Федеральное государственное бюджетное учреждение науки
Институт проблем управления им. В.А. Трапезникова
РОССИЙСКОЙ АКАДЕМИИ НАУК

Научно-производственное объединение
«Информационные и сетевые технологии»

Институт информационных и телекоммуникационных технологий
БОЛГАРСКОЙ АКАДЕМИИ НАУК

РАСПРЕДЕЛЕННЫЕ КОМПЬЮТЕРНЫЕ И ТЕЛЕКОММУНИКАЦИОННЫЕ СЕТИ: УПРАВЛЕНИЕ, ВЫЧИСЛЕНИЕ, СВЯЗЬ (DCCN-2015)

МАТЕРИАЛЫ ВОСЕМНАДЦАТОЙ МЕЖДУНАРОДНОЙ НАУЧНОЙ КОНФЕРЕНЦИИ

(19–22 октября 2015 г., Москва, Россия)

Под общей редакцией д.т.н. В.М. Вишневского

Москва
ИПУ РАН
2015

V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences

Research and development company **“Information and networking technologies”**

Institute of Information and Communication Technologies
Bulgarian Academy of Sciences

DISTRIBUTED COMPUTER AND COMMUNICATION NETWORKS: CONTROL, COMPUTATION, COMMUNICATIONS (DCCN-2015)

PROCEEDINGS OF THE EIGHTEENTH INTERNATIONAL SCIENTIFIC CONFERENCE

(19–22 october 2015 г., Moscow, Russia)

Under the general edition of Dr. of Computer Science V.M. Vishnevskiy

Moscow
ICS RAS
2015

УДК 004.7:004.4].001:621.391:007
ББК 32.973.202:32.968
Р 24

Распределенные компьютерные и телекоммуникационные сети: управление, вычисление, связь (DCCN-2015) = Distributed computer and communication networks: control, computation, communications (DCCN-2015) : материалы Восемнадцатой междунар. науч. конфер, 19–22 окт. 2015 г., Москва: / Ин-т проблем упр. им. В.А. Трапезникова Рос. акад. наук ; под общ. ред. В.М. Вишневского – М.: ИПУ РАН, 2015. – 656 с.
– ISBN 978-5-91450-170-6.

В научном издании представлены материалы Восемнадцатой международной научной конференции «Распределенные компьютерные и телекоммуникационные сети: управление, вычисление, связь» по следующим направлениям:

- Архитектура компьютерных и телекоммуникационных сетей.
- Управление в компьютерных и телекоммуникационных сетях.
- Оценка производительности беспроводных сетей трансляции мультимедийной информации.
- Аналитическое и имитационное моделирование сетевых протоколов.
- Теория очередей и теория надежности.
- Беспроводные сети IEEE 802.11, IEEE 802.15, IEEE 802.16 и UMTS (LTE).
- Технология RFID и ее применение в интеллектуальных транспортных системах.
- Проектирование протоколов (MAC-уровня) сантиметрового и миллиметрового диапазона радиоволн.
- Интернет, веб-приложения и услуги.
- Интеграция приложений в распределенных информационных системах.

В материалах конференции DCCN-2015, подготовленных к выпуску Козыревым Д.В. обсуждены перспективы развития и сотрудничества в этой сфере.

Сборник материалов конференции предназначен для научных работников и специалистов в области теории и практики построения компьютерных и телекоммуникационных сетей.

Текст воспроизводится в том виде, в котором представлен авторами

Утверждено к печати Программным комитетом конференции

ISBN 978-5-91450-170-6

ANALYSIS OF QUEUEING MODELS WITH STATE-DEPENDENT JUMP PRIORITIES

A.Z. Melikov[1], *A. M. Rustamov*[2], *T.I. Jafarzade*[3], *J. Sztrik*[4]
[1]Institute of Control Systems, ANAS, F. Agayev 9,AZ1141, Baku, Azerbaijan,
[2]Qafqaz University, H. Aliyev 120, AZ0101, Khirdalan, Baku, Azerbaijan,
[3]National Aviation Academy, Merdekan pr. 30, AZ1045, Baku, Azerbaijan,
[4]University of Debrecen, 4032, Debrecen, Hungary,
[1]agassi.melikov@gmail.com, [2]anrustemov@qu.edu.az,
[3]turan_jafarzade@hotmail.com, [4]sztrik.janos@inf.unideb.hu

Abstract

In this paper, exact and approximate approaches for studying queuing models with state-dependent jump priorities are developed. Models with finite separate buffers for heterogeneous calls are investigated. It is shown that the investigated models might be described by two-dimensional Markov Chains. One of the main challenge in exact approach for the solution of appropriate system of balance equations for state probabilities becomes big computation for large scale models. To overcome the indicated difficulties an approximate approach based on the state space merging algorithms is developed. This approach allows constructing simple algorithms to calculate the quality of service metrics of the examined models.

Keywords: queuing models, jump priority, Markov chains, space merging, numerical analysis

1. Introduction

Priorities are effective tools to solve the problems of quality of service (QoS) provisioning of heterogeneous calls in queuing systems. By nature the priorities can be broadly divided into two classes: static and dynamic. Static priorities (relative or preemptive) are defined in advance and they do not change during the whole system operation time [1]. In literature relative static priorities in queuing systems with buffers sometimes are called HOL-priorities (Head-Of-Line), .i.e. in static priorities call for service is chosen from the head of line according to the highest priority. Dynamic priorities in turn are divided into two classes: dynamical vs time and dynamical vs state. In dynamical versus time priorities the priority of the calls can be changed according their waiting times (or sojourn time) [2]. In dynamical versus state priorities (they sometimes are called state-dependent priorities) calls can change priority according the state of the system where the state is described by vector whose components indicate the number of heterogeneous calls in the queue (or in the system) [3].

The drawback of static priorities is that when they are used in real systems the delay of low priority calls is too large especially for the system with heavy loads of high priority calls. Dynamic priorities allow avoiding the starvation of low priority calls. Detailed review of priority schemas might be found in [4].

As a rule, classical priorities (static or dynamic) are used to determine type of call from the buffer which must be send to channel for servicing. However, some scientific and practical interest represents the priorities which are introduced to change (either increase or decrease) the priorities of calls in buffer. These changes are realized instantaneously so such kinds of priorities are called jump priorities (JP).

The pioneer work on the analyzing dynamical vs time HOL-priorities with priority jumps (HOL-JP) is [5]. In this paper dynamical vs time HOL-JP was proposed where calls with low priority can jump to another buffer with high priority after waiting some (deterministic) period of time in native buffer. Formulas for calculation of the mean waiting time of the heterogeneous calls were developed in [5].

Dynamical vs states HOL-JP in discrete-time queuing models were proposed in [6-9]. In [5-9] queuing models with infinite buffers are investigated. So, they have little applicability in the real communication networks. In particular, real communication networks have finite buffer capacity. Secondly, investigated HOL-JP is defined by state-independent probabilities. Therefore they cannot be adapted for real situations according to the changes of loads of heterogeneous calls.

Different approach to study queuing models with dynamical versus state HOL-JP can be found in the papers [10-13] and in chapter 5 of the book [14] where new type of randomized state-dependent JP for continuous-time queuing systems with finite buffers was proposed. They make it possible to pass to from the L-queue (queue for low priority calls, L-calls) into the H-queue (queue for high priority calls, H-calls) only at the instants of arrival of the L-calls, but the probability of such transitions depend only on the number of L-calls in the system. In chapter 5 of the book [14] models with separate buffers which jump priorities depending only on the number of H-calls in the system were examined. In the indicated works [10-14] methods of calculation of main QoS metrics of the investigated models are proposed. To the best of our knowledge, models in which JP depends on the number of both types of calls in the system are not examined. In this paper we investigate such kinds of models. Our contribution consist of two parts; 1) we propose novel kind of state-dependent jump priorities, and 2) both exact and approximate methods to calculate the QoS metrics of queuing models with such kind of priorities are developed.

The rest of the paper is organized as follows. In section 2, model with separate buffers is defined and state-dependent JP is introduced. In section 3, exact method of calculation its QoS metrics is developed. In section 4 an approximate method to solving the same problem is developed. Conclusion remarks are given in section 5.

2. Jump Priorities in Model with Separate Buffer

In the single server queueing system two Poisson traffic of heterogeneous calls have different arrival rate $\lambda_i, i = 1, 2$. We determined first type of calls as high priority calls (H-calls) while second type of calls are treated as low priority calls (L-calls). In general, H-calls have relative priority over L-calls while channel is idle, namely in the case of absence of H-call in the buffer, L-call can be served. If there isn't any call in the buffer, then the channel becomes free. Service intensity of the server is the same for both types of the call where it is determined as μ obeying exponential distribution.

Consider the model with separate buffers, i.e. it is assumed that there are two isolated buffers — H-buffer (for waiting H-calls) and L-buffer (for waiting L-calls) with size of R_1 and R_2 $(0 < R_i < \propto, i = 1, 2)$ respectively.

Decision epochs coincide with the arrival moments of L-calls. In this model state-dependent HOL-JP is defined as follows.

- High priority calls are always accepted to the H-buffer with probability 1 if there is a free place in this buffer. If the H-buffer is full then arriving H-call is dropped with probability 1.
- If upon arrival of L-call the number of calls of this type equals $i, i < R_2$, and the number of H-call equals $j, j < R_1$, then L-call joins the H-buffer with probability $\alpha_i(j)$ and in future it will be served as H-call; and arriving L-call joins the L-buffer with probability $1 - \alpha_i(j)$.
- If upon arrival of L-call the number of H-call equals R_1, then L-call joins the L-buffer if there is free place in this buffer; otherwise, arriving L-call is dropped with probability 1.
- If upon arrival of L-call L-buffer is full and the number of H-call equals $j, j < R_1$, then L-call joins the H-buffer with probability $\alpha_{R_2}(j)$; and arriving L-call is dropped with probability $1 - \alpha_{R_2}(j)$.

The problem is finding the QoS metrics for this model. The main QoS metrics are the following: the stationary probability of losing the calls of the i-th type, the mean number of the i-th type calls in the buffers and the mean call transmission delay of the i-th type calls, $i = 1, 2$.

3. Exact Method

The state of the system is defined by 2-D vectors $\mathbf{n} = (n_1, n_2)$ where the first component indicates the number of H-calls and the second one the number of L-calls respectively. So, operation of this system is described by the 2-D Markov Chain (2-D MC) with the following state space:

$$S = \{\boldsymbol{n} : n_i = 0, 1, ..., R_i, i = 1, 2\}. \tag{1}$$

Transition intensity from state $\boldsymbol{n} \in S$ to state $\boldsymbol{n}' \in S$ are denoted by $q(\boldsymbol{n}, \boldsymbol{n}')$. Then nonnegative elements of the generating matrix (Q-matrix) of the given 2-D MC can be calculated as below:

$$q(\boldsymbol{n}, \boldsymbol{n'}) = \begin{cases} \lambda_1 + \lambda_2 \alpha_{n_2}(n_1), & \text{if } n_1 < R_1, \boldsymbol{n'} = \boldsymbol{n} + \boldsymbol{e_1} \\ \lambda_2(1 - \alpha_{n_2}(n_1)), & \text{if } n_1 < R_1, n_2 < R_2, \boldsymbol{n'} = \boldsymbol{n} + \boldsymbol{e_2} \\ \lambda_2, & \text{if } n_1 = R_1, \boldsymbol{n'} = \boldsymbol{n} + \boldsymbol{e_2} \\ \mu, & \text{if } n_1 > 0, \boldsymbol{n'} = \boldsymbol{n} - \boldsymbol{e_1} \text{ or } n_1 = 0, \boldsymbol{n'} = \boldsymbol{n} - \boldsymbol{e_2} \\ 0, & \text{in other cases.} \end{cases} \tag{2}$$

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

The stationary probability of state $n \in S$ is denoted by $p(n)$. Construction and solution of the corresponding system of balance equations (SBE) for the given 2-D MC is the standard way for determining the stationary state probabilities. It is constructed with regard to (2) and here is omitted.

After determining the state probabilities from SBE, one can establish its QoS metrics. As indicated above, H-calls are lost if upon their arrivals H-buffer is full. Hence, the loss probability for H-calls (CLP_1) can be determined as follows:

$$CLP_1 = \sum_{i=0}^{R_2} p(R_1, i). \tag{3}$$

Similarly, we conclude that the loss probability of L-calls (CLP_2) is given by

$$CLP_2 = p(R_1, R_2) + \sum_{i=0}^{R_1 - 1} p(i, R_2)(1 - \alpha_{R_2}(i)). \tag{4}$$

The mean numbers of the H-calls (L_1) and L-calls (L_2) in the queue are determined as the expected values of appropriate discrete random variables:

$$L_1 = \sum_{i=1}^{R_1} i \sum_{j=0}^{R_2} p(i, j); \tag{5}$$

$$L_2 = \sum_{i=1}^{R_2} i \sum_{j=0}^{R_1} p(j, i). \tag{6}$$

Further, formulas (3)-(6) and modified Little's formula can be used to evaluate the mean times of call transmission delay (CTD_i) for the heterogeneous:

$$CTD_1 = \frac{L_1}{\lambda_1^{(c)}}, \tag{7}$$

$$CTD_2 = \frac{L_1 + L_2}{\lambda_1^{(c)} + \lambda_2^{(c)}}, \tag{8}$$

where $\lambda_1^{(c)}$ and $\lambda_2^{(c)}$ are carried loads of H-calls and L-calls, respectively. These parameters are calculated as follows:

$$\lambda_1^{(c)} = \lambda_1 \left(1 - \sum_{j=0}^{R_2} p(R_1, j)\right) + \lambda_2 \sum_{i=0}^{R_1-1} \sum_{j=0}^{R_2} p(i,j)\alpha_j(i),$$

$$\lambda_2^{(c)} = \sum_{i=0}^{R_1} \sum_{j=0}^{R_2-1} p(i,j)\left(1 - \alpha_j(i)\right).$$

By implementation of programming languages it is possible to solve the SBE for the steady-state probabilities $p(\boldsymbol{n}), \boldsymbol{n} \in S$ with a help of numerical methods of the linear algebra. This method of calculation of QoS metrics is called the exact (precise) method. In cases of small dimensions of state space (1) this method is reasonable to calculate QoS metrics of the system. But for large scale system it isn't suitable. Therefore, we need to find out a more efficient method to calculate the QoS metrics of the models with large dimensions of buffers.

4. Approximate Method

Below we consider asymptotic analysis of the QoS metrics for large scale models, i.e. when R_1 and R_2 take large values. The developed approximate method has high accuracy for heavy traffic regime of H-calls. In other words, below we consider asymptotic analysis of the large scale model with heavy loads of H-calls, i.e. it is assumed that $\nu_1 >> \nu_2 >> 1$, where $\nu_i = \lambda_i/\mu, i = 1, 2$. Note that this assumption make sence for the jump priorites for the L-calls in the systems with heavy loads of H-calls.

Consider the following splitting of the state space (1):

$$S = \bigcup_{i=0}^{R_2} S_i, S_i \bigcap S_j = \emptyset, i \neq j, \tag{9}$$

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

We notice that the assumption made about the relation of the loads of the heterogeneous calls enables one to satisfy the condition for correct use of the algorithms of state space merging of the 2-D MC (see [3, Appendix]): transition intensities within classes $S_i, i = 0, 1, ..., R_2$, are essentially higher than those between states of different classes. The classes of microstates S_i are united into individual merged states $< i >$, and in the original state space S the following merge function is defined:

$$U(\boldsymbol{n}) =< i >, \text{if } \boldsymbol{n} \in S_i. \tag{10}$$

The function (10) defines a merged model with the state space $\Omega = \{< i >: i = 0, 1, ..., R_2\}$. Let us consider the problem of calculation of state probabilities inside the splitting models. The stationary probability of the state (k, i) in the split model with the state space S_i is denoted by $\rho_i(k), i = 0, 1, ..., R_2, k = 0, 1, ..., R_1$.

Each split model with state space S_i. is a 1-D birth and death process with the parameters that are calculated as follows:

$$q_i(k_1, k_2) = \begin{cases} \lambda_1 + \lambda_2 \alpha_i(k_1), & \text{if } k_2 = k_1 + 1 \\ \mu, & \text{if } k_2 = k_1 - 1 \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

Consequently, we have

$$\rho_i(k) = \prod_{j=0}^{k-1} (\nu_1 + \nu_2 \alpha_i(j)) \rho_i(0), k = 1, ..., R_1, \tag{12}$$

where $\rho_i(0) = \left(1 + \sum_{k=1}^{R_1} \prod_{j=0}^{k-1} (\nu_1 + \nu_2 \alpha_i(j))\right)^{-1}$.

The elements of the Q-matrix of the merged model are denoted by $q(< k >, < k' >), < k >, < k' > \in \Omega$. According to the algorithm of state space merging of the 2-D MC (see [3, Appendix]) these elements are given by

$$q(< k >, < k' >) = \sum_{\boldsymbol{n} \in S, \boldsymbol{n}' \in S_{k'}} q(\boldsymbol{n}, \boldsymbol{n}') \rho_{n_1}(n_2). \tag{13}$$

So, by using (2), (12) and (13) after some mathematical transformations the following formulae are obtained

$$q_i(< k >, < k' >) = \begin{cases} \lambda_2 \left(\rho_k(R_1) + \sum_{i=0}^{R_1 - 1} (1 - \alpha_k(i)\right), & \text{if } k' = k + 1 \\ \mu \rho_k(0), & \text{if } k' = k - 1 \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

From (14) we can calculate the probabilities of the merged states $\pi(< k >), < k > \in \Omega$ as follows:

$$\pi(< k >) = \nu_2^k \prod_{j=0}^{k-1} \Lambda_j \pi(< 0 >), k = 1, 2, ..., R_2, \tag{15}$$

where $\pi(< 0 >) = \left(1 + \sum_{k=1}^{R_2} \nu_2^k \prod_{j=0}^{k-1} \Lambda_j\right)^{-1}, \Lambda_j = \frac{\rho_j(R_1) + \sum_{i=0}^{R_1 - 1} (1 - \alpha_j(i)) \rho_j(i)}{\rho_{j+1}(0)}, j = 0, 1, 2, ..., R_2 - 1$.

The state probabilities of the initial 2-D MC are determined approximately as follows:

$$p(i,j) \approx \rho_j(i)\pi(< j >). \tag{16}$$

By taking into account (16), (12) and (15) we can calculate approximate values of state probabilities of initial 2-D MC, and omitting the intermediate mathematical calculations the following approximate formulae to calculate the QoS metrics (3)-(6) are obtained:

$$CLP_1 \approx \sum_{i=0}^{R_2} \rho_i(R_1)\pi(< i >). \tag{17}$$

$$CLP_2 \approx \pi(< R_2 >)\left(\rho_{R_2}(R_1) + \sum_{i=0}^{R_1-1} \rho_{R_2}(i)(1-\alpha_{R_2}(i))\right). \tag{18}$$

$$L_1 \approx \sum_{i=1}^{R_1} i \sum_{k=0}^{R_2} \rho_k(i)\pi(< k > . \tag{19}$$

$$L_2 \approx \sum_{k=1}^{R_2} k\pi(< k >). \tag{20}$$

The approximate value of QoS metrics CTD_i (see (7), (8)) are determined from (17)-(20) after the calculation of the parameters CLP_i and $L_i, i = 1, 2$. Here it should be mention that, approximate values of the carried loads of H-calls and L-calls are calculated according to the formula (16).

The developed approximate formulas allow one to carry out an authentic analysis of QoS metrics over any range of change of values of loading parameters of the heterogeneous traffic and also at any buffers sizes. Another goal of performing numerical experiments was the estimation of the proposed approximate formulas accuracy. In order to be short, here the appropriate results are omitted. Let us only note that accuracy of the proposed approximate formulas is acceptable for engineering practice. The bigger the ratio of loads of H-calls to L-calls, the higher the accuracy of approximate value of QoS metrics.

5. Conclusion

This paper proposed a new class of state-dependent JP in queueing systems with finite separate buffers for heterogeneous calls. An exact and approximate approaches for calculating the QoS metrics of heterogeneous calls in such systems are developed. They might be used to investigate the models of queueing systems with finite common buffer for heterogeneous calls as well. The important advantage of approximate 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 use the proposed formulae to find the optimal (in given sense) values of jump priorities. Latest problems are important especially for the threshold-based non-randomized JP-schemas and they are a subject for further study.

Acknowledgments.

The publication was supported by the T'AMOP-4.2.2.C-11/1/KONV-2012-0001 project. The project has been supported by the European Union, co-financed by the European Social Fund.

REFERENCES

1. Jaiswal HK. Priority queues. New York: Academic Press; 1968.
2. Kleinrock L. A delay dependent queue discipline. Naval Research Logistics Quarterly Journal 1964; 11: 329-341.
3. Melikov A, Ponomarenko L, Kim CS. Performance Analysis and Optimization of Multi-Traffic on Communication Networks. Heidelberg: Springer; 2010.
4. Wittevrongel S, De Vuyst S, Sys C, Bruneel H. A reservation-based scheduling mechanism for fair QoS provisioning in packet-based networks. In: Proceeding of the 26th IEEE International Teletraffic Congress, Karlskrona: 2014, p. 55-62.
5. Lim Y, Kobza JE. Analysis of delay dependent priority discipline in an integrated multiclass traffic fast packet switch. IEEE Transactions on Communication 1990; 38(5): 659-665.
6. Maertens T, Walraevens J, Bruneel H. On priority queues with priority jumps. Performance Evaluation 2006; 63(12): 1235-1252.
7. Maertens T, Walraevens J, Bruneel H. A modified HOL priority scheduling discipline: performance analysis. European Journal of Operation Research 2007; 180(3): 1168-1185.
8. Maertens T, Walraevens J, Bruneel H. Performance comparison of several priority schemes with priority jumps. Annals of Operation Research 2008; 162: 109-125.
9. Walraevens J, Steyaert B, Bruneel H. Performance analysis of single-server ATM queue with priority scheduling. Computers and Operation Research 2003; 30(12): 1807-1829.
10. Melikov AZ, Ponomarenko LA., Kim CS. Algorithmic approach to analysis of queuing system with finite buffers and jump priorities. Journal of Automation and Information Sciences 2012; 44(12): 43-54.
11. Kim CS, Oh Y, Melikov AZ. A space merging approach to the analysis of the performance of queueing models with buffers and priority jumps. Industrial Engineering and Management Systems 2013; 12(3): 274-280.
12. Melikov AZ, Ponomarenko LA, Kim CS. Approximate method for analysis of queuing models with jump priorities. Automation and Remote Control 2013; 74(1): 62-75.
13. Melikov AZ, Ponomarenko LA, Kim CS. Numerical method for analysis of queuing models with priority jumps. Cybernetics and System Analysis 2013; 49(1): 55-61.
14. Melikov A., Ponomarenko L. Multidimensional queueing models in telecommunication networks. Heidelberg: Springer; 2014.