On the analysis of the limited resources queuing system under MAP arrivals

Eduard Sopin[1,2*] and *Konstantin* Samouylov[1]

[1]Peoples' Friendship University of Russia (RUDN University), 6 Miklukho-Maklaya St, Moscow, 117198, Russian Federation
[2]Institute of Informatics Problems, FRC CSC RAS, 44-2 Vavilova Str., Moscow, 119333, Russian Federation

Abstract. In the paper, we analyse a multiserver queuing system with discrete limited resources and random resource requirements under MAP arrivals, which can adequately model resource allocation schemes in the contemporary wireless networks. The equilibrium system of equations is derived in the vector form and is solved numerically. With stationary probability distribution, we provide formulas for the average and the variance of the occupied resources, as well as for the blocking probability. The results are illustrated by a numerical example.

1 Introduction

Fast mobile connections over the last decade become more available for business and personal communications. Since then portable devices have dramatically improved the processing capabilities and the battery saving. Both of this factors impact on the mobile data traffic and shift the focus from voice services to the video and data services.

The significant increase in the user traffic demands requires a broadband improvement and advances techniques in the future evolution of mobile systems, often referred to as fifth generation (5G) networks [1-2]. Among the implemented improving in modulation and coding, MIMO techniques, the network solutions are also targeted to improve the wireless network performance. Network densification [3], direct in-band and out-of-band device-to-device communications [4] are considered the popular and cost effective solutions to challenge the spatial frequency reuse and substantial capacity gain.

Application of a classic queuing models fails to cover a specific mechanism of the frequency resources allocation in LTE-Advanced networks. The physical resources is usually represented by a time-frequency resource grid. Each resource element is made of one OFDM subcarrier during the interval of one OFDM symbol. These resource elements can be grouped into resource blocks (RBs) with a total bandwidth of 180 kHz per one block. Based on the imbedded channel quality feedback mechanism each block of coded bits is matched with a target rate then a modulation scheme is selected. In a simple words RBs are scheduled in respect to the channel quality indicator (CQI) and required bitrate that a user equipment communicate with a base station before the beginning of each subframe (1 ms) [5]. The maximum achievable bitrate for a user session i can be provided by a Shannon formula:

$$C_i^{\max} = \omega \log_2(1 + \mathrm{SNR}_i \, p_{\max}),$$

where ω is a spectral bandwidth, $p_{\max}$ is a base station maximum transmit power and SNR_i is a signal-to-noise ratio for an i-th user session. It basically depends on the conditions of the radio signal propagation, the distance to the base station and its transmit power. The new session admission is allowed if $\sum_i \frac{C_i}{C_i^{\max}} \le 1$, where C_i is a required rate at each subframe that depends on a distance. Thus, the resource requirements of each user session depends on a number of model parameters and varies for each session. The ratio $\frac{C_i}{C_i^{\max}}$ may be interpreted as a share of system resources required by a user session.

The described above resource allocation process can be modelled in terms of multiserver queuing systems, in which customers require not only a server, but also a random amount of limited resources. Such queuing systems with random resource requirements, Poisson incoming flow and exponentially distributed service time was investigated in [6-8] and was extended in [9] by implementing signals that trigger resource reallocation process. However the real traffic model could hardly be adequately described by a Poisson arriving process. In this paper we consider a queuing system with random resource requirements under Markovian arrival process that can approximate a wide range of real processes. The proposed method of analysis is applied for only discrete resource requirements, the case of continuous resource requirements is left for further research.

* Corresponding author: sopin_es@rudn.university

The rest of the paper is organized as follows. Section 2 introduces the mathematical model in terms of queuing system with resource requirements of a random volume and mathematical framework for the performance measures evaluation. Numerical results are presented in section 3 while section 4 concludes the paper.

2 Mathematical model

In this section, we describe the queuing system with random resource requirements under MAP arrivals and provide the analysis of its stationary characteristics.

2.1 Model description

Consider a multiserver queuing system with N servers and R resources. Customers arrive according to a Markovian Arrival Process (MAP) governed by a continuous-time finite-state Markov chain with M states, which is named here the underlying Markov chain. The arrival process is characterized by the matrices $\mathbf{A}_0$ and $\mathbf{A}_1$, $\mathbf{A}_0 + \mathbf{A}_1 = \mathbf{A}$, where $\mathbf{A}$ is the generator matrix of the underlying Markov chain [10-11]. Assume that the matrix $\mathbf{A}$ is irreducible and denote θ the stationary probability vector of the underlying Markov chain. Then, the customer arrival rate is given by $\lambda = \theta\mathbf{a}$, where $\mathbf{a} = \mathbf{A1}$ and $\mathbf{1}$ is a vector of ones with appropriate size.

An arriving customer requires not only a server, but also j resources with probability p_j, $j \geq 0$. We assume the resource requirements to be mutually independent identically distributed random variables, independent of arrival and service processes. If upon arrival of a customer there are no available servers of not enough free resources in the system to meet the requirements, then the customer is lost. If upon arrival of a customer there are enough resources, then the customer occupies one server and the required number of resources for whole service duration. Service times are assumed to be exponentially distributed with intensity μ.

In order to simplify the analysis, we do not keep track of number of resources occupied by each customer. Instead of it, we keep track of only total number of occupied resources by all customers together [6]. Thus, we do not know the exact number of resources released on the departure of a customer. If there are k customers in the system just before the departure and r resources are occupied, then j resources are released upon the departure of a customer with probability $\dfrac{p_{r-j}^{(k-1)} p_j}{p_r^{(k)}}$, where $p_r^{(k)}$ is k-fold convolution of probability distribution $\{p_j\}$, $j \geq 0$.

The system behavior is described by a Markov process $X(t) = \{\xi(t), \delta(t), \eta(t)\}$, where $\xi(t)$ denotes number of customers in the system at time t, $\delta(t)$ – total number of occupied resources and $\xi(t)$ is the state of underlying Markov chain. The state set $X = \bigcup_{k=0}^{N} X_k$ can be easily decomposed into subset $X_0 = \{(k,m) : 0 \leq k \leq N, 1 \leq m \leq M\}$, and N subsets $X_k = \{(k,r,m) : 0 \leq k \leq N, 0 \leq r \leq R, p_r^{(k)} > 0, 1 \leq m \leq M\}$, $1 \leq k \leq N$. Let s_k be the number of nonzero probabilities in the k-fold convolutions $p_r^{(k)}$, $0 \leq r \leq R$. Then the number of states in subset X_0 is M, while subset X_k contain $s_k M$ states. Finally, let $I(k,r)$ be the sequence number of $p_r^{(k)}$ at the row of nonzero k-fold convolutions of probability distribution $\{p_j\}$, $j \geq 0$.

2.2 Stationary probability distribution and performance measures

Denote stationary probabilities

$$q_0(m) = \lim_{t\to\infty} P\{\xi(t) = 0, \eta(t) = m\}, 1 \leq m \leq M, \quad (1)$$

$$q_k(r,m) = \lim_{t\to\infty} P\{\xi(t) = k, \delta(t) = r, \eta(t) = m\}, \quad (k,r,m) \in X_k. \quad (2)$$

Then vectors $\mathbf{q}_k = (\mathbf{q}_k(r_1), \ldots, \mathbf{q}_k(r_{s_k}))$, $1 \leq k \leq N$ where $\mathbf{q}_k(r_i) = (q_k(r_i,1), \ldots, q_k(r_i,M))$, $(k, r_i, m) \in X_k$, and $\mathbf{q}_0 = (q_0(1), q_0(2), \ldots, q_0(M))$ satisfy the following system of equilibrium equations.

$$\begin{aligned} &\mathbf{q}_0\mathbf{D}_0 + \mathbf{q}_1\mathbf{M}_1 = \mathbf{0}, \\ &\mathbf{q}_{k-1}\mathbf{\Lambda}_{k-1} + \mathbf{q}_k\mathbf{D}_k + \mathbf{q}_{k+1}\mathbf{M}_{k+1} = \mathbf{0}, 1 \leq k \leq N-1, \\ &\mathbf{q}_{N-1}\mathbf{\Lambda}_{N-1} + \mathbf{q}_N\mathbf{D}_N = \mathbf{0}. \end{aligned} \quad (3)$$

The matrix $\mathbf{D}_0$ is given by

$$\mathbf{D}_0 = \mathbf{A} - \left(\sum_{j=0}^{R} p_j\right)\mathbf{A}_1, \quad (4)$$

while matrices $\mathbf{D}_k, 1 \leq k \leq N$ have block-diagonal structure. The dimension of the matrices $\mathbf{D}_k, 1 \leq k \leq N$ is $s_k M \times s_k M$ and the size of each block is $M \times M$. Diagonal blocks $\mathbf{d}_k(j,j), 1 \leq j \leq s_k$ of $\mathbf{D}_k, 1 \leq k \leq N-1$ are described by formula (5) :

$$\mathbf{d}_k\left(I(k,r), I(k,r)\right) = \mathbf{A} - k\mu\mathbf{I} - \left(\sum_{j=0}^{R-r} p_j\right)\mathbf{A}_1, \quad 0 \leq r \leq R, p_r^{(k)} > 0, \quad (5)$$

where $\mathbf{I}$ is a unit matrix. All diagonal blocks of $\mathbf{D}_N$ are calculated by

$$\mathbf{d}_N\left(I(N,r), I(N,r)\right) = \mathbf{A} - N\mu\mathbf{I}. \quad (6)$$

The matrix $\mathbf{\Lambda}_0$ can be represented as a row of $M \times M$ size blocks, the blocks $\boldsymbol{\lambda}_0(i), 1 \leq i \leq s_1$ have the following form

$$\boldsymbol{\lambda}_0\left(I(1,r)\right) = p_r\mathbf{A}_1, \ 0 \leq r \leq R, \ p_r > 0. \quad (7)$$

The size of matrices $\mathbf{\Lambda}_k$, $1 \le k \le N-1$ is $s_k M \times s_{k+1} M$, their blocks $\boldsymbol{\lambda}_k(i,j)$, $1 \le i \le s_k$, $1 \le j \le s_{k+1}$ are given by

$$\boldsymbol{\lambda}_k\left(I(k,r), I(k+1,s)\right) = p_{s-r}\mathbf{A}_1, \quad 0 \le r \le s \le R,\ p_r^{(k)} > 0,\ p_s^{(k+1)} > 0. \tag{8}$$

The matrix $\mathbf{M}_1$ can be represented as a column of blocks of size $M \times M$, the blocks $\mathbf{m}_1(i), 1 \le i \le s_1$ have the following form

$$\mathbf{m}_1\left(I(1,r)\right) = \mu\mathbf{I},\ \ 0 \le r \le R,\ p_r > 0. \tag{9}$$

The size of matrices $\mathbf{M}_k$, $2 \le k \le N$ is $s_k M \times s_{k-1} M$, their blocks $\mathbf{m}_k(i,j)$, $1 \le i \le s_k$, $1 \le j \le s_{k-1}$ are given by

$$\mathbf{m}_k\left(I(k,r), I(k-1,s)\right) = \frac{p_{r-s} p_s^{(k-1)}}{p_r^{(k)}} k\mu\mathbf{I}, \quad 0 \le s \le r \le R,\ p_r^{(k)} > 0,\ p_s^{(k-1)} > 0. \tag{10}$$

The system of equations (3) along with normalizing condition gives stationary probabilities (1-2).

With the stationary probability distribution, we can evaluate various performance metrics of the system, such as the average b_1 and the variance $\overline{b}_2$ of occupied resources :

$$b_1 = \sum_{(k,r,m)\in X_k} r q_k(r,m)\,, \tag{11}$$

$$\overline{b}_2 = \sum_{(k,r,m)\in X_k} r^2 q_k(r,m) - b\,. \tag{12}$$

According to the properties of MAP, the blocking probability π is given by

$$\pi = 1 - \lambda^{-1}\left(\sum_{\substack{k,r:0\le k\le N-1\\ 0\le r\le R,\, p_r^{(k)}>0}} \mathbf{q}_k(r)\sum_{j=0}^{R-r} p_j\right)\mathbf{a}\,. \tag{13}$$

3 Numerical analysis

In this section, we provide some results of the performance measures analysis. Mapping radio resource allocation policy to the probability distribution $\{p_j\},\ j \ge 0$ of the resource requirements in the queuing system is a quite complicated problem [12], so we left it to the further research.

We assume that resource requirements follow geometric distribution with parameter $p = 0.8$ and

$$p_j = (1-p)p^j, \quad j \ge 0\,.$$

Calculations were held for three arrival flows with equal arrival intensity ($\lambda = 8$) but different variation coefficient v of the interarrival times. The first arrival flow with $v = 0.7$ is defined by

$$A = \begin{bmatrix} -6 & 3 & 3 \\ 3 & -5 & 2 \\ 3 & 5 & -8 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 1 & 3 & 3 \\ 3 & 0 & 8 \\ 2 & 2 & 0 \end{bmatrix}.$$

Second arrival process represents MMPP flow (Markov Modulated Poisson Process) with variation coefficient $v = 1.5$:

$$A = \begin{bmatrix} -4 & 1 & 3 \\ 12 & -24 & 12 \\ 3 & 1 & -4 \end{bmatrix}, \quad A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 15 \end{bmatrix}.$$

And third arrival process is a Poisson flow with variation coefficient of the interarrival times $v = 1$.

Further, we assume that $N = R = 40$ and provide evaluation of performance metrics (11 – 13) for different values of system load $\rho = \dfrac{\lambda}{\mu}$. The results are plotted on figures 1, 2 and 3.

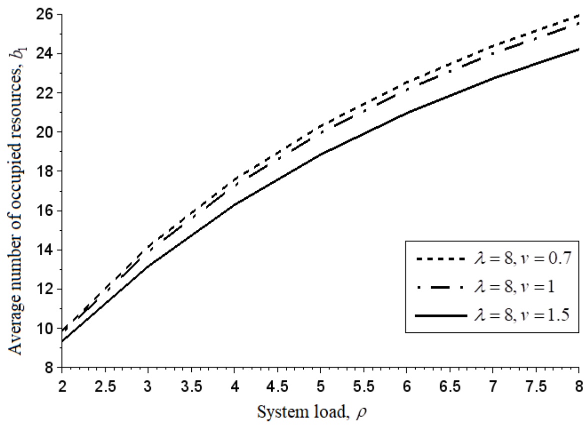

Fig. 1. Average number of occupied resources.

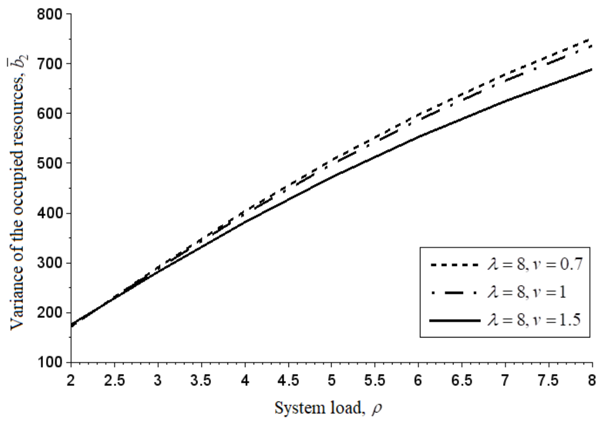

Fig. 2. Variance of the occupied resources.

Figures show that all of the investigated performance indicators improve with the decrease of the variability of interarrival times. Since the majority of request flows in real wireless networks have coefficient of variation less than one, assumption of arrival flow to be distributed according to Poisson law leads to lower bound of performance measures. However, to receive more accurate results one need to approximate the arrival flow with MAP.

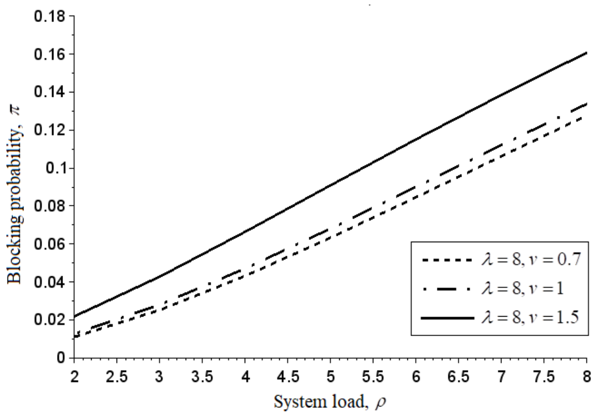

Fig. 3. Blocking probability.

4 Conclusion

In the paper, we analysed a multiserver queuing system with discrete limited resources under MAP arrivals. The described queuing system may be applied for the analysis of performance measures of modern wireless networks. We derived the system of equilibrium equations that can be solved numerically. Formulas for the main performance measures were deduced. Finally, we illustrated our results by numerical example.

In the future studies, we plan to develop an efficient algorithm for approximation of performance measures.

The research was financially supported by the Ministry of Education and Science of the Russian Federation (the Agreement number 02.a03.21.0008), and RFBR according to the research projects No. 16-37-60103 and No. 16-07-00766.

References

1. Ericsson mobility report, Ericsson AB, 2017 https://www.ericsson.com/assets/local/mobility-report/documents/2017/ericsson-mobility-report-june-2017.pdf
2. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021 https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf
3. J. Andrews, H. Claussen, M. Dohler, and S. Rangan, *IEEE Journal on Selected Areas in Communications*, **30**, 3, pp. 497–508 (2012).
4. G. Fodor, S. Parkvall, S. Sorrentino, P. Wallentin, Q. Lu, and N. Brahmi, *IEEE Access*, **2**, 1, pp. 1510–1520 (2014).
5. J. Fan, Q. Yin, G. Li, B. Peng, and X. Zhu, *IEEE Transactions on Wireless Communications*, **10**, 11, pp. 3966–3972 (2011).
6. O.M. Tikhonenko, *Automation and Remote Control*, **58**:6, pp. 969–973 (1997).
7. V.A. Naumov, K.E. Samuilov, A.K. Samuilov, *Automation and Remote Control*, **77**, 8, pp 1419–1427 (2016).
8. V. Naumov, K. Samouylov, N. Yarkina, E. Sopin, S. Andreev and A. Samuylov, *Proc. of 7th Congress on Ultra Modern Telecommunications and Control Systems*, pp. 100–103 (2015).
9. K. Samouylov, E. Sopin and O. Vikhrova, Analysis of the queueing system with resources and signals, *Proc. of 8th Congress on Ultra Modern Telecommunications and Control Systems*, (in print)
10. G. Basharin and V. Naumov, *Proc. of the 3rd International Seminar on Teletraffic Theory "Fundamentals of Teletraffic Theory"*, Moscow, VINITI, pp. 38–44 (1984).
11. D. M. Lucantoni, K. S. Meier-Hellstern, and M. F. Neuts, *Advances in Applied Probability*, **22**, 3, pp. 676–705 (1990).
12. Sopin E., Samouylov K., Vikhrova O., Kovalchukov R., Moltchanov D., Samuylov A., *Lecture Notes in Computer Science*, **9870**, pp. 440–450 (2016).