Numerical Approach to Analysis of Queuing Models with Priority Jumps

Agassi Melikov[1], Turan Jafar-zade[2]
[1]Institute of Cybernetics, Baku, Azerbaijan,
[2]National Aviation Academy, Baku, Azerbaijan
[1]*agassi.melikov@rambler.ru,* [2]*turan_jafarzade@hotmail.com*

***Abstract*— A space merging approach to study the queuing models with finite buffers and jump priorities is proposed. It is assumed that upon arriving a call with low priority one call of such kind might be transfer to the end of queue of high priority calls. Transfer probabilities are depending on state of the queue of heterogeneous calls. The algorithms to calculate the quality of service metrics of such queuing models are developed.**

***Keywords*— queuing model; priority jumps; quality of service; calculation method**

I. INTRODUCTION

In modern packet switching networks heterogeneous traffic make various demands to quality of service (QoS) metrics. So, real time traffic are sensitive to possible delays while non-real time traffic demand as small as possible loss of their packages. An effective way for satisfaction of conflicting requirements of heterogeneous calls (packets) is using priorities. Various kinds of priorities in queuing models are investigated for a long time.

In the last years in [1]-[4] new type of HOL-priorities (Head-Of-Line) are studied. Authors of the mentioned work have entitled the investigated priorities Head-Of-Line with Priority Jumps (HOL-PJ). In these models two types of calls - calls of a high priority (H-calls) and calls of a low priority (L-calls) are offered. For buffering of calls of each type there are infinite queues. In papers [1]-[4] authors have developed formulas for generating functions of length of queues of calls of both types and a waiting time in queue of H-calls, and also their moments.

It is necessary to notice, that mentioned above papers [1]-[4] are devoted research of models with infinite queues which cannot be accepted as adequate models of the realistic telecommunication systems, since as a rule real telecommunication systems have the finite buffers for the temporary storage of heterogeneous calls (packets).

In this paper, the new class of jump priorities in systems with the finite queues which have randomized property is introduced. They allows jump from L-queue to H-queue only upon arrivals L-calls and probability of jump depends on number of L-calls in queue. Introduction of restrictions on the size of buffers for heterogeneous calls leads to necessity of definition of new QoS metric - cell loss probability (CLP). Other differing moment of this paper from papers [1]-[4] consists that here for the system analysis the state space merging (SSM) approach [5] is used. By using the given approach simple computational procedures are developed for a finding of all QoS metrics of the investigated models.

II. DEFINITION OF JUMP PRIORITIES

We consider a continuous time system with two (separate) queues of finite capacity and with one (common) transmission channel. Two types of Poisson traffic arrive at the system: the first traffic represents traffic of cells of real time (H-calls) while the second traffic is traffic of cells of non-real time (L-calls). Intensity of traffic i is equal to λi, i=1,2 and channel occupancy time is a random variable which has exponential distribution with parameter μ for calls of both types.

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

It is clear that the HOL-priority scheme provides the good performance for H-calls and at the same time the performance of L-calls can be severely degraded especially in case when the system is highly loaded by H-calls. So for the purpose of increase the chances of L-calls to be served for comprehensible time jump priorities are introduced. Nevertheless these priorities leads to insignificant worsen the QoS metrics of H-calls. The basic questions in introduction of jump priorities are definition of the moment of jumps from L-queue to H-queue and numbers of the L-calls transferred to H-queue. Here jump priorities are defined as follows. First of all, note that H-calls always are accepted with probability 1 if at the moment of their arrival there is at least one empty place in the H-buffer; otherwise they are lost with probability 1.

If upon arrival of a L-call number of calls of the given type in the buffer equally k and there is an empty place in H-queue then with probability $\alpha(k)$ one L-call immediately jump to tail of H-queue (for definiteness of a statement we will assume that to the H-buffer jump a L-call standing in the head of L-queue); with complementary probability $1-\alpha(k)$ arrived L-call joins queue if there is an empty place. In case of successful jump the L-call becomes an H-call, and further is served as an H-call in accordance to HOL-priorities. If upon arrival of L-call there is no empty place in H-queue then with probability 1 arrived L-call joins L-queue if here there is an empty place; otherwise, with probability 1 it is lost.

III. METHODS FOR CALCULATE THE QOS METRICS

Basic QoS metrics of the investigated system are cell loss probability (CLP_i), an average number of cells of each type in buffers (Qi) and an average cell transfer delay (CTDi), i=1,2.

The state of buffers at any moment of time can be described by means of a two-dimensional vector n = (n1, n2) where ni denotes number of cells of type i in the buffer, i = 1,2. So, functioning of the given system is described by two-dimensional chain Маркова (2-D MC) with state space

S := {n: ni = 0,1,…,Ri, i=1,2}.

Transitions between states of the system occur only upon arrival of calls and their leaving from system after completing of service. By taking into account above statements we conclude that non-negative elements of a Q-matrix of the given 2-D MC is defined from following relations:

$$q(\boldsymbol{n},\tilde{\boldsymbol{n}}) = \begin{cases} \lambda_1+\lambda_2\alpha(n_2), & \text{if } \tilde{\boldsymbol{n}}=\boldsymbol{n}+\boldsymbol{e_1}, \\ \lambda_2, & \text{if } \tilde{\boldsymbol{n}}=\boldsymbol{n}+\boldsymbol{e_2}, \\ \mu, & \text{if } n_1>0, \tilde{\boldsymbol{n}}=\boldsymbol{n}-\boldsymbol{e_1} \text{ or } n_1=0, \tilde{\boldsymbol{n}}=\boldsymbol{n}-\boldsymbol{e_2}, \\ 0 & \text{in other cases,} \end{cases} \quad (1)$$

where $\boldsymbol{e}_1$=(1,0), $\boldsymbol{e}_2$=(0,1).

The given 2-D MC is strictly continuous with respect to the first component while it is weakly continuous with respect to second one (for definitions see Appendix of book [5]). The system of global balance equations (SGBE) for the steady-state probabilities $p(\boldsymbol{n})$, $\boldsymbol{n}\in S$, might be constructed by using the relations (1). Desired QoS metrics are determined via the stationary distribution of the initial model. So, by using the PASTA theorem [6] we obtain:

$$CLP_1=\sum_{k=0}^{R_2} p(R_1,k) \quad (2)$$

$$CLP_2=\sum_{k=0}^{R_1-1} p(k,R_2)\big(1-\alpha(R_2)\big)+p(R_1,R_2) \quad (3)$$

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

$$Q_k=\sum_{i=1}^{R_k} i\xi_k(i), \quad (4)$$

Where

$$\xi_k(i)=\sum_{\boldsymbol{n}\in S} p(\boldsymbol{n})\delta(n_k,i)\,, k=1,2,$$

are marginal probability mass functions and $\delta(x,y)$ represents Kronecker's symbols.

The QoS metrics CTD_k are calculated by using modified Little's formula:

$$CTD_k=\frac{Q_k}{\lambda_k(1-CLP_k)}\,, k=1,2. \quad (5)$$

Stationary distribution is determined as a result of solution of a mentioned above SGBE of the given 2-D MC. However, to solve the last problem one requires laborious computational efforts for large values of R_1 and R_2 since the corresponding SGBE has no explicit solution. To overcome the mentioned difficulties, a new efficient and refined approximate method for calculation of stationary distribution of the given model is suggested below.

The following splitting of state space S is examined:

$$S=\bigcup_{i=0}^{R_2} S_i\,, S_i\cap S_j=\emptyset, i\neq j, \quad (6)$$

Where

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

Furthermore, state classes S_i combine into separate states $<i>$ and the following merging function in state space S is introduced:

$$U(\boldsymbol{n})=<i> \text{ if } \boldsymbol{n}\in S_i. \quad (7)$$

Function (7) determines a merged model which is a 1-D MC with the state space $\Omega=\{<i>: i=0,1,\dots,R_2\}$. Then, according to SSM approach, the stationary distribution of the initial model approximately equals:

$$p(k,i)\approx\rho_i(k)\pi(<i>), \quad (8)$$

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

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

$$q_i(k_1,k_2)=\begin{cases} \lambda_1+\lambda_2\alpha(i), & \text{if } k_2=k_1+1, \\ \mu, & \text{if } k_2=k_1-1, \\ 0 & \text{in other cases.} \end{cases} \quad (9)$$

So, stationary distribution within class S_i is

$$\rho_i(k)=\theta_i^k\cdot\frac{1-\theta_i}{1-\theta_i^{R_1+1}}\,, k=0,1,\dots,R_1 \quad (10)$$

where $\theta_i:=\nu_1+\nu_2\alpha(i)$. In order to be short here we give only formulas for case $\theta_i\neq1$.

Then from (2) and (10) by means of SSM approach the elements of the generating matrix of a merged model are found:

$$q(<i_1>,<i_2>) = \begin{cases} \lambda_2(1-\alpha(i_1))\left(1-\rho_{i_1}(R_1)\right)+\lambda_2\rho_{i_1}(R_1), & \text{if } i_2 = i_1+1, \\ \mu\rho_{i_1}(0), & \text{if } i_2 = i_1-1, \\ 0 & \text{in other cases.} \end{cases} \quad (11)$$

The formula (11) allows determination of the stationary distribution of a merged model. It coincides with an appropriate distribution of state probabilities of a 1-D BDP, for which transition intensities are determined in accordance with (11). Consequently, the stationary distribution of a merged model is determined as

$$\pi(<i>) = \prod_{j=1}^{i} A_j \pi(0)\,,\ i = 1,2,\dots,R_2, \quad (12)$$

Where

$$A_j = \nu_2 \cdot \frac{(1-\alpha(j-1))\left(1-\rho_{j-1}(R_1)\right)+\rho_{j-1}(R_1)}{\rho_j(0)},$$

$$\pi(0) = 1/\left(1+\sum_{k=1}^{R_2}\prod_{i=1}^{k} A_i\right).$$

Then by using (10) and (12) from (8) the stationary distribution of the initial 2-D MC can be found. So, summarizing the above and omitting the complex algebraic transformations the following approximate formulae for calculation of QoS metrics (2)-(4) can be suggested:

$$CLP_1 \approx \sum_{k=0}^{R_2}\rho_k(R_1)\pi(<k>) \quad (13)$$

$$CLP_2 \approx \pi(<R_2>)\left((1-\alpha(R_2))\left(1-\rho_{R_2}(R_1)\right)+\rho_{R_2}(R_1)\right) \quad (14)$$

$$Q_1 \approx \sum_{k=1}^{R_1} k \sum_{i=0}^{R_2}\rho_i(k)\pi(<i>); \quad (15)$$

$$Q_2 \approx \sum_{k=1}^{R_2} k\pi(<k>)\,. \quad (16)$$

After calculation of CLP_k and Q_k from (5) we determine CTD_k, k=1,2.

IV. CONCLUSION

In this paper, the new and effective approach for calculate the QoS metrics of heterogeneous calls in a finite queuing systems with jump priorities is offered. The important advantage of the offered approach consists that it can be used for models of any dimension since QoS metrics are calculated by means of explicit formulas. It is possible to investigate models with the common limited queue of heterogeneous calls and generalize the proposed jump priorities for a case when probabilities $\alpha(k)$ depend also on number of calls of the first type.

REFERENCES

[1] *Walraevens J., Steyaert B., Bruneel H.* Performance analysis of single-server ATM queue with priority scheduling // Computers and Operations Research. 2003. Vol.30, no.12. P.1807-1829.

[2] *Maertens T., Walraevens J., Bruneel H.* On priority queues with priority jumps // Performance Evaluation. 2006. Vol.63, no.12. P.1235-1252.

[3] *Maertens T., Walraevens J., Bruneel H.* A modified HOL priority scheduling discipline: performance analysis // European Journal of Operational Research. 2007. Vol.180, no.3. P.1168-1185.

[4] *Maertens T., Walraevens J., Bruneel H.* Performance comparison of several priority schemes with priority jumps // Annals of Operations Research. 2008. Vol.162. P. 109-125.

[5] *Ponomarenko L., Kim C.S., Melikov A.* Performance analysis and optimization of multi-traffic on communication networks. New York: Springer, 2010.

[6] *Wolff R.W.* Poisson arrivals see time averages // Operations Research. 1992. Vol.30, no.2. P.223- 231