimensional Markov chain (2D MC) with the following state space:

$$S := \{\boldsymbol{n}: n_i = 0, 1, \cdots, R_i, i = 1, 2\}, \quad (1)$$

Transitions between states of the system occur only upon the arrival of the calls and their departure from the system. Thus, we conclude that the non-negative elements of a Q-matrix of the given 2D MC are defined by the following relations:

$$q(\mathbf{n},\tilde{\mathbf{n}})=\begin{cases}\lambda_1+\lambda_2\alpha(n_2), & if\ \tilde{\mathbf{n}}=\mathbf{n}+\mathbf{e_1}\\ \lambda_2\delta(n_1,R_1)+\lambda_2(1-\alpha(n_2))(1-\delta(n_1,R_1)), & if\ \tilde{\mathbf{n}}=\mathbf{n}+\mathbf{e}_2\\ \mu, & if\ n_1>0,\tilde{\mathbf{n}}=\mathbf{n}-\mathbf{e_1}\ or\ n_1=0,\tilde{\mathbf{n}}=\mathbf{n}-\mathbf{e_2}\\ 0, & in\ other\ cases\end{cases} \tag{2}$$

where $\boldsymbol{e}_1=(1,\ 0)$, $\boldsymbol{e}_2=(0,\ 1)$ and $\delta(x,\ y)$ represents the Kronecker delta function.

The given 2D MC is strictly continuous with respect to the first component, but it is weakly continuous with respect to the second one–for the definitions of these terms, see the appendix of book by Ponomarenko *et al.* (2010). The system of global balance equations (SGBEs) for the steady-state probabilities $p(\boldsymbol{n})$, $\boldsymbol{n}\in S$, might be constructed by using relations (2), because the evidence used to construct this SGBE is not shown. The existence of the stationary regime is proven by the fact that all the states of a finite-dimensional state space S are communicating. To this SGBE, the normalizing condition is added:

$$\sum\nolimits_{n\in s}p(\mathbf{n})=1 \tag{3}$$

The desired QoS metrics are determined via the stationary distribution of the initial model. Therefore, by using the PASTA theorem (Wolff, 1992), we obtain:

$$CLP_1=\sum\nolimits_{k=0}^{R_2}p(R_1,\ k) \tag{4}$$

$$CLP_2=\sum\nolimits_{k=0}^{R_1-1}p(k,\ R_2)(1-\alpha(R_2))+p(R_1,\ R_2) \tag{5}$$

The mean numbers of calls in the buffers are also calculated via the stationary distribution as follows:

$$Q_k=\sum\nolimits_{i=1}^{R_k}i\xi_k(i) \tag{6}$$

where $\xi_k(i)=\sum\nolimits_{n\in S}p(\mathbf{n})\delta(n_k,\ i),\ k=1,\ 2$ are the marginal probability mass functions.

The QoS metrics CTD_k are calculated using a modified version of Little's formula:

$$CTD_k=\frac{Q_k}{\lambda_k(1-CLP_k)},\ k=1,\ 2 \tag{7}$$

The stationary distribution is determined as a result of the solution of the abovementioned SGBE of the given 2D MC. Because the corresponding SGBE has no explicit solution, laborious computational efforts for large values of R_1 and R_2 are required. To overcome these difficulties, a new efficient and refined approximate method of calculating a stationary distribution of the given model is suggested below.

For the correct application of the proposed method, which is based on the SSM approach (Ponomarenko *et al.*, 2010), it is assumed that $\lambda_1>>\lambda_2$, i.e., that $\nu_1>>\nu_2$, where $\nu_j:=\lambda_j/\mu$, $j=1,\ 2$. This assumption is not extraordinary for the investigated models because for systems with high loadings of H-calls, the introduction of priority jumps for L-calls is reasonable.

The following splitting of the state space S is examined:

$$S=\bigcup\nolimits_{i=0}^{R_2}S_i,\ \ S_i\cap S_j=\mathbf{0},\ \ i\neq j \tag{8}$$

where, $S_i=\{n\in S:n_2=i\}$, $i=0,\ 1,\ 2,\ \cdots,\ R_2$.

Note that the assumption above meets the major requirement for the correct use of the SSM approach (Ponomarenko *et al.*, 2010), which is as follows. The state space of the initial model must be split into classes, such that transition probabilities within the classes are essentially higher than those between the states of the different classes. Indeed, from (2) the abovementioned major requirement is observed to be fulfilled when using splitting (8).

Furthermore, the each state class S_i combines into unique separate state i, i.e., in the future, state class S_i replaced by merged state i, $i=0,\ 1,\ \cdots,\ R_2$. Subsequently, the following merging function in state space S is introduced:

$$U(\mathbf{n})=i,\ \text{ if }\mathbf{n}\in S_i \tag{9}$$

Function (9) determines a merged model, which is a 1D MC with the state space $\{i:i=0,\ 1,\ \cdots,\ R_2\}$. Then, according to the SSM approach, the stationary distribution of the initial model approximately equals:

$$p(k,\ i)\approx\rho_i(k)\pi(i) \tag{10}$$

where $\{\rho_i(k):\ k=0,\ 1,\ \cdots,\ R_1\}$ is the stationary distribution of a split model with state space S_i, and $\{\pi(i):\ i=0,\ 1,\ \cdots,\ R_2\}$ is the stationary distribution of the merged model.

By using (2), we conclude that each split model with state space S_i represents a 1D birth-death process (BDP), for which the elements of the generating matrix are obtained as follows:

$$q_i(k_1,\ k_2)=\begin{cases}\lambda_1+\lambda_2\alpha(i), & if\ k_2=k_1+1\\ \mu, & if\ k_2=k_1-1\\ 0, & in\ other\,cases\end{cases}. \tag{11}$$

As a result, the stationary distribution within class S_i is

$$\rho_i(k)=\theta_i^k\cdot\frac{1-\theta_i}{1-\theta_i^{R_1+1}},\ k=0,\ 1,\ 2,\ \cdots,\ R_1 \tag{12}$$

where $\theta_i:=\nu_1+\nu_2\alpha(i)$. To be concise, we give the formulas for only the case $\theta_i\neq 1$.

According to the SSM approach for 2D MC (Ponomarenko *et al.*, 2010), we conclude that the elements of the generating matrix of a merged model, which are denoted by $q(i,\ j),\ i,\ j\in\Omega$, are determined as follows:

$$q(i,\ j)=\sum\nolimits_{\substack{\mathbf{n}\in S_i\\ \tilde{\mathbf{n}}\in S_j}}q(\mathbf{n},\ \tilde{\mathbf{n}})p(\mathbf{n}). \tag{13}$$

Then, from (13), by taking into account (2) and after some algebra is performed, we have

$$q(i_1,i_2)=\begin{cases}\lambda_2(1-\alpha(i))\left(1-\rho_{i_1}(R_1)\right)+\lambda_2\rho_{i_1}(R_1), & \textit{if } i_2=i_1+1\\ \mu\rho_{i_1}(0), & \textit{if } i_2=i_1-1\\ 0, & \textit{in other cases}\end{cases} \quad (14)$$

Formula (14) allows the determination of the stationary distribution of a merged model. The formula coincides with an appropriate distribution of the state probabilities of a 1D BDP, for which the transition intensities are determined in accordance with (14). Consequently, the stationary distribution of a merged model is determined to be as follows:

$$\pi(i)=\prod\nolimits_{j=1}^{i}A_j\pi(0),\ i=1,\ 2,\ \cdots,\ R_2 \quad (15)$$

where,

$$A_j=v_2\cdot\frac{\left(1-\alpha(j-1)\right)\left(1-\rho_{j-1}(R_1)\right)+\rho_{j-1}(R_1)}{\rho_j(0)},$$

$$\pi(0)=1\Big/\left(1+\sum\nolimits_{k=1}^{R_2}\prod\nolimits_{i=1}^{k}A_i\right).$$

Next, by using (12) and (15), and based on (10), the stationary distribution of the initial 2D MC can be determined. Thus, in summary, after performing complex algebraic transformations, which are omitted in this paper, the following approximate formulae for the calculation of the QoS metrics (4)–(6) are proposed:

$$CLP_1\approx\sum\nolimits_{k=0}^{R_2}\rho_k(R_1)\pi(k) \quad (16)$$

$$CLP_2\approx\pi(R_2)\Big(\big(1-\alpha(R_2)\big)\big(1-\rho_{R_2}(R_1)\big)+\rho_{R_2}(R_1)\Big) \quad (17)$$

$$Q_1\approx\sum\nolimits_{k=1}^{R_1}k\sum\nolimits_{i=0}^{R_2}\rho_i(k)\pi(i) \quad (18)$$

$$Q_2\approx\sum\nolimits_{k=1}^{R_2}k\pi(k) \quad (19)$$

After calculating CLP_k and Q_k from (7), we determine $CTD_k,\ k=1,\ 2$.

4. NUMERICAL RESULTS

The formulas developed above allow the behavior of the QoS metrics of an investigated system to be determined with respect to the change in its structural and loading parameters. Due to space limitation, only portions of numerical results are shown in Figures 1–6. In these numerical experiments, the initial data are $R_2=15$, $\lambda_1=5$, $\lambda_2=1$, $\mu=10$. We consider two cases.

Case 1: $\alpha(k)$ are state-dependent and are defined in accordance to the threshold scheme, i.e.,

$$\alpha(k)=\begin{cases}0.2, & \text{if } 0\le k\le 5,\\ 0.5, & \text{if } 6\le k\le 10,\\ 0.8, & \text{if } 11\le k<15.\end{cases}$$

Case 2: $\alpha(k)$ are constant (i.e., uniform scheme) and $\alpha(k)=0.3$ for any $k=0, 1, 2, \cdots, 14$.

As shown in Figures 1 and 2, it is clear that the use of state-dependent (variable) jump priorities does not worsen the loss probability of H-calls, and at the same time, the loss probability of L-calls essentially decreases. For both QoS metrics CLP_1 and CLP_2, the state-dependent jump priorities scheme (case 1) appears to perform better, than the uniform scheme (case 2). For the QoS metric Q_1, we observed the same behavior (see Figure 3). However, for QoS metric Q_2, we observed an opposite situation, i.e., for this function, the uniform scheme appears to perform better than the state-dependent jump priorities scheme (see Figure 4). The values of function CTD_1 in both cases are almost the same (see Figure 5), while for the function CTD_2, the uniform scheme (case 1) is more desirable than the threshold one (see Figure 6). Note that the characteristics of the changes of all the QoS metrics are very similar for both schemes (see Figures 1–6).

Both exact and approximate approaches were used in the numerical experiments. The high accuracy of the proposed approximate method should be noted. Thus, a comparative analysis of the approximate values and the exact values obtained using the SGBE solution indicates that their differences are negligible if condition $\lambda_1>>\lambda_2$ is valid. Similar results were observed for the both cases indicated above, but these results are omitted due to space limitations. Note that the high accuracy of the proposed method has been verified for other values of the traffic parameters of the system that satisfy the condition $\lambda_1>>\lambda_2$. Also note that, in terms of simplicity and efficiency, the proposed approximate method is unequivocally superior to the method based on the SGBE solution. This superiority is due to the fact that in the approximation approach, the QoS metrics are calculated by means of simple and obvious formulae.

The proposed method makes it possible to optimize the buffer size subject to QoS requirements. For the sake of brevity, we consider only one such problem. With fixed λ_1, λ_2, μ, and R_2, it is necessary to find the maximal value (if it exists) of R_1^* for which the given requirements of CTD_1 and CTD_2 are met, i.e., it is necessary to solve the following problem:

$$R_1 \to \max \qquad (20)$$

$$s.t. \quad CTD_k \le \tau_k,\ k = 1,\ 2 \qquad (21)$$

where $\tau_k,\ k = 1,\ 2$, are given values.

For the solution of problem (20) and (21), the well-known dichotomy method may be applied by using the monotonic properties of the functions CTD_1 and CTD_2 with respect to the parameter R_1. The results of a solution of the optimization problem (20) and (21) for the hypothetical model with initial data $R_2 = 15$, $\lambda_1 = 8$, $\mu = 12$ are shown in Table 1. In the same way, a similar problem requiring the minimization of R_1 subject to (21) may be solved.

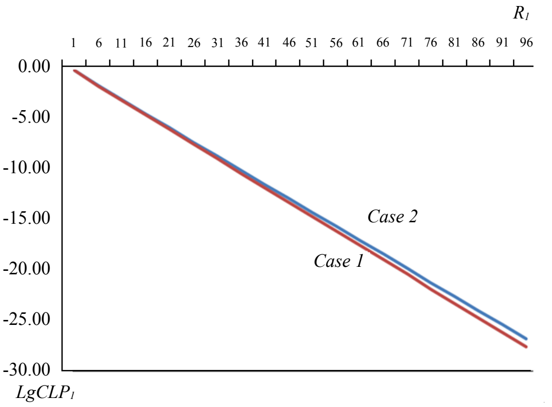

Figure 1. The loss probability vs. the buffer size for calls of type 1.

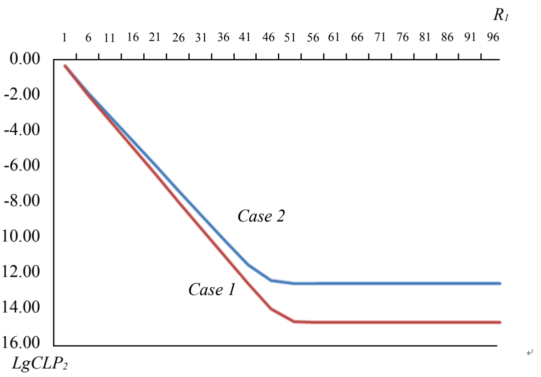

Figure 2. CLP_2 vs. R_1.

Figure 2. The loss probability for calls of type 2 vs. the buffer size for calls of type 1.

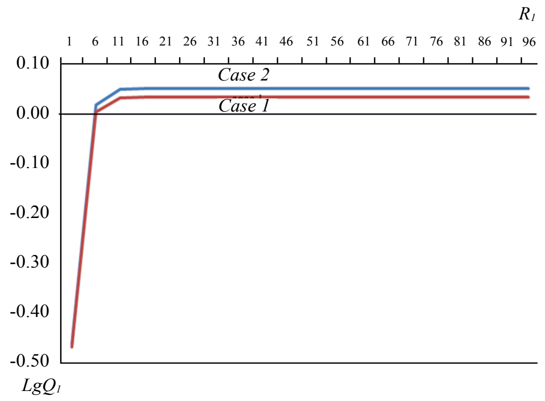

Figure 3. The queue length vs. the buffer size for calls of type 1.

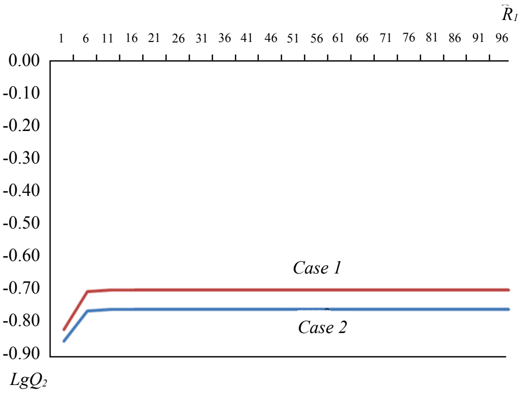

Figure 4. The queue length for calls of type 2 vs. the buffer size for calls of type 1.

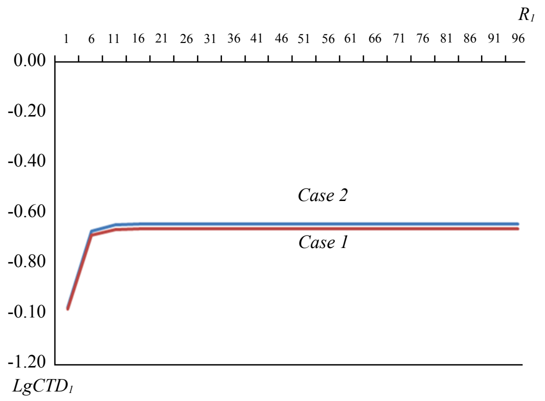

Figure 5. The transfer delay vs. the buffer size for calls of type 1.

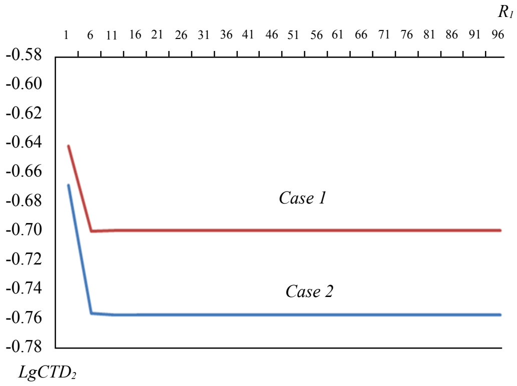

Figure 6. The transfer delay for calls of type 2 vs. the buffer size for calls of type 1.

Table 1. The solution results for the problem (20) and (21)

No. of variant	λ_2	τ_1	τ_2	R_1^*	
				Case 1	Case 2
1	4	1	1	4	4
2	4	1	2	16	97
3	4	1	5	45	98
4	4	1	10	45	98
5	4	5	1	4	4
6	4	10	1	4	4
7	6	4	5	54	100
8	6	4	10	54	100
9	6	8	5	98	100
10	2	10	5	100	100

5. CONCLUSION

This paper proposed a new and effective approach for calculating the QoS metrics of heterogeneous calls in a finite queuing system with HOL-PJ. The important advantage of this approach lies in the use of explicit formulae to calculate the QoS metrics, which enables our approach to be used for models of any dimension. In addition, it is possible to investigate models with the common, limited queue of heterogeneous calls and generalize the proposed HOL-PJ scheme in the event that the probabilities α(k) also depend on the number of H-calls. These problems are a subject for further study.

ACKNOWLEDGMENTS

This work has been supported by a Sangji University Research Grant 2010.

REFERENCES

Bruneel, H., Steyaert, B., Desmet, E., and Petit, G. H. (1994), Analytic derivation of tail probabilities for queue lengths and waiting times in ATM multiserver queues, *European Journal of Operational Research*, **76**(3), 563-572.

Demoor, T., Fiems, D., Walraevens, J., and Brunnel, H. (2011), Partially shared buffers with full or mixed priority, *Journal of Industrial and Management Optimization*, **7**(3), 735-751.

Kim, C. S., Melikov, A. Z., and Ponomarenko, L. A. (2007), Approximation method for performance analysis of queuing systems with multimedia traffics, *Applied and Computational Mathematics*, **6**(2), 218-226.

Kleinrock, L. (1964), A delay dependent queue discipline, *Naval Research Logistics Quarterly*, **11**(3-4), 329-341.

Kleinrock, L. (1976), *Queuing Systems. Volume II: Computer Applications*, John Wiley and Sons, New York, NY.

Lee, Y. and Choi, B. D. (2001), Queuing system with multiple delay and loss priorities for ATM networks, *Information Sciences*, **138**(1-4), 7-29.

Lim, Y. and Kobza, J. E. (1990), Analysis of a delay-dependent priority discipline in an integrated multiclass traffic fast packet switch, *IEEE Transactions on Communications*, **38**(5), 659-665.

Maertens, T., Walraevens J., and Bruneel, H. (2007), A modified HOL priority scheduling discipline: performance analysis, *European Journal of Operational Research*, **180**(3), 1168-1185.

Maertens, T., Walraevens, J., and Bruneel, H. (2006a), On priority queues with priority jumps, *Performance Evaluation*, **63**(12), 1235-1252.

Maertens, T., Walraevens, J., and Bruneel, H. (2008), Performance comparison of several priority schemes with priority jumps, *Annals of Operations Research*, **162**(1), 109-125.

Maertens, T., Walraevens, J., Moeneclaey, M., and Bruneel, H. (2006b), A new dynamic priority scheme: performance analysis, *Proceedings of the 13th International Conference on Analytical and Stochastic Modeling Techniques and Applications*, Bonn, Germany, 77-84.

Melikov, A. Z., Feyziev, V. S., and Rustamov, A. M. (2006), Analysis of model of data packet processing in ATM networks with multiple space and time priorities, *Automatic Control and Computer Sciences*, **40**(6), 38-45.

Ponomarenko, L., Kim, C. S., and Melikov, A. (2010), *Performance Analysis and Optimization of Multi-Traffic on Communication Networks*, Springer, New York, NY.

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

Wolff, R. W. (1992), Poisson arrivals see time averages, *Operations Research*, **30**(2), 223-231.