



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
РОССИЙСКИЙ УНИВЕРСИТЕТ ДРУЖБЫ НАРОДОВ
ИНСТИТУТ ПРОБЛЕМ УПРАВЛЕНИЯ
ИМ. В.А. ТРАПЕЗНИКОВА РАН

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ (ИТММ-2020)

**МАТЕРИАЛЫ
XIX Международной конференции
имени А. Ф. Терпугова
2–5 декабря 2020 г.**



ТОМСК
«Издательство НТЛ»
2021

УДК 519
ББК 22.17
И74

И74 Информационные технологии и математическое моделирование (ИТММ-2020): Материалы XIX Международной конференции имени А. Ф. Терпугова (2–5 декабря 2020 г.). – Томск: Изд-во НТЛ, 2021. – 498 с.

ISBN 978-5-89503-647-1

Сборник содержит избранные материалы XIX Международной конференции имени А. Ф. Терпугова по следующим направлениям: теория массового обслуживания и ее приложения, интеллектуальный анализ данных и визуализация, информационные технологии и программная инженерия, математическое и компьютерное моделирование технологических процессов.

Для специалистов в области информационных технологий и математического моделирования.

УДК 519
ББК 22.17

Редколлегия:

А.А. Назаров, доктор технических наук, профессор,
С.П. Моисеева, доктор физико-математических наук, профессор,
А.Н. Моисеев, доктор физико-математических наук, доцент,
М.П. Фархадов, доктор технических наук, профессор,
Е.Ю. Лисовская, кандидат физико-математических наук.

*Конференция проведена при поддержке
международного научно-методического центра
Томского государственного университета по математике,
информатике и цифровым технологиям в рамках
федерального проекта «Кадры для цифровой экономики»
национальной программы
«Цифровая экономика в Российской Федерации»*

ISBN 978-5-89503-647-1

© Авторы. Текст, 2021
© ООО «Издательство НТЛ».
Оформление. Дизайн, 2021



NATIONAL RESEARCH TOMSK STATE UNIVERSITY
PEOPLES' FRIENDSHIP UNIVERSITY OF RUSSIA
V.A. TRAPEZNIKOV INSTITUTE OF CONTROL SCIENCES
OF RUSSIAN ACADEMY OF SCIENCES

INFORMATION TECHNOLOGIES AND MATHEMATICAL MODELLING (ITMM-2020)

**PROCEEDINGS
of the 19th International Conference
named after A. F. Terpugov
2020 December, 2–5**



TOMSK
"Scientific Technology
Publishing House"
2021

UDC 519
LBC 22.17
I60

160 Information technologies and mathematical modelling (ITMM-2020):
Proceedings of the 19th International Conference named after
A. F. Terpugov (2020 December, 2–5). – Tomsk: Scientific Techno-
logy Publishing House, 2021. – 498 p.

ISBN 978-5-89503-647-1

This volume presents selected papers from the XIX International Conference named after A.F. Terpugov. The papers are devoted to new results in the following areas: queuing theory and its applications, data mining and visualization, information technology and software engineering, mathematical and computer modeling of technological processes.

UDC 519
LBC 22.17

Editors:

A.A. Nazarov, Doctor of Technical Sciences, Professor,
S.P. Moiseeva, Doctor of Physical and Mathematical Sciences, Professor,
A.N. Moiseev, Doctor of Physical and Mathematical Sciences, Associ-
ate Professor,
M.P. Farhadov, Doctor of Technical Sciences, Professor,
E.Yu. Lisovskaya, Candidate of Physical and Mathematical Sciences.

*The conference was supported by
International Computer Science
Continues Professional Development Center
of the Federal project «Human Resources for the Digital Economy»
of the National program
«Digital Economy of the Russian Federation»*

ISBN 978-5-89503-647-1

© Authors. Text, 2021
© Scientific Technology Publishing
House. Design, 2021

ТЕОРИЯ МАССОВОГО ОБСЛУЖИВАНИЯ И ЕЕ ПРИЛОЖЕНИЯ

QUEUING THEORY AND APPLICATION

To queueing system model performance measures analysis under network slicing*

Yves Adou, Ekaterina Markova

Peoples' Friendship University of Russia (RUDN University), Moscow, Russia

In modern wireless telecommunication networks, the network slicing (NS) technology [1] is being widely launched around the world. This technology was developed to address the problem of data traffic volume exponential growth [2, 3]. It permits a more effective utilization of the available radio resources, which are indispensable for data transfer in wireless networks. NS technology allows to represent the whole network infrastructure of mobile network operators (MNOs) in the form of various configurable logical networks called multi-service wireless network slices, each of which can be rented by virtual network operators (VNOs).

This paper considers the operation of a single VNO renting from one MNO a multi-service wireless network slice. We assume that VNO provides to its customers only services which generate best effort traffic with minimum guarantees (BG), characterized by the fact that a minimum bit rate is assigned for customers services. BG-services correspond to elastic or non-real time traffic such as file sharing, web browsing and social networking [4].

We assume that if there is free resource, the elastic session will be established upon arrival, otherwise it will await in the buffer, from where it can depart for buffer's retrial group [5, 6] and return after a moment. Let us consider, as example of physical interpretation, a web page reloading process by one user after awaiting a certain time this web page to load. In this case, user await in our model corresponds to elastic session await in buffer, and web page reloading process – to elastic session departure for buffer's retrial group and return to buffer after a moment.

* The reported study was funded by RFBR, project number 19-07-00933.

Similar model with limited storage capacities of buffer and its retrieval group has already been investigated in [7], where a computed numerical solution of the equilibrium equations system was used to analyze the characteristics. In this paper, we propose to consider a more complex case of that model with unlimited storage capacities of buffer and its retrieval group.

Mathematical model

We assume that a multi-service wireless network slice has total network capacity R_{nc} Mbps. Elastic sessions arrive in the system according to the Poisson process with rate λ and are parametrized by the exponentially distributed average file size θ MB and the minimum assignable bit rate b Mbps. The average service time of established elastic sessions is exponentially distributed with rate $\mu^{-1} = \theta/R_{nc}$ seconds. Capacities of buffer and its retrieval group are unlimited.

We assume that an elastic session is immediately established upon arrival if there is free resource, i.e. quantity of simultaneously established elastic sessions is less than maximum $N_{res} = \lfloor R_{nc}/b \rfloor$, otherwise its establishment is delayed and it awaits free resource in buffer. Note that elastic sessions awaiting in buffer can depart for its retrieval group after an exponentially distributed time β^{-1} and return after an exponentially distributed time α^{-1} . Also, each established elastic session ends successfully and retries are unlimited. The corresponding scheme model is shown in Fig. 1. Note that elastic sessions are characterized by a uniformly distributed channel rate R_{nc} between simultaneously established elastic sessions, thus service rate of single established elastic session is defined as μ/n , where $n \leq N_{res}$.

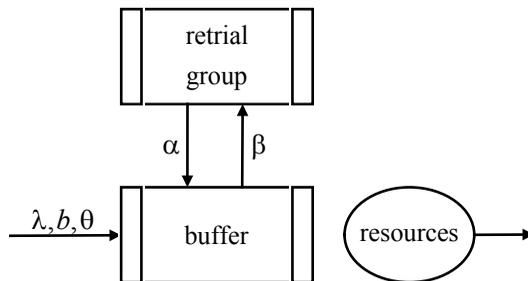


Fig. 1. QS model with unlimited storage capacities of buffer and its retrieval group scheme

According to above described radio admission control scheme and taking into account that system's arrival process is Poisson distributed while average service time is exponentially distributed, the system's behavior could be described using a three-dimensional vector (n_1, n_2, s) over state space $\mathbf{X} = \{(n_1, n_2, s) : n_1 \geq 0, n_2 \geq 0, s = 0, \dots, N_{res}\}$, where n_1 represents the quantity of elastic sessions awaiting free resources in buffer, n_2 – the quantity of elastic sessions awaiting in buffer's retrial group and s – the quantity of simultaneously established elastic sessions. The corresponding state transition and central state transition diagrams are shown respectively in Fig. 2 and Fig. 3.

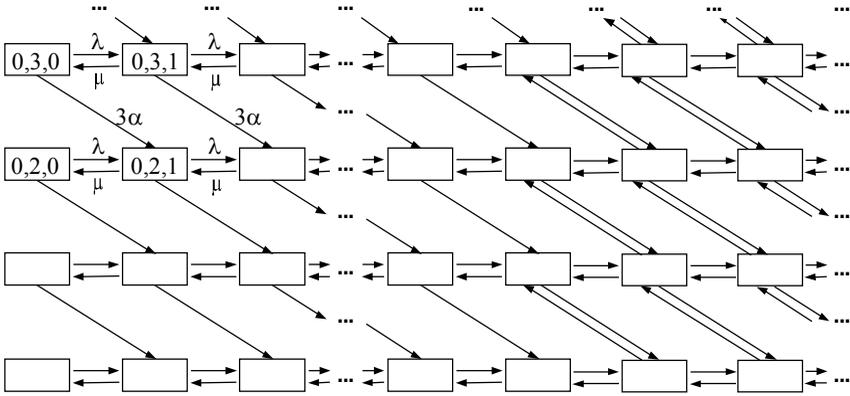


Fig. 2. State transition diagram

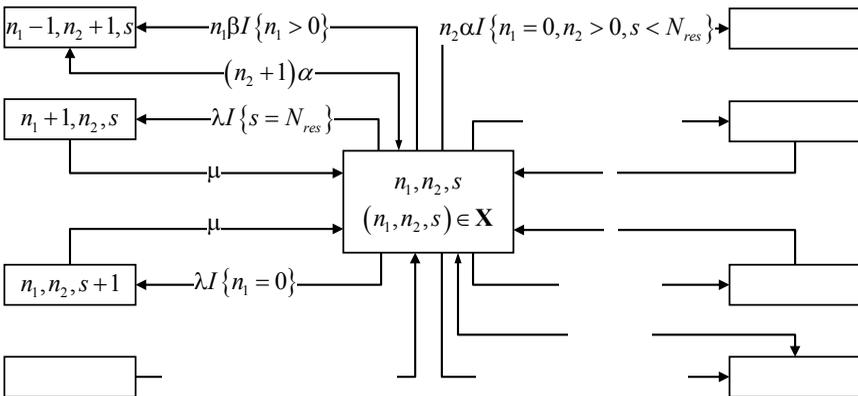


Fig. 3. Central state transition diagram

The corresponding Markov process is described by the following equilibrium equations system according to central state transition diagram (Fig. 3):

$$\begin{aligned}
& (\lambda + \mu I\{s > 0\} + n_1 \beta I\{n_1 > 0\} + n_2 \alpha I\{n_2 > 0\}) p(n_1, n_2, s) = \\
& = \lambda I\{n_1 > 0\} p(n_1 - 1, n_2, s) + \mu I\{s = N_{res}\} p(n_1 + 1, n_2, s) + \\
& + \lambda I\{n_1 = 0, s > 0\} p(n_1, n_2, s - 1) + \mu I\{n_1 = 0, s < N_{res}\} p(n_1, n_2, s + 1) + \quad (1) \\
& + (n_2 + 1) \alpha I\{n_1 = 0, s > 0\} p(n_1, n_2 + 1, s - 1) + (n_2 + 1) \alpha I\{n_1 > 0\} \times \\
& \times p(n_1 - 1, n_2 + 1, s) + (n_1 + 1) \beta I\{n_2 > 0, s = N_{res}\} p(n_1 + 1, n_2 - 1, s),
\end{aligned}$$

where $p(n_1, n_2, s)$, $(n_1, n_2, s) \in \mathbf{X}$ represents the stationary probability distribution of elastic sessions' quantity in system and I – the function indicator that equals 1, when condition is met and 0 otherwise.

Product form solution for stationary probability distribution computation

Since buffer and its retrial group have unlimited storage capacities, the probably only way for finding the stationary probability distribution $p(n_1, n_2, s)$, $(n_1, n_2, s) \in \mathbf{X}$ should be through a generating function-based approach [6], but we avoid its introduction using instead equilibrium and local balance equations systems. Thus, we obtain a product form solution for stationary probability distribution computation:

$$p(n_1, n_2, s) = \begin{cases} p(0, 0, 0) \cdot a(n_1, s), & n_1 \geq 0, n_2 = 0, s = 0, \dots, N_{res}; \\ p(0, 0, 0) \cdot d(n_1, n_2), & n_1 \geq 0, n_2 \geq 0, s = N_{res}; \\ p(0, 0, 0) \cdot b(n_2, s), & n_1 = 0, n_2 \geq 0, s = 0, \dots, N_{res} - 1, \end{cases} \quad (2)$$

with

$$\begin{aligned}
p(0, 0, 0) &= \left(\frac{1}{1 - \rho} + \sum_{k=0}^{N_{res}} \sum_{j=1}^k b(j, k) + \sum_{i=1}^{\infty} \sum_{j=1}^{N_{res}+i} d(i, j) \right)^{-1}; \\
a(i, j) &= \rho^{i+j}, i \geq 0, j \geq 0; \quad d(i, j) = \delta^j \frac{(i+j)!}{i!j!} \rho^{N_{res}+i+j}, i \geq 0, j \geq 0; \\
b(i, j) &= \frac{\delta^i}{j!} \left(\frac{\mu}{\lambda + i\alpha} \right)^{N_{res}-j} \rho^{N_{res}+i}, i \geq 0, j \geq 0,
\end{aligned}$$

where $\rho = \lambda/\mu$ represents the average offered load of incoming elastic sessions and $\delta = \beta/\alpha$ – the ratio coefficient of elastic' sessions average time spent in buffer to average time spent in buffer's retrial group.

Numerical analysis of incoming' elastic sessions immediate establishment probability

Having computed the stationary probability distribution one can compute the system' main performance measure. Here we compute only one of them – the immediate establishment probability of incoming elastic sessions:

$$\text{Prob} = \lim_{n_1 \rightarrow \infty} \sum_{s=0}^{N_{res}-1} \sum_{n_2=0}^{N_{res}+n_1} p(0, n_2, s). \quad (3)$$

Numerical analysis was performed to determine the accuracy of the computed product form (2). Let us calculate percentage error for computing incoming' elastic sessions immediate establishment probability (3) using computed numerical solution in [7] and product form solution.

The results obtained using both product form and numerical solution coincide as illustrated in Fig.1. by the curves and in Table 1 by absolute error calculation values.

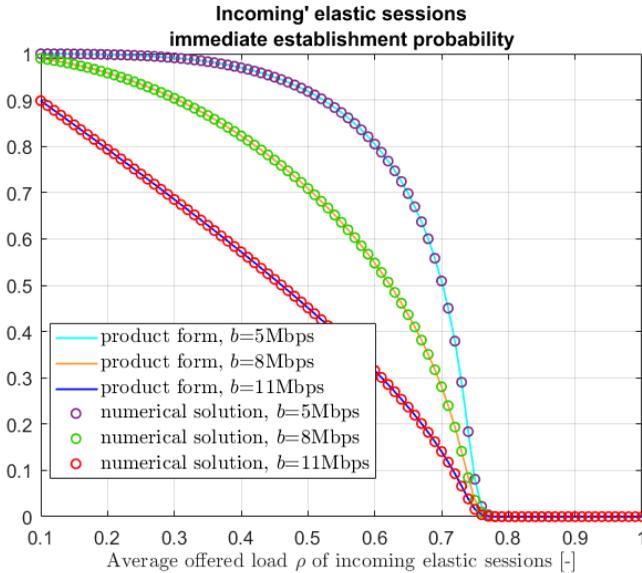


Fig. 1. Immediate establishment probability depending on average offered load ρ of incoming elastic sessions for $R_{nc} = 20$ Mbps, $\theta = 8$ MB, $\alpha^{-1} = 3$ s, $\beta^{-1} = 1$ s, limit of n_1 as it approaches infinity is 1000

Table 1

Absolute error for incoming' elastic session immediate establishment probability calculation, $R_{nc} = 20$ Mbps, $\theta = 8$ MB, $\alpha^{-1} = 3$ s, $\beta^{-1} = 1$ s, limit of n_1 as it approaches infinity is 1000

Parameters	$\rho = 0.1$	$\rho = 0.3$	$\rho = 0.5$	$\rho = 0.7$
$b = 5$ Mbps	0 0%	7.01296489e-14 0%	3.86746059e-15 0%	3.53375158e-14 0%
$b = 8$ Mbps	0 0%	1.74086275e-08 0%	0 0%	1.22617108e-14 0%
$b = 11$ Mbps	0 0%	4.83553803e-06 0%	1.98127869e-14 0%	2.76230925e-15 0%

Conclusions

We continued study on queuing system (QS) model with unlimited storage capacities of buffer and its retrial group. Product form solution for calculating stationary probability distribution was obtained. A comparative analysis of incoming' elastic sessions immediate establishment probability computation using numerical solution of the equilibrium equations system and a proposed product form solution was given. As result, both product form and numerical solution coincide with absolute error value equaling 0%.

Generating function-based approach will be implemented in further investigations, and the operation of multiple VNOs renting from one MNO a multi-service wireless network slice will be considered as last stage of this research subject.

REFERENCES

1. China Mobile, Huawei, and Industry Partners Release "Categories and Service Levels of Network Slice", White Paper, March 2020.
2. Ericsson, "Ericsson mobility report", June 2020.
3. Cisco, "Cisco Annual Internet Report (2018–2023)", White Paper, March 2020.
4. TS 129 512 – V15.3.0 – 5G; 5G System; Session Management Policy Control Service; Stage 3 (3GPP TS 29.512 version 15.3.0 Release 17) ETSI, Sept 2020.
5. Vishnevsky V.M., Dudin A.N., Klimenok V.I. Stochastic systems with correlated flows. Theory and application in telecommunication networks. Moscow: Advertising and publishing center «TECHNOSPHERE», 2018. 564 p. ISBN 978-5-94836-520-6.
6. Bocharov P.P., D'Apice C., Pechinkin A.V., Salerno S. Queueing Theory. De Gruyter, 2003. DOI: 10.1515/9783110936025.
7. Markova E., Adou Y., Ivanova D., Golskaia A., Samouylov K. (2019) Queue with Retrial Group for Modeling Best Effort Traffic with Minimum Bit Rate Guarantee Transmission Under Network Slicing // Vishnevskiy V., Samouylov K., Kozyrev D. (eds.). Distributed Computer and Communication Networks. DCCN 2019. LNCS. V. 11965. Springer, Cham. DOI: 10.1007/978-3-030-36614-8_33.

Analysis of the simplified network slicing model*

Kirill Ageev¹, Eduard Sopin^{1,2}

¹ People's Friendship University of Russia (RUDN University), Russia, Moscow

² Institute of Informatics Problems, Federal Research Center
«Computer Science and Control», Russia, Moscow

Network slicing allows a mobile network operator to provide a part of the radio resource to virtual network operators in the form of network slices. A network slice can be allocated for specific types of services to several virtual network operators that provide similar services, or separately for each virtual operator [1].

A model of radio resources sharing with a dynamic distribution of resources between classes was described in [2], the characteristics were calculated and compared with the results of simulations.

In this article we considered a simplified model for sharing radio resources in the network. The total resource volume is divided into $K + 1$ blocks. $1, \dots, K$ blocks are isolated with each own flow of customers, serving intensities and resource distribution. There is also the shareable block ($\#0$), which accepts customers, that are blocked in $1, \dots, K$.

Model description

K Poisson flows of customers arrive into multiserver queuing system with intensities $\lambda_k, k = 1, \dots, K$. For each type of customers, the required amount of resources is defined by the probability distribution $\{p_{k,r}, r \geq 1\}$. Service times are exponentially distributed with intensity μ_k .

Customers of type k firstly arrive to the block k , which has $R_k, k = 1, \dots, K$ resources. As the simplification, we assume that the resource volumes $R_k, k = 1, \dots, K$ are fixed for all blocks and do not change in time. Maximum number of customers in block k is $N_k = R_k$.

Besides, there is also a shareable block with R_0 resources. If a k -type customer is blocked due to insufficient resources at block k , then it is redirected to the shareable block.

* The reported study was funded by RFBR, projects number 19-07-00933 and 19-37-90147.

Total volume of resources is the following:

$$R = \sum_{k=0}^K R_k . \quad (1)$$

Let $n_k(t)$ be the number of customers in block k at moment t ($t > 0$), and $\gamma_k(t) = (\gamma_{k_1}(t) \dots \gamma_{k_m}(t))$ – the vector of occupied resources by each customer in block. To simplify the description of the set of states we applied the aggregation of occupied volume of resources. $\Gamma_k(t)$ – total occupied volume of resources in block k [4]. Therefore, the set of states can be expressed as

$$X_k = \{(n_k, \Gamma_k) : 0 \leq n_k \leq N_k, 0 \leq \Gamma_k \leq R_k\}, k = 1, 2, \dots, K . \quad (2)$$

Assume that a k -type customer arrives to the system and it requires r resources. Then, if there are not enough resources in block k to meet the resource requirements, $\Gamma_k + r \geq R_k$, the customer is redirected to the common block R_0 . If there are not enough resources for the customer in the common block, $\Gamma_0 + r \geq R_0$, the customer is lost.

Denote $q_{k,n}(j)$ the stationary probability that there are n customers occupying j resources at block k . The exact expression for the probabilities $q_{k,n}(j)$ were derived in [4]. Distribution of resource requirements for the flow of rejected customers in block k can be calculated according to the next formula:

$$\tilde{p}_{k,r} = \frac{1}{\pi_k} p_r \left(\sum_{\substack{1 \leq n \leq N_k - 1 \\ R_{k-r+1} \leq j \leq R_k}} q_{k,n}(j) + \sum_{1 \leq j \leq R_k} q_{k,N_k}(j) \right) . \quad (3)$$

To define the resource requirements distribution of the customers that arrive to the shareable resource block, we combine distributions of resource requirements for each block k :

$$p_{0,r} = \sum_{k=1}^K \frac{P_k \pi_k}{\rho_0} \tilde{p}_{k,r} . \quad (4)$$

The offered load at the shareable block can be calculated according to next formula:

$$\rho_0 = \rho_1 \pi_1 + \rho_2 \pi_2 + \dots + \rho_K \pi_K . \quad (5)$$

The scheme of the described model is presented on Fig. 1.

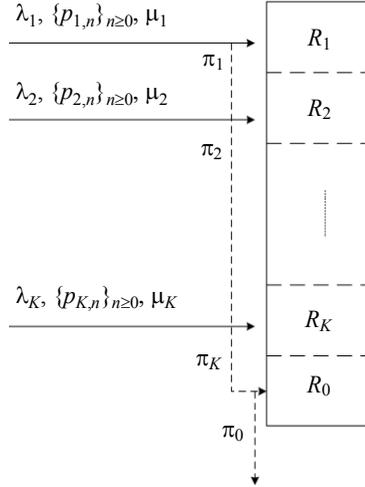


Fig. 1. The scheme of the model

Analysis of the performance measures

Denote $G_k(n, m) = \sum_{i=0}^n \frac{\rho_k^i}{i!} \sum_{0 \leq j \leq m} p_{k,j}^{(i)}$, where $\{p_{k,j}^{(i)}\}$ is the i -fold convolution of the distribution $\{p_{k,j}\}$. According to [4], the stationary probability distributions at block k can be evaluated using the functions $G_k(n, m)$ in the following way:

$$G_k(N_k, R_k)^{-1} G_k(n, m) = \sum_{\substack{0 \leq i \leq n \\ 0 \leq j \leq r}} q_{k,i}(j). \quad (6)$$

Based on recurrent algorithm, presented in [2], we can find G by the recurrent relation

$$G(n, m) = \frac{\rho_k}{n_k} \sum_{0 \leq j \leq m} p_{k,j} (G(n-1, r-m) - G(n-2, r-m)) + G(n-1, m), 2 \leq n \leq N_k. \quad (7)$$

The initial values are as follows:

$$G(1, m) = 1 + \rho_k \sum_{0 \leq j \leq m} p_{k,j}, 0 \leq m \leq R_k, \quad (8)$$

$$G(0, m) = 1, 0 \leq m \leq R_k. \quad (9)$$

We calculated the probability of redirecting customers from block k to the shareable block 0 by next formula:

$$\pi_k = 1 - G_k^{-1}(N_k, R_k) \sum_{0 \leq j \leq R_k} p_{k,j} G_k(N_k - 1, R_k - j). \quad (10)$$

We next used formula to calculate the average volume of occupied resources in block k :

$$b_k = R_k - G_k^{-1}(N_k, R_k) \sum_{r=1}^{R_k} G_k(N_k, R_k - r). \quad (11)$$

The resource requirements distribution of the redirected customers at block 0 can be also obtained with the help of $G_k(n, m)$, by substituting (6) into (3):

$$\begin{aligned} \tilde{p}_{k,r} &= p_{k,r} G_k(N_k, R_k)^{-1} (G_k(N_k - 1, R_k) - G_k(N_k - 1, R_k - r) + \\ &+ G_k(N_k, R_k) - G_k(N_k - 1, R_k)) = p_{k,r} G_k(N_k, R_k)^{-1} (G_k(N_k, R_k) - \\ &- G_k(N_k - 1, R_k - r)). \end{aligned} \quad (12)$$

Blocking probability for k -type customers derived by formula

$$B_k = \pi_k \pi_0. \quad (13)$$

Numeric results

For the experiment we took the number of blocks $K = 3$. The distribution of resource requirements were taken from [5] and is assumed the same for each block. Note that the average resource requirement is 2.99. For blocks 1–3 we set the offered load $\rho_1 = 200$, $\rho_2 = 60$, $\rho_3 = 32$.

Blocks 1–3 are assumed to have the same number of resources. The results were obtained for different capacity of the shareable block R_0 with fixed total number of resources R . The range of the resource volumes in blocks 1–3 is $R_1 = R_2 = R_3 = [60, 100]$. The range for block 0 is $R_0 = [50, 170]$. Total volumes of servers and resources are $R = N = 350$.

Figure 2 show the dependence of the probability of redirecting customers from blocks 1–3 to the shareable block on its capacity. Note that the probability is increasing, which is correct, because we decrease the resources for blocks 1–3.

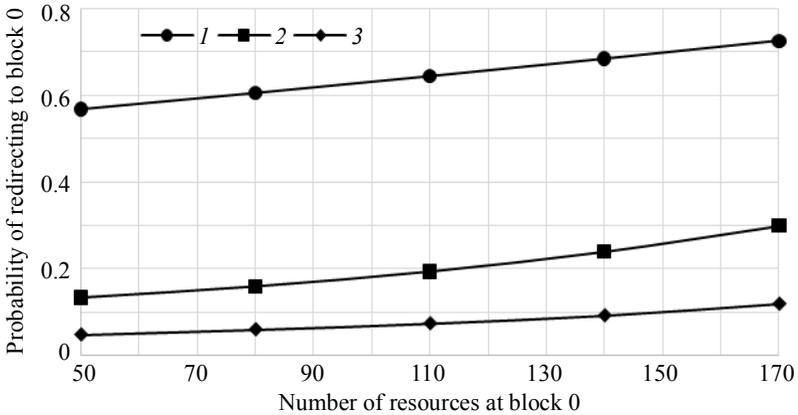


Fig. 2. Probability of redirecting to the shareable block as a function of its capacity

Figure 3 show the dependence of blocking probability of each customer types on the capacity of the shareable block.

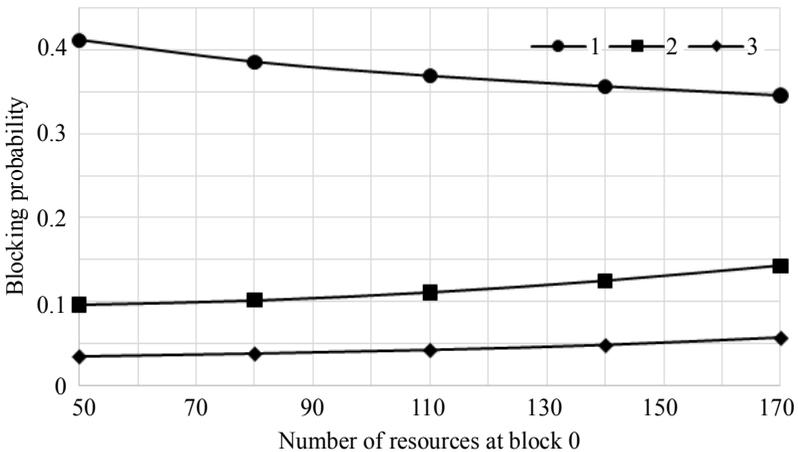


Fig. 3. Blocking probability depending on distribution of resources

On Fig. 3 we can see, that for customers in blocks 2 and 3 blocking probability increase and stay lower than for block 1. While for customers from block 1, the probability of blocking decreases. This behavior is explained by the fact that an equal amount of resources is allocated for blocks 1–3, and the offered load for block 1 is significantly higher than for blocks 2 and 3. Thus, when the volume of resources for blocks 1–3 decreases, the intensity of redirections from block 1 to block 0 increases. That allows customers from block 1 to occupy more resources in block 0, thus reducing the blocking probability.

Conclusions

In this paper, a simplified model of radio resource slicing is studied. The system is described in terms of resource queuing systems. The stationary characteristics of the model are calculated using a recurrent algorithm obtained in early studies. A numerical experiment was performed for three blocks. Based on the results, it can be noted that the proposed algorithms and formulas can be applied to the analysis of the systems that use radio resource slicing mechanisms.

At the next stage, we plan to develop the simulation tool to compare the results with analytical calculations.

REFERENCES

1. 3GPP TS 23.501 V15.4.0 – System architecture for the 5G System. http://www.3gpp.org/ftp/Specs/archive/23_series/23.501/23501-f40.zip.
2. *Sopin E.S., Ageev K.A., Markova E.V., Vikhrova O.G., Gaidamaka Yu.V.* Performance analysis of M2M traffic in LTE network using queuing systems with random resource requirements // Automatic Control and Computer Sciences. 2018. V. 52. No. 5. P. 345–353.
3. *Ageev K.A., Sopin E.S., Yarkina N.V., Samouylov K.E., S Shorgin.Ya.* Analysis of the network slicing mechanisms with guaranteed allocated resources for various traffic types // Informatics and Applications. 2020. V. 14. Iss. 3. P 94–100.
4. *Naumov V.A., Samuilov K.E., Samuilov A.K.* On the total amount of resources occupied by serviced customers // Automation and Remote Control. 2016. V. 77. P. 1419–1427. DOI: 10.1134/S0005117916080087.
5. *Begishev V., Moltchanov D., Sopin E., Samuylov A., Andreev S., Koucheryavy Y., Samouylov K.* Quantifying the impact of guard capacity on session continuity in 3GPP new radio systems // IEEE Transactions on Vehicular Technology. 2019. V. 68. No. 12. P. 12345–12359. DOI: 10.1109/TVT.2019.2948702.

An eigen value approach to a discrete-time queueing model with N -policy on two modes of service

Anilkumar M.P.¹, K.P. Jose²

¹*Dept. of Mathematics, T. M. Govt. College, Tirur-676502, Kerala, India*

²*Dept. of Mathematics, St. Peter's College, Kolenchery-682311, Kerala, India*

This paper analyses a discrete-time queueing model with two modes of service and N -policy. The arrival of customers constitutes a Bernoulli process with parameter p . There are two types of service; mode 1 and mode 2 of which service times are geometrically distributed with parameters q_1 and q_2 respectively, where $q_1 < q_2$. Initially, the service starts with mode 1 and when the number of customers reaches N , the server tries to change the type of the service to mode 2 with probability θ (called (θ, N) -policy). Once the type of the service is changed from mode 1 to mode 2, it will resume the reduced rate when either of two cases happen; i) the number of customers in the system is either less than N and ii) the number of customer is reduced to zero. These two cases are studied in Model I and Model II respectively. The eigenvalue approach is used to analyze Model I and consequently obtain the rate matrix of the model. Using the rate matrix of the Model I, we analyze Model II. On the basis of a suitable cost function, numerical experiments are conducted for the models and computed the optimum value of N .

For the details of theory of generalized eigenvalue, one can refer Gohberg et al. [2]. The method of generalized eigenvalue is successfully applied in queueing theory by Mitrani and Chakka [7] and Haverkort and Ost [6]. The other notable works carried out using generalized eigen value approach by Grassmann and Drekić [3] in tandem queues, Grassmann and Tavokali [4] in tandem queues with a movable server and Drekić and Grassmann [1] in priority queues. Recently, Grassmann and Tavakoli [5] demonstrated the fastest way to find the distribution of the queue length in a discrete GI/G/1 queue with bounded support from the waiting time distribution.

Mathematical modelling – Model I

In this model, we assume that once the mode is changed to mode 2, it will continue until the number of customers in the system is less than N . We assume that in a time slot, arrival takes place at the beginning of a slot which

generalized eigen value of $P(x)$ and g is the corresponding eigen vector (Ghoberg [2]). The equation $|p(x)|=0$ implies that either $pq_2 + t_2x + \bar{p}q_2x^2 = 0$ or $pq_1\bar{\theta} + (t_1\bar{\theta} - 1)x + \bar{p}q_1\bar{\theta}x^2 = 0$. From the first equation $x=1$ or $\frac{pq_2}{\bar{p}q_2}$. Since $|x| < 1$ is necessary condition for the

convergence, we can only consider the value $x_2 = \frac{pq_2}{\bar{p}q_2}$. The other root will

be $x_1 = \frac{\theta(\bar{\theta})^{-1} + \bar{p}q_1 + \bar{p}q_1 - \sqrt{(\theta(\bar{\theta})^{-1} + \bar{p}q_1 + \bar{p}q_1)^2 - 4p\bar{p}q_1\bar{q}_1}}{2\bar{p}q_1}$, which lies

between 0 and 1. In order to simplify the notations, assume that

$$q_1(x) = pq_1\bar{\theta} + (t_1\bar{\theta} - 1)x + \bar{p}q_1\bar{\theta}x^2,$$

$$q_2(x) = pq_2 + t_2x + \bar{p}q_2x^2,$$

$$h(x) = (pq_1 + t_1x + \bar{p}q_1x^2)\theta.$$

Let $g^{(i)} = [g_1^{(i)}, g_2^{(i)}]$ be eigen vector corresponding to the eigen value x_i for $i = 1, 2$. Then $g^{(1)} = \left[1, -\frac{h(x_1)}{q_2(x_1)}\right]$ and $g^{(2)} = [0, 1]$. Let $c = [c_1 \ c_2]$,

$G = \begin{bmatrix} 1 & -\frac{h(x_1)}{q_2(x_1)} \\ 0 & 1 \end{bmatrix}$ and $\wedge = \begin{bmatrix} x_1 & \\ & x_2 \end{bmatrix}$, then $\pi_n = c \wedge^{n-1} G$ for $n > N$,

where $c = \frac{p}{pq_1} \left(\frac{\bar{p}q_1}{pq_1}\right)^{N-2} B_0(G - GA_1 - \wedge GA_2)^{-1} \pi_0$, and π_0 is obtained by

using the normalizing condition $\sum_{i=0}^{\infty} \pi_i = 1$.

Relationship with matrix – analytic method

By analyzing the transition probability matrix, using matrix analytic method in Nuets [9], we get $\pi_n = \pi_N R^{n-N}$ for $n \geq N$, where R is the minimal non negative solution of $R^2 A_2 + RA_1 + A_0 = R$. Then from $\pi_n = c \wedge^{n-N} G$

follows $\pi_n R^{n-N} = c \wedge^{n-N} G$ and $R = G^{-1} \wedge G$. On simplification,

$$R = \begin{bmatrix} x_1 & \frac{\theta x_1}{\theta p q_2 (1-x_1)} \\ 0 & x_2 \end{bmatrix} \text{ similar to the rate matrix in Naishuo Tian et al. [10].}$$

Mathematical modelling – Model II

Assume that the service mode is changed to mode 2. it will continue the same until the number of customers reaches zero. Then $(N(n), J(n))$, $n = 1, 2$, is quasi birth death process with stste space $\{0\} \cup \{(i, j) : i \geq 1, j = 1, 2\}$.

The system is stable if and only if $p < q_2$. Let $(\pi_1^*, \pi_2^*, \dots)$ be the steady state probability vector of P^* . By looking through the structure of P^* , we get $\pi_{i+1}^* = \pi_2^* R$ for $i > N-1$. The boundary probability vectors $(\pi_0^*, \pi_1^*, \dots, \pi_{N-1}^*)$ are calculated using the following equations. Let $\pi_i^* = (\pi_1^*, \pi_2^*)$ for $i \geq 1$

$$\pi_{11}^* = \frac{p}{p q_1} \frac{(1-t_1^{N-3}) p \bar{q}_1 (1-t_1^{N-3})(1-t_1-x_1 \bar{p} q_1)}{(1-t_1^{N-2}) p \bar{q}_1 - (1-t_1^{N-1})(1-t_1-x_1 \bar{p} q_1)} \pi_0^*,$$

$$\pi_{12}^* = \frac{p}{p q_2} \frac{p \bar{q}_1 (t_1^{N-3} - t_1^{N-2})(1-t_1^{N-1}) - (1-t_1-x_1 \bar{p} q_1)(t_1^{N-2} - t_1^{N-3})}{(1-t_1^{N-2}) p \bar{q}_1 - (1-t_1^{N-1})(1-t_1-x_1 \bar{p} q_1)} \pi_0^*,$$

$$\pi_{i1}^* = \frac{(1-t_1^i)}{1-t_1} \pi_{11}^* - \frac{p(1-t_1^{i-1})}{p q_1 (1-t_1)} \pi_0^* \text{ for } 2 \leq i \leq N-1,$$

$$\pi_{i2}^* = \frac{1-t_2^i}{1-t_2} \pi_{12}^* \text{ for } 2 \leq i \leq N-1,$$

where π_0^* is obtained from the normalizing condition $\pi_0^* + (\pi_1^* + \pi_2^* + \dots \pi_{N-1}^*) (I - R)^{-1} e = 1$.

Performance measures

Without loss generality we may assume that $(\psi_0, \psi_1, \dots, \psi_{N-1}, \psi_N, \dots)$ be the steady state probability vector for the models with

$\Psi_i = (\Psi_{i1}, \Psi_{i2})$ for $i \geq 1$ with the assumption that $\Psi_{i2} = 0$ for $1 \leq i \leq N-1$. The following relevant performance measures are considered.

1. Expected number of customers in the system, $EC = \sum_{i=1}^{\infty} i\Psi_i$.
2. Probability that the server is busy with q_1 service, $PBq_1 = \sum_{i=1}^{\infty} \Psi_{i1}$.
3. Probability that the server is busy with q_2 service, $EBq_2 = \sum_{i=N}^{\infty} \Psi_{i2}$.
4. Probability of starting a q_2 service in an epoch

$$Pstartq_2 = \theta \left(\sum_{i=N}^{\infty} \Psi_{i1} + p\Psi_{(N-1)1} \right).$$

5. Expected departure after completing the service is given by

$$EDS = q_1 \sum_{i=1}^{\infty} \Psi_{i1} + q_2 \sum_{i=1}^{\infty} \pi_{i2}.$$

Cost analysis and optimisation

On the basis of the above performance measures, we define expected total cost per unit time as

$$ETC = C_1 EC + C_2 PBq_1 + C_3 EBq_2 + C_4 Pstartq_2 + C_5 EDS.$$

Where the individual costs C_1, C_2, C_3, C_4 and C_5 are given by

C_1 : holding cost of customers/unit/unit time,

C_2 : running q_1 service/unit time,

C_3 : running q_2 service/unit time,

C_4 : starting cost q_2 service/unit time,

C_5 : service cost of customers/unit/unit time.

Figure 1 compares the expected total cost for both model. It is obvious from the figure that for each value of n , the expected total cost per unit time for Model II is less than that of Model I. This is due to the low switching rate of Model II compared to Model I. From the figure, it is also evident that the value N at which ETC is minimum for Model II is greater than that of Model I.

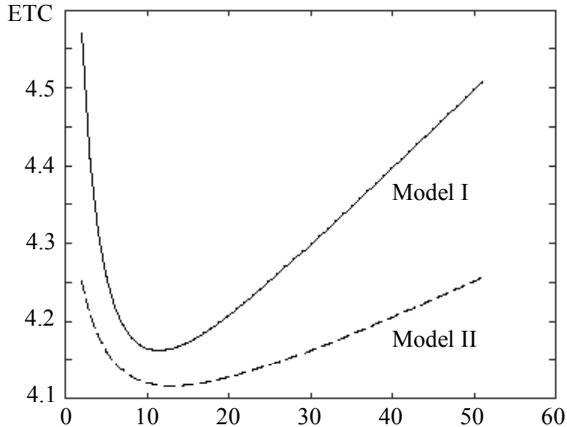


Fig. 1. ETC vs. N
 $(p = 0.55; q_1 = 0.5; q_2 = 0.7; \theta = 0;$
 $C_1 = 0.01; C_2 = 1; C_3 = 2; C_4 = 1; C_5 = 5)$

Concluding remarks

This paper analyzed an eigenvalue approach for finding the optimum value of the number of customers in the system at which the service rate is to be changed. We obtained the rate matrix and a closed-form solution to this model. To minimize the starting cost for changing the mode of service, a modified model is also discussed. This paper can be extended by considering the arrival process as a discrete Markovian arrival process and service time as discrete phase-type distribution.

REFERENCES

1. *Drekic S., Grassmann W.K.* An eigenvalue approach to analyzing a Finite source priority queueing model // *Annals of Operations Research*. 2002. V. 112(1-4). P. 139–152.
2. *Gohberg I., Lancaster P., Rodman L.* Matrix Polynomials. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), 1982.
3. *Grassmann W.K., Drekic S.* An analytical solution for a tandem queue with blocking // *Queueing Systems*. 2000. V. 36(1-3). P. 221–235.
4. *Grassmann W.K., Tavakoli J.* A tandem queue with a movable server: an eigenvalue approach // *SIAM Journal on Matrix analysis and Applications*. 2002. V. 24(2). P. 465–474.
5. *Grassmann W.K., Tavakoli J.* The distribution of the line length in a discrete time GI/G/1 queue // *Performance Evaluation*. 2019. V. 131. P. 43–53.

6. *Haverkort B.R., Ost A.* Steady-state analysis of infinite stochastic petri nets: Comparing the spectral expansion and the matrix-geometric method // Proceedings of the Seventh International Workshop on Petri Nets and Performance Models. IEEE, 1997. P. 36–45.
7. *Mitrani I., Chakka R.* Spectral expansion solution for a class of markov models: Application and comparison with the matrix-geometric method // Performance Evaluation. 1995. V. 23(3). P. 241–260.
8. *Morse P.M. et al.* Queues, inventories, and maintenance. 1958.
9. *Neuts M.F.* Matrix-geometric solutions in stochastic models: an algorithmic approach. Courier Corporation, 1994.
10. *Tian N., Ma Zh., Liu M.* The discrete time geom/geom/1 queue with multiple working vacations // Applied Mathematical Modelling. 2008. V. 32(12). P. 2941– 2953.

A MAP/PH(1), PH(2)/2 inventory system with production, multiple servers and vacations

P. Beena¹, K.P. Jose²

¹ Govt. Sanskrit College, Pattambi, Palakkad-679303, Kerala, India

² Dept. of Mathematics, St. Peter's College, Kolenchery-682311, Kerala, India

This article considers a production inventory system with multiple servers, each of which takes multiple vacations under Bernoulli vacation scheduling service. Customer's arrival constitutes a Markovian Arrival Process (MAP) and heterogeneous servers provide phase-type service. When the production process begins, the time required to produce an item and the duration of vacation of servers are exponentially distributed. The production process begins when the inventory level reaches a prefixed point and it stops when the inventory level reaches the maximum S . The QBD process is analyzed in the steady-state using Matrix Analytic Method (MAM). We develop an appropriate cost function based on the performance measures and compare the total expected cost in the different combinations of arrivals and service processes under the Bernoulli vacation schedule.

In a queueing system, vacation means that the period at which servers are not convenient for providing the service in a frequent period. Krishnakumar et al. [6] studied a MAP/Ph(1), Ph(2)/2 queue with multiple vacations under the Bernoulli vacation scheduling service. The study on the Bernoulli vacation model was initiated by Keilson and Servi [5]. Ayyappan and Shyamala [1] investigated a single server Bernoulli vacation model in which the server provides two stages of mandatory service to customers who arrived in batches. Beena and Jose [2] discussed a production inventory system with the production rate depend on the stock level. Tyagi et al. [11] analyzed a 1-limited, multiple server, multiple vacation queueing model in which vacation time follows a phase-type distribution. Suganya and Sivakumar [10] studied a MAP/Ph(2) inventory system in which the arrival of buyers is according to a MAP. Krishnamoorthy and Viswanath [8] discussed a production inventory model with the production process follow a Markovian production scheme. Krishnakumar and Pavai Madheswari [7] analyzed a multiple server queueing model in which each server takes the Bernoulli vacation schedule. A single server vacation model in which the primary customers can opt for additional service with some probability or quit from the system with its complementary probability was analyzed by

Chakravarthy and Ozkar [3]. Jose and Beena [4] analyzed a multiple server production inventory system with the retrial of customers.

Mathematical model

In this paper, a production inventory system where only one customer comes into the system at a time and servers extend the heterogeneous service to the customers. The arrivals of customers are characterized by MAP with representation $(D_0, D_1)_l$ and service rates of server 1 and server 2 are phase-type distributed having representations $(\alpha, S)_m$ and $(\beta, T)_n$ respectively. Vacation duration of servers 1 and 2 are exponentially distributed with rate θ_1 and θ_2 . If there are no customers in the system or the inventory level is zero or both, the servers will always take a vacation. Upon completion of the service, the servers can choose a vacation with probability $p_i, i=1,2$, or continue the service with its complementary probability $q_i = 1 - p_i, i=1,2$, if there are a positive inventory level and customers in the waiting area. When the servers return after a vacation, they again go back to the vacation if the system is empty or the inventory level is zero, or both. The servers will continue to do this until they find the system nonempty with a positive inventory level. Items are only available after a certain period and are distributed according to an exponential distribution with parameter $\gamma > 0$. The notations used in this model are

- $N(t)$ and $I(t)$ denote the number of customers in the system and inventory level at time t . $C(t)$ and $J(t)$ indicate server status and production status, where

$$C(t) = \begin{cases} 0, & \text{if both the servers are on vacation,} \\ 1 & \text{if server 1 is busy and server 2 is on vacation,} \\ 2, & \text{if server 1 is on vacation and server 2 is busy,} \\ 3, & \text{if both the servers are busy;} \end{cases}$$

$$J(t) = \begin{cases} 0, & \text{if production process gets switched OFF,} \\ 1, & \text{if production process gets switched ON;} \end{cases}$$

- $e_1 = e_{l(2S-s)+lm(2S-s-1)+\ln(2S-s-1)+lmm(2S-s-2)}$, $e_3 = e_{l(2S-s)}$,
 $e_2 = e_{l(2S-s)+lm(2S-s-1)+\ln(2S-s-1)}$ and e is a column vector of 1's.

The normalizing equation is

$$x_0 e_3 + x_1 e_2 + x_2 (I - R)^{-1} e_1 = 1. \quad (5)$$

R is the minimal nonnegative solution of the matrix quadratic equation $R^2 A_2 + R A_1 + A_0 = 0$ where the spectral radius of R is less than one. The subvectors x_0, x_1, x_2 and $x_i, i \geq 3$, can be calculated using equations (1) – (5).

System performance measures

1) The expected number of customers in the system is

$$E_{EC} = x_1 e_2 + x_2 \left[2(I - R)^{-1} + R(I - R)^{-2} \right] e_1.$$

2) The expected switching rate is

$$E_{SWR} = \sum_{i=1}^{\infty} y_{(i,1,0,s+1)} [I_l \otimes S^0] e + \sum_{i=1}^{\infty} y_{(i,2,0,s+1)} [I_l \otimes T^0] e + \\ + \sum_{i=2}^{\infty} y_{(i,3,0,s+1)} [I_l \otimes (S^0 \oplus T^0)] e.$$

3) The expected number of departures after completing service is

$$E_{EDS} = \sum_{i=1}^{\infty} \sum_{k=s+1}^S y_{(i,1,0,k)} [I_l \otimes S^0] e + \sum_{i=1}^{\infty} \sum_{k=1}^{S-1} y_{(i,1,1,k)} [I_l \otimes S^0] e + \\ + \sum_{i=1}^{\infty} \sum_{k=s+1}^S y_{(i,2,0,k)} [I_l \otimes T^0] e + \sum_{i=1}^{\infty} \sum_{k=1}^{S-1} y_{(i,2,1,k)} [I_l \otimes T^0] e + \\ + \sum_{i=2}^{\infty} \sum_{k=s+1}^S y_{(i,3,0,k)} [I_l \otimes (S^0 \oplus T^0)] e + \sum_{i=2}^{\infty} \sum_{k=2}^{S-1} y_{(i,3,1,k)} [I_l \otimes (S^0 \oplus T^0)] e.$$

4) The mean production rate is

$$E_{EPR} = \gamma \left\{ \sum_{i=0}^{\infty} \sum_{k=0}^{S-1} y_{(i,0,1,k)} + \sum_{i=1}^{\infty} \sum_{k=1}^{S-1} y_{(i,1,1,k)} + \sum_{i=1}^{\infty} \sum_{k=1}^{S-1} y_{(i,2,1,k)} + \sum_{i=2}^{\infty} \sum_{k=2}^{S-1} y_{(i,3,1,k)} \right\}.$$

5) Expected inventory level is

$$E_{EI} = \sum_{i=0}^{\infty} \sum_{j=s+1}^S \sum_{j_0=1}^l j y_{(i,0,0,j,j_0)} + \sum_{i=0}^{\infty} \sum_{j=1}^{S-1} \sum_{j_0=1}^l j y_{(i,0,1,j,j_0)} +$$

$$\begin{aligned}
& + \sum_{i=1}^{\infty} \sum_{j=s+1}^S \sum_{j_0=1}^l \sum_{j_1=1}^m j \gamma_{(i,1,0,j,j_0,j_1)} + \sum_{i=1}^{\infty} \sum_{j=s+1}^S \sum_{j_0=1}^l \sum_{j_1=1}^m j \gamma_{(i,1,1,j,j_0,j_1)} + \\
& + \sum_{i=1}^{\infty} \sum_{j=s+1}^S \sum_{j_0=1}^l \sum_{j_2=1}^n j \gamma_{(i,2,0,j,j_0,j_2)} + \sum_{i=1}^{\infty} \sum_{j=1}^{S-1} \sum_{j_0=1}^l \sum_{j_2=1}^n j \gamma_{(i,2,1,j,j_0,j_2)} + \\
& + \sum_{i=2}^{\infty} \sum_{j=s+1}^S \sum_{j_0=1}^l \sum_{j_1=1}^m \sum_{j_2=1}^n j \gamma_{(i,3,0,j,j_0,j_1,j_2)} + \sum_{i=2}^{\infty} \sum_{j=2}^{S-1} \sum_{j_0=1}^l \sum_{j_1=1}^m \sum_{j_2=1}^n j \gamma_{(i,3,1,j,j_0,j_1,j_2)}.
\end{aligned}$$

Optimization problem

The expected total cost of the system per unit per unit time is given by

$$T_{\text{cost}} = c_1 E_{SWR} + c_2 E_{EI} + c_3 E_{EC} + c_4 E_{EDS},$$

where c_1 : the procurement cost per unit per unit time; c_2 : the holding cost of inventory per unit per unit time; c_3 : the holding cost of customers per unit per unit time and c_4 : the cost due to service per unit per unit time. The long-term expected cost behaviour based on variations in the values of γ under Bernoulli vacation schedule in different service distributions for the negative and positive correlated arrivals are displayed in Tables 1 and 2.

MAP with positive correlation:

$$D_0 = \begin{bmatrix} -4.05 & 1.55 \\ 3.5 & -5.5 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 2.05 & 0.45 \\ 1 & 1 \end{bmatrix}.$$

MAP with negative correlation:

$$D_0 = \begin{bmatrix} -5.5 & 3.5 \\ 1 & -3.5 \end{bmatrix}, \quad D_1 = \begin{bmatrix} 1 & 1 \\ 1 & 1.5 \end{bmatrix}.$$

Three different PH-distributions are considered for each server's service time distribution, and by normalizing these processes, servers 1 and 2 will have specific service rates.

Erlang distribution (ER):

$$\begin{aligned}
S &= \begin{bmatrix} -5.1613 & 5.1613 \\ 0 & -5.1613 \end{bmatrix}, \quad T = \begin{bmatrix} -5.0746 & 5.0746 \\ 0 & -5.0746 \end{bmatrix}, \\
\alpha &= [1 \ 0], \quad \beta = [1 \ 0].
\end{aligned}$$

Hyper Exponential distribution (HX):

$$S = \begin{bmatrix} -4.2665 & 0 \\ 0 & -1 \end{bmatrix}, T = \begin{bmatrix} -7.437 & 0 \\ 0 & -1 \end{bmatrix},$$

$$\alpha = [0.8 \quad 0.2], \beta = [0.7 \quad 0.3].$$

Exponential distribution (EX):

$$\alpha = [1], \beta = [1], S = [-2.5806], T = [-2.5373],$$

$$S = 20, s = 5, c_1 = 1250, c_2 = 4, c_3 = 100, c_4 = 1,$$

$$\theta_1 = 30, \theta_2 = 50, p_1 = 0.5, p_2 = 0.6.$$

Table 1

Expected total cost in MAP(C-)

γ	EX/ER	EX/EX	EX/HX	ER/ER	ER/EX	ER/HX	HX/ER	HX/EX	HX/HX
6	62.19	61.94	62.75	62.37	61.99	55.02	70.30	62.59	63.21
7	61.24	60.98	61.73	61.43	61.05	54.55	68.81	61.58	62.15
8	60.63	60.37	61.68	60.82	60.44	54.26	67.85	60.94	61.47
9	60.20	59.93	60.62	60.40	60.01	54.05	67.18	60.49	61.00
10	59.88	59.61	60.28	60.08	59.69	53.90	66.68	60.16	60.65

Table 2

Expected total cost in MAP(C+)

γ	EX/ER	EX/EX	EX/HX	ER/ER	ER/EX	ER/HX	HX/ER	HX/EX	HX/HX
6	53.19	53.40	53.58	53.03	53.19	50.13	56.44	53.52	53.63
7	52.78	52.99	53.14	52.63	52.79	49.93	55.79	53.09	53.17
8	52.52	52.72	52.85	52.37	52.53	49.80	55.37	52.81	52.87
9	52.34	52.54	52.65	52.19	52.35	49.72	55.08	52.62	52.66
10	52.20	52.40	52.51	52.06	52.21	49.65	54.86	52.47	52.51

Concluding remarks

In this paper we analyzed a production inventory system and consider different combinations of arrival and service processes for the multiple servers in the Bernoulli vacation schedule. We can extend the work to the production inventory model with phase-type distributed server relaxation times and lead time.

REFERENCES

1. *Ayyappan G., Shyamala S.* $M^{[x]}/G_1, G_2/1$ with setup time, Bernoulli vacation, break down, and delayed repair // *International Journal of Stochastic Analysis*. 2014. V. 2014. DOI: 10.1155/2014/892867.
2. *Beena P., Jose K.P.* A MAP/Ph(1),Ph(2)/2 production inventory model with inventory dependent production rate and multiple servers // *AIP Conference Proceedings*. 2020. V. 2261(1). P. 030052.
3. *Chakravarthy S.R., Ozkar S.* Map/Ph/1 queueing model with working vacation and crowdsourcing // *Mathematica Applicanda*. 2016. V. 44(2). P. 263.
4. *Jose K.P., Beena P.* On a retrial production inventory system with vacation and multiple servers // *International Journal of Applied and Computational Mathematics*. 2020. V. 6(4). P. 1–17.
5. *Keilson J., Servi L.* Oscillating random walk models for GI/G/1 vacation systems with Bernoulli schedules // *Journal of Applied Probability*. 1986. V. 23(3). P. 790–802.
6. *Kumar B.K., Madheswari S.P., et al.* Analysis of an m/m/n queue with Bernoulli service schedule // *International Journal of Operational Research*. 2009. V. 5(1). P. 48–72.
7. *Kumar B.K., Rukmani R., Thangaraj V.* Analysis of MAP/Ph(1),Ph(2)/2 queue with Bernoulli vacations // *International Journal of Stochastic Analysis*. 2008.
8. *Krishnamoorthy A., Narayanan V.C.* Production inventory with service time and vacation to the server // *IMA Journal of Management Mathematics*. 2011. V. 22(1). P. 33–45.
9. *Neuts M.F.* *Matrix-Geometric Solutions to Stochastic Models* // DGOR. Springer, 1984. P. 425
10. *Suganya C., Sivakumar B.* MAP/Ph(1),Ph(2)/2 finite retrial inventory system with service facility, multiple vacations for servers // *International Journal of Mathematics in Operational Research*. 2019. V. 15(3). P. 265–295.
11. *Tyagi A., Harms J., Kamal A.* Analysis of M/M/S queues with 1-limited service // *Journal of the Operational Research Society*. 2002. V. 53(7). P. 752–767.

Optimization of task offloading thresholds in the fog computing system*

Anastasia Daraseliya¹, Eduard Sopin^{1,2}

¹ Peoples' Friendship University of Russia (RUDN University), Moscow, Russia

² Institute of Informatics Problems FRC CSC RAS, Moscow, Russia

In recent years, fog computing has received attention from the scientific and industrial community. Many papers were related to opportunities and challenges of fog, focusing primarily on the networking context of the Internet of Things (IoT) [1]. Another one of the most popular topics and pressing research issue is the compromise between the energy-efficiency and the response time in offloading of mobile application tasks to fog computing infrastructure. The paper [2] presents the results of a study on energy consumption, execution delay and payment cost of offloading processes in a fog computing network in terms of queuing theory. Research in [3] focuses on energy-efficient task offloading, whose main idea is taking into account both energy consumption and schedule delay under fog devices.

In our previous paper [4], we developed an analytical framework for response time analysis that takes into account the variation of tasks in terms of processing volume. Then in the paper [5], we analyze the two-parameter offloading mechanism that takes into account both the computing complexity and the data size to be transferred in case of offloading. In [6] we derived the cumulative distribution function of the response time in terms of Laplace-Stieltjes Transform. In the current work, we solve the optimization problem by minimizing energy efficiency, subject to the average time constraint and taking into account the probability that the time exceeds a given threshold.

Mathematical model

We consider a distributed computing system that consist of mobile devices (MDs), a fog node and a remote cloud. MDs run real-time applications that require significant amount of computational resources. For each task, a MD makes a decision, whether it will be offloaded to the fog node or processed locally. The capacity of the fog node is limited, which means if there are too many tasks offloaded, then some of the offloaded tasks

* The reported study was funded by RFBR, project numbers 18-07-000576 (recipient E.Sopin) and 20-07-01052 (recipient A. Daraseliya).

are redirected to the remote cloud to prevent the fog node congestion. In terms of queuing theory, the considered system can be represented as shown in the on Fig. 1

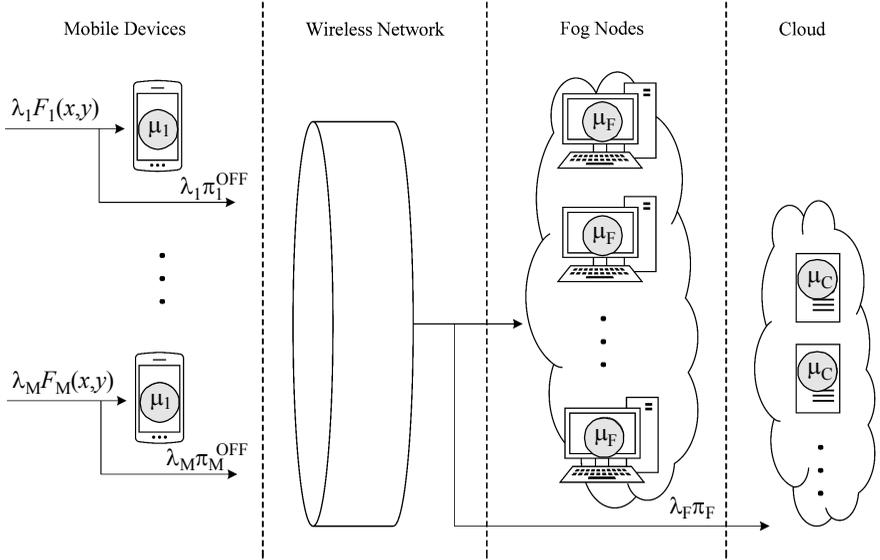


Fig. 1. Mathematical model in terms of queuing network

Assume there are M MDs, each of them generating a flow of tasks that is exponentially distributed according to Poisson's law with intensity $\lambda_i, i = 1, \dots, M$. Each task is characterized by the amount of processing volume required and the data size to be transferred in case of offloading. We assume that the processing volume (measured in millions of instructions, MI) and the data size (measured in MB) are independent random variables with CDFs $W_i(x)$ and $S_i(x)$, probability density functions (PDF) $w_i(x)$ and $s_i(x)$ respectively. MDs process locally served tasks in the FCFS mode with constant serving rate $\mu_i, i = 1, \dots, M$ (measured in MIPS).

We propose the offloading mechanism that implies offloading tasks that are «heavy» in terms of processing volume and «light» in terms of data size. Splitting to «heavy» and «light» tasks are done by the threshold O_w on the processing volume and the threshold O_s on the data size. Hence, the

offloading probability $\pi_{i,O}$ on the i -th MD is evaluated [5, 6] according to the following formula:

$$\pi_{i,O} = \int_{O_w}^{+\infty} w_i(x) dx \int_0^{O_s} s_i(y) dy = (1 - W_i(O_w) S_i(O_s)). \quad (1)$$

If a task is processed locally, then the response time consists of processing time on an MD only. If a task is offloaded to the fog node, then the total response time is the sum of task transmission time to the fog node through wireless network, the processing time on the fog node and the transmission time back to the MD. If the fog node is overloaded and an offloaded task is sent to the remote cloud, then the processing time at the fog is replaced by the transmission time between the fog node and the remote cloud, the processing time on the cloud and the transmission time back to the fog node.

We assume that the wireless network provides total bitrate R , which is used to transmit the data of tasks one-by-one in FCFS order, so the transmission time is obtained as the fraction of the data size of a task and total bitrate R . On the other side, the transmission time between the fog node and the cloud is assumed constant.

The fog node provide computational resources by means of virtual machines (VMs), each of them having the constant serving rate λ_F . The total number of VMs at the fog node is N . The constant serving rate μ_C of VMs at the cloud is greater than μ_F , and amount of computational resources (VMs) at the remote cloud is assumed to be large enough, so that it is cannot be overloaded.

The energy consumption analysis

In this section, we present the formulas for the average power consumption of MD obtained at an earlier stage of research [5].

The distribution function of the processing volume on a MD can be determined by conditional CDF $W_{MD,i}(x)$ as follows.

$$W_{MD,i}(x) = \begin{cases} \frac{W_i(O_w) + (W_i(x) - W_i(O_w))(1 - S_i(O_s))}{1 - \pi_{i,O}}, & x > O_w, \\ \frac{W_i(x)}{1 - \pi_{i,O}}, & x \leq O_w. \end{cases} \quad (2)$$

Having obtained the distribution function of the processing volume $W_{MD,i}(\mu_i x)$, we can find the serving time at a MD. By virtue of the fact that the serving rate on the i -th MD is constant and being μ_i , its CDF is easily obtained as $T_{MD,i}(x) = W_{MD,i}(\mu_i x)$. The average serving time at MD i $t_{MD,i}(x)$ can be found through integration using the CDF $T_{MD,i}(x)$.

The energy consumption for tasks processed on MD is proportional to the processing volumes of tasks, therefore the average energy consumption $E_{pr,i}$ during locally executing on i -th MD can be evaluated as follows:

$$E_{pr,i} = P_{pr,i} t_{MD,i}, \quad (3)$$

where $P_{pr,i}$ is the power consumption (W) during the processing of the i -th MD, which is considered constant for simplicity of calculations.

In order to estimate the energy consumption during transmitting, we obtain the CDF $S_{tr,i}(x)$ of the file size to be transmitted:

$$S_{tr,i}(x) = \begin{cases} \frac{1}{\pi_{i,0}}(1 - W_i(O_w))S_i(x), & x \leq O_s, \\ 1, & x > O_s. \end{cases} \quad (4)$$

The average file size transmitted by i -th MD, can be calculated through integration using CDF $S_{tr,i}(x)$.

The energy consumption during transmitting is also proportional to the transmission time, so the average energy consumption $E_{tr,i}$ of the i -th MD during task transmission is

$$E_{tr,i} = P_{tr,i} \frac{\theta_i}{R}. \quad (5)$$

Then the average energy consumption for any i -th MD is the weighted sum of processing and transmission energies:

$$E_i = (1 - \pi_{i,0})E_{pr,i} + \pi_{i,0}E_{tr,i}. \quad (6)$$

At the end, we can evaluate the average energy consumption for a task from an arbitrary MD as follows

$$E = \sum_{i=0}^M \lambda_i / \left(\sum_{j=1}^M \lambda_j \right) E_i. \quad (7)$$

The response time analysis

In [6] we derived the Laplace-Stieltjes Transforms (LST) for all delay components, and obtain the LST of the total response time. First, we derived the LST $\tilde{T}_{MD,i}(s)$ of the service time at the MD i , with the the LST $\phi_{MD,i}$ of the sojourn time distribution on mobile device i . Then we obtained the the LST $\tilde{T}_{F,i}(s)$ and $\tilde{T}_{C,i}(s)$ of the service time distribution at the fog node and cloud, respectively. Having obtained the LST of all these delay component distributions, we made use of the convolution formula and obtain the LST $\tilde{\tau}(s)$ of the response time distribution of a task from MD i :

$$\begin{aligned} \tilde{\tau}_i(s) = & (1 - \pi_{i,O})\phi_{MD,i}(s) + \pi_{i,O}(1 - \pi_F)\tilde{T}_{F,i}(s)\phi_{F,i}^2(s) + \\ & + \pi_{i,O}\pi_F\tilde{T}_{C,i}(s)\phi_{C,i}^2(s)\tilde{T}_{FC}(s). \end{aligned} \quad (8)$$

After this we used numerical Reverse LST to the $\tilde{\tau}(s)$ to evaluate the CDF $\tau(x)$ of the response time.

Actually, the average response time can be calculated as

$$\tau = \sum_{i=0}^M \lambda_i / \left(\sum_{j=1}^M \lambda_j \right) \tau_i. \quad (9)$$

The resulting expressions allow to get the probability $\Pi(T)$ that the response time is lower than a threshold T

$$\Pi(T) = \tau(T). \quad (10)$$

Optimization problem

In order to find the minimum energy consumption E under constraints on the average response time and the probability $\Pi(T)$ that the response time is lower than a threshold T , we formulate the optimization problem as follows:

$$\begin{cases} E = \sum_{i=0}^M \lambda_i / \left(\sum_{j=1}^M \lambda_j \right) E_i \rightarrow \min, \\ \tau = \sum_{i=0}^M \lambda_i / \left(\sum_{j=1}^M \lambda_j \right) \tau_i \leq T, \\ \Pi(T) \leq \Pi^*. \end{cases} \quad (11)$$

Due to the large amount of information, a table with input parameters values for the numerical analysis and numerical results will be presented in an extended version of the paper. The main metric of interest here is the minimum power consumption E for a task from an arbitrary MD under the constraints from the optimization problem.

Conclusions

In the paper, we focused on the analysis of the reducing energy consumption of MD's that use fog computing infrastructure to increase the performance and to improve the battery life of mobile devices. We have formulated and solved the problem of energy consumption optimization using constraints on the average response time and the probability that the response time is lower than a certain threshold, on the basis of which we offer some recommendations for offloading the system.

REFERENCES

1. *Chiang M., Zhang T.* Fog and IoT, an overview of research opportunities // IEEE Internet of Things Journal. 2016. V. 3(6). P. 854–864.
2. *Chang Z., Zhou Z., Ristaniemi T., Niu Z.*, Energy efficient optimization for computation offloading in fog computing system // GLOBECOM 2017 – 2017 IEEE Global Communications Conference. IEEE, Singapore, 2017. P. 1–6.
3. *Jiang Y., Chen Y., Yang S., Wu C.* Energy-efficient task offloading for time-sensitive applications in fog computing // IEEE Systems Journal. 2019. V. 13. Iss. 3. P. 2930–2941.
4. *Sopin E., Daraseliya A., Correia M.*, Performance analysis of the offloading scheme in a fog computing system // 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). IEEE, Moscow, 2018. P. 1–5.
5. *Sopin E., Samouylov K., Shorgin S.* The analysis of the computation offloading scheme with two-parameter offloading criterion in fog computing // Internet and Distributed Computing Systems. IDCS 2019. Lecture Notes in Computer Science. V. 11874. Springer, Cham. DOI: 10.1007/978-3-030-34914-1_2.
6. *Sopin E., Zolotous N., Ageev K., Shorgin S.* Analysis of the response time characteristics of the fog computing enabled real-time mobile applications // 20th International Conference NEW2AN 2020, Lecture Notes in Computer Science.

A queueing system with probabilistic joining strategy for priority customers*

Dhanya Babu, Varghese C. Joshua,
Achyutha Krishnamoorthy

Department of Mathematics, CMS College Kottayam, Kerala, India

Classical priority queueing systems consider two or more parallel queues, formed by customers of distinct priorities 1 (highest), 2, ..., m (least), respectively. We can see different types of queueing models in [1] to [4]. White and Christie were the first to consider priority queues. In [7], they considered the simple 2-priority (M,M)/M/1 queue and analyzed both non-preemptive and preemptive cases of services to derive the system state probability distribution, thereby deriving all important system performance measures.

This paper is an extension of [5], a single server queueing model with an optimal joining strategy for priority customers. In that paper we introduced an entirely different priority queue: two types of priority waiting lines are formed at a counter manned in a single server queue. One is called ordinary customers queue (OC) with infinite capacity waiting space and the second one is referred to as queue of priority customers (PC) with finite capacity waiting space. The problems addressed in that paper were what is the optimal joining strategy of a priority customer upon arrival whether to join the PC queue or OC queue to reduce his waiting time assuming there is no additional benefit (reward) for joining the PC queue, except that he is offered service as per the K -policy, and the other is what is the optimal K value in the sense of reducing the waiting of PC customers and minimizing their loss due to finite capacity of the PC queue? In this paper, along with the optimal joining strategy we introduced probabilistic joining strategy for priority customers to join in PC queue or in OC queue by assuming additional benefit (reward). The case of $K = 1$ leads to classical 2-priority queue; $K = 2$ results in alternating queue and K infinity leads to exhaustive service discipline of OC queue before the PC queue is attended.

A customer gets priority either by paying a cost or by any other means. Priority customers receive service on the basis of a token system which

* Research work of the first author was supported by Maulana Azad National Fellowship [F1-17.1/2015-16/MANF-2015-17-KER-65493] of University Grants Commission, India.

works according to the following rule: $K - 1$ lower priority customers are served consecutively, and the K^{th} one is from the priority queue, if there is any. Priority customers have the right to take the strategic decision in choosing the queue on arrival, if such a customer joins ordinary queue then he loses the special benefit that he would have got otherwise. We introduce the joining strategy for the priority customers as follows:

$$\text{Let } S_1 = \left\lfloor \frac{i + r - (K + 1)}{K - 1} \right\rfloor (K - 1), S_2 = (K - 1)j, \text{ and } S_1^* = \left\lfloor \frac{i}{K - 1} \right\rfloor (K - 1).$$

In effect with K -policy for taking for service, the priority customer arriving at the station decides to join PC or OC queue according to the following conditions:

- If an ordinary customer is in service,

An arriving priority customer decide to join PC queue if $S_1 > S_2$.

An arriving priority customer decide to join OC queue if $S_1 \leq S_2$.

- If a priority customer is in service,

An arriving priority customer decide to join PC queue if $S_1^* > S_2$.

An arriving priority customer decide to join OC queue if $S_1^* \leq S_2$.

Steady-state analysis of the model is done. Some system characteristics are evaluated, a social optimisation problem is discussed and numerical illustrations are provided.

Mathematical model

We consider a single server queueing system with two parallel queues of which one is a finite buffer for priority customers and the other infinite for ordinary customers. The two types of customers arrive according to a Marked Markovian Arrival Process (MMAP) with representation (D_0, D_1, D_2) with order n . The arrival process is governed by a continuous time Markov chain $\{A(t): t \geq 0\}$ with state space $\{0, 1, 2, \dots, n\}$. Even if the priority customer has the opportunity to get the chance to enter into service immediately as an ordinary customer than to stand in priority queue, we assume that some of priority customers take probabilistic decision whether to give up their additional benefit (reward) or not. Instead of joining in the ordinary queue according to the optimal joining strategy, let the priority customer join in PC queue itself with a probability p , where $0 < p < 1$. Service time of an ordinary customer follows phase type distribution with representation $\text{PH}(\alpha, T)$ of order s_1 and that of priority customer follows phase type distribution with representation $\text{PH}(\beta, S)$ of order s_2 .

Let $N_1(t)$, $N_2(t)$, $R(t)$, $A(t)$ and $S(t)$ denote the number of customers at time t in ordinary queue, the number of customers at time t in priority queue, the token numbers from 1, 2, ..., $K-1$, K ; the arriving phase of customer, and the service phase of customer in service respectively.

Let $\{(N_1(t), N_2(t), R(t), S(t), A(t)); t \geq 0\}$ be the Markov Process on the state space $\Omega = I^* \cup \left(\bigcup_{i=1}^{\infty} l(i) \right)$, where $I^* = \{(0, 0, 0, 0, v) : v = 1, 2, \dots, n\}$ and for $i \geq 1$, $l(i) = \{(i, j, r, u, v) : i \geq 1; 0 \leq j \leq M; 1 \leq r \leq K; 1 \leq u \leq m; 1 \leq v \leq n\}$ depending on the joining strategy where, $m = \delta_{[r=1 < K]}s_1 + \delta_{[r=K]}s_2$,

$$\delta_{[condition]} = \begin{cases} 1, & \text{if condition is true,} \\ 0, & \text{otherwise.} \end{cases}$$

This model is a level independent quasi birth and death process (LIQBD). This can be conveniently and efficiently solved by the classical matrix Geometric method (see[6]).

Mathematical analysis

Stability condition of the model is obtained. Steady-state vector is obtained by using Matrix Analytic Method.

The steady-state vector x_i is obtained as

$$x_i = x_{i-1}L_i; \text{ where } L_i = -A_{0(i-1)}[A_{1i} + L_{i+1}A_{2(i+1)}]^{-1}; 1 \leq i \leq N$$

$$\text{and } x_{M(K-1)+i} = x_{M(K-1)}R^i; i \geq 1$$

Several Performance Measures are obtained of which some important performance measures are expected number of ordinary customers in the system, expected Number of priority customers in the system, expected number of customers in the system, probability that the server is idle, probability that the server is busy with an ordinary customer, probability that the server is busy with a priority customer, probability that a priority customer is blocked from entering the system upon arrival, expected waiting time of a tagged ordinary customer who joins the queue as r^* th customer, expected waiting time of a tagged priority customer who joins the queue as s^* th customer. A social optimisation problem is discussed as below.

- A reward or benefit R_1 monetary units for a priority customer joining in the ordinary queue.

- A reward or benefit R_2 monetary units for a priority customer joining in the priority queue.

- A waiting cost h_1 monetary units for each unit of time that a priority customer waiting in ordinary queue.
 - A waiting cost h_2 monetary units for each unit of time that a priority customer waiting in priority queue.
 - A cost of c_1 monetary units for each priority customer being served in ordinary queue.
 - A cost of c_2 monetary units for each priority customer being served in priority queue.
 - W_{Q_1} be the expected waiting time of an ordinary customer.
 - W_{Q_2} be the expected waiting time of priority customer.
- Net Benefit for a priority customer, if he joins in the ordinary queue Q_1 according to joining strategy is

$$B_1 = R_1 - (1 - p) h_1 W_{Q_1} - c_1.$$

Net Benefit for a priority customer if he joins in the priority queue Q_2 is

$$B_2 = R_2 - p h_2 W_{Q_2} - c_2.$$

- If $B_1 > 0$ and $B_1 > B_2$, then the priority customer join in ordinary queue Q_1 .
 - If $B_1 > 0$ and $B_1 < B_2$, then the priority customer join in priority queue Q_2 .
 - If $B_1 < 0$ and $B_2 > 0$, then the priority customer join in priority queue Q_2 .
 - If $B_1 < 0$ and $B_2 < B_1$, then the priority customer join in ordinary queue Q_1 .
 - If $B_1 = 0$ and $B_2 = 0$, then the priority customer either join in ordinary queue Q_1 or in priority queue Q_2 .
 - If $B_1 = 0$ and $B_1 > B_2$, then the priority customer join in ordinary queue Q_1 .
 - If $B_1 = 0$ and $B_1 < B_2$, then the priority customer join in priority queue Q_2 .
- Similarly,
- If $B_2 > 0$ and $B_2 > B_1$, then the priority customer join in priority queue Q_2 .
 - If $B_2 > 0$ and $B_2 < B_1$, then the priority customer join in ordinary queue Q_1 .
 - If $B_2 < 0$ and $B_1 > 0$, then the priority customer join in ordinary queue Q_1 .
 - If $B_2 < 0$ and $B_1 < B_2$, then the priority customer join in priority queue Q_2 .
 - If $B_2 = 0$ and $B_1 < B_2$, then the priority customer join in priority queue Q_2 .
 - If $B_2 = 0$ and $B_1 > B_2$, then the priority customer join in ordinary queue Q_1 .

Numerical examples

Here we illustrate the social optimisation problem of priority customers on the basis of net benefits B_1 and B_2 with some examples. In Fig. 1 and Fig. 2 x-axis is taken as probability of joining strategy p and y-axis is taken as net benefits for a priority customer if he joins in the ordinary queue or in priority queue. Fig.1 shows that as p increases, the net benefit for a priority

customer to join in ordinary queue increases and is positive and is greater than B_2 so priority customer join in ordinary queue Q_1 . Fig. 2 shows that the net benefit for a priority customer to join in priority queue decreases as p increases and is positive and is greater than B_1 , so priority customer join in Q_2 .

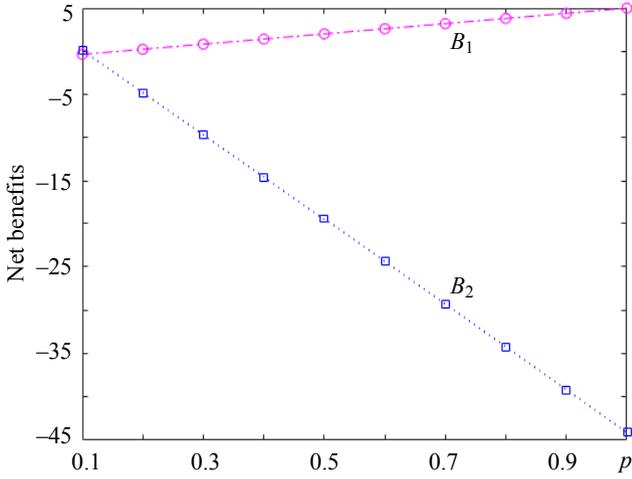


Fig. 1. Effect of p on B_1 and B_2

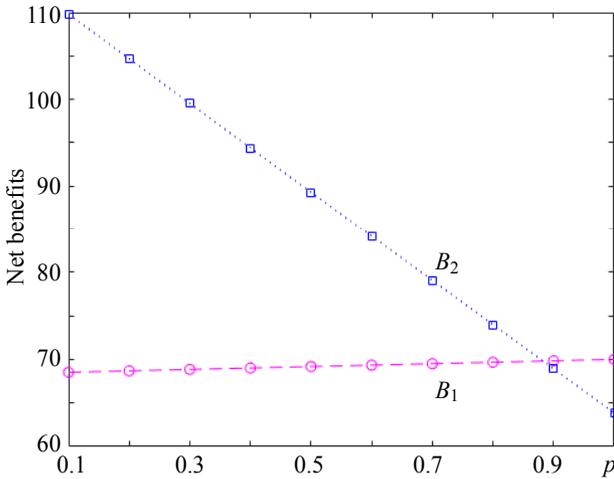


Fig. 2. Effect of p on B_1 and B_2

Conclusions

In this paper we considered a single server queueing system with non-preemptive priority and a joining strategy for priority customers. Even if an optimal joining strategy exists for a priority customer upon arrival whether to join the priority queue or ordinary queue to reduce his waiting time, priority customers take probabilistic decision whether to give up their additional benefit(reward) for joining the priority queue. We illustrated a social optimisation problem and we conclude that priority customers take decision depending on the values of net benefits.

REFERENCES

1. *Chakravarthy S., Thiagarajan S.* Two parallel queues with simultaneous services and Markovian arrivals // Journal of Applied Mathematics and Stochastic Analysis. 1997. V. 10. No.4. P. 383–405.
2. *Dudin S., Kim Ch., Dudina O.* MMAP/M/N queueing system with impatient heterogeneous customers as a model of a contact center // Computers and Operations Research. 2013. V. 40. No.7. P. 1790–1803.
3. *Krishnamoorthy A., Deepak T.G., Joshua V.C.* Queues with postponed work // TOP. 2004. V. 12. P. 375–398.
4. *Krishnamoorthy A., Manjunath A.S.* On priority queues generated through customer induced service interruption // Neural, Parallel, and Scientific Computations. 2015. V. 23. P. 459–486.
5. *Krishnamoorthy A., Joshua V. C., Babu Dhanya.* A token based parallel processing queueing system with priority // 20th International Conference (DCCN), Proceedings, CCIS. 2017. V.700. P. 231–240.
6. *Neuts M.F.* Matrix-Geometric Solutions in Stochastic Models – An Algorithmic Approach. Baltimore: The Johns Hopkins University Press, 1981.
7. *White H., Christie L.S.* Queueing with preemptive priorities or with breakdown // Operations Research. 1958. V. 6. P. 79–96.

On convergence of queuing network with changing structure to stationary distribution*

Elmira Kalimulina

Institute of Control Sciences of Russian Academy of Sciences, Russia, Moscow

Mathematical model

This paper is a continuation of previous author's research on ergodicity of some models for queuing networks [2, 3]. The set of random graphs and the sequence of matrices describing the changing network structure has been used instead of the fixed graph here. The main results about an ergodicity and bounds for rate of convergence to stationary distribution are formulated under more general assumptions on intensity rates.

The standard queueing network model [1] has been modified by an assumption about changing structure: each node may be blocked and restored with intensities $\alpha_i, \beta_i, i = 1, \dots, m$, where $m (m < \infty)$ – the number of nodes. This assumption results adding some component to standard state space of the network process. The standard approach implies the network matrix to be fixed, the network model considered here is quite different. We denote the state space of the graph transformation process as the set G . It is formed in the following way: the node i is «removed» with some intensity α_i (failure rate for this node) or it can be restored with some intensity β_i . So, the state space for our network process is the following, it is extended by adding the component G :

$$\mathbf{n} = (G, n_1, n_2, \dots, n_m) \in |G| \times Z_+^m =: E,$$

where G is a component describing the graph (or transition matrix) transformation. We can find a degree distribution for the process from state space G :

The average number of vertices of degree k at time t : $\{E(k, t)\} = E P(k, t)$ can be described by the equation

$$\{E(k, t+1)\} = \{E(k, t)\} - \frac{\alpha_k}{E \sum_k P(k) \alpha_k} \{E(k, t)\} +$$

* The publication has been prepared with the support of the Russian Foundation for Basic Research according to the research project No.20-01-00575 A.

$$+ \frac{\alpha_{k-1}}{E \sum_{k-1} P(k-1) \alpha_{k-1}} \{E(k-1, t)\} + \frac{\alpha_{k+1}}{E \sum_{k+1} P(k+1) \alpha_{k+1}} \{E(k+1, t)\}.$$

It describes the evolution of graph of our network structure in time and for the continuous time takes the form

$$E \frac{\partial P(k, t)}{\partial t} = -\alpha_k P(k, t) + \alpha_{k-1} E(k-1, t) + P(k+1, t) + \alpha_{k+1} P(k+1, t).$$

The following Lemma holds for this equation:

Lemma the equation is linear homogeneous equation (under assumption of constant failure and recovery rates) and has a stationary solution:

$$P(k) = \lim_{t \rightarrow \infty} P(k, t).$$

The network process $X(t)$ state is described by the following vector:

$$\mathbf{n} = ((n_1, s_1), (n_2, s_2), \dots, (n_m, s_m)),$$

where n_i – the number of customers at the i -th node, the s_i components characterize the i th node works or not.

The behaviour of the process $X(t)$ is a Markov chain in continuous time. It includes an embedded homogeneous Markov chain with positive probabilities of single jumps for s_i and n_i components separately.

We are interested in the process for s_i (denote it as $X_{R_i}(t)$) and the whole process $X(t)$.

Convergence of process $X_{R_i}(t)$. The behaviour of the process $X_{R_i}(t)$ is defined by following transition probabilities:

$$P\{X_{S_i}(t+1) = X_{S_i}(t)\} = \frac{\sum_{j=1}^m \gamma_j - \gamma_i}{\sum_{j=1}^m \gamma_j},$$

$$P\{X_{S_i}(t+1) = 1 - X_{S_i}(t)\} = \frac{\gamma_i}{\sum_{j=1}^m \gamma_j},$$

where

$$\gamma_i = \alpha_i I\{s_i = 0\} + \beta_i I\{s_i = 0\}.$$

The ergodicity and the convergence rate of $R(t)$ may be estimated from the Markov-Doebelin condition (see, e.g. Doebelin, 1938 [4]). So the number

of nodes each customer visited within the network is less than some geometrically distributed random variable and has a finite expectation.

The process $X_R(t)$ takes values from the finite set ($\mathbf{R} = \|r_{ij}(t)\|$), so $X_R(t)$ has the stationary distribution and converges to it *exponentially*.

Conclusions

The results on exponential rate of convergence for some queueing networks have been presented.

It is impossible to obtain the stationary distribution in an explicit form by standard techniques, for example, by matrix methods, operational calculus, and the Laplace transform even in a very simple case when the flows in the system are simple, but the number of parameters and nodes is large (even in the finite case).

In these situations, the only thing we can do is to prove the ergodicity of the system, but for practical application it is not enough and we have to build a simulation model. The simulation results allow us to obtain some estimate, but for practical use it is necessary to be sure that the numerical estimate is close enough to the limiting value.

The convergence rate guarantees that simulation results are close enough to the limit values. One of the results for network convergence rate has been presented in this paper.

REFERENCES

1. *E. van Doorn*. Many papers on the rate of convergence of birth-death processes.
2. *Kalimulina E.Yu*. Analysis of unreliable open queueing network with dynamic routing. DCCN, Springer, 2015.
3. *Kalimulina E.Yu*. Rate of convergence to stationary distribution for unreliable Jackson-type queueing network with dynamic routing. DCCN, Springer, 2016.
4. *Doebelin W*. Mathématique de l'Union Interbalkanique, 1938

Analysis of the processor sharing systems with random serving rate coefficients*

Maksim Korshikov¹, Eduard Sopin^{1,2}

¹ People's Friendship University of Russia (RUDN University), Moscow, Russia

² Institute of Informatics Problems, Federal Research Center «Computer Science and Control», Moscow, Russia

In the modern wireless networks, the bitrate achieved by elastic sessions depends not only on the number of current sessions, but also on the state of the radio channel. Based on signal-to-noise ratio, a base station chooses one of the Modulation and Coding Schemes (MCS) [1]. Each MCS is characterized by a certain value of spectral efficiency, which determines the bitrate achieved on the unit spectrum bandwidth. Consequently, even if the spectrum bandwidth is divided equally between all current sessions, the bitrates of the sessions vary significantly [2]. So, to analyze the performance measures of a base station that serves elastic sessions these peculiarities should be taken into account.

In this paper, we develop the queuing system model with processor sharing (PS) discipline with random serving rate coefficients. Each arriving customer is characterized by its length (job volume) and the serving rate coefficient with certain distributions. The coefficients remain constant during the service process of customers. First, we describe the model and then derive formulas for the main performance metrics.

Mathematical model

Consider a single server queuing system that serves customers according to the PS discipline, but no more than N customers simultaneously. Customers arrive according to a Poisson process with intensity λ , the length of customers is exponentially distributed with parameter μ . Besides, a customer is characterized by a serving rate coefficient v , which may take values from the set $V = \{v_1, v_2, \dots, v_L\}$, $v_1 < v_2 < \dots < v_L$. The distribution of serving rate coefficients is denoted by $\{p_i\}, i \geq 0$, where $p_i = P\{v = v_i\}$.

So, if there are n customers in the system and serving rate coefficient is represented by the vector (u_1, u_2, \dots, u_n) , then each customer is allocated $1/n$

* The reported study was funded by RSF, project no. 20-71-00124.

of the processing time. The service rate of the i -th customer is a product of the allocated processing time ratio and its serving rate coefficient, that is u_i / n . Once the job associated with a customer is done, the customer leaves the system. If an arriving customer finds that there are N customers in the system, then it is lost.

The behavior of the system may be described by Markov process $X_1(t) = \{\xi(t), (v_1(t), v_2(t), \dots, v_{\xi(t)}(t))\}$, where $\xi(t)$ is the number of customers in the system at time t , while $(v_1(t), v_2(t), \dots, v_{\xi(t)}(t))$ is the vector of serving rate coefficients. The set of states is expressed by the following formula:

$$S_1 = \{(n, u_1, u_2, \dots, u_n) : 0 \leq n \leq N, u_i \in V, i = 1, 2, \dots, n\}. \quad (1)$$

Denote $q_n(u_1, u_2, \dots, u_n)$ the stationary probability that the system is in state $(n, u_1, u_2, \dots, u_n)$. Then the system of equilibrium equations can be written as follows:

$$\lambda q_0 = \mu \sum_{l=1}^L v_l q_1(v_l), \quad (2)$$

$$\left(\lambda + \frac{\mu}{n} \sum_{i=1}^n u_i \right) q_n(u_1, u_2, \dots, u_n) = \frac{\lambda}{n} \sum_{i=1}^n p_{u_i} q_{n-1}(u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_n) + \frac{\mu}{n+1} \sum_{l=1}^L \sum_{i=1}^{n+1} \left(v_l + \sum_{j=1}^n u_j \right) q_{n+1}(u_1, \dots, u_{i-1}, v_l, u_i, \dots, u_n), n = 1, 2, \dots, N-1, \quad (3)$$

$$\left(\frac{\mu}{N} \sum_{i=1}^N u_i \right) q_N(u_1, u_2, \dots, u_N) = \frac{\lambda}{N} \sum_{i=1}^N p_{u_i} q_{N-1}(u_1, \dots, u_{i-1}, u_{i+1}, \dots, u_N). \quad (4)$$

Computational complexity of the numerical solution of the system (2) – (4) increase very fast, since the number of states in set V is $\frac{L^{N+1} - 1}{L - 1}$. So, in the next section we propose a simplified model based on the state aggregation technique [3].

Analysis of the simplified model

To simplify the model, we propose to trace only the sum of the serving rate coefficients of all customers in the system. Consider the simplified

process $X_2(t) = \{\xi(t), v(t)\}$, where $\xi(t)$ is the number of customers in the system at time t , as before, and $v(t)$ is the sum of their serving rate coefficients. The set of states is given by

$$S_2 = \{(n, u) : 0 \leq n \leq N, p_u^{(n)} > 0\}, \quad (5)$$

where distribution $\{p_u^{(n)}\}$ is the n -fold convolution of the distribution $\{p_u\}$.

However, as a result of the simplification, we cannot exactly determine the decrease of the $v(t)$ after a departure from the system. To handle the problem, we introduce the probabilities $\varphi_i(n, u)$ that a customer's serving rate coefficient is equal to i given that the sum of the n customers' coefficients is u . The latter can be obtained by Bayes formula:

$$\varphi_i(n, u) = \frac{p_i p_{u-i}^{(n-1)}}{p_u^{(n)}}, \quad i \leq u, (n, u) \in S_2. \quad (6)$$

The stationary probabilities $q_n(u)$ of the process $X_2(t)$ satisfy the following equilibrium equations:

$$\lambda q_0 = \mu \sum_{l=1}^L v_l q_1(v_l), \quad (7)$$

$$\begin{aligned} \left(\lambda + \frac{\mu u}{n} \right) q_n(u) &= \lambda \sum_{l=1}^L p_{v_l} q_{n-1}(u - v_l) + \\ &+ \frac{\mu}{n+1} \sum_{l=1}^L (u + v_l) q_{n+1}(u + v_l) \varphi_{v_l}(n+1, u + v_l), \\ n &= 1, 2, \dots, N-1, (n, u) \in S_2, \end{aligned} \quad (8)$$

$$\frac{\mu u}{n} q_N(u) = \lambda \sum_{l=1}^L p_{v_l} q_{N-1}(u - v_l), (N, u) \in S_2. \quad (9)$$

The system (7) – (9) along with the normalizing condition can be solved using any appropriate numerical method to obtain the stationary probabilities.

Since a customer is blocked upon arrival if there already N customers in the system, then the blocking probability π is

$$\pi = \sum_{(N, u) \in S_2} q_N(u). \quad (10)$$

Analysis of the mean sojourn time

According to [4], customers' sojourn time in the system has a phase-type cumulative distribution function (CDF) $F(x)$. To analyze it, we introduce a Markov process $Y_v(t)$, $v \in V$, that describes the behaviour of the system from arrival to departure of a customer with the serving rate coefficient v . The set of states of $Y_v(t)$ is given by

$$S_v = \{(n, u) : 1 \leq n \leq N, p_{u-v}^{(n-1)} > 0\} \cup \omega, \quad (11)$$

where ω is the absorbing state, which is reached at the departure of the considered customer.

The generator matrix of $Y_v(t)$ has the following form:

$$\mathbf{A} = \begin{pmatrix} \mathbf{G} & \mathbf{g} \\ \mathbf{0} & 0 \end{pmatrix}, \quad (12)$$

where \mathbf{g} is an exit vector to the absorbing state. The matrix \mathbf{G} has block-tridiagonal structure with diagonal \mathbf{D}_n , $n=1,2,\dots,N$, upper diagonal $\mathbf{\Lambda}_n$, $n=1,2,\dots,N-1$, and lower diagonal blocks \mathbf{M}_n , $n=2,3,\dots,N$. The block \mathbf{D}_1 consist of only one element, $\mathbf{D}_1 = (-\lambda - \mu v)$, while other diagonal blocks are diagonal matrices

$$\mathbf{D}_n = \text{diag} \left(-\left(\lambda + \frac{\mu(v + (n-1)v_1)}{n} \right), \dots, -\left(\lambda + \frac{\mu(v + (n-1)v_L)}{n} \right) \right), \\ n = 2, 3, \dots, N-1, \quad (13)$$

with no λ in \mathbf{D}_N .

Block $\mathbf{\Lambda}_1$ is a row vector $(\lambda p_{v_1}, \dots, \lambda p_{v_L})$, while other upper diagonal blocks are given by

$$\mathbf{\Lambda}_n = \begin{pmatrix} \lambda p_{v_1} & \dots & \lambda p_{v_L} & 0 & \dots & 0 \\ 0 & \lambda p_{v_1} & \dots & \lambda p_{v_L} & 0 & \dots & 0 \\ 0 & 0 & \ddots & \ddots & \ddots & 0 & \dots \\ 0 & \dots & & 0 & \lambda p_{v_1} & \dots & \lambda p_{v_L} \end{pmatrix}, \\ n = 2, 3, \dots, N-1. \quad (14)$$

Block \mathbf{M}_2 is a column vector $\left(\frac{\mu v_1}{2}, \dots, \frac{\mu v_L}{2}\right)^T$, and other lower diagonal blocks have the following form:

$$\mathbf{M}_n = \begin{pmatrix} \frac{\mu(n-1)v_1}{n} & 0 & \dots & 0 \\ \frac{\mu((n-2)v_1 + v_2)}{n} \varphi_{v_2} \times & \frac{\mu((n-2)v_1 + v_2)}{n} \varphi_{v_1} \times & 0 & \dots & 0 \\ \times(n-1, (n-2)v_1 + v_2) & \times(n-1, (n-2)v_1 + v_2) & \ddots & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & 0 & \dots & 0 \\ 0 \dots 0 & \frac{\mu(N-1)v_L}{n} & 0 & \dots & 0 \end{pmatrix},$$

$n = 3, 4, \dots, N.$ (15)

The exit vector \mathbf{g} can be expressed as

$$\mathbf{g} = \left(\mu v, \frac{\mu v}{2}, \dots, \frac{\mu v}{2}, \frac{\mu v}{3}, \dots, \frac{\mu v}{3}, \dots, \frac{\mu v}{N}, \dots, \frac{\mu v}{N} \right)^T.$$

According to the PASTA property [5], the stationary distribution of $X_2(t)$ is equal to the stationary distribution of the Markov chain $X_2(t_n - 0)$ embedded at the moments $t_n, n = 1, 2, \dots$, just before arrivals. Since the initial distribution θ of the Markov process $Y_v(t)$ is equal to the distribution of the Markov chain $X_2(t_n + 0)$, we obtain

$$\theta_{n,u} = \begin{cases} \frac{q_0}{1 - \pi}, & n = 1, u = v; \\ \frac{q_{n-1}(u - v)}{1 - \pi}, & n = 2, 3, \dots, N, p_{u-v}^{(n-1)} > 0. \end{cases} \quad (16)$$

Again according to [4], the mean sojourn time \bar{w}_v for a customer with serving rate coefficient v is

$$\bar{w}_v = -\theta \mathbf{G}^{-1} \mathbf{1}. \quad (17)$$

Finally, averaging over all possible serving rate coefficients, we obtain the formula for the mean sojourn time of an arbitrary customer:

$$\bar{w} = \sum_{l=1}^L p_{v_l} \bar{w}_{v_l}. \quad (18)$$

Conclusions

In the paper, we developed a queuing system model with PS discipline and random serving rate coefficients. We also proposed a simplification of the model that allowed to decreased the dimension of the set of states. For the simplified model, we obtained the equilibrium equations for the stationary distribution, as well as formulas for the blocking probability and the mean sojourn time. In our future work, we plan to develop efficient approximate methods for the performance measures evaluation.

REFERENCES

1. *ETSI TS 136.213*. Evolved Universal Terrestrial Radio Access (E-UTRA). Physical layer procedures. Release 13, 2016
2. *ReyhaniMasoleh A.* Resource Allocation in Uplink Long Term Evolution, 2013. Electronic Thesis and Dissertation Repository. P. 1627.
3. *Naumov V.A., Samuilov K.E., Samuilov A.K.* On the total amount of resources occupied by serviced customers // *Automation and Remote Control*. 2016. V. 77. P. 1419–1427.
4. *Asmussen S.* Applied Probability and Queues. New York: Springer, 2003.
5. *Wolff R.W.* Poisson arrivals see time averages // *Operations Research*. 1982. V. 30. No. 2. P. 223–231.

A reliability problem with Interdependent Lifetimes

Achyutha Krishnamoorthy, Varghese C. Joshua,
Ambily P. Mathew

Department of Mathematics, CMS College, Kottayam, Kerala, India

We consider a two-unit system. The lifetimes of these components, when separately considered have independent(not necessarily identical) Phase type distributions. While forming part of the system, the components are interconnected and hence their lifetimes are correlated. By considering the product space of the two Markov chains generating the two phase type distributions, we compute the lifetime distribution of the system. For this we impose a semi-Markov structure in the transitions in this product space. The cases of both serial and parallel systems are examined. Numerical illustrations are provided to have a glimpse of how the system performs.

Introduction

Many physical as well as virtual systems may be viewed as assemblies of many interacting units or components. Consider a system with n units. If such a system operates as long as atleast k among the n units are operational, it is called a k -out-of- n system. k -out-of- n systems include both serial as well as parallel systems. Serial systems are $n-1$ out-of- n systems and parallel systems are 1-out-of- n systems. In [3], Krishnamoorthy and Ushakumari investigated reliability of a k -out-of- n system with repair and retrial of failed units. A survey of reliability studies of consecutive- k -out-of- n :F and related systems can be found in [2]. More on reliability can be found in [6]. Most of the works in literature deal with independent units of systems and the analysis is much easier. Research reports on system reliability, when components have correlated life times, are also investigated in the literature.

In [1], Buchholz and Kreige considered a queue where the inter-arrival times are correlated and, additionally, service times are also correlated with inter-arrival times. The major result of [1] is the presentation of approaches to fit the parameters of the model, namely the MMAP, the PH distribution and the parameters introducing correlation between inter-arrival and service times, according to some trace of inter-arrival and corresponding service times. Bhaskar Sengupta in [5] studied a first-come-first-served single server

semi-Markovian queue in which both the arrival and service mechanisms are semi-Markov processes. In [5] the interarrival time and service times may depend on one another and the marginal distribution of the service times is assumed to be phase-type. In [7], Van Houdt investigated a broad class of semi-Markovian queues introduced by Sengupta. A thorough approach to the semi-Markov processes and their applications in reliability can be found in [4].

However, we introduce in this note, a dependence of component life times as determined by a semi-Markov process. To the best of our knowledge, this type of component life time dependencies are not investigated in the literature. This has a very great significance: The sojourn time in a state (the dimension of the state space depends on the number of units in the system) depends on the state in which it is in and the state to be visited next. We assume the sojourn times to have exponential distribution with parameter depending on the present state and the state to be visited next, as determined by the Markov chain rule (one-step transition probabilities).

We consider a two-unit system in which the two components have lifetimes C_i , $i=1,2$. It is assumed that C_1 follows phase type distribution with representation $\text{PH}(\tau, T)$ having m phases and C_2 follows phase type distribution with representation $\text{PH}(\beta, S)$ with n phases respectively. $S_0 = -Se$ and $T_0 = -Te$; e is a column vector of 1's of appropriate dimension. This system is serial if failure of C_1 or C_2 results in failure of the system whereas it is parallel if the failure of both C_1 and C_2 is needed for the system to fail (both do not fail simultaneously, being continuous in time). Assume that the evolution of the system is according to a two-dimensional SMP (Semi-Markov Process) on the product space of the two Markov chains generating the PH distributions. The state space of the process is $\{(i_1, i_2) / 1 \leq i_1 \leq m+1, 1 \leq i_2 \leq n+1\}$. (i_1, i_2) is the state of the combined process at time t . In the time interval $(t, t+h)$, it can stay there itself or can have one component change: $(i_1, i_2) \rightarrow (i_1', i_2)$ or $(i_1, i_2) \rightarrow (i_1, i_2')$.

For a serial system either a visit to the state $\{(m+1, i_2) / 1 \leq i_2 \leq n\}$ or $\{(i_1, n+1) / 1 \leq i_1 \leq m\}$ results in system failure. In the case of a parallel system, we can have the state $(m+1, n+1)$ and the system failure occurs when the state $(m+1, n+1)$ is realized. The absorption state $m+1$ of the

lifetime C_1 and the absorption state $n+1$ of the lifetime C_2 are realized at different time points as the system being continuous in time. Once in failed state, it stays there forever unless a repair mechanism is available. In case a repair mechanism is available, then we speak about the system in $t \geq 0$; else we expect the system fail in a finite time and it stays there forever. Designate by $\lambda_{(i_1, i_2), (i'_1, i'_2)}$ the rate of phase change in the first component and by $\mu_{(i_1, i_2), (i_1, i'_2)}$ the rate of phase change in the second component in a small interval $(t, t+h)$:

$$\begin{aligned} P[(C_1(t+h), C_2(t+h)) = (i'_1, i'_2) / (C_1(t), C_2(t)) = (i_1, i_2)] = \\ = \lambda_{(i_1, i_2), (i'_1, i'_2)} h + o(h); \end{aligned} \quad (1)$$

$$\begin{aligned} P[(C_1(t+h), C_2(t+h)) = (i_1, i'_2) / (C_1(t), C_2(t)) = (i_1, i_2)] = \\ = \mu_{(i_1, i_2), (i_1, i'_2)} h + o(h). \end{aligned} \quad (2)$$

For no change in states the probability is

$$1 - \sum \left[\lambda_{(i_1, i_2), (i_1, i_2)} + \mu_{(i_1, i_2), (i_1, i_2')} \right] h + o(h), \quad (3)$$

whereas the summation ranges over all $i_1 \neq i'_1 \in \{1, 2, \dots, m+1\}$ and $i_2 \neq i'_2 \in \{1, 2, \dots, n+1\}$ with the condition that change in both the components do not occur simultaneously. In the following section, we find out the life time distribution of both the serial as well as the parallel systems.

Distribution of the lifetime of a two component serial system

The lifetime of a two component serial system follows phase type distribution with representation $\text{PH}(\mathbf{a}, U)$ where $U = (U_{ij})$ is a block matrix with m rows and n columns. Let \mathbf{a} denote the initial probability vector corresponding to the above PH distribution. Each block U_{ij} in the i^{th} row and the j^{th} column where $i \neq j \in \{1, 2, \dots, m\}$ is of the form $U_{ij} = \text{diag}(\lambda_{(i,1),(j,1)}, \lambda_{(i,2),(j,2)}, \dots, \lambda_{(i,n),(j,n)})$. For $i = j \in \{1, 2, \dots, m\}$ the matrix $U_{ii} = (\mu_{(i,k),(i,k')})$. $\mu_{(i,k),(i,k')}$ represents the element in the k^{th} row and k'^{th}

column of U_{ii} when $k, k' \in \{1, 2, \dots, n\}$. $U_0 = -Ue$ The mean of the above PH distribution is given by $-\alpha U^{-1}e$.

Reliability of the system in the interval $[0, t]$, assuming it started operation at time zero is the probability that it provides failure free operation in that interval. The reliability function is given by $R_s(t) = \alpha \exp(Ut)e$.

Theorem 1: For the serial system, using (1), (2), and (3) we get a finite system (of mn) of difference-differential equations. Solving this under the initial condition α we get the probability that the system survives beyond the time t . The complement of this gives the lifetime distribution of the system.

Distribution of the lifetime of a two component parallel system

The distribution of lifetime of a two component parallel system follows phase type distribution with representation $PH(\gamma, V)$, where $V = (V_{ij})$ is a block matrix with $m+1$ rows and n columns. Let γ denote the initial probability vector corresponding to the above PH distribution. Each block V_{ij} in the i^{th} row and the j^{th} column where $i \neq j \in \{1, 2, \dots, m\}$ is of the form

$$V_{ij} = \text{diag}(\lambda_{(i,1),(j,1)}, \lambda_{(i,2),(j,2)}, \dots, \lambda_{(i,n+1),(j,n+1)})$$

and $V_{m+1,j} = \text{diag}(\lambda_{(m+1,1),(j,1)}, \lambda_{(m+1,2),(j,2)}, \dots, \lambda_{(m+1,n),(j,n)})$

for $j \neq m+1 \in \{1, 2, \dots, m\}$. For $i = j \in \{1, 2, \dots, m\}$ the matrix $V_{ii} = (\mu_{(i,k),(i,k')})$.

$\mu_{(i,k),(i,k')}$ represents the element in the k^{th} row and k'^{th} column of V_{ii} , when

$k, k' \in \{1, 2, \dots, n+1\}$ and $V_{m+1,i} = (\mu_{(m+1,k),(m+1,k')})$. when $k, k' \in \{1, 2, \dots, n\}$,

$V_0 = -Ve$. The mean of the above PH distribution is given by $-\gamma V^{-1}e$.

Assuming that the system started operation at time zero, the reliability function is given by $R_p(t) = \gamma \exp(Vt)e$.

Theorem 2: For the serial system, using (1), (2), and (3) we get a finite system of difference-differential equations. Solving this under the initial condition γ we get the probability that the parallel system survives beyond the time t . The complement of this gives the lifetime distribution of the system.

Numerical Example

For the serial system, the graph of the reliability function, when the lifetime is distributed according to a phase type distribution having representation $\text{PH}(\boldsymbol{\alpha}, U)$ with $\boldsymbol{\alpha} = [0.15 \ 0.09 \ 0.06 \ 0.35 \ 0.21 \ 0.14]$;

$$U = \begin{pmatrix} -2.04 & 0.30 & 0.30 & 0.48 & 0 & 0 \\ 0.90 & -3.72 & 0.90 & 0 & 0.72 & 0 \\ 0.30 & 0.30 & -2.04 & 0 & 0 & 0.96 \\ 0.64 & 0 & 0 & -3.24 & 0.50 & 0.50 \\ 0 & 0.96 & 0 & 1.50 & -5.96 & 1.50 \\ 0 & 0 & 1.28 & 0.50 & 0.50 & -3.08 \end{pmatrix}; \quad U_0 = \begin{pmatrix} 0.96 \\ 1.20 \\ 0.48 \\ 1.60 \\ 2.00 \\ 0.80 \end{pmatrix}$$

is given by Fig. 1.

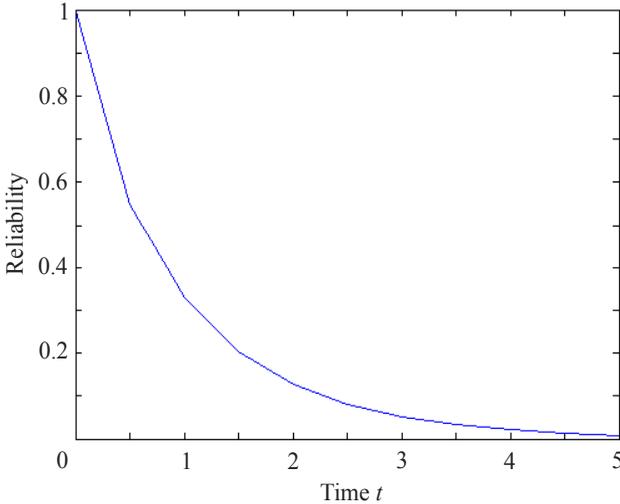


Fig. 1. Reliability in a serial system

Conclusions

We studied the interdependence among the lifetimes of a two component system under the assumption that the lifetimes of these components when considered separately have independent phase type distributions. The probability that the system survives beyond the time t can be found in both serial and parallel systems.

REFERENCES

1. *Buchholz P., Kriege J.* Fitting correlated arrival and service times and related queueing performance // *Queueing Systems*. 2017. V. 85. P. 337–359.
2. *Chao M.T., Fu J.C., Koutras M.V.* Survey of reliability studies of consecutive- k -out-of- n :F and related systems // *IEEE Transactions on Reliability*. 1995. V. 44. No. 1. P. 120–127
3. *Krishnamoorthy A., Ushakumari P.V.* Reliability of a k -out-of- n system with repair and retrieval of failed units // *TOP*. 1999. V. 7. P. 293–304.
4. *Limnios N., Oprisan G.* Semi-Markov processes and reliability. New York: Springer Science+Business Media, 2001.
5. *Sengupta Bhaskar.* The semi-Markovian queue: theory and applications // *Stochastic Models*. 1990. V. 6. No. 3. P. 383–413.
6. *Sivazlian B.D., Stanfel L.E.* Analysis of systems in operations research. NJ, Prentice-Hall, 1975.
7. *Van Houdt B.* A matrix geometric representation for the queue length distribution of multitype semi-Markovian queues // *Performance Evaluation*. 2012. V. 69. Iss. 7-8. P. 299–314.

The exact formulas for state-dependent Markov retrial queues

Eugene Lebedev¹, Vadim Ponomarov¹,
Oksana Pryshchepa²

¹Taras Shevchenko National University of Kyiv, Ukraine

²National University of Water and Environmental Engineering, Rivne, Ukraine

In this paper, we consider Markov models for two types of multiserver retrial queues with an input flow rate that depends on a number of calls in an orbit: classical state-dependent models and state-dependent queues with limited number of retrials. For both systems, the conditions for the existence of the stationary regime are defined and formulas for steady-state probabilities are presented. The investigative technique is based on approximation of the input system by the system with truncated state space.

Mathematical model of the classical state-dependent retrial queue

Consider a continuous time Markov chain $X(t) = (X_1(t); X_2(t))$, $X_1(t) \in \{0, 1, \dots, c\}$, $X_2(t) \in \{0, 1, \dots\}$, that describes a service process in the following system. The service facility contains c identical servers. Service rate of each server is $\mu > 0$, $\nu > 0$ – rate of retrial calls flow, $\lambda_j > 0$ – input flow rate when there are j sources of retrial calls (j calls in an orbit). The first component $X_1(t) \in \{0, 1, \dots, c\}$ indicates the number of busy servers at the instant $t \geq 0$ and the second one $X_2(t) \in \{0, 1, \dots\}$ is the number of retrial sources.

Let's write up the ergodicity conditions for $X(t)$, $t \geq 0$.

Lemma 1. Let $\lambda = \overline{\lim}_{j \rightarrow \infty} \lambda_j < \infty$. Then under $\lambda < c\mu$ the chain $X(t)$ is ergodic and its limit distribution is the same as the single stationary one.

For the construction of calculating schemes and explicit formulas, we will use a system with a truncated state space. Such a model operates similarly to the original queue but has a restriction on the size of an orbit: all new calls are lost when all servers are occupied and there are N calls in the orbit already. Formally, the service process in such queue is described by the Markov chain $X(t, N) = (X_1(t, N), X_2(t, N))$, $X_1(t, N) \in \{0, 1, \dots, c\}$, $X_2(t, N) \in \{0, 1, \dots, N\}$. A state space $S(N)$ of the Markov chain $X(t, N)$ is finite: $S(N) = \{0, 1, \dots, c\} \times \{0, 1, \dots, N\}$. Therefore for $X(t, N)$ there always exists a stationary regime, and via $\pi_{ij}(N)$, $(i, j) \in S(X, N)$ we will designate its stationary probabilities.

Next we introduce matrices which are given by the model parameters:

$$A(j) = \| a_{ik}(j) \|_{i,k=1}^c \text{ is a matrix with entries } a_{i-1}(j) = (j+1)v, i = 1, \dots, c-1;$$

$$a_{cc-1}(j) = \frac{(j+1)v c \mu}{\lambda_j}, k \neq c-1; a_{cc-1}(j) = \frac{(j+1)v(\lambda_j + c\mu)}{\lambda_j} \text{ and all other}$$

entries are equal to 0;

$$B(j) = \| b_{ik}(j) \|_{i,k=1}^c \text{ is a three-diagonal matrix with entries } b_{i-1}(j) = -\lambda_j,$$

$$i = 2, \dots, c; b_{ii}(j) = \lambda_j + jv + (i-1)\mu, i = 1, \dots, c; b_{i+1}(j) = -i\mu, i = 1, \dots, c-1;$$

$$C(j) = \| c_{ik}(j) \|_{i,k=1}^c, \text{ where } (c_{11}(N), c_{12}(N), \dots, c_{1c}(N)) = e_1^T = (1, 0, \dots, 0);$$

$$\text{and for } i = 2, \dots, c \text{ } (c_{i1}(N), c_{i2}(N), \dots, c_{ic}(N)) = (b_{i-11}(N), b_{i-12}(N), \dots, b_{i-1c}(N)).$$

For $\pi_{ij}(N)$ Kolmogorov equations may be solved in an explicit vector-matrix form.

Theorem 1. Let

$$\pi_j^T(N) = (\pi_{0j}(N), \pi_{1j}(N), \dots, \pi_{c-1j}(N)),$$

$$\Delta_j(N) = \frac{\left(\prod_{i=j}^{N-1} B^{-1}(i) A(i) \right) C^{-1}(N) e_1}{e_1^T \left(\prod_{i=0}^{N-1} B^{-1}(i) A(i) \right) C^{-1}(N) e_1}, j = 0, 1, \dots, N,$$

where we use the commonly accepted agreement $\prod_{i=N}^{N-1} B^{-1}(i) A(i) = E$.

Then

$$\pi_j(N) = \pi_{00}(N) \Delta_j(N), j = 0, 1, \dots, N,$$

$$\pi_{cj}(N) = \pi_{00}(N) \frac{(j+1)v}{\lambda_j} \bar{1}^T(c) \Delta_{j+1}(N), j = 0, 1, \dots, N-1,$$

$$\pi_{cN}(N) = \pi_{00}(N) \frac{1}{c\mu} (\lambda_N e_c^T + Nv \bar{1}^T(n)) \Delta_N(N),$$

$$\pi_{00}(N) = \left\{ \bar{1}^T(c) \Delta_0(N) + \sum_{j=1}^N \left(1 + \frac{jv}{\lambda_{j-1}} \right) \bar{1}^T(c) \Delta_j(N) + \right.$$

$$\left. + \frac{1}{n\mu} (\lambda_N e_n^T + Nv \bar{1}^T(n)) \Delta_N(N) \right\}^{-1}.$$

Steady state probabilities $\pi_{ij}(N)$ approach to the probabilities π_{ij} when the truncation level N increases. This fact is a direct corollary of the stochastic ordering of the migration processes ([1, p. 111–116]).

In case of $c = 1, 2$ the results of theorem 1 turn into the explicit formulas of a scalar type.

Corollary 1. Let for the state-dependent M|M|1-queue the conditions of Lemma 1 hold. Then a stationary regime exists and the stationary probabilities take the form

$$\begin{aligned}\pi_{0j} &= \pi_{00} \prod_{i=1}^j \rho_{i-1} \left[1 + \frac{1}{i} (\lambda_{i-1} / \nu - 1) \right], \\ \pi_{1j} &= \pi_{00} \frac{j+1}{\lambda_j / \nu} \prod_{i=1}^{j+1} \rho_{i-1} \left[1 + \frac{1}{i} (\lambda_{i-1} / \nu - 1) \right], j = 0, 1, \dots, \\ \pi_{00} &= \left\{ 1 + \sum_{j=1}^{\infty} \left(1 + \frac{j}{\lambda_{j-1} / \nu} \right) \prod_{i=1}^{j+1} \rho_{i-1} \left[1 + \frac{1}{i} (\lambda_{i-1} / \nu - 1) \right] \right\}^{-1},\end{aligned}$$

where $\rho_i = \lambda_i / \mu$ is a queue load provided that the number of calls in an orbit is equal to i .

Corollary 2. Let for the state-dependent M|M|2-queue the conditions of Lemma 1 hold. Then a stationary regime exists and the stationary probabilities take the form:

$$\begin{aligned}\pi_{0j} &= \pi_{00} \frac{\lambda_0^2}{(\lambda_j + j\nu)^2 + j\nu\mu} \delta_j, \quad \pi_{1j} = \pi_{00} \frac{\lambda_0^2}{\mu} \frac{\lambda_0^2}{(\lambda_j + j\nu)^2 + j\nu\mu} \delta_j, \\ \pi_{2j} &= \pi_{00} \frac{\lambda_0^2}{\mu} \frac{(j+1)\nu}{\lambda_j} \frac{\mu + \lambda_{j+1} + (j+1)\nu}{(\lambda_{j+1} + (j+1)\nu)^2 + (j+1)\nu\mu} \delta_{j+1}, j = 0, 1, \dots,\end{aligned}$$

where

$$\pi_{00} = \left\{ 1 + \frac{\lambda_0}{\mu} + \frac{\lambda_0^2}{\mu} \sum_{j=1}^{\infty} \left(1 + \frac{j\nu}{\lambda_{j-1}} \right) \frac{\mu + \lambda_j + j\nu}{(\lambda_j + j\nu)^2 + j\nu\mu} \delta_j \right\}^{-1},$$

$$\delta_j = \prod_{i=1}^j \rho_{i-1} \frac{(\lambda_i + i\nu)^2 + i\nu\mu}{i\nu(\lambda_i + i\nu + \mu(1 + \rho_{i-1}))}, j = 0, 1, \dots, \rho_i = \lambda_i / 2\mu \text{ is a load for the}$$

M|M|2-queue.

Steady state analysis of systems with limited number of retrials

Consider an $(m+1)$ -dimensional Markov chain $Q^m(t) = (Q_0^m(t), Q_1^m(t), \dots, Q_m^m(t))$, $t \geq 0$ in the phase space $S(Q^m) = \{0, 1, \dots, c\} \times Z_+^m$, where $Z_+ = \{0, 1, \dots\}$. The component $Q_0^m(t)$ indicates the number of busy servers at the instant $t \geq 0$ and $Q_k^m(t)$, $k = 1, \dots, m$ is equal to the number of retrial calls that have made k unsuccessful attempts to get a service. In the queue the number of retrials is limited by m .

The system is given by the following parameters: $\lambda_{j_1, \dots, j_k, \dots, j_m}$ – input flow rate, $j_k \in Z_+$, $k = 1, 2, \dots, m$; μ – service rate; v_k – k -th retrial rate, $k = 1, 2, \dots, m$.

Ergodicity conditions for the process $Q^m(t)$ are presented by the following statement.

Lemma 2. If $\overline{\lim}_{j_m \rightarrow \infty} j_m^{-1} \lambda_{j_1 \dots j_k \dots j_m} < v_m$ and $\lambda_{j_1 \dots j_k \dots j_m}, \mu, v_k > 0, j_k \in Z_+, k = 1, 2, \dots, m$, then $Q^m(t)$ is ergodic and its ergodic distribution $\pi_{ij_1 \dots j_k \dots j_m}, (i, j_1, \dots, j_m) \in S(Q^m)$ is the same as the single stationary one.

In order to obtain representation of $\pi_{ij_1 \dots j_k \dots j_m}, (i, j_1, \dots, j_m) \in S(Q^m)$ in terms of system parameters, we again use approximation with truncated system technique. This system has a fixed number of N places in each set of retrials. This means that the call leaves the system after k -th attempt to get service if all servers are busy and there are N calls that have made k attempts already. The service process for $m \geq 2$ becomes complicated and explicit formulas of the steady state probabilities have not been found so far. However, for $m = 1$ we give representation of $\pi_{ij_1 \dots j_k \dots j_m}, (i, j_1, \dots, j_m) \in S(Q^m)$ in terms of system parameters in an explicit form.

Theorem 2. If for the one-channel state-dependent queue with one retrial ($m = 1$) the conditions of Lemma 2 are realized, then for the service process $Q^1(t) = (Q_0^1(t), Q_1^1(t))$ there exists a stationary regime and stationary probabilities are of a form:

$$\pi_{0j} = \frac{1}{j! v_1^j} \prod_{i=1}^j \frac{\lambda_{i-1} (\lambda_{i-1} + (i-1)v_1)}{\lambda_i + i v_1 + \mu} \pi_{00}, j = 1, 2, \dots,$$

$$\pi_{1j} = \frac{\lambda_0}{\mu} \frac{1}{j! v_1^j} \prod_{i=1}^j \frac{\lambda_{i-1} (\lambda_i + i v_1)}{\lambda_i + i v_1 + \mu} \pi_{00}, j = 0, 1, \dots,$$

where

$$\pi_{00}^{-1} = \sum_{j=0}^{\infty} \left\{ \frac{1}{j!v_1^j} \prod_{i=1}^j \frac{\lambda_{i-1}(\lambda_{i-1} + (i-1)v_1)}{\lambda_i + iv_1 + \mu} + \frac{\lambda_0}{\mu} \frac{1}{j!v_1^j} \prod_{i=1}^j \frac{\lambda_{i-1}(\lambda_i + iv_1)}{\lambda_i + iv_1 + \mu} \right\}.$$

This result is similar to Corollary 1. In order to formulate an analog of Corollary 2 we need the following notations. Let $x_j = \sum_{n=1}^{\infty} \frac{\alpha_{j+n}}{\beta_{j+n}}$ be the

continued fraction where for $j = 1, 2, \dots$,

$$\alpha_j = -\frac{\lambda_{j-1}((\lambda_{j-1} + (j-1)v_1)^2 + (j-1)v_1\mu)}{j(j+1)v_1^2\mu},$$

$$\beta_j = -\frac{jv_1((\lambda_j + \mu + jv_1)^2 + \mu(\lambda_{j-1} + \mu + jv_1))}{j(j+1)v_1^2\mu}.$$

Theorem 3. If for the two-channel state-dependent queue with one retrial the conditions of Lemma 2 are realized, then for the service process there exists a stationary regime and stationary probabilities are of a form:

$$\pi_{0j} = \left(\prod_{k=0}^{j-1} x_k \right) \pi_{00},$$

$$\pi_{1j} = \frac{\lambda_j + jv_1}{\mu} \left(\prod_{k=0}^{j-1} x_k \right) \pi_{00},$$

$$\pi_{2j} = \frac{1}{2\mu^2} \left((\lambda_j + jv_1)^2 + jv_1\mu - (j+1)v_1\mu x_j \right) \left(\prod_{k=0}^{j-1} x_k \right) \pi_{00},$$

$$j = 0, 1, \dots,$$

where

$$\pi_{00}^{-1} = \frac{1}{2\mu^2} \sum_{j=0}^{\infty} \left((\lambda_j + \mu + jv_1)^2 + \mu(\mu + jv_1 - (j+1)v_1x_j) \right) \prod_{k=0}^{j-1} x_k.$$

We note that in contrast to the classical models now the explicit formulas contain the continued fractions as components. The obtained results can be used to solve optimization problems in the class of threshold strategies (see, for example, [2–5]).

REFERENCES

1. *Falin G.I., Templeton J.G.C.* Retrial Queues. London: Chapman and Hall, 1997.
2. *Dudin A.N., Chakravarthy S.R.* Multi-threshold control of the BMAP|SM|1|K queue with group services // J. of Appl. Mathematics and Stoch. Analysis. 2003. V. 16. No. 4. P. 327–347.
3. *Che Soong Kim, Klimenok V.I., Birukov A., Dudin A.N.* Optimal multi-threshold control by the BMAP|SM|1 retrial system // Ann. Oper. Res. 2006. V. 141(1). P. 193–210.
4. *Lebedev E.A., Ponomarev V.D.* Retrial queues with variable service rate // Cybernetics and Syst. Analysis. 2011. V. 47. No. 3. P. 434–441.
5. *Atencia I., Lebedev E., Ponomarov V., Livinska H.* Special retrial queues with state-dependent input rate // Communications in Computer and Inf. Sci. 2019. V. 1109. P. 73–85.

Gaussian approximation and reducing of dimension for a general-type multichannel network

Eugene Lebedev, Hanna Livinska

Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

One of significant problems in simulation and studying of stochastic networks is connected with the large dimension of descriptive processes for the networks and complexity of the phase space of a stochastic model. To overcome such problems, an approach of asymptotic merging of nodes set is proposed for investigation of a general-type multichannel queueing network.

The merging approach of phase space for stochastic systems was pioneered by works of V.S. Korolyuk ([3, 4]) and V.V. Anisimov ([1]), who developed different methodologies in this direction. Nowadays, many works are dedicated to the solution of the phase merging problems (see, for example, [2, 5, 10]). This approach provides the possibility to reduce dimension of considered stochastic processes under some assumptions on transient probabilities of connected semi-Markov processes, and therefore it simplifies notably network analysis.

In the work we use the approach of asymptotic merging applied to the set of nodes for multichannel queueing networks operating under heavy traffic conditions. For such networks we consider a multidimensional service process that is a stochastic process indicating the number of customers being processed at the nodes of the network. Considering the network processing under assumptions of heavy traffic regime yields the possibility to approximate the service process by a Gaussian process (see, for example, [7–9]).

However, for multichannel stochastic networks, the service process can be a vector-process with high dimension. An additional condition for the switching matrix of the network provides an effect of consolidation for some network nodes to a similar-type node, and it allows to merge set of the network nodes. As a result, for the service process in the merged network, we have a form of limit processes, simpler than for the original one.

Note that we prove the convergence of the corresponding service processes in a uniform topology. This can be used to the calculating of the quality functionals of network operation and to the solving optimization problems ([6]).

In the work we consider a multichannel network of a $[G|GI|\infty]^r$ type. Such a network consists of r service nodes. The i -th node receives calls from the outside for servicing at the network, we denote the number of calls arrived in the period of time $[0, t]$ by $v_i(t)$. There are no restrictions on the structure of the input flow $v(t) = (v_1(t), \dots, v_r(t))'$.

Each of the r nodes is a multi-channel queueing system. If a call arrives in such a system, its service begins immediately. For the i -th node distribution function of service time will be denoted by $G_i(t)$, and its Laplace-Stieltjes transformation by

$$G_i(s) = \int_0^{\infty} e^{-st} dG_i(t), \quad i = 1, \dots, r.$$

The direction of movement of a call within the network is controlled by the switching matrix $P = \|p_{ij}\|_1^r$. For any $i = 1, \dots, r$, if a call has its service finished in the i -th node, $p_{ir+1} = 1 - \sum_{j=1}^r p_{ij}$ represents the probability of the call's exit from the network.

The service process in the $[G|GI|\infty]^r$ -network is defined as a r -dimensional process $Q(t) = (Q_1(t), \dots, Q_r(t))'$, where $Q_i(t)$ is the number of calls in the i -th node at an instant t , $t \geq 0$.

Let a call be in the node numbered m at the initial instant $t = 0$, and its service does not begin before $t = 0$. We connect the service process $Q(t)$ with a semi-Markov process $x^{(m)}(t) \in \{1, \dots, r, r+1\}$ which describes the service path of the call within the $[G|GI|\infty]^r$ -network. Based on the algorithm of call service at the $[G|GI|\infty]^r$ -network, we conclude that $x^{(m)}(t)$ has a semi-Markov matrix $\|G_{ij}(t)\|_1^{r+1}$ of the form

$$G_{ij}(t) = \begin{cases} p_{ij}G_i(t), & i = 1, 2, \dots, r, \quad j = 1, 2, \dots, r, r+1, \\ \delta_{r+1j}G_{r+1}(t), & i = r+1, \quad j = 1, 2, \dots, r, r+1, \end{cases}$$

$$G_{r+1}(t) = \begin{cases} 0, & t < 1, \\ 1, & t \geq 1, \end{cases}$$

where δ_{ij} is the Kronecker delta. At the initial instant $t=0$ we have $x^{(m)}(0) = m$ and distribution function of the sojourn time in the initial state coincides with $G_m(t)$.

The state $r+1$ for the process $x^{(m)}(t)$ is absorbing. Absorption in $r+1$ is interpreted as an exit from the network. We denote the transient probabilities of the semi-Markov process $x^{(m)}(t)$ by $p_i^m(t) = P\{x^{(m)}(t) = i\}$, $p_{ij}^m(s, t) = P\{x^{(m)}(s) = i, x^{(m)}(t) = j\}$, $s < t$, $P(t) = \|p_j^i(t)\|_r^r$.

Now let us introduce four conditions that we need for proving our limit theorems. The first condition is connected with the starting load of the network, the second and the third are heavy traffic conditions, and the fourth is a condition for the switching matrix to be splittable.

So, firstly, during this work it is assumed that at the initial time $t=0$ the network is empty:

$$1) Q_i(0) = 0, \quad i = 1, \dots, r.$$

The main purpose of this work is to study the service process $Q(t)$ in heavy traffic regime. Heavy traffic regime means that the network parameters depend on the series number ($Q(t) = Q^{(n)}(t)$) so that the following two conditions for the input flow and service time at the network nodes are met.

$$2) \text{ There exist constants } \lambda_i \geq 0, \quad i = 1, \dots, r, \quad \lambda_1 + \dots + \lambda_r \neq 0, \text{ such that}$$

$$n^{-1/2} (v_1^{(n)}(nt) - \lambda_1 nt, \dots, v_r^{(n)}(nt) - \lambda_r nt) \xrightarrow[n \rightarrow \infty]{U} W(t) = (W_1(t), \dots, W_r(t))',$$

where $W(t)$ is an r -dimensional Brownian motion process with a zero vector of mean values $EW(1) = 0$ and with a correlation matrix

$EW(1)W'(1) = \sigma^2 = \|\sigma_{ij}^2\|_r^r$, the symbol \xrightarrow{U} means weak convergence in the uniform topology.

3) For some numerical sequence $\{g_n\}_1^n$ such that $g_n \rightarrow \infty$ when $n \rightarrow \infty$, there exist limits for Laplace-Stilties transformations $G_i(s) = G_i^{(n)}(s)$:

$$\lim_{n \rightarrow \infty} g_n (1 - G_i^{(n)}(s/n)) = \mu_i(s), \quad i = 1, \dots, r,$$

And, at least, in order to the set $I = \{1, \dots, r\}$ of $[G | GI | \infty]^r$ -network nodes be able to be asymptotically merged, we require the following condition:

4) The set of serving nodes I can be divided into classes I_1, \dots, I_{r_0} ($I_i \cap I_j = \emptyset$, $i \neq j$, $i, j = 1, \dots, r_0$) in such a way that

$$\text{i) } P^{(n)} = P_0 + g_n^{-1} B_0 + o(g_n^{-1}),$$

$$p_{ir+1}^{(n)} = g_n^{-1} b_{ir+1} + o(g_n^{-1}), \quad i = 1, \dots, r;$$

$$\text{ii) } a_{\alpha\alpha} = - \sum_{\beta=1, \beta \neq \alpha}^{r_0+1} a_{\alpha\beta} \neq 0;$$

iii) The spectral radius of the matrix $\hat{P} = \|(1 - \delta_{\alpha\beta}) a_{\alpha\beta} / (-a_{\alpha\alpha})\|_1^{r_0}$ is strictly less than 1,

where

$P_0 = \|\delta_{\alpha\beta} P(\alpha)\|_1^{r_0}$, $P(\alpha) = \|p_{ij}(\alpha)\|_{i,j \in I_\alpha}$ is an indecomposable stochastic matrix with stationary distribution $\rho_i(\alpha)$, $i \in I_\alpha$;

$B_0 = \|b_{ij}\|_1^r = \|B(\alpha, \beta)\|_1^{r_0}$, $B(\alpha, \beta) = \|b_{ij}\|_{i \in I_\alpha, j \in I_\beta}$ are rectangular matrices of size $|I_\alpha| \times |I_\beta|$, $\alpha, \beta = 1, \dots, r_0$;

$a_{\alpha\beta} = \rho'(\alpha) \cdot B_0 \cdot 1(\beta)$, $\alpha, \beta = 1, \dots, r_0$, $\alpha \neq \beta$, $a_{\alpha\alpha} = \rho'(\alpha) \cdot b_{r+1}$, $\alpha = 1, \dots, r_0$, $b'_{r+1} = (b_{1r+1}, \dots, b_{rr+1})$;

$\rho(\alpha)$ is an r -dimensional vector with its i -th entry equal to $\rho_i(\alpha)$ if $i \in I_\alpha$, and equal to zero otherwise, $\alpha = 1, \dots, r_0$;

$1(\beta)$ is an r -dimensional vector with its i -th entry equal to 1 if $i \in I_\beta$, and equal to zero otherwise, $\beta = 1, \dots, r_0$.

Let $\hat{x}^\alpha(t) \in \{1, \dots, r_0, r_0 + 1\}$ be a semi-Markov process, for which the state $r_0 + 1$ is an absorbing one, $\hat{x}^\alpha(0) = \alpha$, $\alpha = 1, \dots, r_0$, the matrix of transient probabilities between states $\{1, \dots, r_0\}$ for the embedded chain is equal to \hat{P} , and the sojourn time in the state α has the following Laplace-Stiljes transformation:

$$\hat{G}_\alpha(s) = \frac{(-a_{\alpha\alpha})}{(-a_{\alpha\alpha}) + \hat{\mu}_\alpha(s)}, \quad \hat{\mu}_\alpha(s) = \sum_{i \in I_\alpha} \rho_i(\alpha) \mu_i(s), \quad \alpha = 1, 2, \dots, r_0.$$

The corresponding probability distributions of $\hat{x}^\alpha(t)$ are denoted by

$$\hat{p}_\beta^\alpha(t) = P(\hat{x}^\alpha(t) = \beta), \quad \hat{P}(t) = \|\hat{p}_\beta^\alpha(t)\|_{\alpha, \beta=1}^{r_0},$$

$$\hat{p}_{\beta\gamma}^\alpha(s, t) = P(\hat{x}^\alpha(s) = \beta, \hat{x}^\alpha(t) = \gamma), \quad s < t.$$

In order to describe the limit of the sequence of stochastic processes

$$\hat{\xi}^{(n)}(t) = n^{-1/2} (Q^{(n)'}(nt) - n \lambda' \left(\int_0^t P^{(n)}(nu) du \right)) \hat{E}, \quad n \geq 2,$$

where $\lambda' = (\lambda_1, \dots, \lambda_r)$, $\hat{E} \|1(1) \dots 1(r_0)\|$ is a rectangular matrix of size $r \times r_0$, it is necessary to introduce two independent Gaussian processes $\hat{\xi}^{(1)}(t)$ and $\hat{\xi}^{(2)}(t)$ which have zero mean values and correlation matrices

$$\hat{R}^{(1)}(t) = \int_0^t \hat{P}'(u) \hat{\sigma}^2 \hat{P}(u) du,$$

$$\hat{R}^{(1)}(s, t) = \int_0^s \hat{P}'(u) \hat{\sigma}^2 \hat{P}(u+t-s) du, \quad s < t,$$

$$\hat{R}^{(2)}(t) = \int_0^t [\Delta(\hat{\lambda}' \hat{P}(u)) - \hat{P}'(u) \Delta(\hat{\lambda}) \hat{P}(u)] du,$$

$$\hat{R}^{(2)}(s, t) = \sum_{\alpha=1}^{r_0} \hat{\lambda}_\alpha \int_0^s [\Delta(\hat{p}_\alpha(u)) - \hat{p}_\alpha(u) \hat{p}_\alpha'(u)] \hat{E}^{(\alpha)}(u, u+t-s) du, \quad s < t,$$

where $\Delta(x) = \|\delta_{\alpha\beta} x_\alpha\|_1^{r_0}$ for any vector $x' = (x_1, \dots, x_{r_0})$, $\hat{p}_\alpha(t)$ is the line of the matrix $\hat{P}(t)$ with the number α , $\hat{\sigma}^2 = \|\hat{\sigma}_{\alpha\beta}\|_1^{r_0} = \hat{E}' \sigma^2 \hat{E}$, $\hat{\lambda}' = (\hat{\lambda}_1, \dots, \hat{\lambda}_{r_0}) = \lambda' \hat{E}$, $\hat{E}^{(\alpha)}(s, t) = \|\hat{E}_{\beta\gamma}^\alpha(s, t)\|_{\beta, \gamma=1}^{r_0}$, $\hat{E}_{\beta\gamma}^\alpha(s, t) = \hat{p}_{\beta\gamma}^\alpha(s, t) / \hat{p}_\beta^\alpha(s)$ if $\hat{p}_\beta^\alpha(s) \neq 0$ and $\hat{E}_{\beta\gamma}^\alpha(s, t) = 0$ otherwise.

For the sequence $\hat{\xi}^{(n)}(t)$, $n \geq 2$, such a result is valid.

Theorem 1. *Let the queueing network of $[G|GI|\infty]^r$ type satisfy conditions 1) – 4). Then, on any finite interval $[0, T]$, the sequence of stochastic processes $\hat{\xi}^{(n)}(t)$, $n \geq 2$, weakly converges to $\hat{\xi}^{(1)}(t) + \hat{\xi}^{(2)}(t)$ in the uniform topology.*

Note, that if in each node of the $[G|GI|\infty]^r$ -network there exists a mean service time of calls and distribution function of service time does not depend on the series number n , then condition 3) is fulfilled for $g_n = n$, $n \geq 1$.

In this case, the semi-Markov process $\hat{x}^\alpha(t)$, $t \geq 0$, is a Markov chain with continuous time and the limit process $\hat{\xi}(t) = \hat{\xi}^{(1)}(t) + \hat{\xi}^{(2)}(t)$ is a Gaussian diffusion process. In the general case, $\hat{\xi}(t)$ can be a non-Markovian Gaussian process with continuous sample functions.

REFERENCES

1. *Anisimov V.* Asymptotic enlargement of the states of random processes // *Cybernetics*. 1973. V. 9. P. 129–132
2. *Anisimov V.* Switching Processes in Queueing Models. ISTE Ltd, 2008.
3. *Gusak D., Korolyuk V.* Asymptotic behaviour of semi-Markov processes with a splittable state set // *Probability Theory and Mathematical Statistics*. 1971. V. 5. P. 43–50.
4. *Korolyuk V., Turbin A.* Mathematical Foundation of the State Lumping of Large Systems. Dordrecht, Netherlands: Kluwer Acad. Press, 1993.
5. *Korolyuk V., Limnios N.* Stochastic Systems in Merging Phase Space. Singapore: World Scientific, 2005.
6. *Lebedev E., Makushenko I.* Profit maximization and risk minimization in semi-Markovian networks // *Cybernetics and Systems Analysis*. 2007. V. 43/2. P. 213–224.
7. *Lebedev E., Livinska G.* Gaussian approximation of multi-channel networks in heavy traffic // *Communications in Computer and Information Science*. 2013. V. 356. P. 122–130.
8. *Lebedev E., Livinska H.* On Gaussian approximation of queueing networks with different starting load // *Communications in Computer and Information Science*. 2018. V. 912. P. 27–38.
9. *Lebedev E., Livinska H.* On transient and stationary regimes for multi-channel networks with periodic inputs // *Appl. Statistics and Computation*. 2018. V. 319. P. 13–23.
10. *Samoilenko I.* Large deviations for random evolutions with independent increments in the scheme of Lévy approximation with split and double merging // *Random Operators and Stochastic Equations*. 2015. V. 22(2). P. 137–149.

On A single server queueing inventory system with common life time for inventoried items*

Khamis Abdullah Khamis AL Maqbali,
Varghese C. Joshua, Achyutha Krishnamoorthy

CMS College Kottayam, India

We consider a single server queueing inventory model. The customers arrive according to Markovian Arrival Process (MAP). The service is assumed to follow Phase type (PH) distribution. An inventory of commodities is attached to the service station. The Common Life Time (CLT) for inventoried items is assumed to follow PH distribution. The inventoried items perish all together. In this case, the supply of items is immediately in local purchase to bring the inventory level to maximum inventory level S . The inventory is not allowed to go down to zero because of local purchase. Each service require a unit of commodity for service. This unit is instantaneously taken at the beginning of the service. The replenishment of inventory follows (s,S) policy with lead time positive. The lead time follows exponential distribution. In the case of local purchase, the outstanding order of the normal purchase (wait until replenishment) is cancelled. Service of a customer begins only when the server is free. Otherwise, the arriving customer joins the buffer. Steady state analysis of the model is performed. Some performance measures are estimated.

Introduction

In many real life situations, customer, who need inventoried items to complete his service, may arrive to service station according to Markovian arrival process. After that, he may go through different phases to complete his service in order to get the inventory. Moreover, common life time for inventoried items may go through different phases until perishing, before they are taken by customers. The Markovian arrival process is more general than Poisson process. MAP keeps the memoryless property of the Poisson process (partial memoryless) [2].

* Supported by the Indian Council for Cultural Relations (ICCR) and Ministry of Higher Education in Sultanate of Oman.

Many papers studied queueing inventory models. For example, Krishnamoorthy et al. [4] studied a PH/PH/1 queueing inventory system under (s,S) policy when the lead time is zero. AL Maqbal, Joshua and Krishnamoorthy [3] studied M/PH/1 queueing inventory system under (s,S) policy with lead time positive. Also, Krishnamoorthy and Shajin [5] studied the MAP/PH/1 queueing inventory system under (s,S) policy with lead time positive. In addition, Divya et al. [1] studied MAP/PH/1 queueing inventory system with processing of service items under vacation and N-policy with impatient customers. In their study, customers arrive according to MAP and service time follows two different phase type distribution. The inventory processing time follows phase type distribution. Moreover, Nair and Jose [8] studied the MAP/PH/1 production inventory model with varying service rates under (s,S) policy with lead time positive.

Some papers studied queueing inventory systems with common life time. For instance, Shajin et al. [10] studied a MAP/PH/1 queueing inventory system with Markovian lead time to bring the inventory level to its maximum. In their study, the common life time of inventoried items follows independent exponential distribution. Besides this, their study provided an interesting application of queueing inventory model with common life time as medicines with the same expiry date. Moreover, Shajin et al. [9] studied a MAP/M/1 and M/M1 queueing inventory system with advanced reservation and cancellation for the next K time frames a head in the case of overbooking. In their study, the CLT of inventoried items follows Phase type distribution.

Some papers studied queueing inventory systems with local purchase. Local purchase was introduced by Krishnamoorthy and Raju [6]. Krishnamoorthy, Varghese and Lakshmy [7] studied an (s,S) production inventory model with positive service time under local purchase.

As mentioned above, Krishnamoorthy and Shajin [5] and Divya et al. [1] studied MAP/PH/1 queueing inventory system. Then Shajin et al. [10] studied a MAP/PH/1 queueing inventory system with common life time. In this paper, we consider an MAP/PH/1 queueing inventory model under (s,S) policy with lead time positive. Besides this, we consider PH distributed CLT for inventoried items. According to the CLT, the inventoried items perish all together. The supply of items is immediately in local purchase to bring the inventory level to the maximum inventory level S .

This model can be described as follows: customers arrive according to MAP with representation (D_0, D_1) of order y . The service is assumed to follow PH-distribution with representation (β, T) of order m . An inventory of commodities is attached to the service station. The inventoried items have CLT which follows PH-distribution with representation (α, W) of order l . We assume that the inventoried items perish all together. In this model, the inventory is not allowed to go down to zero because of local purchase. In order to keep customer goodwill during stock out, the supply of items is immediately in local purchase to bring inventory level to S . Local purchases are purchased at a higher cost than the regular order (wait until replenishment) procedure.

Each service require a unit of commodity for service. This unit is instantaneously taken at the beginning of the service. The replenishment of inventory follows (s, S) policy with lead time positive. The lead time follows exponential distribution with rate θ . When $1 \leq i \leq s$, the replenishment occurs to bring the inventory level i to S according to the rate of lead time. In the case of local purchase, the outstanding order of the normal purchase (wait until replenishment) is cancelled. According to MAP, the first arriving customer instantaneously takes one item of the inventory at the beginning of his service. Then, he can immediately go through different phases of the service. According to MAP, the next arriving customer can directly go through different phases of service when the previous customer completed his service. Otherwise, he must wait in the buffer until the previous customer completes his service. This keeps going.

According to types of blood group, blood bank has store for each blood group. The motivation for the model comes from the inventory management of one store in bank blood. For example, patients deal with one type of blood group in this store. They arrive according to MAP. When the service station is available, the patient may go through different phases in the hospital and one blood bag is immediately taken from store to the patient at the beginning of his service. The CLT for blood bags may go through different phases until perishing, before patients take the blood bags. In this case, the supply of blood bag is immediately in local purchase.

Mathematical description of the model

For the analysis of the model, we introduce the following notations. Let $N(t)$ be the number of customers in the system at time t .

$I(t)$ be the number of items, which are the same type, in the inventory at time t .

$L(t)$ be the phase of common life time at time t .

$M(t)$ be the phase of service at time t .

$Y(t)$ be the phase of the arrival process at time t .

$X(t) = \{(N(t), I(t), L(t), M(t), Y(t)); t \geq 0\}$ is a continuous time Markov Chain on the state space.

Therefore, this model can be studied as a Level Independent Quasi-Birth-Death (LIQBD) process with state space is given by

$$\Omega = \{(0, i, l_1, y_1); 1 \leq i \leq S; 1 \leq l_1 \leq l; 1 \leq y_1 \leq y\} \cup \\ \cup \{(n, i, l_1, m_1, y_1); n \geq 1; 1 \leq i \leq S; 1 \leq l_1 \leq l; 1 \leq m_1 \leq m; 1 \leq y_1 \leq y\}.$$

The terms of transitions of the states are given in the Table 1.

Table 1

Intensities of transitions

From	To		Transition Rate
$(0, 1, l_1, y_1)$	$(1, S, l_1, m_1, y_1')$	$1 \leq l_1 \leq l; 1 \leq m_1 \leq m$	$d_{y_1 y_1'}(1)\beta_{m_1}$
$(0, i, l_1, y_1)$	$(1, i-1, l_1, m_1, y_1')$	$2 \leq i \leq S$	$d_{y_1 y_1'}(1)\beta_{m_1}$
(n, i, l_1, m_1, y_1)	$(n+1, i, l_1, m_1, y_1')$	$1 \leq n; 1 \leq i \leq S$	$d_{y_1 y_1'}(1)$
$(0, i, l_1, y_1)$	$(0, i, l_1, y_1')$	$1 \leq i \leq S, y_1 \neq y_1'$	$d_{y_1 y_1'}(0)$
(n, i, l_1, m_1, y_1)	(n, i, l_1, m_1, y_1')	$1 \leq n; 1 \leq i \leq S; y_1 \neq y_1'$	$d_{y_1 y_1'}(0)$
$(0, i, l_1, y_1)$	$(0, S, l_1, y_1)$	$1 \leq i \leq s$	θ
$(1, i, m_1, y_1)$	(n, S, l_1, m_1, y_1)	$1 \leq n; 1 \leq i \leq s$	θ
(n, i, l_1, m_1, y_1)	$(0, i, l_1, y_1)$	$1 \leq i \leq S$	$\tau_{m_1}^0$
(n, i, l_1, m_1, y_1)	$(n-1, i-1, l_1, m_1', y_1)$	$2 \leq i \leq S; 2 \leq n$	$\tau_{m_1}^0 \beta_{m_1'}$

From	To		Transition Rate
$(n, 1, l_1, m_1, y_1)$	$(n-1, S, l_1, m_1', y_1)$	$2 \leq n$	$\tau_{m_1}^0 \beta_{m_1}$
(n, i, l_1, m_1, y_1)	$(n, i-1, l_1, m_1', y_1)$	$1 \leq n ; m_1 \neq m_1'; 1 \leq i \leq S$	$\tau_{m_1 m_1'}$
$(0, i, l_1, y_1)$	$(0, S, l_1', y_1)$	$1 \leq i \leq S$	$w_{l_1}^0 \alpha_{l_1}$
(n, i, l_1, m_1, y_1)	(n, S, l_1', m_1, y_1)	$1 \leq n ; 1 \leq i \leq S$	$w_{l_1}^0 \alpha_{l_1}$
$(0, i, l_1, y_1)$	$(n, 0, l_1', y_1)$	$1 \leq n ; 1 \leq i \leq S; l_1 \neq l_1'$	$w_{l_1 l_1'}$
(n, i, l_1, m_1, y_1)	$(n, 0, l_1', m_1, y_1)$	$1 \leq n ; 1 \leq i \leq S; l_1 \neq l_1'$	$w_{l_1 l_1'}$

The infinitesimal generator Q of the Level Independent Quasi-Birth-Death (LIQBD) process with state space is of the form

$$Q = \begin{pmatrix} B_{00} & B_{01} & & & \\ B_{10} & A_1 & A_0 & & \\ & A_2 & A_1 & A_0 & \\ & & A_2 & A_1 & A_0 \\ & & & \ddots & \ddots & \ddots \end{pmatrix}.$$

Steady-state analysis

Stability condition

Theorem 1. The Markov chain with the infinitesimal generator Q of the Level Independent Quasi-Birth-Death (LIQBD) is stable, if and only if $\lambda < \mu$.

Where

$$\lambda = \left(\sum_{i=0}^{(mS)} \pi_i \right) D_1 e_y ;$$

π_i is a row vector of order (y) and

$$\mu = \left(\sum_{i=i}^{(S+1)} \pi_i \right) \Lambda e_{(yml)}.$$

Where $\Lambda = [I_l \otimes [T_m^0 \otimes (\beta \otimes I_y)]]$ and π_i are row vectors of order (yml) .

Stationary distribution

The stationary distribution of the Markov chain under consideration can be obtained by solving [6] $XQ = 0$ and $Xe = 1$.

Let X be decomposed with Q as following: $X = (X_0, X_1, \dots)$, where $X_0 = (X_{01}, \dots, X_{0S})$;

$$X_{0k} = (X_{0k1}, X_{0k2}, \dots, X_{0kl}) \text{ for } k = 1, 2, \dots, S;$$

$$X_{0kr} = (x_{0kr1}, x_{0kr2}, \dots, x_{0krj}) \text{ for } r = 1, 2, \dots, l;$$

$$X_i = (X_{i1}, \dots, X_{iS}) \text{ for } i = 1, 2, \dots;$$

$$X_{ik} = (X_{ik1}, X_{ik2}, \dots, X_{ikl}) ; X_{ikr} = (X_{ikr1}, X_{ikr2}, \dots, X_{ikrm}) ;$$

$$X_{ikrj} = (x_{ikr1j}, x_{ikr2j}, \dots, x_{ikrjy}) \text{ for } j = 1, 2, \dots, m .$$

Performance measures

Under steady state, we obtain some performance measures of the system as following:

1. Expected number of customers in the system: $E[N] = \sum_{i=0}^{\infty} i X_i e.$

2. Expected number of items in inventory: $E[I] = \sum_{i=0}^{\infty} \sum_{k=1}^S k X_{ik} e.$

3. Probability that the server is idle: $b_0 = \sum_{k=1}^S X_{0k} e .$

Conclusions

In this paper, we analyse an MAP/PH/1 system with inventory level under (s, S) policy with lead time positive and with common life time for inventoried items. In the case of expiry of the common life time for the inventoried items, the supply of items is immediately in local purchase to bring the inventory level to the maximum inventory level S . Different performance measures are estimated under Steady state condition.

REFERENCES

1. Divya V., Krishnamoorthy A., Vishnevsky V.M., Kozyrev D.V. On a queueing system with processing of service items under vacation and N-policy with impatient customers // *Queueing Models and Service Management*. 2020. V. 3. No. 2. P. 167–201.
2. *He Qi-Ming*. Fundamentals of matrix-analytic methods. New York: Springer, 2014.
3. K.A.K AL Maqbali, Joshua V.C., Krishnamoorthy A. On a single server queueing inventory system // *Distributed Computer and Communication Networks: Control, Computation, Communications*. Conference, DCCN2020, Moscow, 14–18 September 2020.
4. Krishnamoorthy A., Jose K.P., Narayanan V.C. Numerical investigation of a PH/PH/1 inventory system with positive service time and shortage // *Neural Parallel Sci. Comput*. 2008. V. 16. P. 579–592.
5. Krishnamoorthy A., Dhanya S. MAP/PH/1 retrial queueing-inventory system with orbital search and renegeing of customers // *International Conference on Analytical and Computational Methods in Probability Theory*. Springer, Cham, 2017. P. 158–171.
6. Krishnamoorthy A., Raju N. N-Policy for (s,S) perishable inventory system with positive lead time // *Korean J. Comput. Appl. Math*. 1998. V. 5(1). P. 253–261.
7. Krishnamoorthy A., Resmi V., Balakrishna L. Production inventory system with positive service time under local purchase // *International Conference on Information Technologies and Mathematical Modelling*. Springer, Cham, 2019. P. 243–256.
8. Salini S. Nair, Jose K.P. A MAP/PH/1 production inventory model with varying service rates // *International Journal of Pure and Applied Mathematics*. 2017. V. 117. No. 12. P. 373–381.
9. Dhanya S., Krishnamoorthy A., Dudin A.N., Joshua V.C., Jacob V. On a queueing-inventory system with advanced reservation and cancellation for the next K time frames ahead: the case of overbooking // *Queueing Systems*. 2020. V. 94. No. 1-2. P. 3–37.
10. Dhanya S., Krishnamoorthy A., Manikandan R. On a queueing-inventory system with common life time and Markovian lead time process // *Operational Research*. 2020. P. 1–34.
11. Stewart W.J. *Probability, Markov Chains, Queues, and Simulation*. Princeton University Press, 2009.

Analyses of feedback queue with positive server setup time and impatient calls

Agassi Melikov¹, V. Divya², Sevinc Aliyeva³

¹*Institute of Control Systems, Azerbaijan*

²*N.S.S. College, Cherthala-688556, Kerala, India*

³*Baku State University, Azerbaijan*

In this paper, generalization of the model of queuing systems with feedback in Melikov *et al.* [1] is proposed by introducing the effect of impatience of calls when the server is in setup mode. First of all, note that many real systems are described by such kind of models. Classical examples are communication networks where erroneously transmitted data is retransmitted. The second example is production systems in which the defective part requires re-processing. The third example is modern service integrators, in which customers who have received quality service once again return to them for service.

In this field the pioneer are Takacs' works [2, 3] in which models of queuing systems with instantaneous [2] and delayed feedback [3] are investigated. Such kind of models are intensively investigated in the last two decades. Kumari [4] gives a detailed review of work done until 2011. Further developments could be accessed from Koroliuk *et al.* [5] and Melikov *et al.* [6].

To our best knowledge, there are no works devoted to the feedback queuing model with a setup time for servicing of calls that required repeated processing (feedback calls, f -calls). Here the model of a queuing system with an infinite buffer, instantaneous feedback and positive server setup time is proposed. It is assumed that calls in the buffer are impatient if the server is in the setup period. To calculation of the steady-state probabilities and performance measures of the system the matrix-geometric method (MGM) [7] is used.

Description of the system and construction of the generating matrix

Consider a single-server queuing system with an infinite buffer, which receives a Poisson flow of primary calls (p -calls) from the outside. The service times are independent and identically distributed (iid) r.v. and have a common exponential cumulative distribution function (cdf) with an average value μ^{-1} . After the completion of the servicing process, p -calls,

independently of each other and according to the Bernoulli scheme, either immediately require repeated servicing with probability (w.p) σ , or leave the system with a complementary probability $1 - \sigma$. The service times of f -calls are also iid r.v. with the same average μ^{-1} .

To start the process of servicing f -calls, the server needs some positive random time, which is called the server setup time. It has an exponential cdf with an average value θ^{-1} , i.e. at any time, the server can be in one of two status: in operational mode or setup mode. It is not allowed to interrupt the server setup time, i.e. the f -call, which initiates the server setup process, remains on it for the entire setup period.

The p -calls are considered to have information about the server status. This means that if the server is in setup mode, then their intensity is equal to λ_0 , otherwise, it is equal to λ_1 . Calls in the buffer are impatient during the server setup time and only the call at the head of the queue being impatient. In other words, during the server setup period, the call that is at the head of the queue leaves the system after a random time, which has an exponential distribution with an average value τ^{-1} .

It is required to find the joint distribution of the number of calls in the system and the state of the server, as well as the following performance measures of the system: the average number of calls in the system (L_s), fraction of the time in which the system is in operation mode (T_{wr}) and setup mode (T_{sr}), as well as the intensity of the server switching from setup mode to operating one (R_{sw}).

The state of the system at an arbitrary moment of time is determined by the two-dimensional vector (n, k) , where n is the number of calls in the system, k is the state of the server, i.e. $k = 0$ if the server is in setup, $k = 1$ if the server is operation mode. The system operation is described by a two-dimensional Markov Chain (2D MC) with the following infinitesimal matrix:

$$Q = \begin{bmatrix} B_0 & C_0 & \dots & \dots \\ B_1 & A_1 & A_0 & \dots \\ \dots & A_2 & A_1 & A_0 \\ \dots & \dots & \dots & \dots \end{bmatrix},$$

where block matrices are defined as follows: $B_0 = (-\lambda_1)$ is the matrix of dimension 1×1 ; $C_0 = (0 \quad \lambda_1)$ is the matrix of dimension 1×2 ; A_0, A_1 and

A_2 are matrices of dimension 2×2 , i.e.

$$A_2 = \begin{pmatrix} \tau & 0 \\ 0 & \mu(1-\sigma) \end{pmatrix}; A_1 = \begin{pmatrix} -(\lambda_0 + \theta + \tau) & \theta \\ \mu\sigma & -(\lambda_1 + \mu) \end{pmatrix}; A_0 = \begin{pmatrix} \lambda_0 & 0 \\ 0 & \lambda_1 \end{pmatrix}.$$

Let $\mathbf{v} = (v_0, v_1)$ is a vector of state probabilities of an MC with two states and an infinitesimal generator A , where

$$A = A_0 + A_1 + A_2 = \begin{pmatrix} -\theta & \theta \\ \mu\sigma & -\mu\sigma \end{pmatrix},$$

i.e. indicated above vector $\mathbf{v} = (v_0, v_1)$ is found from the following system of equations:

$$\mathbf{v}A = \mathbf{0}, \mathbf{v}\mathbf{e} = 1, \quad (8)$$

where $\mathbf{e} = (1, 1)^T$.

From (8) we find that $v_0 = \mu\sigma/(\theta + \mu\sigma)$, $v_1 = \theta/(\theta + \mu\sigma)$. The quasi birth-death process is stable if and if the following condition is satisfied [7]:

$$\mathbf{v}A_0\mathbf{e} < \mathbf{v}A_2\mathbf{e}. \quad (9)$$

From (9) we find the following condition for the ergodicity of the system:

$$\lambda_1\theta + \lambda_0\mu\sigma < \mu(\theta(1-\sigma) + \tau\sigma). \quad (10)$$

When the ergodicity condition (10) is fulfilled, the steady-state probabilities of the system are calculated as follows.

Let us express the steady-state probabilities $p(n, k), (n, k) \in E$ in vector form $\mathbf{p} = (\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2, \dots)$, where \mathbf{p}_0 has dimension 1, i.e. $\mathbf{p}_0 = (p(0, 1))$, while $\mathbf{p}_n, n \geq 1$, are two-dimensional vectors, i.e. $\mathbf{p}_n = (p(n, 0), p(n, 1))$. According to the algorithm of the MGM [7], the indicated quantities satisfy the equation $\mathbf{p}_n = \mathbf{p}_1 R^{n-1}, n \geq 2$, where R is the minimal non-negative solution of the following quadratic matrix equation:

$$R^2 A_2 + R A_1 + A_0 = 0.$$

Probabilities of boundary states \mathbf{p}_0 and \mathbf{p}_1 are found from the following system of equations:

$$\begin{aligned} \mathbf{p}_0 B_0 + \mathbf{p}_1 B_1 &= 0; \\ \mathbf{p}_0 C_0 + \mathbf{p}_1 (A_1 + R A_2) &= 0. \end{aligned} \quad (11)$$

System of equations (11) is solved by taking into account the normalization condition:

$$p(0,1) + p_1(I - R)^{-1}e = 1,$$

where I is the identity matrix of dimension 2×2 .

After finding the steady-state probabilities, the desired performance measures of the system are calculated as follows:

$$L_s = \sum_{n=1}^{\infty} \sum_{k=0}^1 np(n, k);$$

$$T_{sr} = \sum_{n=1}^{\infty} p(n, 0);$$

$$T_{wr} = \sum_{n=0}^{\infty} p(n, 1);$$

$$R_{sw} = \theta \sum_{n=1}^{\infty} p(n, 0).$$

Note that for certain ratios of the system parameters, the space merging method [5, 6] can be used to solve this problem as well.

Conclusions

The paper studies a model of a single-server queuing system with an infinite buffer, instantaneous feedback and impatient calls. Upon completion of the service, some of the calls, according to the Bernoulli scheme, either leave the system or immediately require repeated service. To start the service of f -call, the server needs some random setup time, which has an exponential distribution function. It is considered that when the server is in the setup mode, it cannot handle call, and interruption of the setup period is not allowed. Calls in the buffer are impatient when the server is in the setup mode. The rate of incoming calls depends on the server status.

It is shown that the mathematical model of the system under study is a 2D MC with an infinite-dimensional state space. The ergodicity condition of the model is found and to study the corresponding 2D MC the matrix-geometric method is used.

REFERENCES

1. *Melikov A.Z., Aliyeva S.H., Shahmaliyev M.O.* Methods to calculate the system with instantaneous feedback and varying arrival rate // Automation and Remote Control. 2020. V. 81. No. 9. P. 1637–1648.
2. *Takacs L.* A single-server queue with feedback // Bell Syst. Tech. J. 1963. V. 42. P. 505–519.
3. *Takacs L.* A queuing model with feedback // Operations Research. 1977. V. 11. P. 345–354.
4. *Kumari N.* An analysis of some continuous-time queueing systems with feedback. Ph.D. thesis. Punjabi University, 2011. 235 p.
5. *Koroliuk V.S., Melikov A.Z., Ponomarenko L.A., Rustamov A.M.* Methods for analysis of multi-channel queuing models with instantaneous and delayed feedbacks // Cybernetics and Systems Analysis. 2016. V. 52. No. 1. P. 58–70.
6. *Melikov A.Z., Aliyeva S.H., Sztrik J.* Analysis of queuing system MMPP/M/K/K with delayed feedback // Mathematics. 2019. V.7. Art. 1128.
7. *Neuts M.F.* Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach. Baltimore: John Hopkins Univer. Press, 1981. 332 p.

A two-class service system for performance analysis of network slicing with QoS Isolation*

Faina Moskaleva¹, Ekaterina Lisovskaya¹,
Yuliya Gaidamaka^{1,2}

¹ Peoples' Friendship University of Russia (RUDN University), Moscow, Russia

² Federal Research Center «Computer Science and Control»
of the Russian Academy of Sciences (FRC CSC RAS), Moscow, Russia

The new 5G networks and their operators have to manage a wide range of services with very varied connection requirements, targeting new market segments and vertical industries. Network slicing is a key technology that allows network operators to provide their physical infrastructure to support various services with different requirements [1]. Different sets of services may be associated with logically independent end-to-end networks, i.e. slices. The slice is a logical network that provides functional capabilities and network characteristics [2]. Slices are configured and managed by tenants, to whom operators delegate control over resource utilization and service performance. Network slicing properties [3] are the automation of creating and setting slice, the slice isolation (slice independence from traffic in other slices, but also security, etc.), the elasticity of slicing (fair [4] and efficient use of resources, adaptation to conditions), management (self-management in a slice) and ability to assign priority slices.

The key feature of network slicing for ensuring performance and high quality of service is isolation, which limits the influence of slices on each other. Using isolation and resource sharing strategies on the radio interface is a rather entangled process [5].

However, as shown in [6], it is possible to achieve an optimal state between isolation and efficiency by setting some network parameters. It allows us to prioritize and configure slices in accordance with the specific tasks for which it is used. Nevertheless, resource management inside the slice and between different slices should guarantee not only slice isolation but also the fair sharing of resources between users, as noted in [7].

* The publication has been prepared with the support of the “RUDN University Program 5-100” (recipient Yu.Gaidamaka). The reported study was funded by RFBR, project number 18-07-00576 and 20-07-01064.

This paper proposes the use of queueing theory and optimization theory methods. So it is a combination and extension of the research conducted in [8] and [9]. In [8] the authors consider one of the potential radio resource allocation schemes for multiservice wireless networks with Network Slicing technology. They describe it by using the retrial queue with the orbit to wait for sessions. In [9], the authors propose a policy for bandwidth-greedy communication service, which provides efficient resource usage, fairness and slice isolation.

The paper organizes as follows. In the next section, we construct a mathematical model of a network with two active slices. The third section is a detailed description of the mathematical model for network slicing with isolation. In the fourth section, we present an algorithm for the sharing of resources, ensuring their fair and efficient use. Finally, we conclude with some remarks on open issues.

Mathematical model

Let two Poisson processes, corresponding to requests for data transmission from users of two different slices, arrive at the multilinear queueing system (QS). The intensities of arrivals are constant and equal to λ_1 and λ_2 , respectively. Service durations per resource unit are independent random variables distributed exponentially with the parameters μ_1 and μ_2 for the first and second slices, respectively. Let the total amount of QS resources for customers servicing be equal to C . The amount of resources allocated to the request depends on the system load and varies in the ranges $[b_{min}, b_{max}]$ and $[d_{min}, d_{max}]$ for first and second slices respectively.

We define a stochastic process $\mathbf{X}(t) = \{M(t), N(t), t > 0\}$, where $M(t) = m$ is the number of customers the first slice at the time t , $N(t) = n$ is the number of customers in the second slice at the time t . Moreover

$$m \in \left\{ 0, 1, \dots, \left\lfloor \frac{C}{b_{min}} \right\rfloor \right\}, \quad n \in \left\{ 0, 1, \dots, \left\lfloor \frac{C}{d_{min}} \right\rfloor \right\}.$$

Then the states space of the two-dimensional process has the form

$$\mathbb{X} = \{(m, n) : mb_{min} + nd_{min} \leq C\}.$$

Denote the amount of the allocated resource to one request in the first slice in the state (m, n) by $b(m, n)$, and in the second slice by $d(m, n)$. Then the service intensities for the first and second slices are defined as $mb(m, n)\mu_1$ and $nd(m, n)\mu_2$. To ensure the isolation of slices, we introduce the state of the

process $M(t) > \bar{M}$, where \bar{M} is the number of customers that can be guaranteed to be accepted in the first slice, and $N(t) > \bar{N}$, where \bar{N} is the number of customers that can be guaranteed to be accepted in the second slice. In the case of resource bottleneck, if one of the slices exceeds a predetermined threshold value, then the arrival request of the second slice can crowd out one or more requests of the first to get service (see Fig. 1),

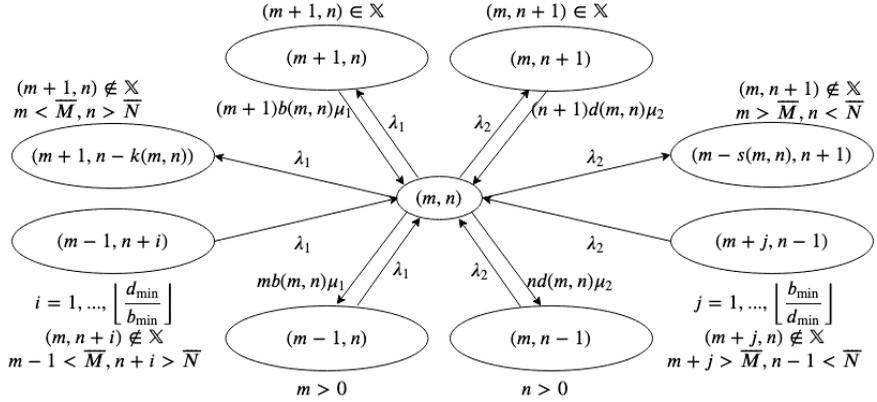


Fig. 1. The central state and the transition rates for the stochastic process $\mathbf{X}(t)$

the number of which is calculated as

$$\left[k(m, n) = \frac{(m+1)b_{\min} + nd_{\min} - C}{d_{\min}} \right], \quad \left[s(m, n) = \frac{mb_{\min} + (n+1)d_{\min} - C}{b_{\min}} \right].$$

If none of the slices exceeds the threshold value or exceeds both in case of a resource bottleneck, the arrival customer will be dropped.

Thus, we write down the set of conditions for dropping requests and service interruptions for the first and second slices. The states space of customers dropping:

$$\mathbb{B}_1^{arr} = \{(m, n) \in \mathbb{X} : (m+1, n) \notin \mathbb{X}, m \geq \bar{M}\},$$

$$\mathbb{B}_2^{arr} = \{(m, n) \in \mathbb{X} : (m, n+1) \notin \mathbb{X}, n \geq \bar{N}\}.$$

The states space of service interruption:

$$\mathbb{B}_1^{pr} = \{(m, n) \in \mathbb{X} : (m, n+1) \notin \mathbb{X}, m > \bar{M}, n < \bar{N}\},$$

$$\mathbb{B}_2^{Pr} = \{(m, n) \in \mathbb{X} : (m+1, n) \notin \mathbb{X}, m < \bar{M}, n > \bar{N}\}.$$

To find the probabilities state vector $p = [p(m, n)]_{(m, n) \in \mathbb{X}}$, it is necessary to solve the system of linear equations

$$\begin{cases} \mathbf{p}\mathbf{Q} = \mathbf{0}, \\ \mathbf{p}\mathbf{e} = \mathbf{1}, \end{cases}$$

where \mathbf{Q} is the generator matrix of the two-dimensional stochastic process $\mathbf{X}(t)$, and \mathbf{e} is the unit column vector.

We write the generator matrix, its elements will have the form:

$$q((m, n), (m', n')) = \begin{cases} \lambda_1, m' = m+1, n' = n, (m', n') \in \mathbb{X}, \\ \lambda_1, m' = m+1, n' = n-k, (m+1, n) \notin \mathbb{X}, m < \bar{M}, n \geq \bar{N}, \\ \lambda_2, m' = m, n' = n+1, (m', n') \in \mathbb{X}, \\ \lambda_2, m' = m-s, n' = n+1, (m, n+1) \notin \mathbb{X}, m \geq \bar{M}, n < \bar{N}, \\ mb(m, n)\mu_1, m' = m-1, n' = n, m' > 0, \\ nd(m, n)\mu_2, m' = m, n' = n-1, n' > 0, \\ Q, m' = m, n' = n, \\ 0, \text{ otherwise,} \end{cases}$$

where

$$Q = -\lambda_1 \cdot I((m+1, n) \in \mathbb{X} \cup n \geq \bar{N}) - \lambda_2 \cdot I((m, n+1) \in \mathbb{X} \cup m \geq \bar{M}) - mb(m, n)\mu_1 \cdot I(m > 0) - nd(m, n)\mu_2 \cdot I(n > 0).$$

Having obtained the probability distribution, we can find some metrics characterizing the system performance for customers:

- probabilities of customers dropping and service interruption:

$$B_s^{arr} = \sum_{(m, n) \in \mathbb{B}_s^{arr}} p(m, n), \quad B_s^{Pr} = \sum_{(m, n) \in \mathbb{B}_s^{Pr}} p(m, n), \quad s = 1, 2;$$

- average service time:

$$S_1 = \frac{N_1}{\lambda_1(1 - B_1^{arr}) - \lambda_2 \sum_{(m, n) \in \mathbb{B}_1^{arr}} s(m, n) p(m, n)},$$

$$S_2 = \frac{N_2}{\lambda_2(1 - B_2^{arr}) - \lambda_1 \sum_{(m, n) \in \mathbb{B}_2^{arr}} k(m, n) p(m, n)};$$

- probability of violation:

$$V_1 = \sum_{m > \bar{M}} p(m, n), \quad V_2 = \sum_{n > \bar{N}} p(m, n).$$

Now, we consider a subspace of states where each request in both slices can be allocated the maximum required resource amount, i.e. $\Omega_1 = \{(m, n): mb_{max} + nd_{max} \leq C\}$, which we will call the states subspace of excess resources. Note that the states subspace $\Omega_0 = X \setminus \Omega_1$ are the states of limited resources, for which the task is to propose an algorithm for the efficient and fair allocation of radio resources.

Resource sharing

In order to the resource sharing algorithm to satisfy the requirement of fairness and efficient resource use, we will solve the optimization problem. Let the data transfer rate to the user of the first slice corresponding to the allocated resource $b(m, n)$ in the state (m, n) have the utility functions

$$U(b(m, n)) = \ln b(m, n) \text{ or } U(b(m, n)) = \frac{\ln b(m, n) - \ln b_{min}}{\ln b_{max} - \ln b_{min}}.$$

Similarly, for the second slice,

$$U(d(m, n)) = \ln d(m, n) \text{ or } U(d(m, n)) = \frac{\ln d(m, n) - \ln d_{min}}{\ln d_{max} - \ln d_{min}}.$$

The optimal resource amounts correspond to solving the following optimization problem

$$\begin{aligned} w_1(m, n)mU(b(m, n)) + w_2(m, n)nU(d(m, n)) &\rightarrow \max \\ \text{s.t. } mb(m, n) + nd(m, n) &= C, \\ \text{over } \begin{cases} b_{min} \leq b(m, n) \leq b_{max}, \\ d_{min} \leq d(m, n) \leq d_{max}. \end{cases} \end{aligned}$$

The weights $w_1(m, n)$ and $w_2(m, n)$ will be calculated using the following formulas:

$$w_1(m, n) = \begin{cases} 1, & m \leq \bar{M}, \\ \frac{1}{m - \bar{M} + 1}, & m > \bar{M}, \end{cases} \quad w_2(m, n) = \begin{cases} 1, & n \leq \bar{N}, \\ \frac{1}{n - \bar{N} + 1}, & n > \bar{N}. \end{cases}$$

Thus, the stationary point of the objective function is located at the intersection of lines $mb(m, n) + nd(m, n) = C$ and $w_2(m, n)b(m, n) = w_1(m, n)d(m, n)$.

Conclusions

This paper presents a model of resource sharing for network slicing to ensure the isolation of slices. The developing algorithm for sharing resources takes into account the features of the presented mathematical model. The resource sharing problem is formulated as an optimization problem and its solution is obtained. The resource sharing algorithm allows us to separate network resources fairly and effectively. The objective of further tasks is to compare the proposed algorithm with complete sharing and complete isolation algorithms.

REFERENCES

1. *NGMN*, 5G White Paper, Technical Report, 2015.
2. *ETSI*, 5G: System Architecture for the 5G System (Release 15), 3GPP TS 23.501 V 15.2.0, 2018.
3. *GSMA*, Network Slicing – Use Case Requirements, Technical Report, 2018.
4. *Jain R., Chiu D.M., WRR H.* A quantitative measure of fairness and discrimination for resource allocation in shared computer systems // *CoRR* cs.NI/9809099, 1998.
5. *Richart M., Baliosian J., Serrat J., Gorricho J.* Resource slicing in virtual wireless networks: A survey // *IEEE Transactions on Network and Service Management*. 2016. V. 13. P. 462–476.
6. *Marabissi D., Fantacci R.* Highly flexible RAN slicing approach to manage isolation, priority, efficiency // *IEEE Access*. 2019. V. 7. P. 97130–97142.
7. *Alfoudi A.S.D., Shah Newaz H. S., Otebolaku A., Lee G. M., Pereira R.* An efficient resource management mechanism for network slicing in a LTE network // *IEEE Access*. 2019. V. 7. P. 89441–89457.
8. *Markova E., Adou K.Y.B., Ivanova D., Golskaia A., Samouylov K.* Queue with retrial group for modeling best effort traffic with minimum bit rate guarantee transmission under network slicing // *Lecture Notes in Computer Science*. 2019. V. 11965. P. 432–442.
9. *Yarkina N., Gaidamaka Y., Correia L. M., Samouylov K.* An analytical model for 5G network resource sharing with flexible SLA-oriented slice isolation // *Mathematics*. 2020. V. 8. No. 7. Art. no. 1177.

Asymptotic-diffusion analysis of multiserver retrial queueing system with priority customers^{*}

Anatoly Nazarov¹, Tuan Phung-Duc^{2,3}, Yana Izmailova¹

¹ Tomsk State University, Tomsk, Russia

² Faculty of Engineering Information and Systems, University of Tsukuba, Tsukuba, Japan

³ VNU Vietnam Japan University, Hanoi, Vietnam

This paper considers a priority multi-server retrial queue with two classes of customers. Primary customers have preemptive priority over secondary users. The dynamics of primary customers is the same as that of an Erlang loss system with Poisson input and exponential service time distribution. Secondary users can cognitively use the channels when they are not used by primary users. Secondary users that see all the channels occupied upon arrival join the orbit and retry later. Arriving primary users that see all the channels occupied by other primary users are lost. If all the channels are occupied but some channels are occupied by secondary users, one of these ongoing secondary users is interrupted and enters the orbit. Secondary users from the orbit retry to occupy an idle server until they are successfully occupying one. For this model, we consider an asymptotic regime in which the retrial rate is extremely low. While the number of secondary users in the orbit explodes in this regime, we prove that a scaling version of the number of users in the orbit weakly converges to a diffusion process.

1. Introduction

Priority queue is one of the main streams in queueing theory because the model naturally arises in various applications from services to telecommunication and computer systems. In practice, service differentiation is required in various situations. In service systems such as airline, first class customer has priority over economy class. In telecommunication system, voice packets have higher priority than other data packets such as email etc.

Recently, mobile traffic has explosively increased leading to the shortage of wireless spectrums. Cognitive radio networks are promising technologies for this spectrum shortage problem. In cognitive radio networks, primary users and secondary users are available. Primary users are granted some

^{*} The publication was financially supported by RFBR according to the research project No.18-01-00277.

spectrum bands. Secondary users can cognitively utilize these bands when they are not used by primary users. To this end, secondary users must use these band in such a way that does not interfere primary users. Once primary users arrive, ongoing secondary users must evacuate and then sense to occupy a channel in a later time. Motivated by this application, we consider a multi server priority retrial queue. Primary users and secondary users arrive at the system of multiple servers according to Poisson processes. Service time distributions of primary and secondary users are exponential distributions with distinct parameters. From the viewpoint of primary users, the dynamics of the number of primary users is the same as that of an Erlang loss system.

Upon arrival, if there are some idle servers, the secondary user can use one of these idle ones, otherwise they join the orbit and sense to find an idle one in an exponentially distributed time. Upon the arrival of a primary user, if there are some idle servers, the primary user chooses one of them. Otherwise, the primary user is lost if all the channels are used by other primary users, or interrupts one of ongoing secondary users. The interrupted ongoing secondary user returns to the orbit and sense to occupy an idle server. The sensing process of blocked secondary users and the interrupted ones is the same as the orbit in retrial queues.

As a related work, Akutsu and Phung-Duc [1] study a similar model where secondary users must sense before accessing the channel using simulation. In the study [1], analytical solution is derived for the single server case and simulation is carried out for the multi-server case. Furthermore, Phung-Duc et al. [6] analytically study the same model and derive the stability condition. Morozov et al. [4] study the model where secondary can wait at a buffer in case an idle server is not available or in case of being interrupted. In paper [7] consider the effect of retrial phenom on in performance modeling of radio byusing a finite-source queueing model. Nemouchi H. and Sztrik J. [8] used stochastic simulation for performance evaluation of cognitive radio network. In this paper, in contrast to the models in [1, 4, 6, 7, 8], we consider a new model where fresh secondary users can occupy an empty channel upon arrival without sensing.

In this paper, rather than finding a numerical solution, we consider asymptotic analysis under the condition that the delay in the orbit is extremely long using the asymptotic diffusion method [2, 3, 5]. Under that condition, the number of customers in the orbit explosively increases. However, using an appropriate scaling, we prove that the scaled number of customers in the orbit weakly converges to a diffusion process. The results

allow us to derive an approximation for the distribution of the number of customers in the orbit.

The rest of the paper is organized as follows. In section 2, we present the model in detail. Section 3 presents the first order asymptotics while second order asymptotic analysis is presented in Section 4. An approximation of the distribution of the stationary number of customers in the orbit is presented in Section 5. Concluding remarks are presented in Section 6.

2. Mathematical model

We consider multiserver retrial queueing system with two incoming streams of primary users and secondary users. We assume that arrival flows to the system follow the stationary Poisson processes with intensity λ_1 and λ_2 , respectively. The first arrival flow of primary users has preemptive priority over that of secondary users. The system has N servers. Service times of customers of the first and the second flows are exponentially distributed with rate μ_1 and μ_2 , respectively. If upon arrival, the customer (of either the first or the second flow) finds free servers, then he gets on one of these free servers and starts the service. If all servers are busy with priority customers, the incoming primary customer is lost otherwise, the primary customer will preempt a non-priority customer and the preempted customer joins the orbit. Non-priority customers that see all the servers busy upon arrival join the orbit. From the orbit non-priority customers retry in to occupy a free server after an exponentially distributed time with rate σ .

Let $n_1(t)$, $n_2(t)$ denote the numbers of servers busy serving priority and non-priority customers at the time t , respectively and let $i(t)$ denote the number of customers in the orbit at the time t .

We can see that three-dimensional random process $\{n_1(t), n_2(t), i(t)\}$ is a continuous time Markov chain.

We can write the probability distribution. The system of Kolmogorov differe $P(n_1, n_2, i, t) = P\{n_1(t) = n_1, n_2(t) = n_2, i(t) = i\}$, $n_1 + n_2 = n, n \leq N$ of the process $\{n_1(t), n_2(t), i(t)\}$ rental equations for the probability distribution $P(n_1, n_2, i, t)$, $n_1 + n_2 = n, n \leq N$ is the following:

$$\frac{\partial P(0, 0, i, t)}{\partial t} = -(\lambda_1 + \lambda_2 + i\sigma)P(0, 0, i, t) + \mu_1 P(1, 0, i, t) + \mu_2 P(0, 1, i, t), \quad n = 0,$$

$$\begin{aligned}
\frac{\partial P(n_1, n_2, i, t)}{\partial t} &= -(\lambda_1 + \lambda_2 + \mu_1 n_1 + \mu_2 n_2 + i\sigma)P(n_1, n_2, i, t) + \\
&+ \lambda_1 P(n_1 - 1, n_2, i, t) + (i + 1)\sigma P(n_1, n_2 - 1, i + 1, t) + \lambda_2 P(n_1, n_2 - 1, i, t) + \\
&+ \mu_1 (n_1 + 1)P(n_1 + 1, n_2, i, t) + \mu_2 (n_2 + 1)P(n_1, n_2 + 1, i, t), \quad 1 \leq n < N, \\
\frac{\partial P(n_1, n_2, i, t)}{\partial t} &= -(\lambda_1 + \lambda_2 + \mu_1 n_1 + \mu_2 n_2)P(n_1, n_2, i, t) + \lambda_2 P(n_1, n_2, i - 1, t) + \\
&+ \lambda_1 P(n_1 - 1, n_2 + 1, i - 1, t) + \lambda_2 P(n_1, n_2 - 1, i, t) + \lambda_1 P(n_1 - 1, n_2, i, t) + \\
&+ (i + 1)\sigma P(n_1, n_2 - 1, i + 1, t), \quad n = N, n_1 < N, \\
\frac{\partial P(N, 0, i, t)}{\partial t} &= -(\lambda_2 + N\mu_1)P(N, 0, i, t) + \lambda_1 P(N - 1, 0, i, t) + \\
&+ \lambda_1 P(N - 1, 1, i - 1, t) + \lambda_2 P(N, 0, i - 1, t), \quad n_1 = N.
\end{aligned} \tag{1}$$

We introduce partial characteristic function, denoting $j = \sqrt{-1}$

$$H(n_1, n_2, u, t) = \sum_{i=0}^{\infty} e^{jui} P(n_1, n_2, i, t).$$

We rewrite system (1) as

$$\begin{aligned}
\frac{\partial H(0, 0, u, t)}{\partial t} &= -(\lambda_1 + \lambda_2)H(0, 0, u, t) + j\sigma \frac{\partial H(0, 0, u, t)}{\partial u} + \\
&+ \mu_1 H(1, 0, u, t) + \mu_2 H(0, 1, u, t), \quad n = 0, \\
\frac{\partial H(n_1, n_2, u, t)}{\partial t} &= -(\lambda_1 + \lambda_2 + \mu_1 n_1 + \mu_2 n_2)H(n_1, n_2, u, t) + j\sigma \frac{\partial H(n_1, n_2, u, t)}{\partial u} + \\
&+ \lambda_1 H(n_1 - 1, n_2, u, t) - j\sigma e^{-ju} \frac{\partial H(n_1, n_2 - 1, u, t)}{\partial u} + \lambda_2 H(n_1, n_2 - 1, u, t) + \\
&+ \mu_1 (n_1 + 1)H(n_1 + 1, n_2, u, t) + \mu_2 (n_2 + 1)H(n_1, n_2 + 1, u, t), \quad 1 \leq n < N, \\
\frac{\partial H(n_1, n_2, u, t)}{\partial t} &= -(\lambda_1 + \lambda_2 + \mu_1 n_1 + \mu_2 n_2)H(n_1, n_2, u, t) + \lambda_2 e^{ju} H(n_1, n_2, u, t) + \\
&+ \lambda_1 e^{ju} H(n_1 - 1, n_2 + 1, u, t) + \lambda_2 H(n_1, n_2 - 1, u, t) + \lambda_1 H(n_1 - 1, n_2, u, t) - \\
&- j\sigma \frac{\partial H(n_1, n_2 - 1, u, t)}{\partial u}, \quad n = N, n_1 < N, \\
\frac{\partial H(N, 0, u, t)}{\partial t} &= -(\lambda_2 + N\mu_1)H(N, 0, u, t) + \lambda_1 H(N - 1, 0, u, t) + \\
&+ \lambda_1 e^{ju} H(N - 1, 1, u, t) + \lambda_2 e^{ju} H(N, 0, u, t), \quad n_1 = N.
\end{aligned} \tag{2}$$

The main contribution of the paper is to find the probability distribution of the number of customers in the orbit and to find the stationary probability distribution of the states of the servers, i.e., the number n_1 of servers busy with serving priority customers and the number n_1 of servers busy with servicing non-priority customers and to find the blocking probability of priority customers.

System of equations (2) can not be solved in a direct way, so we will solve under condition of long delay customers in the orbit, i.e., when $\sigma \rightarrow 0$.

3. First stage of the asymptotic-diffusion analysis

In system of equations (2), we introduce the following replacements:

$$\sigma = \varepsilon, \quad \tau = \varepsilon t, \quad u = \varepsilon w, \quad H(n_1, n_2, u, t) = F(n_1, n_2, w, \tau, \varepsilon).$$

We can transform system (2) to the form

$$\begin{aligned} \varepsilon \frac{\partial F(0, 0, u, t)}{\partial \tau} &= -(\lambda_1 + \lambda_2)F(0, 0, w, \tau, \varepsilon) + j \frac{\partial F(0, 0, w, \tau, \varepsilon)}{\partial w} + \\ &+ \mu_1 F(1, 0, w, \tau, \varepsilon) + \mu_2 F(0, 1, w, \tau, \varepsilon), \quad n = 0, \\ \frac{\partial F(n_1, n_2, w, \tau, \varepsilon)}{\partial \tau} &= -(\lambda_1 + \lambda_2 + \mu_1 n_1 + \mu_2 n_2)F(n_1, n_2, w, \tau, \varepsilon) + \\ &+ \lambda_1 F(n_1 - 1, n_2, w, \tau, \varepsilon) + \frac{\partial F(n_1, n_2, w, \tau, \varepsilon)}{\partial w} - \\ &- j e^{-j\varepsilon w} \frac{\partial F(n_1, n_2 - 1, w, \tau, \varepsilon)}{\partial w} + \lambda_2 F(n_1, n_2 - 1, w, \tau, \varepsilon) + \\ &+ \mu_1 (n_1 + 1)F(n_1 + 1, n_2, w, \tau, \varepsilon) + \mu_2 (n_2 + 1)F(n_1, n_2 + 1, w, \tau, \varepsilon), \quad 1 \leq n < N, \\ \frac{\partial F(n_1, n_2, w, \tau, \varepsilon)}{\partial \tau} &= -(\lambda_1 + \lambda_2 + \mu_1 n_1 + \mu_2 n_2)F(n_1, n_2, w, \tau, \varepsilon) + \\ &+ \lambda_2 e^{j\varepsilon w} F(n_1, n_2, w, \tau, \varepsilon) + \lambda_1 e^{j\varepsilon w} F(n_1 - 1, n_2 + 1, w, \tau, \varepsilon) + \\ &+ \lambda_2 F(n_1, n_2 - 1, w, \tau, \varepsilon) + \lambda_1 F(n_1 - 1, n_2, w, \tau, \varepsilon) - \\ &- j \frac{\partial F(n_1, n_2 - 1, w, \tau, \varepsilon)}{\partial w}, \quad n = N, \quad n_1 < N, \end{aligned} \quad (3)$$

$$\begin{aligned} \frac{\partial F(N, 0, w, \tau, \varepsilon)}{\partial \tau} &= -(\lambda_2 + N\mu_1)F(N, 0, w, \tau, \varepsilon) + \lambda_1 F(N - 1, 0, w, \tau, \varepsilon) + \\ &+ \lambda_1 e^{j\varepsilon w} F(N - 1, 1, w, \tau, \varepsilon) + \lambda_2 e^{j\varepsilon w} F(N, 0, w, \tau, \varepsilon), \quad n_1 = N. \end{aligned}$$

We denote $F(n_1, n_2, w, \tau) = \lim_{\varepsilon \rightarrow 0} F(n_1, n_2, w, \tau, \varepsilon)$.

We state the following theorem.

Theorem 1. *Let $F(n_1, n_2, w, \tau)$ denote the solution of the system equation (3) by taking the limit as $\varepsilon \rightarrow 0$. Then function $F(n_1, n_2, w, \tau)$ has the following form:*

$$F(n_1, n_2, w, \tau) = \exp\{jw\tau\} R(n_1, n_2, x).$$

Here the probabilities $R(n_1, n_2, x)$ are a solution to the system of equations

$$-(\lambda_1 + \lambda_2 + x)R(0, 0, x) + \mu_1 R(1, 0, x) + \mu_2 R(0, 1, x) = 0:$$

for $n = 0$:

$$\begin{aligned} & -(\lambda_1 + \lambda_2 + \mu_1 n_1 + \mu_2 n_2 + x)R(n_1, n_2, x) + \lambda_2 R(n_1, n_2 - 1, x) \\ & + \lambda_1 R(n_1 - 1, n_2, x) + xR(n_1, n_2 - 1, x) + \mu_1 (n_1 + 1)R(n_1 + 1, n_2, x) + \\ & + \mu_2 (n_2 + 1)R(n_1, n_2 + 1, x) = 0; \end{aligned}$$

for $1 \leq n < N$:

$$\begin{aligned} & -(\lambda_1 + \mu_1 n_1 + \mu_2 n_2)R(n_1, n_2, x) + (\lambda_2 + x)R(n_1, n_2 - 1, x) + \\ & + \lambda_1 R(n_1 - 1, n_2 + 1, x) + \lambda_1 R(n_1 - 1, n_2, x) = 0; \end{aligned}$$

for $n = N, n_1 < N$:

$$-N\mu_1 R(N, 0, x) + \lambda_1 (R(N - 1, 0, x) + R(N - 1, 1, x)) = 0; \quad (4)$$

for $n_1 = N$ and $x(\tau)$ is a solution of the ordinary differential equation

$$x'(\tau) = -x(\tau) \sum_{n \leq N-1} R(n_1, n_2, x) + (\lambda_1 + \lambda_2) \sum_{n=N} R(n_1, n_2, x) - \lambda_1 R(N, 0, x). \quad (5)$$

Taking into account that $\sum_{n_1+n_2=n \leq N} R(n_1, n_2, x) = 1$, we denote

$$r(x) = \sum_{n < N} R(n_1, n_2, x), \quad 1 - r(x) = \sum_{n=N} R(n_1, n_2, x).$$

Denoting the right-hand side of equation (5) as $a(x)$, we can write following equality

$$a(x) = -xr(x) + (\lambda_1 + \lambda_2)(1 - r(x)) - \lambda_1 R(N, 0, x). \quad (6)$$

In stationary regime, equation (5) transforms to the following equation

$$-xr(x) + (\lambda_1 + \lambda_2)(1 - r(x)) - \lambda_1 R(N, 0, x) = 0.$$

Here x is the point of rest. Let us denote the solution of this equation as κ . Substituting $x = \kappa$ into the system equation (4), we obtain stationary probabilities $R(n_1, n_2, \kappa) = R(n_1, n_2)$ of the states of the servers. Probability $R(N, 0)$ is the blocking probability of priority customers.

4. Second stage of the asymptotic-diffusion analysis

In the system equation (2) let us denote

$$H(n_1, n_2, u, t) = e^{\frac{u}{\sigma} x(\sigma t)} H^{(2)}(n_1, n_2, u, t).$$

The functions $H^{(2)}(n_1, n_2, u, t)$ are the partial characteristic functions of the values of centered random process $i(t) - \frac{x(\sigma t)}{\sigma}$.

By putting $\sigma = \varepsilon^2$, $\tau = \varepsilon t$, $u = \varepsilon w$, $H^{(2)}(n_1, n_2, u, t) = F^{(2)}(n_1, n_2, w, \tau, \varepsilon)$, we get

$n = 0$:

$$\begin{aligned} & \varepsilon^2 \frac{\partial F^{(2)}(0, 0, u, t)}{\partial \tau} + j\varepsilon w a(x) F^{(2)}(0, 0, w, \tau, \varepsilon) = \\ & = \mu_1 F^{(2)}(1, 0, w, \tau, \varepsilon) + j\varepsilon \frac{\partial F^{(2)}(0, 0, w, \tau, \varepsilon)}{\partial w} + \mu_2 F^{(2)}(0, 1, w, \tau, \varepsilon) - \\ & \quad - (\lambda_1 + \lambda_2 + x) F^{(2)}(0, 0, w, \tau, \varepsilon); \end{aligned}$$

$1 \leq n < N$:

$$\begin{aligned} & \varepsilon^2 \frac{\partial F^{(2)}(n_1, n_2, w, \tau, \varepsilon)}{\partial \tau} + j\varepsilon w a(x) F^{(2)}(n_1, n_2, w, \tau, \varepsilon) = \\ & = -(\lambda_1 + \lambda_2 + \mu_1 n_1 + \mu_2 n_2 + x) F^{(2)}(n_1, n_2, w, \tau, \varepsilon) + \lambda_1 F^{(2)}(n_1 - 1, n_2, w, \tau, \varepsilon) + \\ & \quad + j\varepsilon \frac{\partial F^{(2)}(n_1, n_2, w, \tau, \varepsilon)}{\partial w} - j\varepsilon e^{-j\varepsilon w} \frac{\partial F^{(2)}(n_1, n_2 - 1, w, \tau, \varepsilon)}{\partial w} + \\ & \quad + e^{-j\varepsilon w} x F^{(2)}(n_1, n_2 - 1, w, \tau, \varepsilon) + \lambda_2 F^{(2)}(n_1, n_2 - 1, w, \tau, \varepsilon) + \\ & \quad + \mu_1 (n_1 + 1) F^{(2)}(n_1 + 1, n_2, w, \tau, \varepsilon) + \mu_2 (n_2 + 1) F^{(2)}(n_1, n_2 + 1, w, \tau, \varepsilon); \end{aligned}$$

$n = N, n_1 < N$:

$$\varepsilon^2 \frac{\partial F^{(2)}(n_1, n_2, w, \tau, \varepsilon)}{\partial \tau} + j\varepsilon w a(x) =$$

$$\begin{aligned}
&= -(\lambda_1 + \lambda_2 + \mu_1 n_1 + \mu_2 n_2 + x + \lambda_2 e^{j\varepsilon w}) F(n_1, n_2, w, \tau, \varepsilon) - \\
&- j\varepsilon e^{-j\varepsilon w} \frac{\partial F^{(2)}(n_1, n_2 - 1, w, \tau, \varepsilon)}{\partial w} + \lambda_1 F^{(2)}(n_1 - 1, n_2, w, \tau, \varepsilon) + \\
&+ \lambda_1 e^{j\varepsilon w} F^{(2)}(n_1 - 1, n_2 + 1, w, \tau, \varepsilon) + (\lambda_2 + e^{-j\varepsilon w} x) F^{(2)}(n_1, n_2 - 1, w, \tau, \varepsilon); \quad (7)
\end{aligned}$$

$n_1 = N$:

$$\begin{aligned}
&\varepsilon^2 \frac{\partial F^{(2)}(N, 0, w, \tau, \varepsilon)}{\partial \tau} + j\varepsilon w a(x) = \\
&= \lambda_1 F^{(2)}(N - 1, 0, w, \tau, \varepsilon) + \lambda_1 e^{j\varepsilon w} F^{(2)}(N - 1, 1, w, \tau, \varepsilon) - \\
&- (\lambda_2 (1 - e^{j\varepsilon w}) + N\mu_1) F^{(2)}(N, 0, w, \tau, \varepsilon).
\end{aligned}$$

We denote $F^{(2)}(n_1, n_2, w, \tau) = \lim_{\varepsilon \rightarrow 0} F^{(2)}(n_1, n_2, w, \tau, \varepsilon)$.

For the solution of the system of equations (7), we write in the form $F^{(2)}(n_1, n_2, w, \tau) = \Phi(w, \tau) R(n_1, n_2, x)$.

We state the following theorem.

Theorem 2. *Function $\Phi(w, \tau)$ is a solution of differential equation:*

$$\frac{\partial \Phi(w, \tau)}{\partial \tau} = a'(x) w \frac{\partial \Phi(w, \tau)}{\partial w} - \frac{w^2}{2} b(x) \Phi(w, \tau), \quad (8)$$

where $a(x)$ is determined by the equality (6), $b(x)$ has the form

$$b(x) = a(x) + 2 \left(x r(x) - x \sum_{n < N} g(n_1, n_2, x) + (\lambda_1 + \lambda_2) \sum_{n=N} g(n_1, n_2, x) - g(N, 0, x) \right).$$

Here functions $g(n_1, n_2, x)$ are defined by the following system of equations:

$$\begin{aligned}
&-(\lambda_1 + \lambda_2 + x)g(0, 0, x) + \mu_1 g(1, 0, x) + \mu_2 g(0, 1, x) = a(x)R(0, 0, x), \quad n = 0; \\
&-(\lambda_1 + \lambda_2 + \mu_1 n_1 + \mu_2 n_2 + x)g(n_1, n_2, x) + (\lambda_2 + x)g(n_1, n_2 - 1, x) + \\
&+ \lambda_1 g(n_1 - 1, n_2, x) + \mu_1 (n_1 + 1)g(n_1 + 1, n_2, x) + \mu_2 (n_2 + 1)g(n_1, n_2 + 1, x) = \\
&= a(x)R(n_1, n_2, x) + xR(n_1, n_2 - 1, x), \quad 1 \leq n < N; \\
&-(\lambda_1 + \mu_1 n_1 + \mu_2 n_2)g(n_1, n_2) + (\lambda_2 + x)g(n_1, n_2 - 1, x) + \lambda_1 g(n_1 - 1, n_2 + 1, x) + \\
&+ \lambda_1 g(n_1 - 1, n_2, x) = (a(x) - \lambda_2)R(n_1, n_2, x) - \\
&- \lambda_1 R(n_1 - 1, n_2 + 1, x) + xR(n_1, n_2 - 1, x), \quad n = N, n_1 < N;
\end{aligned}$$

$$\begin{aligned}
 & -N\mu_1 g(N, 0, x) + \lambda_1 (g(N-1, 0, x) + g(N-1, 1, x)) = \\
 & = a(x)R(N, 0, x) - \lambda_1 R(N-1, 1, x) - \lambda_2 R(N, 0, x), \quad n_1 = N.
 \end{aligned}$$

Applying the inverse Fourier transform in the system (8), we obtain Fokker-Plank equation for probability density function

$$P(y, \tau) = \frac{\partial P\{y(\tau) < y\}}{\partial y} \text{ of process } y(\tau) :$$

$$\frac{\partial P(y, \tau)}{\partial \tau} = -\frac{\partial \{ya'(x)P(y, \tau)\}}{\partial y} + \frac{1}{2} \frac{\partial^2 \{b(x)P(y, \tau)\}}{\partial y^2}.$$

Hence, stochastic process $y(\tau)$ is a diffusion process with drift coefficient $ya'(x)$ and diffusion coefficient $b(x)$.

5. Third stage of the asymptotic-diffusion analysis

We consider stochastic process: $y(\tau) = \lim_{\sigma \rightarrow 0} \sqrt{\sigma} \left(i(\sigma\tau) - \frac{x(\tau)}{\sigma} \right)$.

Let us consider the following stochastic process: $z(\tau) = x(\tau) + \sqrt{\sigma}y(\tau)$.

We consider stationary probability density function for the process $z(\tau)$:

$$\pi(z) = \frac{\partial P\{z(\tau) < z\}}{\partial z}.$$

Let us state the following statement.

Theorem 3. *Stationary probability density function of the process $z(\tau)$ has the form*

$$\pi(z) = \frac{C}{b(z)} \exp \left\{ \frac{2}{\sigma} \int_0^z \frac{a(x)}{b(x)} dx \right\},$$

where C is normalizing constant.

We build the approximation of the probability distribution of the number of customers in the orbit using expression

$$Pd(i) = \frac{\pi(i\sigma)}{\sum_{n=0}^{\infty} \pi(n\sigma)}.$$

6. Numerical example

We consider particular case $N = 2$ servers.

Table 1

The model of parameters

	The rate of arrival flow, $\lambda_n, n = 1, 2.$	The service rate, $\mu_n, n = 1, 2.$	The rate of a delay in the orbit
Primary flow	1	2	–
Secondary flow	2	3	0.1

The graph of a diffusion approximation $Pd(i)$ of the probability distribution of the number of customers in the orbit is given in Fig. 1.

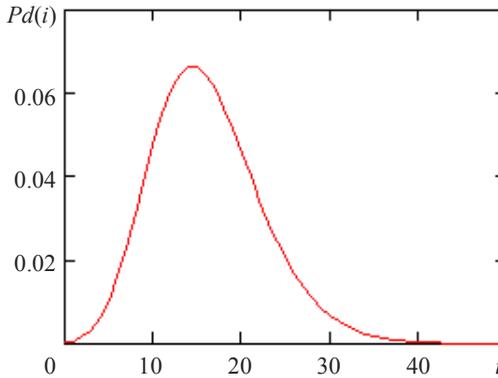


Fig. 1. Graph of a approximation $Pd(i)$

Conclusions

In this paper, we considered a multiserver retrial queueing system with priority customers. Using method asymptotic-diffusion analysis under limit condition of long delay customers in the orbit, we proved that the scaled number of customers in the orbit converges to a diffusion process which was used to build an approximation for stationary probability distribution of the number of customers in the orbit. The blocking probability of priority customers was obtained.

REFERENCES

1. *Akutsu K., Phung-Duc T.* Analysis of retrial queues for cognitive wireless networks with sensing time of secondary users // *Lecture Notes in Computer Science*. 2019. V. 11688. P. 77–91.
2. *Fedorova E., Nazarov A., Moiseev A.* Asymptotic analysis methods for multi-server retrial queueing systems // *Applied Probability and Stochastic Processes*. 2020. P. 159–177.
3. *Moiseev A., Nazarov N., Paul S.* Asymptotic diffusion analysis of multi-server retrial queue with hyper-exponential service // *Mathematics*. 2020. V. 8. No. 4. P. 531–531-16.
4. *Morozov E., Rogozin S., Nguyen Q.H., Phung-Duc T.* Modified Erlang loss system for cognitive wireless networks. Submitted, 2020.
5. *Nazarov A., Phung-Duc T., Paul S., Lizura O.* Asymptotic-diffusion analysis for retrial queue with batch poisson input and multiple types of outgoing calls // *Lecture Notes in Computer Science*. 2019. V. 11965. P. 207–222.
6. *Phung-Duc T., Akutsu K., Kawanishi K., Salameh O., Wittevrongel S.* Queueing models for cognitive wireless networks with sensing time of secondary users. 2020 (submitted).
7. *Sztrik J., Bérczes T., Nemouchi H., Melikov A.Z.* Performance modeling of finitesource cognitive radio networks using simulation // *Communications in Computer and Information Science*. 2016. V. 678. P. 64–73.
8. *Nemouchi H., Sztrik J.* Performance evaluation of finite-source cognitive networks with non-reliable services using simulation // *Proceedings of the 10th International Conference on Applied Informatics Eger, Hungary, January 30–February 1, 2017*. P. 225–234. DOI: 10.14794/ICAI.10.2017.225.

Asymptotic analysis of Markovian retrial queue with unreliable server and two-way communication under low rate of retrials condition *

Anatoly Nazarov¹, Tuan Phung-Duc^{2,3}, Svetlana Paul¹,
Olga Lizyura¹, Ksenia Shulgina¹

¹ Tomsk State University, Tomsk, Russia

² University of Tsukuba, Tsukuba, Japan

³ VNU Vietnam Japan University, Hanoi, Vietnam

Retrial queues with two-way communication arose as models of blended call centers, where the operator can provide both inbound and outbound calls. The research area of retrial queues is a branch of queueing theory and literature on this topic is rich and vast [1].

The main feature of retrial queues is retrial behaviour of the customers: if the operator is busy at the instant of the call, the customers retry to occupy the server after some time [2, 3]. During the retrying process, the customers stay in a virtual waiting room that we call the orbit.

The idea of call blending is to improve the productivity of classical call centers by reducing the idle time of an operator [4]. The unreliability of the server is also of interest as a common phenomenon in communication and service systems [5].

Mathematical model

We consider a single-server retrial queue with two-way communication. We have two classes of calls in the system: incoming calls and outgoing calls. Primary incoming calls form a Poisson process with rate λ . Idle server starts the service immediately. The durations of service times for incoming calls are exponentially distributed with parameter μ_1 . If the server is busy upon arrival, the incoming call joins the orbit and retries to occupy the server after a random delay, whose duration follows the exponential distribution with parameter σ .

In its idle time, the server can make outgoing calls. There are multiple types of outgoing calls in the system. The server makes an outgoing call of

* The publication was financially supported by RFBR according to the research project No. 18-01-00277

type n with rate α_n and serves it for an exponentially distributed time with parameter μ_n . We number the types of outgoing calls from 2 to N .

The unreliability of the server is defined by three parameters: γ_0 is the rate of breakdowns when the server is idle, γ_1 is the rate of breakdowns when an incoming call is in service and γ_2 is the rate of restorations. If at the instant of breakdown, the server provides an incoming call, then the call joins the orbit. We assume that there are no breakdowns during service of outgoing calls, since the server itself initiates the call.

Problem Definition

Let $k(t)$ denote the state of the server at time t as follows: 0, if the server is idle; 1, if the server busy with an incoming call; n , if the server is busy with an outgoing call of type n , $n = 2, \dots, N$; $N+1$, if the server is broken. Let $i(t)$ denote the number of calls in the orbit at instant t .

The two-dimensional process $\{k(t), i(t)\}$ is Markovian. Thus, for the stationary probability distribution of this process

$$P\{k(t) = k, i(t) = i\} = P_k(i, t),$$

we can derive Kolmogorov system of equations in stationary regime

$$\begin{aligned} -\left(\lambda + \sum_{n=2}^N \alpha_n + i\sigma + \gamma_0\right)P_0(i) + \sum_{k=1}^N \mu_k P_k(i) + \gamma_2 P_{N+1}(i) &= 0, \\ -(\lambda + \mu_1 + \gamma_1)P_1(i) + \lambda P_0(i) + \sigma(i+1)P_0(i+1) + \lambda P_1(i-1) &= 0, \\ -(\lambda + \mu_n)P_n(i) + \lambda P_n(i-1) + \alpha_n P_0(i) &= 0, \quad n = \overline{2, N}, \\ -(\lambda + \gamma_2)P_{N+1}(i) + \lambda P_{N+1}(i-1) + \gamma_0 P_0(i) + \gamma_1 P_1(i-1) &= 0. \end{aligned} \quad (1)$$

The problem is to obtain the probability distribution of the number of calls in the orbit. We introduce partial characteristic functions

$$H_k(u) = \sum_{i=0}^{\infty} e^{ju_i} P_k(i), \quad k = \overline{0, N+1},$$

where $j = \sqrt{-1}$.

Rewriting the system (1) for the partial characteristic functions, we obtain

$$-\left(\lambda + \sum_{n=2}^N \alpha_n + \gamma_0\right)H_0(u) + j\sigma H_0'(u) + \sum_{k=1}^N \mu_k H_k(u) + \gamma_2 H_{N+1}(u) = 0,$$

$$\begin{aligned}
& (\lambda(e^{ju} - 1) - \mu_1 - \gamma_1)H_1(u) + \lambda H_0(u) - j\sigma e^{-ju} H'_0(u) = 0, \\
& (\lambda(e^{ju} - 1) - \mu_n)H_n(u) + \alpha_n \overline{H}_n(u) = 0, \quad n = \overline{2, N}, \\
& (\lambda(e^{ju} - 1) - \gamma_2)H_{N+1}(u) + \gamma_0 H_0(u) + \gamma_1 e^{ju} H_1(u) = 0. \tag{2}
\end{aligned}$$

Summing up equations of the system (2), we write the additional equation

$$j\sigma e^{-ju} H'_0(u) + (\lambda + \gamma_1)H_1(u) + \lambda \sum_{n=2}^{N+1} H_n(u) = 0, \tag{3}$$

which we will use in further analysis. We will solve the system (2) using asymptotic analysis method under low rate of retrials condition ($\sigma \rightarrow 0$).

Asymptotic analysis under low rate of retrial condition

We denote $\sigma = \varepsilon$ and make the following substitutions in the system (2) and equation (3)

$$u = \varepsilon w, H_k(u) = F_k(w, \varepsilon),$$

in order to obtain the system

$$\begin{aligned}
& -\left(\lambda + \gamma_0 + \sum_{n=2}^N \alpha_n\right)F_0(w, \varepsilon) + j \frac{\partial F_0(w, \varepsilon)}{\partial w} + \sum_{k=1}^N \mu_k F_k(w, \varepsilon) + \gamma_2 F_{N+1}(w, \varepsilon) = 0, \\
& (\lambda(e^{j\varepsilon w} - 1) - \mu_1 - \gamma_1)F_1(w, \varepsilon) + \lambda F_0(w, \varepsilon) - j e^{-j\varepsilon w} \frac{\partial F_0(w, \varepsilon)}{\partial w} = 0, \\
& (\lambda(e^{j\varepsilon w} - 1) - \mu_n)F_n(w, \varepsilon) + \alpha_n \overline{F}_n(w, \varepsilon) = 0, \quad n = \overline{2, N}, \\
& (\lambda(e^{j\varepsilon w} - 1) - \gamma_2)F_{N+1}(w, \varepsilon) + \gamma_0 F_0(w, \varepsilon) + \gamma_1 e^{j\varepsilon w} F_1(w, \varepsilon) = 0, \\
& j e^{-j\varepsilon w} \frac{\partial F_0(w, \varepsilon)}{\partial w} + (\lambda + \gamma_1)F_1(w, \varepsilon) + \lambda \sum_{n=2}^{N+1} F_n(w, \varepsilon) = 0. \tag{4}
\end{aligned}$$

Solving the system (4) in the limit by $\varepsilon \rightarrow 0$, we prove Theorem 1.

Theorem 1. *Suppose that $i(t)$ is the number of calls in the orbit in Markovian retrial queue with unreliable server and two-way communication, then the following limit equality holds:*

$$\lim_{\sigma \rightarrow 0} E \exp\{jw\sigma i(t)\} = \exp\{jw\kappa_1\},$$

where κ_1 is given by

$$\kappa_1 = \lambda \frac{\gamma_2(\lambda + \gamma_1) + \lambda\gamma_1 + \gamma_0(\mu_1 + \gamma_1) + \gamma_2(\mu_1 + \gamma_1) \sum_{n=2}^N \frac{\alpha_n}{\mu_n}}{\gamma_2\mu_1 - \lambda(\gamma_1 + \gamma_2)}.$$

The value of κ_1 defines the asymptotic mean κ_1/σ of the number of calls in the orbit. On the next stage of analysis, we will obtain the characteristic function of the process $i(t)$.

We introduce the following notations in the system (2) and equation (3)

$$H_k(u) = \exp\left(ju \frac{\kappa_1}{\sigma}\right) H_k^{(2)}(u), \quad k = \overline{0, N+1},$$

to derive the system of equations

$$\begin{aligned} & -\left(\lambda + \sum_{n=2}^N \alpha_n + \gamma_0 + \kappa_1\right) H_0^{(2)}(u) + j\sigma \frac{dH_0^{(2)}(u)}{du} + \sum_{k=1}^N \mu_k H_k^{(2)}(u) + \gamma_2 H_{N+1}^{(2)}(u) = 0, \\ & (\lambda(e^{ju} - 1) - \mu_1 - \gamma_1) H_1^{(2)}(u) + \lambda H_0^{(2)}(u) + \kappa_1 e^{-ju} H_0^{(2)}(u) - j\sigma e^{-ju} \frac{dH_0^{(2)}(u)}{du} = 0, \\ & (\lambda(e^{ju} - 1) - \mu_n) H_n^{(2)}(u) + \alpha_n H_0^{(2)}(u) = 0, \quad n = \overline{2, N}, \\ & (\lambda(e^{ju} - 1) - \gamma_2) H_{N+1}^{(2)}(u) + \gamma_0 H_0^{(2)}(u) + \gamma_1 e^{ju} H_1^{(2)}(u) = 0, \\ & -\kappa_1 e^{-ju} H_0^{(2)}(u) + j\sigma e^{-ju} \frac{dH_0^{(2)}(u)}{du} + (\lambda + \gamma_1) H_1^{(2)}(u) + \lambda \sum_{n=2}^{N+1} H_n^{(2)}(u) = 0. \quad (5) \end{aligned}$$

Then we make the substitutions

$$\sigma = \varepsilon^2, \quad u = \varepsilon w, \quad H_k^{(2)}(u) = F_k^{(2)}(w, \varepsilon), \quad k = \overline{0, N+1},$$

and obtain the system of equations

$$\begin{aligned} & -\left(\lambda + \sum_{n=2}^N \alpha_n + \gamma_0 + \kappa_1\right) F_0^{(2)}(w, \varepsilon) + j\varepsilon \frac{\partial F_0^{(2)}(w, \varepsilon)}{\partial w} + \sum_{k=1}^N \mu_k F_k^{(2)}(w, \varepsilon) + \\ & \quad + \gamma_2 F_{N+1}^{(2)}(w, \varepsilon) = 0, \\ & (\lambda(e^{j\varepsilon w} - 1) - \mu_1 - \gamma_1) F_1^{(2)}(w, \varepsilon) + \lambda F_0^{(2)}(w, \varepsilon) + \kappa_1 e^{-j\varepsilon w} F_0^{(2)}(w, \varepsilon) - \\ & \quad - j\varepsilon e^{-j\varepsilon w} \frac{\partial F_0^{(2)}(w, \varepsilon)}{\partial w} = 0, \end{aligned}$$

$$\begin{aligned}
& (\lambda(e^{j\varepsilon w} - 1) - \mu_n)F_n^{(2)}(w, \varepsilon) + \alpha_n F_0^{(2)}(w, \varepsilon) = 0, \quad n = \overline{2, N}, \\
& (\lambda(e^{j\varepsilon w} - 1) - \gamma_2)F_{N+1}^{(2)}(w, \varepsilon) + \gamma_0 F_0^{(2)}(w, \varepsilon) + \gamma_1 e^{j\varepsilon w} F_1^{(2)}(w, \varepsilon) = 0, \\
& -\kappa_1 e^{-j\varepsilon w} F_0^{(2)}(w, \varepsilon) + j\varepsilon e^{-j\varepsilon w} \frac{\partial F_0^{(2)}(w, \varepsilon)}{\partial w} + (\lambda + \gamma_1)F_1^{(2)}(w, \varepsilon) + \\
& \quad + \lambda \sum_{n=2}^{N+1} F_n^{(2)}(w, \varepsilon) = 0.
\end{aligned} \tag{7}$$

Considering the system (7) in the limit by $\varepsilon \rightarrow 0$, we prove Theorem 2.

Theorem 2 *In the context of Theorem 1, the following limit equality is true:*

$$\lim_{\sigma \rightarrow 0} M \exp \left\{ jw\sqrt{\sigma} \left(i(t) - \frac{\kappa_1}{\sigma} \right) \right\} = \exp \left\{ \frac{jw^2}{2} \kappa_2 \right\},$$

where

$$\begin{aligned}
\kappa_2 = & \kappa_1 + \frac{\lambda}{\gamma_2 \mu_1 - \lambda(\gamma_1 + \gamma_2)} \times \\
& \times \left[(\gamma_2(\lambda + \gamma_1) + \gamma_1(\lambda + \mu_1 + \gamma_1)) \frac{\lambda + \kappa_1}{\mu_1 + \gamma_1} + \lambda \gamma_2 (\mu_1 + \gamma_1) \sum_{n=2}^N \frac{\alpha_n}{\mu_n^2} + \right. \\
& \left. + \frac{\lambda}{\gamma_2} (\gamma_0(\mu_1 + \gamma_1) + \gamma_1(\lambda + \kappa_1)) \right].
\end{aligned}$$

Theorem 2 defines the asymptotic variance κ_2/σ of the number of calls in the orbit. Thus, we can conclude that asymptotic probability distribution of the process $i(t)$ is Gaussian with mean κ_1/σ and variance κ_2/σ .

Conclusions

We studied the number of calls in the orbit in Markovian retrial queue with unreliable server and two-way communication. We have obtained the steady-state characteristic function of this process. Asymptotic probability distribution of the process is Gaussian with mean κ_1/σ and variance κ_2/σ . The values of κ_1 and κ_2 are given in Theorem 1 and Theorem 2, respectively.

REFERENCES

1. *Artalejo J.* Accessible bibliography on retrial queues: progress in 2000–2009 // *Mathematical and computer modelling.* 2010. V. 51. No. 9-10. P. 1071–1081.
2. *Aguir S., Karaesmen F., Aksin O., Chauvet F.* The impact of retrials on call center performance // *OR Spectrum.* 2004. V. 26. No. 3. P. 353–376.
3. *Bernett H., Fischer M., Masi D.* Blended call center performance analysis // *IT Professional.* 2002. V. 4. No. 2. P. 33–38.
4. *Aissani A., Phung-Duc T.* Optimal analysis for M/G/1 retrial queue with two-way communication // *Lecture Notes in Computer Science.* 2015. V. 9081. P. 1–14.
5. *Kumar M., Dadlani A., Kim K.* Performance analysis of an unreliable M/G/1 retrial queue with two-way communication // *Operational Research.* 2018. P. 1–14.

Asymptotic waiting time analysis of a M/M/1 retrial queueing system

Anatoly Nazarov, Maria Samorodova

Tomsk State University, Tomsk, Russia

The aim of our paper is to present the waiting time analysis of a M/M/1 retrial queueing system by the use of asymptotic analysis method under a heavy load condition. The analysis was carried out using the distribution of the number of returns of the tagged customer to the orbit, which was also obtained.

Mathematical model

Let us consider a M/M/1 retrial queueing (RQ) system. The system input receives a Poisson flow of requests which is given by a scalar intensity λ . If the server is idle, an incoming request occupies the server; the service of this request starts immediately and the service time has exponential distribution with parameter μ . Served request leaves the system. If the server is busy, an incoming request joins to the orbit and after a random delay, that has exponential distribution with parameter σ , each request generates a retrial attempt in order to capture the server. If the server is idle at the moment of the retrial, this request occupies the server for a random service time; otherwise, it instantly returns to the orbit for a next random delay.

We assume the system being in stationary mode. Let's define W – waiting time of the tagged customer in the orbit as the length of the interval from the moment the request arrives in the system till the start of the service. Let's denote by v the number of transitions of the tagged customer to the orbit. Also we denote by r the probability that the server is busy at the moment the customer arrives at the system. Obviously, $v=0$ with the probability $(1-r)$, that the customer finds the server idle at the moment of the arrival to the system. In addition, we denote by $v(t)$ the number of returns of the tagged customer to the orbit from the moment t until the start of the service. Using above notations:

$$v = \begin{cases} 0, & \text{with probability } (1-r), \\ 1 + v(t), & \text{with probability } r. \end{cases}$$

The characteristic function for W can be written as follows

$$\begin{aligned}
 G(u) &= E\{e^{juW}\} = (1-r) + r \sum_{n=0}^{\infty} E\{e^{juW} / v = 1+n\} P\{v(t) = n\} = \\
 &= (1-r) + r \sum_{n=0}^{\infty} \left(\frac{\sigma}{\sigma - ju} \right)^{1+n} P\{v(t) = n\}. \quad (1)
 \end{aligned}$$

The aim of our study is to find the asymptotic distribution of W the waiting time of the tagged customer in the orbit. As can be seen from (1), for that purpose it is enough to find the probability r and the probability distribution $P\{v(t) = n\}$.

Kolmogorov's equations

Steady-state characteristic functions for $v(t)$ can be written in the following form:

$$G(u) = E\{e^{juv(t)}\} = \sum_{i=0}^{\infty} [G_0(i, u)P_0(i) + G_1(i, u)P_1(i)].$$

Let's consider:

$$G_k(i, u, t) = E\{e^{juv(t)} / k(t) = k, i(t) = i\},$$

where $G_k(i, u, t)$ – conditional characteristic functions for $v(t)$, $P_k(i)$ – stationary probability distribution of a markov chain $\{k(t), i(t)\}$ that describes the system state at time t ; $i(t)$ is the number of customers in the orbit at time t and $k(t)$ – the state of the server at time t .

$$k(t) = \begin{cases} 0, & \text{if the server is idle,} \\ 1, & \text{if the server is busy.} \end{cases}$$

Taking into account that the system is in a stationary mode, for $G_k(i, u)$ we obtain the system of inverse Kolmogorov equations:

$$-(\lambda + i\sigma)G_0(i, u) + \lambda G_1(i, u) + (i-1)\sigma G_1(i-1, u) + \sigma = 0, \quad (2)$$

$$-(\lambda + \mu + \sigma)G_1(i, u) + \mu G_0(i, u) + \lambda G_1(i+1, u) + e^{ju}\sigma G_1(i, u) = 0. \quad (3)$$

**Asymptotic analysis of the number of returns
of the tagged customer to the orbit**

Denote $\rho = \frac{\lambda}{\mu}$, $\varepsilon = 1 - \rho$, making substitutions $u = \varepsilon w$, $i\varepsilon = x$,

$G_k(i, u) = g_k(x, w, \varepsilon)$ and multiplying (2) by ε we obtain:

$$\begin{aligned} & -(\varepsilon\lambda + x\sigma)g_0(x, w, \varepsilon) + \varepsilon\lambda g_1(x, w, \varepsilon) + \\ & \quad + (x - \varepsilon)\sigma g_1(x - \varepsilon, w, \varepsilon) + \varepsilon\sigma = 0, \\ & -(\lambda + \mu + \sigma(1 + e^{j\varepsilon w}))g_1(x, w, \varepsilon) + \\ & \quad + \mu g_0(x, w, \varepsilon) + \lambda g_1(x + \varepsilon, w, \varepsilon) = 0. \end{aligned} \quad (4)$$

Denote $\lim_{\varepsilon \rightarrow 0} g_k(x, w, \varepsilon) = g_k(x, w)$. Passing to the limit $\varepsilon \rightarrow 0$ in (4) we obtain the following system for $g_k(x, w)$:

$$\begin{aligned} & -x\sigma g_0(x, w) + x\sigma g_1(x, w) = 0, \\ & -(\mu + \sigma)g_1(x, w) + (\mu + \sigma)g_0(x, w) = 0. \end{aligned} \quad (5)$$

This system consist of two equivalent equations and $g_0(x, w) = g_1(x, w) = g(x, w)$.

Let us write the solution $g_k(x, w, \varepsilon)$ of the system (4) in the form of a decomposition:

$$g_k(x, w, \varepsilon) = g(x, w) + \varepsilon f_k(x, w) + o(\varepsilon), \quad (6)$$

and rewrite (4) as follows

$$\begin{aligned} & -(\varepsilon\lambda + x\sigma)g_0(x, w, \varepsilon) + (\varepsilon\lambda + x\sigma)g_1(x, w, \varepsilon) - \\ & \quad - \varepsilon \frac{\partial [x\sigma g_1(x, w, \varepsilon)]}{\partial x} + \varepsilon\sigma = O(\varepsilon^2), \\ & -(\mu + \sigma)g_1(x, w, \varepsilon) + \mu g_0(x, w, \varepsilon) + \\ & \quad + \varepsilon \frac{\partial [\lambda g_1(x, w, \varepsilon)]}{\partial x} + e^{j\varepsilon w} \sigma g_1(x, w, \varepsilon) = O(\varepsilon^2). \end{aligned} \quad (7)$$

Substituting decomposition (6) in (7) and after performing some actions on the equations of the system, adding the equations in the system (7) and passing to the limit $\varepsilon \rightarrow 0$ we get the following system:

$$\begin{aligned}
x[f_1(x, w) - f_0(x, w)] &= g(x, w) + x \frac{\partial g(x, w)}{\partial x} - 1, \\
[f_1(x, w) - f_0(x, w)] &= jw \frac{\sigma}{\mu} g(x, w) + \frac{\partial g(x, w)}{\partial x} - \frac{\sigma}{\mu} f_1(x, w), \\
\left(x - \frac{\mu}{\sigma}\right)[f_1(x, w) - f_0(x, w)] &= (1 - jw)g(x, w) + \left(x - \frac{\mu}{\sigma}\right) \frac{\partial g(x, w)}{\partial x} - 1.
\end{aligned}$$

Eliminating $f_1(x, w), f_0(x, w)$ from this system, we obtain the equation for the function $g(x, w)$, solving which we find:

$$g(x, w) = \frac{\frac{\mu}{x\sigma}}{\frac{\mu}{x\sigma} - jw}, \quad (8)$$

where $g(x, w)$ is the conditional characteristic function of the asymptotic normalized number of returns of the tagged customer to the orbit under condition that the asymptotic number of customers in the orbit takes the value x . As shown in [4], $\gamma(x)$ the asymptotic probability density of the number of customers in the orbit is the density of the gamma distribution with parameters $\beta = \frac{\mu + \sigma}{\sigma}$ and $\alpha = 1$.

Let's pass from the conditional characteristic function $g(x, w)$ to the characteristic function $g(w)$:

$$g(w) = \int_0^{+\infty} g(x, w) \gamma(x) dx.$$

It is easy to show that the inverse Fourier transform has the form of the probability distribution density of the limiting value of $v(t)$:

$$f(z) = \int_0^{\infty} \gamma(x) \frac{\mu}{x\sigma} \exp\left\{-\frac{\mu}{x\sigma} z\right\} dx. \quad (9)$$

Asymptotic distribution of the waiting time in the orbit

Using the found distribution density (9), let's get an approximation of the discrete probability distribution of the number of returns of the tagged

customer to the orbit:

$$P_1(n) = f((1-\rho)n) \cdot \left(\sum_{m=0}^{\infty} f((1-\rho)m) \right)^{-1}.$$

Substituting above distribution into (1), we obtain:

$$\begin{aligned} G(u) &= E\{e^{juW}\} = (1-r) + r \sum_{n=0}^{\infty} E\{e^{juW} / v = 1+n\} P\{v(t) = n\} = \\ &= (1-r) + r \sum_{n=0}^{\infty} \left(\frac{\sigma}{\sigma - ju} \right)^{1+n} P_1(n). \end{aligned}$$

Thus, we have found the limiting characteristic function of the waiting time of the customer in the RQ system M/M/1 under a heavy load condition. Applying the inverse Fourier transform of the obtained $G(u)$, we find the asymptotic distribution of the waiting time of the customer in the orbit.

Conclusions

In this paper was presented an analysis of the waiting time and the number of returns of a M/M/1 retrial queueing system. The research was conducted using asymptotic analysis method under a heavy load condition.

REFERENCES

1. *Artalejo J.R., Falin J.I.* Standard and retrial queueing systems: A comparative analysis // *Revista Matematica Complutense*. 2002. V. 15. P. 101–129.
2. *Gomez-Corral A., Ramalhoto M.* On the waiting time distribution and the busy period of a retrial queue with constant retrial rate // *Stochastic Modelling and Applications*. 2000. V. 3. P. 37–47.
3. *Lee S.W., Kim B., Kim J.* Analysis of the waiting time distribution in M/G/1 retrial queues with two way communication // *Annals of Operations Research*. 2020. (Accepted/In press). DOI: 10.1007/s10479-020-03717-2.
4. *Moiseeva E., Nazarov A.* Asymptotic analysis of RQ-systems M/M/1 on heavy load condition // *IV International Conference Problems of Cybernetics and Informatics (PCI)*. 2012. P. 1–3.
5. *Nobel R., Tijms H.* Waiting-time probabilities in the M/G/1 retrial queue // *Statistica Neerlandica*. 2006. V. 60. No. 3. P. 73–78.
6. *Sudyko E., Nazarov A., Sztrik J.* Asymptotic waiting time analysis of a finite-source M/M/1 retrial queueing system // *Probability in the Engineering and Informational Sciences*. 2018. V. 33. No. 3. P. 387–403.

Simulation analysis in cognitive radio networks with unreliability and abandonment*

Hamza Nemouchi, Mohamed Hedi Zaghouni, János Sztrik

*The University of Debrecen, Doctoral School of Informatics, Faculty of Informatics,
Debrecen, Hungary*

The current paper presents a Cognitive Radio Network with impatient customers and unreliable servers by the help of a finite-source retrial queueing system. Our queueing system consists of two not independent, interconnected sub-systems. The first part is allocated to primary requests, with N_1 the number of sources. These sources will be responsible for generating a high priority requests with an inter-request and service times exponentially distributed, using parameters λ_1 and μ_1 , respectively. All the produced requests are directed to a single server unit (PCS) with a FIFO queue.

The second subsystem is devoted to the low-priority requests with number of sources denoted by N_2 , the inter-arrival times and service times in this subsystem are assumed to be exponentially distributed as well, with parameter λ_2 and μ_2 , respectively. Based on the state of both server (idle or busy), the generated primary packet goes to the primary server (if the server is idle) or joins the FIFO queue (if it is busy with a PU). However, if an unlicensed user occupies the PCS, its service is instantly stopped, and will be sent back to the Secondary unit.

Depending on the secondary unit's availability, the aborted task is addressed either to the server or the retrial queue from which retries to get served from the beginning after an exponentially distributed time with parameter ν .

On the other hand, requests from SUs are directed to SCS. If it is idle, the service begins. If not, this unlicensed task will sens the PCS. In case of an idle status for PCS, this service may opportunistically join the high priority channel. If the PCS is engaged, the request goes to orbit. More details can be found in [1–3].

It should be noted that Secondary Users in orbit are obliged to leave the system once their total waiting time exceeds a random abandonment time

* The research work of János Sztrik is supported by the EFOP-3.6.1-16-2016-00022 project. The project is co-financed by the European Union and the European Social Fund. Mohamed Hedi Zaghouni is supported by the Stipendium Hungaricum Scholarship.

generally distributed (Hyper, Hypo, Gamma, Lognormal and Pareto) using τ as a parameter. A random breakdowns and repairs will occur on the secondary server using the mentioned-above distributions with parameters γ_2 and σ_2 , respectively.

Several studies have examined the Abandonment and/or Unreliability on the basis of different scenarios and systems.

At [6] as an example, the authors have presented a retrial queueing system with a single server which is subject to random breakdowns and assuming that collisions may occur when a customer arrives at a busy server which forces both jobs to join the orbit.

However, to get closer to real-life situations and involving more servers to the system, the authors of [7] examined the abandonment concept on a Cognitive Radio Network by setting a constant value for the maximum waiting time (abandonment time) of secondary users. In an extended work [8], the same authors of the above-mentioned paper assumed that the abandonment time is random, using various distributions to investigate their influence on the main performance measures of such a system.

Other probes analysed the abandonment in other types of networks and showed that customers can leave systems from queues, server units while receiving services and while waiting; more details are given in [4, 5]. However, in the current paper, we assume that impatient users (secondary) are forced to leave the system only from the orbit while waiting.

Unreliability of servers was investigated in [9, 10], without taking in consideration that customers have the opportunity to leave the system.

Several figures will show the effects of the abandonment and unlicensed server unreliability on the performance measures of the system using simulation.

Simulation examples

The batch mean value method was used in the simulation to estimate the average response times of the requests. This method is a common confidence interval technique used for the analysis of the steady-state simulation output. See for more details [1–3].

We have treated several scenarios in order to investigate the possible effect of the maximum waiting time of secondary customers (impatience) on the average and variance estimations of the performance measures when it is exponentially distributed, furthermore, the impact of the distribution of the involved random variables (impatience time) when it is generally distributed in case their squared coefficient of variation is less and greater than one.

These investigations are dealt when primary server is reliable and secondary one is non-reliable (Scenario 1) and when both servers are non-reliable (Scenario 2).

Due to of the limitation of pages, in this short paper we present two figures from scenario 1 that illustrate the impact of general distribution while their $C_x^2 > 1$ on the average arbitrary time of secondary customers.

Table 1

Numerical values of model parameters

Fig No.	N_1	N_2	λ_1	λ_2	μ_1	μ_2	ν	τ	γ_2	σ_2
Fig. 1	100	100	1.5	x-axis	1	1	0.1	0.002	0.1	0.1
Fig. 2	100	100	x-axis	1.5	1	1	0.1	0.002	0.1	0.1

Table 2

Numerical values of distributions parameters, $C_x^2 > 1$

Distribution	Hyper	Gamma	Lognormal	Pareto
Mean	500	500	500	500
Variance	640000	640000	640000	640000
Parameters	$P = 0.078$ $\gamma_1, \theta_1 = 0.078$ $\gamma_2, \theta_2 = 0.09219$	$\alpha = 0.9756$ $\beta = 0.048$	$\sigma = 0.8399$ $m = 0.3527$	$\alpha = 1.4055$ $k = 5.7708$

Table 1 shows the value of input parameters for the simulation program, while Table 2 defines the numerical values of the parameters for the distributions illustrated in Fig. 1 and Fig. 2. They have the same mean and variance knowing that the squared coefficient of variation is greater than one.

Let us put some comments on the obtained results. The average sojourn time of different types of customers depending on the arrival intensity is shown in Fig. 1, which illustrates how the mean response time of an arbitrary customer. This mean value can be determined using the law of total expectation. Differences can be observed, the results clearly illustrate the effect of different distributions. Highest values have experience with Pareto distribution in this case. Despite the increasing arrival intensity, the maximum characteristic of the maximum value of the mean value can be shown.

Figure 2 illustrates the effect of different distributions on the mean response time of secondary arbitrary customers in the function of the primary generation rate λ_1 . The mean sojourn time of arbitrary users depends on the average of impatience, probability of abandonment, and success. Despite the highest value of the mean that is obtained when the secondary waiting time is

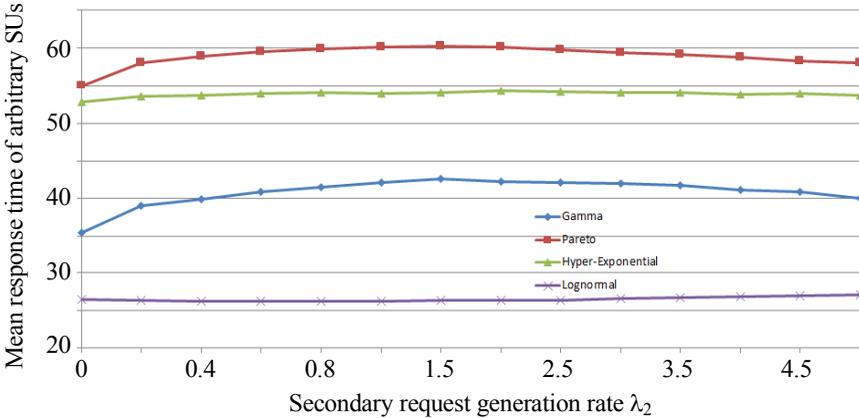


Fig. 1. The effect of the impatience distribution on the mean response time of an arbitrary customer vs secondary request generation rate

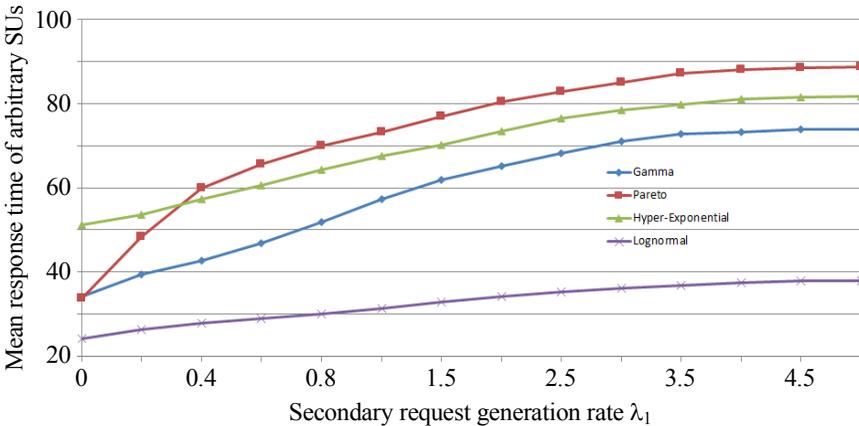


Fig. 2. The effect of the impatience distribution on the mean response time of an arbitrary customer vs primary request generation rate

Pareto distributed, it is clearly seen that when the primary arrival rate is very small (very large inter-arrival time, λ_1 less than 0.4), the distributions have a different impact where we obtain the highest value of the mean when the secondary waiting time is hyper-exponentially distributed. Our Explanation for both figures is the following when the squared coefficient of variation is greater than one, the lognormal distribution gives big values where the

gamma distribution gives small values comparing to Pareto and hyper-exponential. So in the case of gamma distribution customers leave the system more often and stay in the system in the case of the lognormal distribution. Furthermore, when λ_1 is less than one, Fig. 2 shows the characteristic of cognitive technology as more secondary customers access the licensed channel and getting served rather than leave the system.

Conclusions

In this short paper a finite-source retrial queueing system is used to model a cognitive radio network with non-reliable servers and impatient customers. The results obtained have demonstrated the importance of the distribution of the maximum waiting time of a customer (impatient) as it has a big impact on the system performance measures of a such complex system, although the mean and variance of the distributions are the same depending on their squared coefficient of variation.

REFERENCES

1. *Akyildiz I.F., Lee W.Y., Vuran M.C., Mohanty S.* Next generation/dynamic spectrum access/cognitive radio wireless networks // A survey, *Computer networks* 50.13. 2006. P. 2127–2159.
2. *Almási B., Bérczes T., Kuki A., Sztrik J., Wang J.* Performance modeling of finite-source cognitive radio networks // *Acta Cybernetica* 22.3. 2016. P. 617–631.
3. *Gunawardena S., Zhuang W.* Modeling and Analysis of Voice and Data in Cognitive Radio Networks. Springer, 2014.
4. *He Q.M., Zhang H., Ye Q.* An M/PH/K queue with constant impatient time // *Mathematical Methods of Operations Research* 87.1. 2018. P. 139–168.
5. *Ibrahim R.* Managing queueing systems where capacity is random and customers are impatient // *Production and Operations Management* 27.2. 2018. P. 234–250.
6. *Kuki A., Bérczes T., Tóth Á., Sztrik J.* Numerical analysis of finite source Markov retrial system with non-reliable server, collision, and impatient customers // *Annales Mathematicae et Informaticae*. Liceum University Press, 2020. V. 51. P. 53–63.
7. *Zaghouani M.H., Sztrik J.* Performance evaluation of finite-source cognitive radio networks with impatient customers // *Annales Mathematicae et Informaticae*. Liceum University Press, 2020. V. 51. P. 89–99.
8. *Zaghouani M.H., Sztrik J.* Performance simulation of finite-source cognitive radio networks with impatient calls in the orbit // *ISSPSM*. 2020.
9. *Sztrik J., Zaghouani M.H., Uka A.* Reliability analysis of cognitive radio networks // 18th International Conference named after A.F. Terpugov, Information Technologies and Mathematical Modelling, ITMM-2019, Saratov, Russia, 2019.
10. *Nemouchi H., Zaghouani M.H., Sztrik J.* The impact of servers reliability on the characteristics of cognitive radio systems // *The 1st Conference on Information Technology and Data Science CITDS-2020*.

On a MMAP/(PH,PH)/1/(∞ ,N) queueing-inventory system

Nisha Mathew, Varghese Joshua,
Achyutha Krishnamoorthy

Department of Mathematics, CMS College Kottayam, Kerala, India

Inventory models with positive service time (queueing inventory model) were introduced by Sigman and Simchi-Levi in [1]. A survey of inventory with positive service time is given by Krishnamoorthy et al. in [2]. A multi-server queueing system with two types of customers and an infinite buffer is investigated by Dudin et al. in [3]. The arrival flow of customers in [3] is described by a marked Markovian arrival process. A MAP/PH/1 queueing model in which the server is subject to taking vacations is given by C. Sreenivasan et al. in [4]. A classical queueing model with two servers in which the inter arrival times follow Markovian arrival process, the service times are phase type distributed is studied by Ayyappan G. and Gowthami R. in [5]. A single server retrial queueing model with two types of customers with service time distribution following phase type is studied by Krishnamoorthy et al. in [6].

The modern day retail business is driven by multiple platforms, which are either offline or online. Long gone are the days where customers had to be physically present for shopping. Sometimes customers give orders online through various virtual platforms and on a few occasions go out for shopping. The sellers move into all the possibilities of customer interaction so that the sales are always boosted. This requires algorithms which could determine the stock pile required by the sellers well in advance, so that the customer demands are met without delay or failure.

This model is motivated by two types of demands that arrive at shops: one is physically arriving customers and the other is online customers. The physically present customers are attended by the system on a FIFO basis. Online customer demands are attended only when the accumulated number of such demands reaches a threshold N or a random clock realizes whichever occurs first.

The model description and mathematical formulation of the model is done. Several Performance Measures are obtained at steady state. A numerical illustration is provided.

Model description

Our model includes one product, which is been sold through two different platforms – a physical shop and an online platform. We consider a single server queueing inventory system with two types of customers, namely physical customers (type-I) and online customers (type-II). We assume that type-I and type-II customers arrive according to a stationary Marked Markovian Arrival Process(MMAP) having m phases. Let (D_0, D_1, D_2) denote the matrix representation of the MMAP guiding the arrival process. We also assume that the service time is positive. Type-I customers can form an infinite queue, while type-II customer joins a finite buffer of size N . Type-II customers are served in batches with maximum batch size N . A clock is also set, which starts ticking with the first arrival of type-II customer in every cycle. An (s, S) inventory policy is used. It is assumed that N is less than s . Service of a customer requires an inventory item. Each customer (both type-I and type-II) demands one unit of item, having a random duration of service time. Service time distribution of both type-I and type-II customers are assumed to be of phase type with irreducible representations $\text{PH}(\boldsymbol{\alpha}, T)$ with m_1 phases and $\text{PH}(\boldsymbol{\beta}, U)$ with m_2 phases respectively. The vectors T_0 and U_0 are given by $T_0 = -Te$ and $U_0 = -Ue$. Lead time is assumed to be exponential with parameter γ . If the server is idle and at least one item is available in the inventory, type-I customer enters into the service, provided there is at least one type-I customer in the system. A clock is set for type-II customers. Let θ be the rate of realization (parameter) of the exponential clock. The clock starts at the arrival of the first type-II customer. The type-II customers are served only when number of type-II customers reaches N or the clock time expires, whichever occurs first. They are served in bulk, provided two or more type-II customers have joined before the expiry/realization of the clock.

When the service of type-I begins, the inventory level drops by one unit. But when the service of type-II begins, the inventory level drops by n_2 , where n_2 is the number of type-II customers present at that time. When the clock expires or when number of type-II customers reaches N , the service of all these type-II customers is provided in a batch, provided the server was idle at that time and required inventory is available. On the other hand, if the server is busy at that time, type-II customers are served immediately on

completion of service of the current type-I customer. No type II customer is allowed to join the system once the clock expires/ N type II are in the system.

Mathematical formulation

Let

- $N_1(t)$ be the number of type-I customers in the queue at time t ;
- $N_2(t)$ be the number of type-II customers in the finite buffer at time t ;
- $B(t)$ be the server status at time t ;
- $C(t)$ be the clock status at time t ;
- $I(t)$ be the number of items in the inventory at time t ;
- $J(t)$ be the phase of the service process at time t ;
- $A(t)$ be the phase of the arrival process at time t .

Then

$$\{(N_1(t), N_2(t), B(t), C(t), I(t), J(t), A(t)); t \geq 0\}$$

is a continuous time Markov chain on the state space to be described below. This model can be considered as a Level Independent Quasi-Birth-Death (LIQBD) process and a solution is obtained by Matrix Analytic Method. We define the state space of the QBD under consideration and analyze the structure of its infinitesimal generator.

The state space

$$\Omega = \Omega_1 \cup \Omega_2 \cup \Omega_3 \cup \Omega_4 \cup \Omega_5,$$

where

$$\Omega_1 = \{(0, 0, 0, 0, i, 0^*, a) / 0 \leq i \leq S; a = 1, 2, \dots, m\};$$

$$\Omega_2 = \{(n_1, 0, 0, 0, 0, 0^*, a) \cup (n_1, n_2, 0, 1, 0, 0^*, a) / n_1 \geq 1; 1 \leq n_2 \leq N-1; a = 1, 2, \dots, m\};$$

$$\Omega_3 = \{(n_1, n_2, 0, 0, i, 0^*, a) / n_1 \geq 0; 1 \leq n_2 \leq N; 0 \leq i \leq n_2 - 1; a = 1, 2, \dots, m\};$$

$$\Omega_4 = \{(n_1, n_2, 1, c, i, j, a) / n_1 \geq 0; 0 \leq n_2 \leq N; c = 0, 1; 0 \leq i \leq S; \\ j = 1, 2, \dots, m_1; a = 1, 2, \dots, m\};$$

$$\Omega_5 = \{(n_1, n_2, 2, 1, i, j, a) \cup (n_1, 0, 2, 0, i, j, a) \cup (n_1, N, 2, 0, i, j, a) / n_1 \geq 0; \\ 0 \leq n_2 \leq N-1; 0 \leq i \leq S; j = 1, 2, \dots, m_2; a = 1, 2, \dots, m\},$$

0^* represents the phase of an idle server.

The infinitesimal generator Q of the LIQBD describing the above single server queuing inventory system is of the form

$$\begin{pmatrix} B_{00} & B_{01} & 0 & \dots & \dots & \dots \\ B_{10} & A_1 & A_0 & 0 & \dots & \dots \\ 0 & A_2 & A_1 & A_0 & \dots & \dots \\ 0 & 0 & A_2 & A_1 & A_0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix},$$

where B_{00} , B_{01} , B_{10} , A_0 , A_1 , A_2 are all square matrices of appropriate order whose entries are block matrices.

Performance measures

1. Expected number of type-I customers in the system

$$E[N_1] = \sum_{i=0}^{\infty} i \mathbf{x}_i \mathbf{e}.$$

2. Expected number of type-II customers in the system

$$E[N_2] = \sum_{i=0}^{\infty} \sum_{j=1}^N j \mathbf{x}_{ij} \mathbf{e}.$$

3. Expected number of items in the inventory

$$E[I] = \sum_{i=0}^{\infty} \sum_{j=0}^N \sum_{k=0}^2 \sum_{l=0}^1 \sum_{r=0}^S r \mathbf{x}_{ijkl} \mathbf{e}.$$

4. Expected number of customers waiting in the system due to lack of inventory

$$E[W] = \sum_{i=0}^{\infty} \sum_{j=0}^{N-1} i \mathbf{x}_{ij010} \mathbf{e} + \sum_{i=0}^{\infty} \sum_{j=1}^N \sum_{r=0}^{j-1} j \mathbf{x}_{ij00r} \mathbf{e}.$$

5. Probability, that the server is idle:

$$b_0 = \sum_{i=0}^{\infty} \sum_{j=0}^N \mathbf{x}_{ij0} \mathbf{e}.$$

6. Probability, that the server is busy with type-I customer:

$$b_1 = \sum_{i=0}^{\infty} \sum_{j=0}^N \mathbf{x}_{ij1} \mathbf{e}.$$

7. Probability, that the server is busy with type-II customer:

$$b_2 = \sum_{i=0}^{\infty} \sum_{j=0}^N \mathbf{x}_{ij2} \mathbf{e}.$$

8. Probability, that the clock is on:

$$c_1 = \sum_{i=0}^{\infty} \sum_{j=1}^N \sum_{k=0}^2 \mathbf{x}_{ijk1} \mathbf{e}.$$

9. Expected rate at which replenishment of inventory occurs

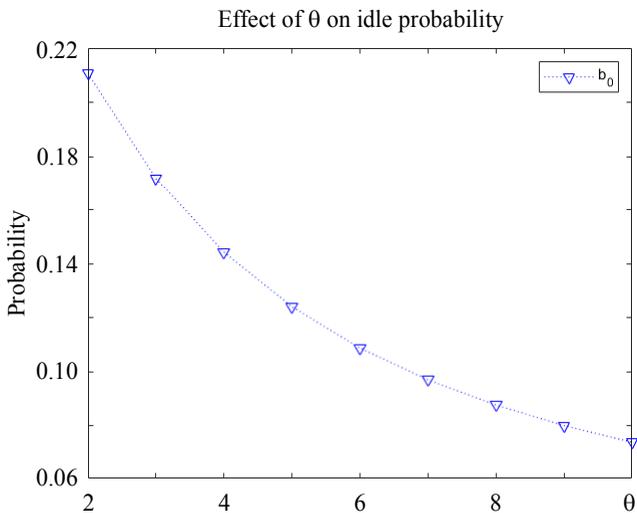
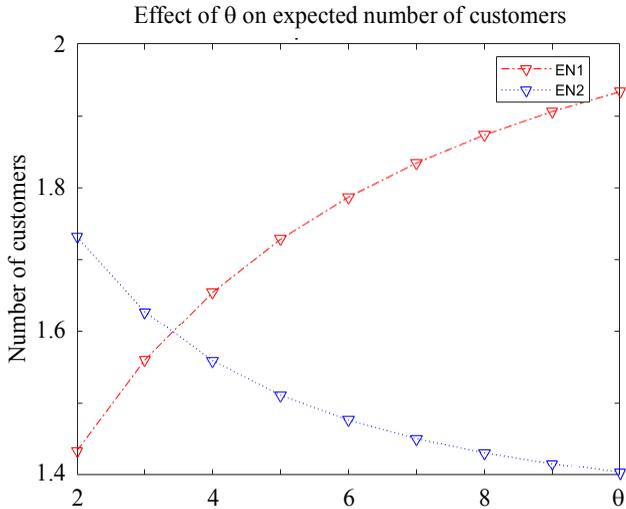
$$E_R = \sum_{i=0}^{\infty} \sum_{j=0}^N \sum_{k=0}^2 \sum_{l=0}^1 \sum_{r=0}^s \gamma \mathbf{x}_{ijklr} \mathbf{e}.$$

10. Probability, that the type-II customer is blocked from entering the service:

$$p_b = \sum_{i=0}^{\infty} \sum_{j=1}^{N-1} \sum_{k=1}^2 \mathbf{x}_{ijk0} \mathbf{e} + \sum_{i=0}^{\infty} \mathbf{x}_{iN} \mathbf{e}.$$

Numerical Examples

In the following example, we illustrate the effect of the parameter θ on the expected number of type-I and type-II customers in the system and also the effect on idle probability. We consider a MMAP in which the arrivals of the type-I and type-II customers are described by matrices (D_0, D_1, D_2) . We keep all variables except θ fixed. As the rate of the exponential clock θ increases, expected number of type-I customers in the system increases while the expected number of type-II customer in the system decreases. As θ increases, idle probability decreases. This is because as θ increases, more type-II customers get service and so expected number of type-I customers in the system increases and expected number of type-II customers decreases.



Conclusions

In this paper we considered a single server queueing inventory model with two types of customers. Service to both types of customers is provided by a single server. Various performance measures are evaluated at steady state conditions. We plan to analyse the problem for cost effectiveness.

REFERENCES

1. *Sigman K., Simchi-Levi D.* Light traffic heuristic for an M/G/1 queue with limited inventory // *Annals of Operation Research*. 1992. V. 40. P. 371–380.
2. *Krishnamoorthy A., Dhanya Shajin, Narayanan V.C.* Inventory with Positive Service Time: a Survey, *Advanced Trends in Queueing Theory: Series of Books “Mathematics and Statistics”*, Sciences. Editors: V. Anisimov and N. Limnios. London: ISTE & Wiley, 2019.
3. *Dudin A., Dudin S., Dudina O.* Analysis of an MMAP/PH1, PH2/N/infty queueing system operating in a random environment // *International Journal of Applied Mathematics and Computer Science*. 2014. V. 24. P. 485–501.
4. *Sreenivasan C., Chakravarthy S.R., Krishnamoorthy A.* MAP/PH/1 queue with working vacations, vacation interruptions and N policy // *Applied Mathematical Modelling*. 2013. V. 37. Iss. 6. P. 3879–3893.
5. *Ayyappan G., Gowthami R.* Analysis of MAP/PH(1),PH(2)/2 queue with Bernoulli schedule vacation, Bernoulli feedback and renege of customers // *International Journal of Applied and Computational Mathematics*. 2019. V. 5. P. 159.
6. *Krishnamoorthy A., Joshua V.C., Ambily P.M.* A retrial queueing system with abandonment and search for priority customers // *DCCN-2017, Springer, CCIS*. 2017. V. 700. P. 98–107.

Analysis of a PH/PH/1 queue with interdependence

K.R. Ranjith¹, Achyutha Krishnamoorthy², B. Gopakumar³

¹Research Scholar, Government Engineering College, Thrissur, India

²Centre for Research in Mathematics, CMS College, Kottayam, India

³Associate Professor, Government Engineering College, Thrissur, India

In this paper we consider queueing system in which both arrival and service time distributions are of Phase type with the important additional assumption that service time distribution is dependent on arrival time distribution.

In paper [4] a numerical investigation of a PH/PH/1 inventory system is considered. The paper [2] is a study on MAP/(PH/PH)/1 queue with self-generation of priorities and non-preemptive service.

Here we provide a brief description of the idea of interdependence among random processes, evolving continuously in time, as introduced by Krishnamoorthy through a series of webinars (SMARTY (Karelian Republic) August 2020; DCCN (Trapeznikov Institute of Control Sciences, Moscow) September 2020; Professor C.R. Rao Birth Centenary Talk (Sree Venkateswara University, Tirupati), September 2020 and a few others).

Consider two phase type processes (α, T) of order n and (β, S) of order m (say type1 and type2 arrivals). We also consider two sets of probabilities p_1, p_2, \dots, p_n and q_1, q_2, \dots, q_m . Consider the Markov chain $X(t) = \{(i, j) : 1 \leq i \leq n; 1 \leq j \leq m\}$. The following transitions are considered:

1. $P[X(t + \Delta t) = (i, j) / X(t) = (i, k)] = p_i s_{kj} \Delta t + O(\Delta t)$,
 $1 \leq i \leq n, j \neq k$, without Type1 or Type 2 arrivals.
2. $P[X(t + \Delta t) = (l, j) / X(t) = (i, k)] = q_j t_{il} \Delta t + O(\Delta t)$,
 $1 \leq j \leq m, i \neq l$, without Type1 or Type 2 arrivals.
3. $P[X(t + \Delta t) = (i, j) / X(t) = (i, k)] = p_i s_k^0 \Delta t \beta_j + O(\Delta t)$,
 $1 \leq i \leq n, 1 \leq j, k \leq m$, with a Type 2 arrival.
4. $P[X(t + \Delta t) = (i, j) / X(t) = (i, k)] = q_j t_i^0 \Delta t \alpha_l + O(\Delta t)$,
 $1 \leq j \leq m, 1 \leq i, l \leq n$, with a Type1 arrival.

The interdependent process will have representation

$$(\alpha \otimes \beta, T \otimes J_q + J_p \otimes S),$$

where J_q and J_p are diagonal matrices of the probabilities (p_1, p_2, \dots, p_n) and (q_1, q_2, \dots, q_m) .

Mathematical model

The system is described as under. Customers arrive to a single server counter according to a phase type distribution with representation (T, β) of order n . Service time duration also follows a phase type distribution with representation (S, α) of order m . Unlike in the usual PH/PH/1 queue, where the inter arrival time distribution and service time distribution, are independent here we make the important assumption that service time distribution depends on inter arrival time distribution. The above assumption is motivated by the following present day scenario. In the early stages of the covid19 outbreak, the health care system in our country took utmost care in the treatment of the patients; they were kept in isolation and were sent home only after several tests. But as the number of cases increased the system had to accommodate more and more patients and so the way of treatment was changed. Here the treatment given (service) depends on the number of patients (arrival)

At time t let $N(t)$ be the number of customers in the system including the one being served, $S_1(t)$ be phase of arrival and $S_2(t)$ be the phase of service.

Then $\Omega = \{X(t) : t \geq 0\} = \{(N(t), S_1(t), S_2(t)), t \geq 0\}$ is a Markov chain $E = \{(0, k) : 1 \leq k \leq n\} \cup \{(i, k, l) : i \geq 1, 1 \leq k \leq n, 1 \leq l \leq m\}$. The state space of the Markov chain can be partitioned into levels \tilde{i} defined as $\tilde{0} = \{(0, 1), (0, 2), \dots, (0, n)\}$ and $\tilde{i} = \{(i, 1, 1), (i, 1, 2), \dots, (i, 1, m), (i, 2, 1), (i, 2, 2), \dots, (i, 2, m), \dots, (i, n, 1), (i, n, 2), \dots, (i, n, m)\}$ for $i \geq 1$. Let (p_1, p_2, \dots, p_n) be a probability vector. In the sequel J_n stands for a diagonal matrix of order n with the diagonal entries p_1, p_2, \dots, p_n ; I_n denotes an identity matrix of order n and e denotes a column matrix of 1's of appropriate order.

Now the infinitesimal generator matrix of the process is

$$T = \begin{bmatrix} B_0 & B_1 & 0 & 0 & 0 & 0 \\ B_2 & A_1 & A_0 & 0 & 0 & 0 \\ 0 & A_2 & A_1 & A_0 & 0 & 0 \\ 0 & 0 & A_2 & A_1 & A_0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & \cdot \end{bmatrix},$$

where $B_0 = T; B_1 = T^\circ \otimes \beta \otimes \alpha; B_2 = J_n \otimes S^\circ$

and $A_1 = T \otimes I_m \otimes J_n \otimes S; A_0 = T^\circ \otimes \beta \otimes I_m; A_2 = J_n \otimes S^\circ \otimes \alpha$.

Stability condition

Let π be the stationary probability vector of A , where

$$\begin{aligned} A &= A_0 + A_1 + A_2 = T^\circ \otimes \beta \otimes I_m + T \otimes I_m + J_n \otimes S + J_n \otimes S^\circ \otimes \alpha = \\ &= (T + T^\circ \otimes \beta) \otimes I_m + J_n \otimes (S + S^\circ \otimes \alpha). \end{aligned}$$

Then the system is stable, if and only if, $\pi A_0 e < \pi A_2 e$.

Computation of steady state vector

The stationary probability vector z of T is of the form $(z_0, z_1, z_1 R, z_1 R^2, \dots)$, where R is the minimal solution of the matrix quadratic equation $R^2 A_2 + R A_1 + A_0 = 0$. The stationary probability vector is obtained by solving the equations

$$\begin{aligned} z_0 B_0 + z_1 B_2 &= 0, \\ z_0 B_1 + z_1 A_1 + z_1 R A_2 &= 0, \\ z_0 + z_1 (I - R)^{-1} e &= 1. \end{aligned}$$

Hence z_1 can be determined upto a multiplicative constant by the equation $z_1 [B_2 e \otimes \beta \otimes \alpha + A_1 + R A_2] = 0$.

Some performance measures

The performance measures of interest are:

1. The probability that there are n of customers in the system, $P_n = z_n e$.
2. Expected number of customers in the system, $E(C) = \sum n P_n$.

Numerical Illustration

As an example, we take

$$\begin{aligned} T &= \begin{bmatrix} -9 & 5 \\ 3 & -4 \end{bmatrix}, \beta = (0.4, 0.6), S = \begin{bmatrix} -7 & 1 & 2 \\ 2 & -11 & 2 \\ 2 & 2 & -10 \end{bmatrix}, \\ \alpha &= (0.2, 0.3, 0.5) \text{ and } p = (0.4, 0.6). \end{aligned}$$

We have the following results:

1. Probability, that there are no customers in the system, $P_0 = 0.3459$.
2. Expected number of customers in the system, $E(C) = 2.0102$.
3. Probability, that the number of customers in the system is less than $10 = 0.99$.

Thus with these arrival and service rates, we are 99% sure that the number of customers in the system will be less than 10. In a health care system under a pandemic, if the capacity of the system is C , if the requirement rises above C , the system collapses. So using this model we can find the probability of such a collapse and can adjust the service rate by altering the vector p and can to an extent help the system to sustain.

Conclusions

In our future work we intend to fit a PH/PH/1 distribution with service dependent on arrival for some available health care data and estimate the parameters involved in such a manner that $P[\text{number of patients in the system} > c]$ is less than a pre assigned number. We also intend to calculate several other performance measures which will help to understand the system in more detail.

REFERENCES

1. *Krishnamoorthy A.* Evolution of several processes with interdependence, governed by multi-dimensional semi-markov process // Webinar talks at SMARTY (Karelian Republic) 16–20 August 2020 – Plenary talk; DCCN (Trapeznikov Institute of Control Sciences, Moscow) 14–18 September 2020; Professor C.R. Rao Birth Centenary Talk (SreeVenkateswara University, Tirupati), 7–10 September 2020 and a few others).
2. *Krishnamoorthy A., Babu S., and Narayanan V.C.* The MAP/(PH/PH)/1 queue with self-generation of priorities and non-preemptive service // European Journal of Operational Research. 2009. V. 195. No. 1. P. 174–185.
3. *Krishnamoorthy A., Viswanath C.N., Joshua V.C., Deepak T.G.* An MAP/PH/1 retrial queue with service interruption and orbital search International // Journal of Applied Mathematics.
4. *Krishnamoorthy A., Jose K.P., Viswanath C.N.* Numerical investigation of a PH/PH/1 inventory system with positive service time with shortage // Neural, Parallel and Scientific Computations. 2008. V. 16. No. 4.
5. *Neuts M.F.* Matrix-geometric Solutions in Stochastic Models. Baltimore: The Johns Hopkins University Press, 1981. 352 p.

Stability condition of a modified Erlang loss system with different service rates*

Stepan Rogozin^{1,2}, Evsey Morozov^{1,2,3}

¹*Institute of Applied Mathematical Research Karelian Research Centre RAS, Petrozavodsk, Russia*

²*Petrozavodsk State University, Petrozavodsk, Russia*

³*Moscow Center for Fundamental and Applied Mathematics, Moscow State University, Moscow, Russia*

In this work we consider the extension of the following system with two-priority classes of customers and several identical servers. The first priority customers are lost if find all servers busy, while the second priority customers form an infinite capacity queue if find all servers busy. All classes of customers follow Poisson input and have general independent identically distributed (iid) service times. The service discipline for class-2 customers is assumed to be FCFS (first-come-first-served). Such a modified Erlang loss system is motivated and studied in the paper [1] in which in particular the stability criterion (1) of this system is obtained.

A new feature of this system considering in this work is that a multi-type server assignment for the priority customers is assumed. Each server can serve only a limited set of subclasses of class-1 customers and different subclasses of customers can have different arrival rates. At the same time, class-2 customers belong to only one type and can be served by any server. In previous work [2] we show that this system has the same stability condition (1) provided identical service rates for all servers. In present work we proof this condition under assumption of Poisson inputs and server-dependent service rates for the sets of priority subclasses. The present model is much more general than that has been considered in the paper [1]. Namely, in the present model each server can serve a limited set of class-1 customers and service rates are class- and server-dependent, while in the model from [1] all servers are homogeneous and service rate of class-1 customers is constant.

This system can be applied in wireless networks, where secondary users (un-licensed users) can cognitively use the bandwidths that are originally allocated to primary users (licensed users). Secondary users should use the bandwidths in such a way that does not interfere primary users [4].

* The research is supported by Russian Foundation for Basic Research, projects 18-07-00147, 18-07-00156 and 19-07-00303.

Description of the model

We consider the extension of the following system with two-priority classes of customers and J identical servers. The first priority customers (class-1) are lost if find all servers busy, while the second priority customers (class-2) form an infinite capacity queue if find all servers busy. Customers of class- i follow Poisson input with rate $\lambda^{(i)}$. Also we assume that class- i customers have independent identically distributed (iid) service times $\{S_n^{(i)}\}$ with generic element $S^{(i)}$. The service discipline for class-2 customers is assumed to be FCFS (first-come-first-served). A new feature of this system considering in this work is that a multi-type server assignment for the priority customers is assumed. Each server can serve only a limited set of subclasses of class-1 customers and different subclasses of customers can have different arrival rates. We assume Poisson inputs and server-dependent service rates for the sets of priority subclasses. At the same time, class-2 customers belong to only one type and can be served by any server. In Fig.1 we present the scheme of this extended system.

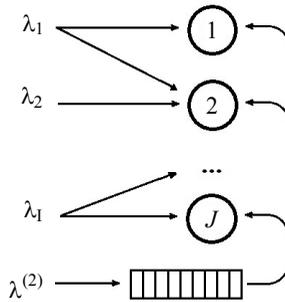


Fig. 1. Extended system

Stability analysis

Denote by $\rho_2 = \lambda^{(2)}ES^{(2)}$ the traffic intensity of class-2 customers and let P_i be the stationary probability that i servers are occupied by class-1 customers. Because the processes related to class-1 customers are positive recurrent and the inputs are Poisson, these stationary probabilities exist [5]. It is assumed that the first customer arrives in an empty system at instant $t_1 = 0$, and if the system is empty at this instant, we call it initially empty system. We note that in this case, the instant $T_0 = 0$ is a regeneration point and then $T_1 =_{st} T$ that is the first regeneration period is stochastically equivalent to generic period.

In this case positive recurrence means that $ET < \infty$. In this section we partly present the proof of the following statement.

Theorem 1. *If condition*

$$\rho_2 + \sum_{i=1}^J iP_i < J, \quad (1)$$

holds then $ET < \infty$, that is initially empty system is positive recurrent.

Proof. In the interval of time $[0, t)$, denote: $\hat{V}_1(t)$ the total time which need to serve class-1 customers accepted by the system (that is $\hat{V}_1(t)$ does not include the lost work); $V_2(t)$ the received work of class-2 customers; $B(t)$ the aggregated busy time of the servers which equals departed work in $[0, t)$.

Also denote $I(t) = \sum_{i=1}^J I_i(t)$, where $I_i(t)$ is the idle time of server i in $[0, t)$.

Obviously the following balance equations hold:

$$\hat{V}_1(t) + V_2(t) = W_1(t) + W_2(t) + B(t) = W(t) + Jt - I(t). \quad (2)$$

One can show that it implies

$$\lim_{t \rightarrow \infty} \frac{EW(t)}{t} = \lim_{t \rightarrow \infty} \frac{E\hat{V}_1(t)}{t} + \lim_{t \rightarrow \infty} \frac{EV_2(t)}{t} - J, \quad (3)$$

where all limits exist.

We denote $Q_1(t)$ the number of busy servers by class-1 customers at instant t , and note that the weak limit $Q_1(t) \Rightarrow Q_1$ exists, since the input is Poisson. Now we show that the following equality holds:

$$\lim_{t \rightarrow \infty} \frac{E\hat{V}_1(t)}{t} = EQ_1 = \sum_{i=1}^J iP_i, \quad (4)$$

where Q_1 is the stationary number of servers occupied by class-1 customers.

It is easy to see that

$$\hat{V}_1(t) = \int_0^t \sum_{i=1}^J i1(Q_1(u) = i)du + W_1(t), \quad (5)$$

where integral means the completed work or the total spent time for service class-1 customers in interval $[0, t)$.

Now we define the completed work of class-1 customers for our extended system. We can split indicators into $2^J - 1$ non intersection indicators of

servers subsets:

$$\begin{aligned}
 1(Q_1(t) = 1) &= \sum_{i=1}^J i 1 \text{ (only the } i\text{-th server is busy at instant } t\text{)}, \\
 1(Q_1(t) = 2) &= \sum_{i,j \in \{1, \dots, J\}, i \neq j} i 1 \text{ (only the } i\text{-th and the } j\text{-th servers} \\
 &\quad \text{are busy at instant } t\text{)}, \\
 &\dots \\
 1(Q_1(t) = J) &= 1 \text{ (all servers are busy at instant } t\text{)}.
 \end{aligned} \tag{6}$$

The following limit exists because, all processes related to class-1 customers are positive recurrent regenerative:

$$\begin{aligned}
 \lim_{t \rightarrow \infty} \frac{E\hat{V}_1(t)}{t} &= \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t \sum_{i=1}^J i 1(Q_1(u) = i) du = \\
 &= \sum_{i=1}^J i \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t 1(Q_1(u) = i) du = \sum_{i=1}^J i P_i.
 \end{aligned} \tag{7}$$

The key observation is that the stationary probability that i servers are occupied by class-1 customers is also the limiting fraction of the corresponding busy time:

$$P_i = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t 1(Q_1(u) = i) du. \tag{8}$$

Also we can replace P_i by the following:

$$P_i = \sum_{j_1, \dots, j_{J-k} \in \{1, \dots, J\}} \pi(\{j_1, \dots, j_{J-k}\}) \quad \text{for all } k \in \{1, \dots, J\}, \tag{9}$$

where $\pi(S)$ is the stationary probabilities that servers S are not busy by class-1 customers that introduced in [3].

From (3) – (7) one can obtain the inequality,

$$\lim_{t \rightarrow \infty} \frac{E\hat{V}_1(t)}{t} = \rho_2 + \sum_{i=1}^J i P_i - J \geq 0, \tag{10}$$

$$\text{or} \quad \rho_2 + \sum_{i=1}^J i P_i \geq J. \tag{11}$$

This contradicts with assumption (1) and implies that

$$Q_2(t) \neq \infty. \tag{12}$$

One can show that this implies positive recurrence. Also we can conclude that the following statement holds.

Theorem 2. *The initially empty system under consideration is positive recurrent if and only if condition (1) holds.*

Remark 1. Following [6] one can prove that, provided condition (1), the first regeneration period length $T_1 < \infty$ w.p.1 under arbitrary fixed values $W(0)$ and $Q(0)$.

Remark 2. Note that $\Delta = J - \sum_{i=1}^J i P_i$ is the mean stationary number of servers available for class-2 customers. Then condition (1) can be rewritten as

$$\rho_2 < \Delta.$$

This condition has the following intuitive probabilistic interpretation: the traffic intensity of class-2 customers must be less than the mean number of available servers Δ .

Conclusions

We consider a modified Erlang loss system with two-priority classes of customers and multi-class and multi-server first priority customers. We prove the stability condition of this system with different service rates. In a future work it is assumed to obtain numerical results related to special cases of this system based on the proved theoretical results.

REFERENCES

1. *Morozov E., Rogozin S., Nguyen H.Q., Phung-Duc T.* Modified Erlang loss system for cognitive wireless networks // Journal of Mathematical Sciences. 2020 (submitted).
2. *Rogozin S.* Simulation a modified Erlang loss system with multi-type servers and multi-type customers // Proceedings of The Second International Workshop on Stochastic Modeling and Applied Research of Technology. 2020 (submitted).
3. *Adan I., Hurkens C., Weiss G.* A reversible erlang loss system with multitype customers and multitype servers // Probability in the Engineering and Informational Sciences. 2010. V. 24(4). P. 535–548.
4. *Ostovar A., Keshavarz H., Quan Z.* Cognitive radio networks for green wireless communications: an overview // Telecommun. Syst. 2020. P. 1–10.
5. *Asmussen S.* Applied probability and Queues. 2nd ed. New York: Springer Verlag, 2003. 438 p.
6. *Morozov E., Delgado R.* Stability analysis of regenerative queues // Autom. Remote Control. 2009. V. 70. P. 1977–1991.

An explicit solution for an inventory model with positive lead time and backlogs

Sandhya E.¹, C. Sreenivasan², Sajeev S. Nair³

¹ *Research Scholar, Government College, Chittur, University of Calicut, India*

² *Government Victoria College, Palakad, India*

³ *Government Engineering College, Thrissur, India*

The first study on queueing inventory systems is due to (Melikov and Molchano 1992) and (Sigman and Simchi-Levi 1992). Later inventory systems where a processing time is required for serving the inventory was done by Berman and et al. [1]. This was a deterministic model. Berman and Kim [2] and Berman and Sapna [3] were the first to discuss inventory with exponential distribution and with arbitrary distribution respectively. Krishnamoorthy and his co-authors used Matrix Analytic Methods to study a inventory models [4–12], where service time for providing the inventoried item is assumed.

Padmavathi I. et al. [14] studied a finite source (s,S) inventory system with postponed demands and server vacation. Krishnamoorthy and Islam [15] considered an (s,S) inventory system with postponed demands in which arrival follows a Poisson distribution and lead time exponential distribution. Sivakumar and Arivarignam [16] studied a perishable inventory system with postponed demands in which the demands that occur during the stock out period enter a pool with independent Bernoulli trial.

Mathematical model

The system under consideration is described as follows. Customers arrive to a queueing system which has only one server according to a Poisson process of rate λ where inventory is served. Inventory is replenished according to (s,S) policy, the replenishment time being an exponential random variable with parameter δ . We assume negligible service time for this model. Customers, who join the queue when the inventory level drops to zero form a queue and remains in the system until inventory replenishment, is realized.

Let $N(t)$ and $S(t)$ denote the number of customers in the system and the inventory level respectively at time t . Then

$$\Omega = \{X(t) : t \geq 0\} = \{(N(t), S(t)) : t \geq 0\}$$

is a Markov chain with state space

$$E = \{(0, k) : 0 \leq k \leq S\} \cup \{(i, 0) : i \geq 1\}.$$

The state space of the Markov chain can be partitioned into levels \tilde{i} defined as $\tilde{0} = \{(0, 0), (0, 1), \dots, (0, S)\}$ and $\tilde{i} = \{((i-1)Q+1, 0), ((i-1)Q+2, 0), \dots, ((i-1)Q+Q, 0)\}; i \geq 1$. Thus the Markov chain under study is a level independent Quasi Birth Death (QBD) process. Here, $S-s=Q$, I_n denotes an identity matrix of order n and e denotes a column vector of 1's of appropriate order. The infinitesimal generator matrix of the process Ω is

$$H = \begin{bmatrix} B_0 & B_1 & 0 & 0 & . & . & . \\ B_2 & A_1 & A_0 & 0 & 0 & . & . \\ 0 & A_2 & A_1 & A_0 & 0 & 0 & . \\ 0 & 0 & A_2 & A_1 & A_0 & 0 & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \end{bmatrix}.$$

Here $B_0 = [b_{ij}]_{(S+1) \times (S+1)}$, where

$$b_{ij} = \begin{cases} -(\lambda + \delta) & : j = i; 1 \leq i \leq s+1, \\ -\lambda & : j = i; s+1 \leq i \leq S+1, \\ \lambda & : j = i-1; 2 \leq i \leq S+1, \\ \delta & : j = Q+i; 1 \leq i \leq s+1, \\ 0 & : otherwise; \end{cases}$$

$$B_1 = [b_{ij}]_{(S+1) \times Q}, \quad b_{ij} = \begin{cases} \lambda : i = j = 1, \\ 0 : otherwise; \end{cases}$$

$$B_2 = [b_{ij}]_{Q \times (S+1)}, \quad b_{ij} = \begin{cases} \delta : i+j=Q; 1 \leq i \leq Q, \\ 0 : otherwise; \end{cases}$$

$$A_2 = \delta I_Q; \quad A_0 = [a_{ij}]_{Q \times Q}; \quad a_{ij} = \begin{cases} \lambda : i = Q, j = 1, \\ 0 : otherwise; \end{cases}$$

$$A_1 = [a_{ij}]_{Q \times Q}; \quad a_{ij} = \begin{cases} -(\lambda + \delta) : i = j, \\ \lambda : j = i+1; 1 \leq i \leq Q-1. \end{cases}$$

Analysis of the model

Define $A = A_0 + A_1 + A_2$ and let $\pi = (\pi_1, \pi_2, \dots, \pi_Q)$ be the steady state vector of the generator matrix A . Then $\pi A = 0$ gives the following equations: $-\lambda\pi_1 + \lambda\pi_Q = 0$; $-\lambda\pi_{i+1} + \lambda\pi_i = 0; 1 \leq i \leq Q-1$. Hence $\pi_1 = \pi_2 = \dots = \pi_Q$. The QBD process with generator matrix A is stable if and only if $\pi A_0 e < \pi A_2 e$ (see Neuts) that is, if and only if, $\frac{\lambda}{Q\delta} < 1$. Thus we have the following theorem for the stability of the system under study.

Theorem: The Markov chain Ω is stable if and only if $\frac{\lambda}{Q\delta} < 1$.

Computation of steady state vector

We find the steady state vector of Ω explicitly. Let $\pi = (\pi_0, \pi_1, \dots)$ be the steady state vector where $\pi_0 = (\pi_0(0,0), \pi_0(0,1), \dots, \pi_0(0,S))$ and $\pi_i = (\pi_i((i-1)Q+1,0), \pi_i((i-1)Q+2,0), \dots, \pi_i((i-1)Q+Q,0)); i \geq 1, \pi H = 0$. Since Ω is a level independent QBD process its steady state vector is given by $\pi_{i+1} = \pi_1 R^i; i \geq 1$ (see Neuts), where R is the minimal non negative solution of the matrix quadratic equation $R^2 A_2 + R A_1 + A_0 = 0$. Since all the rows of A_0 except the last are zeros so are that of R . Assume that

$$R = [r_{ij}]_{Q \times Q}; r_{ij} = \begin{cases} r_j, & i = Q, 1 \leq j \leq Q, \\ 0, & \text{otherwise.} \end{cases}$$

Now from the equations of $R^2 A_2 + R A_1 + A_0 = 0$ we easily see that $r_j = r^j$ for every $j; r = r_1$.

Then all of the equations reduces to $\lambda - (\lambda + \delta)r + \delta r^{Q+1} = 0$.

Dividing by $r-1$ we get $\delta r^Q + \delta r^{Q-1} + \dots + \delta r - \lambda = 0$.

Let $f(x) = \delta x^Q + \delta x^{Q-1} + \dots + \delta x - \lambda$.

Since $f(0)$ and $f(1)$ have opposite sign $f(x)$ has a root r between 0 and 1.

Hence $R = [r_{ij}]_{Q \times Q}; r_{ij} = \begin{cases} r^j, & i = Q, 1 \leq j \leq Q \\ 0, & \text{otherwise} \end{cases}$.

Now from $\pi_{i+1} = \pi_i R^i; i \geq 1$ we get $\pi_2 = (r, r^2, \dots, r^Q) \pi_1(Q, 0);$
 $\pi_3 = (r^{Q+1}, r^{Q+2}, \dots, r^{2Q}) \pi_1(Q, 0)$ and so on.

It remains to get π_0 and π_1 in terms of $\pi_1(Q, 0)$

$$\pi_0(0, 0) = \left\{ (\psi)^Q - r \frac{\delta \left(\frac{1-r^Q (\psi)^Q}{1-r(\psi)} \right) \right\} \pi_1(Q, 0);$$

$$\pi_1(i, 0) = \left\{ (\psi)^{Q-i} - r^{i+1} \frac{\delta \left(\frac{1-r^{Q-i} (\psi)^{Q-i}}{1-r(\psi)} \right) \right\} \pi_1(Q, 0); 1 \leq i \leq Q-1;$$

$$\pi_0(0, 1) = (\psi) \pi_0(0, 0) - \frac{\delta}{\lambda} \pi_1(Q, 0);$$

$$\pi_0(0, i+1) = (\psi) \pi_0(0, i) - \frac{\delta}{\lambda} \left\{ (\psi)^i - r^{Q-i+1} \frac{\delta \left(\frac{1-r^i (\psi)^i}{1-r(\psi)} \right) \right\} \pi_1(Q, 0)$$

$1 \leq i \leq s;$

$$\pi_0(0, i+1) = \pi_0(0, i) - \frac{\delta}{\lambda} \left\{ (\psi)^i - r^{Q-i+1} \frac{\delta \left(\frac{1-r^i (\psi)^i}{1-r(\psi)} \right) \right\} \pi_1(Q, 0),$$

$s+1 \leq i \leq Q-1;$

$$\pi_0(0, Q+i+1) = \pi_0(0, Q+i) - \frac{\delta}{\lambda} \pi_0(0, i); \quad 0 \leq i \leq s-1.$$

Now $\pi_1(Q, 0)$ is got from the condition $\pi_0 e_1 + \left(\sum_{i=1}^{\infty} \pi_i \right) e_2 = 1$ where e_1 and e_2 are column vector of one's of appropriate order.

System performance measures

The expected waiting time of a general customer is given by

$$E(W_L) = \left(\sum_{k=1}^{\infty} \frac{k}{\delta} \pi_k \right) e = \frac{\lambda}{\delta^2} \left(\frac{2}{1-r^Q} + \frac{r^Q}{(1-r^Q)^2} \right) \pi_1(Q, 0) + \frac{\lambda}{\delta^2} \pi_0(0, 0).$$

In a similar manner, we can find the second moment of the waiting time of an customer as

$$E(W_L^2) = \sum_{k=1}^{\infty} W_2^k \pi_k = \frac{2\lambda}{\delta^3} \left[\frac{(1-r^Q)^{-3} - 1}{r^Q} \right] \pi_1(Q, 0) + \frac{2\lambda}{\delta^3} \pi_0(0, 0),$$

$$W_2^k = 2\alpha(T^{-2})e = \frac{k(k+1)}{\delta^2}.$$

Other performance measures such as the expected number of customers in the system, expected inventory level, expected replenishment rate, probability that inventory level is zero and probability that inventory level is greater than s are also calculated.

Numerical illustration

Effect of s on various performance measures $\lambda = 1, \delta = 2, S = 25$

s	E(I)	E(N)	P(I=0)	ERR	P(I>s)	E(W _L)	V(W _L)
5	13.52575	0.02635	0.01318	0.1	0.9	0.00879	0.017503
6	14.0176	0.0185	0.00925	0.10526	0.89474	0.00617	0.012312
7	14.51167	0.01303	0.00651	0.11111	0.88889	0.00435	0.008681
8	15.00717	0.00921	0.0046	0.11765	0.88235	0.00307	0.006141
9	15.50347	0.00653	0.00327	0.125	0.875	0.00218	0.004365
10	16.00004	0.00466	0.00233	0.13333	0.86667	0.00156	0.003118
11	16.49637	0.00335	0.00167	0.14286	0.85714	0.00112	0.002249

Effect of S on various performance measures $\lambda = 1, \delta = 2, s = 5$

S	E(I)	E(N)	P(I=0)	ERR	P(I>s)	E(W _L)	V(W _L)
11	6.37214	0.10702	0.05289	0.3333	0.66667	0.03984	0.090783
12	6.94287	0.08514	0.04226	0.28571	0.71429	0.03056	0.065576
13	7.4823	0.07122	0.03544	0.25	0.75	0.02494	0.051838
14	8.00485	0.06156	0.03069	0.22222	0.77778	0.0212	0.043301
15	8.51776	0.05444	0.02716	0.2	0.8	0.01854	0.037476
16	9.02499	0.04893	0.02443	0.18182	0.81818	0.01655	0.033236
17	9.5288	0.04452	0.02224	0.16667	0.83333	0.01498	0.029986

Effect of δ on various performance measures $\lambda = 1, s = 5, S = 11$

δ	E(I)	E(N)	P(I=0)	ERR	P(I>s)	E(W _L)	V(W _L)
1	6.37214	0.10702	0.05289	0.3333	0.6667	0.03984	0.090783
1.2	6.76228	0.05091	0.03025	0.3333	0.72222	0.01717	0.030885
1.4	7.02939	0.02659	0.01847	0.3333	0.7619	0.00825	0.012362
1.6	7.22392	0.0149	0.01185	0.3333	0.79167	0.0043	0.005552
1.8	7.37208	0.00882	0.00789	0.3333	0.81481	0.00238	0.002704
2	7.48883	0.00545	0.00543	0.3333	0.83333	0.00139	0.001408
2.2	7.58329	0.00349	0.00383	0.3333	0.84848	0.00084	0.000779

Conclusions

We studied a single server queueing model with negligible service time and backlogs. We wish to extend this model to one with positive service time which may have many practical applications.

REFERENCES

1. *Berman O., Kim E., Shimshack D.G.* Deterministic approximations for inventory management at service facilities // IEEE Transactions. 1998. V. 25(5). P. 98–104.
2. *Berman O., Kim E.* Stochastic models for inventory management at service facilities // Communication in Statistics // Stochastic Models. 1999. V. 15. No. 4. P. 695–718.
3. *Berman O., Sapna K.P.* Inventory management at service facilities for systems with arbitrarily distributed service times // Communications in Statistics. Stochastic Models. 2000. V. 16. No. 3-4. P. 343–360.
4. *Deepak T.G., Krishnamoorthy A., Viswanath C.N., Vineetha K.* Inventory with service time and transfer of customers and inventory // Annals of Operations Research. 2008. V. 160. No. 1. P. 191–213.
5. *Krishnamoorthy A., Deepak T.G., Viswanath C.N., Vineetha K.* Effective utilization of idle time in an (s,S) inventory with positive service time // Journal of Applied Mathematics and Stochastic Analysis (JAMSA). 2006. V. 2006. Article ID 69068. DOI: 10.1155/JAMSA/2006/69068.
6. *Krishnamoorthy A., Islam M.E., Viswanath C.N.* Retrial inventory with batch Markovian arrival and positive service time // Stochastic Modelling and Applications. 2006. V. 9. No. 2. P. 38–53.
7. *Krishnamoorthy A., Jose K.P., Viswanath C.N.* Numerical investigation of a PH/PH/1 inventory system with positive service time and shortage // Neural, Parallel & Scientific Computations. 2008. V. 16. No. 4.
8. *Krishnamoorthy A., Sajeev S.N., Viswanath C.N.* An inventory model with retrial and orbital search // Bulletin of Kerala Mathematical Association. Guest Editor: S.R.S Varadhan FRS. October 2009. Special Issue. P. 47–65.
9. *Krishnamoorthy A., Viswanath C.N., Deepak T.G., Vineetha K.* Control policies for inventory with service time; stochastic analysis and applications. 2006. V. 24. No. 4. P. 889–899.
10. *Krishnamoorthy A., Viswanath C.N., Islam M.E.* On a production inventory with service time and retrial of customers // Proceedings of the 11th International Conference on Analytical and Stochastic Modelling Techniques and Analysis (ASMTA). Magdeburg, 2004. P. 238–247.
11. *Krishnamoorthy A., Viswanath C.N., Islam M.E.* Retrial production inventory with MAP and service times // Queues, Flows, Systems, Networks. Proceedings of the International Conference “Modern Mathematical Methods of Analysis and Optimization of Tele-communication Networks”. Gomel, 2003. P.148–156.
12. *Viswanath C.N., Deepak T.G., Krishnamoorthy A., Krishnakumar B.* On an (s,S) inventory system with service time vacation on the server an correlated lead time // Quality Technology and Quantitative Management (QTQM). 2008. P. 129–143. V. 5. No. 2.

13. *Deepak T.G., Viswanath C.N., Joshua V.C.* On an Inventory with positive service time having optional processing time // Bulletin of Kerala Mathematics Association. 2007. V. 4. No. 2. P. 75–86
14. *Padmavathi A., Shophia Lawrence, Sivakumar B.* A finite-source inventory system with postponed demands and modified M vacation policy // OPSEARCH. V. 53(1). 2016. P. 41–62. DOI: 10.1007/s12597-015-0224-7.
15. *Krishnamoorthy A., Islam M.E.* (s,S) inventory system with postponed demands // Stoch. Anal. Appl. 2004. V. 22(3). P. 827–842.
16. *Sivakumar B., Arivarignan G.* An inventory system with postponed demands // Stoch. Anal. Appl. 2007. V. 26(1). P. 84–97.

Transient analysis of an inventory model with instantaneous replenishment and catastrophes

Smija Skaria, Sajeev S. Nair

*Government Engineering College, Thrissur,
APJ Abdul Kalam Technological University, Kerala, India*

Below we present a review of the work on the studies of transient analysis of queueing systems. In paper [1] a transient solution of an M/M/1 queue with catastrophes is considered. In paper [2] a transient analysis of a single server queue with catastrophes, failures and repairs is considered by the authors. In paper [3] a closed form solution for a queueing network model with catastrophes is obtained. We refer to paper [4] for a complete survey on various inventory queueing models.

Mathematical model

The system is described as under. Customers arrive to a counter according to a Poisson process of rate λ , where inventory is served. Service is exponentially distributed with rate μ , but when it ends the whole customers are emptied. Inventory is replenished according to (s, S) policy, replenishment being instantaneous. Further no shortage is permitted.

At time t let $N(t)$ be the number of customers in the system including the one being served and $L(t)$ be the inventory level.

Then $\Omega = \{X(t) : t \geq 0\} = \{(N(t), L(t)), t \geq 0\}$ will be a Markov chain with state space $E = \{(0, k) : s \leq k \leq S-1\} \cup \{(i, k) : i \geq 1, s+1 \leq k \leq S\}$. The state space of the Markov chain can be partitioned into levels \tilde{i} defined as $\tilde{0} = \{(0, s), (0, s+1), \dots, (0, S-1)\}$ and $\tilde{i} = \{(i, s+1), (i, s+2), \dots, (i, S)\}$ for $i \geq 1$. In the following sequel Q stands for $S-s$, I_n denotes an identity matrix of order n and e denotes a column matrix of 1's of appropriate order. Now the infinitesimal matrix of the process is

$$T = \begin{bmatrix} B_0 & B_1 & 0 & 0 & 0 & 0 \\ A_2 & A_1 & A_0 & 0 & 0 & 0 \\ A_2 & 0 & A_1 & A_0 & 0 & 0 \\ A_2 & 0 & 0 & A_1 & A_0 & 0 \\ & & \cdot & \cdot & \cdot & \cdot \\ & & & \cdot & \cdot & \cdot \end{bmatrix},$$

where $B_0 = -\lambda I_Q$; $A_2 = \mu I_Q$; $A_1 = -(\lambda + \mu)I_Q$; $A_0 = \lambda I_Q$,

$$B_1 = \begin{bmatrix} 0 & 0 & \cdot & \cdot & \lambda \\ \lambda & 0 & \cdot & \cdot & 0 \\ 0 & \lambda & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \lambda & 0 \end{bmatrix}.$$

Transient solution

The Chapman – Kolmogorov equations are given by

$$\pi'(0, i)(t) = -\lambda\pi(0, i)(t) + \mu[\pi(1, i+1)(t) + \pi(2, i+1)(t) + \dots],$$

$$s \leq i \leq S-1; \quad (1)$$

$$\pi'(j, i)(t) = -(\lambda + \mu)\pi(j, i)(t) + \lambda\pi(j-1, i)(t), \quad j \geq 2, s+1 \leq i \leq S;$$

$$\pi'(1, S)(t) = -(\lambda + \mu)\pi(1, S)(t) + \lambda\pi(0, S)(t);$$

$$\pi'(1, i)(t) = -(\lambda + \mu)\pi(1, i)(t) + \lambda\pi(0, i)(t), \quad s+1 \leq i \leq S-1. \quad (2)$$

Applying Laplace transform to (1) and (2) we obtain the following.

$$-\lambda\tilde{\pi}(0, i)(a) + \mu[\tilde{\pi}(1, i+1)(a) + \tilde{\pi}(2, i+1)(a) + \dots] = a\tilde{\pi}(0, i)(a), \quad s+1 \leq i \leq S-1;$$

$$-\lambda\tilde{\pi}(0, s)(a) + \mu[\tilde{\pi}(1, s+1)(a) + \tilde{\pi}(2, s+1)(a) + \dots] = a\tilde{\pi}(0, s)(a) - 1;$$

$$\lambda\tilde{\pi}(j, i)(a) = (\lambda + \mu + a)\tilde{\pi}(j+1, i)(a), \quad j \geq 1, \quad s+1 \leq i \leq S;$$

$$\lambda\tilde{\pi}(0, i)(a) = (\lambda + \mu + a)\tilde{\pi}(1, i)(a), \quad j \geq 1, \quad s+1 \leq i \leq S-1;$$

$$\lambda\tilde{\pi}(0, s)(a) = (\lambda + \mu + a)\tilde{\pi}(1, S)(a),$$

where $\tilde{\pi}(j, i)(a)$ stands for the Laplace transform of $\pi(j, i)(t)$.

By taking inverse Laplace transform we obtain the following.

$$\pi(i+1, j)(t) = \pi(i, j)(t) * \lambda e^{-(\lambda+\mu)t}, \quad i \geq 1, \quad s+1 \leq j \leq S;$$

$$\pi(1, j)(t) = \pi(0, j)(t) * \lambda e^{-(\lambda+\mu)t}, \quad s+1 \leq j \leq S-1;$$

$$\pi(1, S)(t) = \pi(0, S)(t) * \lambda e^{-(\lambda+\mu)t};$$

$$\frac{\lambda\mu}{\mu - \lambda} [e^{-\lambda t} - e^{-\mu t}] * \pi(0, s+1)(t) = \pi(0, s)(t) - e^{-\lambda t};$$

$$\frac{\lambda\mu}{\mu - \lambda} [e^{-\lambda t} - e^{-\mu t}] * \pi(0, s)(t) = \pi(0, S-1)(t);$$

$$\frac{\lambda\mu}{\mu - \lambda} [e^{-\lambda t} - e^{-\mu t}] * \pi(0, i+1)(t) = \pi(0, i)(t); \quad s+1 \leq i \leq S-2.$$

Finally $\pi(0,s)(t)$ is the inverse Laplace transform of

$$\frac{\mu + a}{\lambda\mu} \frac{x^{Q-1}}{x^Q - 1}; x = \frac{(\lambda + a)(\mu + a)}{\lambda\mu}.$$

Here * stands for the convolution of two functions and we assume the initial distribution is $(1,0,0,\dots,0)$.

Numerical Analysis

Even though we have obtained explicit expressions for the transient probabilities we compute the above probabilities numerically for various times. Also the following performance measures are evaluated numerically.

1. Expected number of customers in the system:

$$E[N(t)] = \sum_{j=1}^{\infty} \sum_{i=s+1}^S j\pi(j,i)(t).$$

2. Variance of number of customers in the system:

$$E[N(t)^2] - \{E[N(t)]\}^2.$$

3. Expected Inventory Level (EIL) in the system:

$$E[L(t)] = \sum_{j=1}^{\infty} \sum_{i=s+1}^S i\pi(j,i)(t) + \sum_{i=s}^{S-1} i\pi(0,i)(t).$$

4. Effective Flush out Rate:

$$\text{EFR} = \sum_{j=1}^{\infty} \sum_{i=s+1}^S \mu\pi(j,i)(t).$$

5. Probability, that the system is empty:

$$P[N(t) = 0] = 1 - \sum_{j=1}^{\infty} \sum_{i=s+1}^S \pi(j,i)(t).$$

6. Probability, that the system is not empty:

$$P[N(t) > 0] = \sum_{j=1}^{\infty} \sum_{i=s+1}^S \pi(j,i)(t).$$

Steady state analysis

Stability condition

Define $A = A_0 + A_1 + A_2$ and let $\pi = (\pi_1, \pi_2, \dots, \pi_Q)$ be the steady state vector of the generator matrix A . The QBD process with generator matrix T is stable, if and only if, the rate of drift to the left is larger than rate of drift to the right; that is $\pi A_0 e < \pi A_2 e$ (see Neuts) that is, if and only if, $\frac{\lambda}{\mu} < 1$.

Computation of steady state vector

We find the steady state vector of Ω explicitly. Let $\pi = (\pi_0, \pi_1, \dots)$ be the steady state vector, where

$$\pi_0 = (\pi_0(0, s), \pi_0(0, s + 1), \dots, \pi_0(0, S - 1))$$

and $\pi_i = (\pi_i(i, s + 1), \pi_i(i, s + 2), \dots, \pi_i(i, S)), i \geq 1$.

$$\pi T = 0 \Rightarrow \pi_0 B_0 + \pi_1 A_2 + \pi_2 A_2 + \dots = 0;$$

$$\pi_0 B_1 + \pi_1 A_1 = 0; \pi_i A_0 + \pi_{i+1} A_1 = 0, i \geq 1.$$

From the above equations we easily obtain

$$\pi_0(0, s) = \pi_0(0, s + 1) = \dots = \pi_0(0, S - 1)$$

and $\pi_i(i, s + 1) = \pi_i(i, s + 2) = \dots = \pi_i(i, S)$

for all $i \geq 1$.

For solving the above system of equations we first consider a system where units arrive according to a Poisson Process with rate λ and where service is exponentially distributed with rate μ , but when it ends the whole buffer is emptied. Let $x = (x(0), x(1), \dots)$ be the steady state of this system. Then

$$x(j) = p(1 - p)^j, \text{ where } p = \frac{\lambda}{\lambda + \mu}.$$

Now $\pi_0(0, k) = \frac{1}{Q} x(0), s \leq k \leq S - 1$

and $\pi_i(i, k) = \frac{1}{Q} x(i), s + 1 \leq k \leq S, i \geq 1$.

Numerical Illustration

Variation of various performance measures with arrival rate

$$\mu = 2, s = 2, S = 8, t = 1$$

λ	$E[N(t)]$	$V[N(t)]$	EIL	$P[N(t)] = 0$	$P[N(t)] > 0$
3	1.295	2.2723	5.1353	0.404	0.596
4	1.7139	3.3825	5.3816	0.3347	0.6653
5	2.0981	4.5106	5.4983	0.2851	0.7149
6	2.4164	5.5319	5.5053	0.2468	0.7532
7	2.6434	6.3825	5.4051	0.2151	0.7849
8	2.7692	7.058	5.2043	0.1875	0.8125

$$\mu = 2, s = 2, S = 8, t = 5$$

λ	$E[N(t)]$	$V[N(t)]$	EIL	$P[N(t)] = 0$	$P[N(t)] > 0$
3	1.386	3.1113	3.8756	0.3841	0.6159
4	1.6093	4.1062	3.8713	0.2918	0.7082
5	1.6169	4.7523	3.4414	0.2125	0.7875
6	1.4434	4.9308	2.7573	0.1468	0.8534
7	1.1662	4.587	2.0209	0.096	0.904
8	0.8646	3.8304	1.3758	0.0595	0.9405

$$\mu = 2, s = 2, S = 8, t = 10$$

λ	$E[N(t)]$	$V[N(t)]$	EIL	$P[N(t)] = 0$	$P[N(t)] > 0$
3	1.3027	3.0328	3.6708	0.361	0.639
4	1.3382	3.7773	3.1544	0.2427	0.7573
5	1.0988	3.7988	2.2496	0.1444	0.8556
6	0.7415	3.0534	1.3617	0.0754	0.9246
7	0.4211	1.9702	0.7145	0.0342	0.9653
8	0.2054	1.0453	0.3283	0.0141	0.9859

$$\mu = 2, s = 2, S = 8, t = 15$$

λ	$E[N(t)]$	$V[N(t)]$	EIL	$P[N(t)] = 0$	$P[N(t)] > 0$
3	1.2244	2.9464	3.4617	0.3395	0.6607
4	1.1128	3.3919	2.6062	0.2018	0.7982
5	0.7467	2.8444	1.5234	0.0981	0.9019
6	0.3809	1.7059	0.7022	0.0387	0.9613
7	0.1521	0.7524	0.2591	0.0125	0.9875
8	0.0488	0.2559	0.0779	0.0034	0.9966

Variation of performance measure with service rate

$$\lambda = 2, s = 2, S = 8, t = 1$$

μ	$E[N(t)]$	$V[N(t)]$	EIL	$P[N(t)] = 0$	$P[N(t)] > 0$
3	0.6334	0.9437	6.0913	0.6027	0.3973
4	0.4908	0.704	3.6973	0.6675	0.3325
5	0.3973	0.5465	3.4208	0.7145	0.2855
6	0.3325	0.4403	3.2182	0.7501	0.2499
7	0.2855	0.366	3.0645	0.7778	0.2222
8	0.2499	0.3121	2.9442	0.8	0.2

$$\lambda = 2, s = 2, S = 8, t = 5$$

μ	$E[N(t)]$	$V[N(t)]$	EIL	$P[N(t)] = 0$	$P[N(t)] > 0$
3	0.6665	1.1001	3.3558	0.5994	0.4006
4	0.4997	0.7482	3.2042	0.6665	0.3335
5	0.3999	0.5596	3.0482	0.7142	0.2858
6	0.3333	0.4443	2.9151	0.75	0.25
7	0.2857	0.3673	2.808	0.7778	0.2222
8	0.25	0.3125	2.7222	0.8	0.2

$$\lambda = 2, s = 2, S = 8, t = 10$$

μ	$E[N(t)]$	$V[N(t)]$	EIL	$P[N(t)] = 0$	$P[N(t)] > 0$
3	0.6639	1.0991	3.3761	0.5985	0.4015
4	0.4995	0.748	3.174	0.6663	0.3337
5	0.3995	0.5596	3.0057	0.7142	0.2858
6	0.3333	0.4443	2.8771	0.75	0.25
7	0.2857	0.3673	2.7781	0.7778	0.2222
8	0.25	0.3125	2.6998	0.8	0.2

Conclusions

In this paper we could derive an explicit expression for the transient probability vector of an inventory queueing model. We wish to extend this paper by considering positive lead time as well which may have several applications in real life situations.

REFERENCES

1. *Krishna Kumar B., Arivudainambi D.* Transient solution of an M/M/1 queue with catastrophes // *Comput. Math. Appl.* 2000. V. 40. P. 1233–1240.
2. *Krishna Kumar B., Krishnamoorthy A., Pavai Madheswari S., Sadiq Basha S.* Transient analysis of a single server queue with catastrophes, failures and repairs // *Queueing Systems.* 2007.
3. *Chao X.* A queueing network model with catastrophes and product form solution // *OR Lett.* 1995. V. 18. P. 75–79.
4. *Krishnamoorthy A., Dhanya Shajin, Viswanath C. Narayanan.* Inventory with Potive Service Time: A Survey Queueing Theory 2 / coordinated by Vladimir Anisimov, Nikolaos Limnios. ISTE Editions, 2020. P. 201–238.

Simulation of retrial queueing system M/G/1 with impatient customers, collisions and unreliable server*

János Sztrik¹, Ádám Tóth¹,
Elena Danilyuk², Svetlana Moiseeva²

¹University of Debrecen, Debrecen, Hungary

²Tomsk State University, Tomsk, Russia

In this paper, we consider a retrial queueing system of M/G/1 type with an unreliable server, collisions, and impatient customers. The novelty of our work is to carry out a sensitivity analysis applying different distributions of service time of customers on significant performance measures for example on the probability of abandonment, the mean waiting time of an arbitrary, successfully served, impatient customer, etc. The service, retrial, impatience, operation, and repair times are supposed to be independent of each other.

In the paper [1] a retrial queueing system of M/M/1 type with Poisson flow of arrivals, impatient customers, collisions, and unreliable service device is presented. In that, an asymptotic analysis method is used to define the stationary distribution of the number of customers in the orbit. We investigate the same model as in [1], but the results are gathered by our simulation program package. With this approach, it is possible to calculate performance measures that can not be determined or almost impossible to give exact formulas using numerical or asymptotic analysis. Various software packages exist which are capable to describe and perform an evaluation of complex systems if all the random variables are exponentially distributed but undoubtedly the usage of simulation has a tremendous advantage: besides exponential, any other distribution can be integrated into the code.

In real life in many cases customers encountering the service units in busy state may make a decision to attempt to be served after some random time remaining in the system. Instead of residing in a queue these customers

* The work of Dr. János Sztrik is supported by the EFOP-3.6.1-16-2016-00022 project. The project is co-financed by the European Union and the European Social Fund. The work of Ádám Tóth is supported by the ÚNKP-20-4 new national excellence program of the ministry for innovation and technology from the source of the national research, development and innovation fund.

are located in a virtual waiting room called orbit and can be modeled with retrial queues. Queuing systems with retrial queues are widely used tools modelling emerging problems in major telecommunication systems, such as telephone switching systems or call centres. Many papers dealt with these types of systems which can be viewed in the following works like in [2–4].

Models with customers impatience in queues like the process of reneging and balking have been studied by various authors in the past. Most recent results about systems having the impatience property can be found for example in [5–7].

System model

We consider a queueing system of M/G/1 with collisions, impatience of the customers, and an unreliable server which is shown in Fig. 1. The system arrival process is characterized by the Poisson process with rate λ . The arriving customer occupies instantly the service unit in idle state and the distribution of its service is according to exponentially, gamma, Pareto, lognormal, hypo-exponentially, and hyper-exponentially distributed random variable with the same mean value and variance but with different parameters. Otherwise, it is forwarded toward the orbit. The retrial time of the requests is assumed to be exponentially distributed with a rate of σ . In the case of a busy server an arriving customer brings about a collision and both requests enter the orbit. It is supposed that the server is unreliable so it breaks down from time to time according to an exponential distribution with parameter γ_0 when the server is idle and with parameter γ_1 when it is busy. In that period generation of new requests continues but each of them is sent

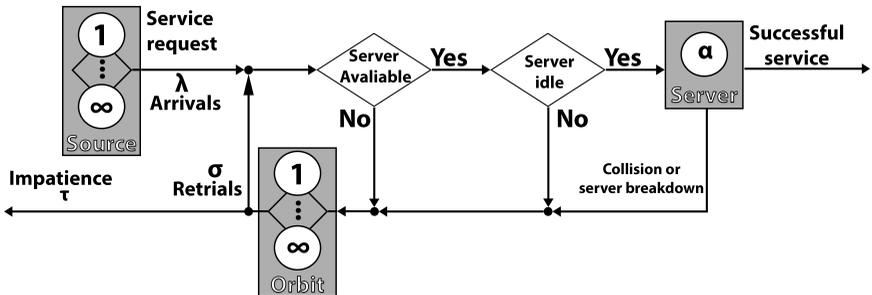


Fig. 1. System model

to orbit. After a breakdown, it is immediately sent for repair and the recovery process is also an exponential random variable with the rate γ_2 . Every customer possesses an «impatience» property meaning that a customer may depart from the system earlier after waiting a random time in the orbit. The distribution of the impatient time follows an exponential distribution with parameter τ . In this unreliable model after interruption or breakdown, it is supposed that requests immediately are placed in the orbit. Every service is independent of the other service including the interrupted ones, too.

Simulation results

To obtain the results of our simulation program a statistic package is used that was developed by Andrea Francini in 1994 [8]. With the help of this tool, it is possible to make a quantitative estimation of the mean and variance values of the desired variables using the method of batch means. There are n observations in every batch and the useful run is divided into a predetermined number of batches. In order for the estimation to work correctly, the batches are necessary to be long enough and approximately independent. It is one of the most popular confidence interval techniques for a steady-state mean of a process. The following works contain more detailed information about this method in [9]. The simulations are performed with a confidence level of 99.9%. The relative half-width of the confidence interval required to stop the simulation run is 0.00001.

The realization of the sensitivity analysis includes four different distributions of service time to compare the performance measures with each other. In every case, the parameters are selected in a way that the mean and variance would be equal. To accomplish that we applied a fitting process that is required to be done and [10] contains detailed information about the whole process describing every used distribution. Two scenarios are developed to investigate the effect of the various distributions. Table 2 shows the chosen parameters of the distribution of service time while Table 1 the values of other parameters. In the first one, the squared coefficient of variation is greater than one and the following distributions are used: hyper-exponential, gamma, Pareto, and lognormal. Results in connection with the second scenario (when the squared coefficient of variation is less than one) were also examined but because of the page limitation, these will be intended to be published in the extended version of the paper.

Table 1

Numerical values of model parameters

σ	γ_0	γ_1	γ_2	τ
0.01; 0.001	0.1	0.2	1	0.02; 0.002

Table 2

Parameters of service time of incoming customers

Distribution	Gamma	Hyper-exponential	Pareto	Lognormal
Parameters	$\alpha = 0.0816326$	$p = 0.4607213$	$\alpha = 2.0400156$	$m = -1.291998$
	$\beta = 0.0816326$	$\lambda_1 = 0.9214426$	$k = 0.5098076$	$\sigma = 1.6074817$
		$\lambda_2 = 1.0785573$		
Mean	1			
Variance	12.25			
Squared coefficient of variation	12.25			

In Figures 2 and 3 the comparison of steady-state distribution can be seen when the distribution of service time of the incoming customers is different. It demonstrates the probability ($P(i)$) of how many customers (i) residing in the orbit. Taking a closer look at the curves in more detail they coincide with normal distribution regardless of the used parameter setting. The figures also show the case of exponential distribution with the same mean as the other applied distributions. The mean number of customers in the orbit significantly differs from each other, at gamma distribution customers spend the fewest at Pareto distribution the highest time for waiting which is quite interesting.

The mean waiting time of an arbitrary customer is presented in the function of the arrival intensity of incoming customers in Figures 4 and 5. Even though the mean and the variance are identical huge gaps develop among the applied distributions. With the increment of the arrival intensity, the mean waiting time of an arbitrary customer increases as well. The same tendency is observable when we use other values of retrial and impatience time. The usage of gamma distribution results in lower mean waiting time compared to the others, especially versus gamma and Pareto distributions.

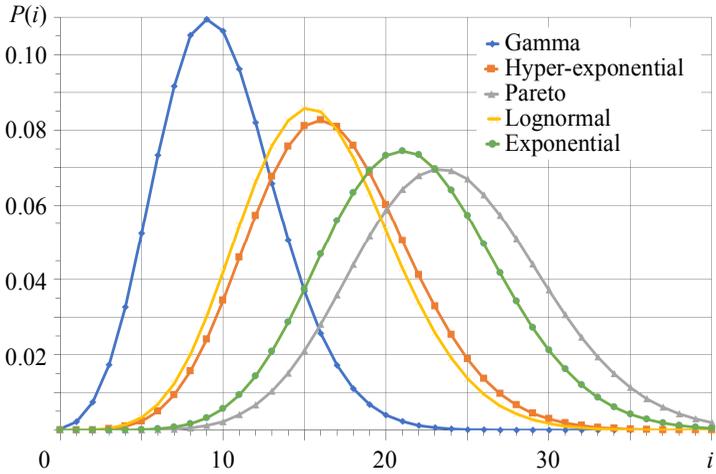


Fig. 2. Distribution of the number of customers in the orbit using various distributions, $\sigma = 0.01$, $\tau = 0.02$, $\lambda = 0.7$

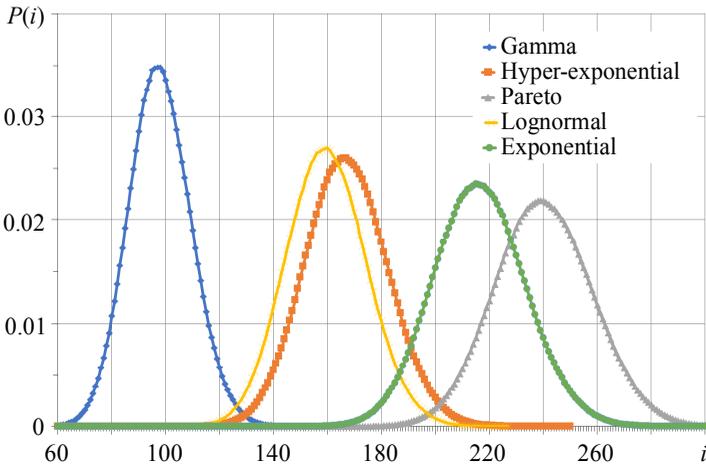


Fig. 3. Distribution of the number of customers in the orbit using various distributions, $\sigma = 0.001$, $\tau = 0.002$, $\lambda = 0.7$

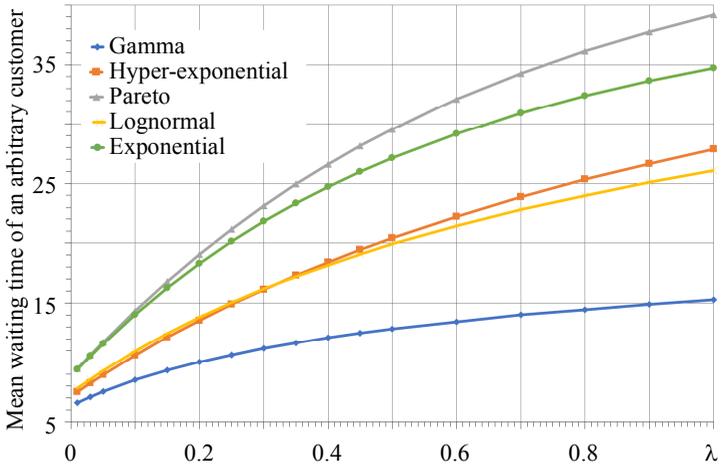


Fig. 4. Mean waiting time of an arbitrary customer vs arrival intensity using various distributions, $\sigma = 0.01$, $\tau = 0.02$

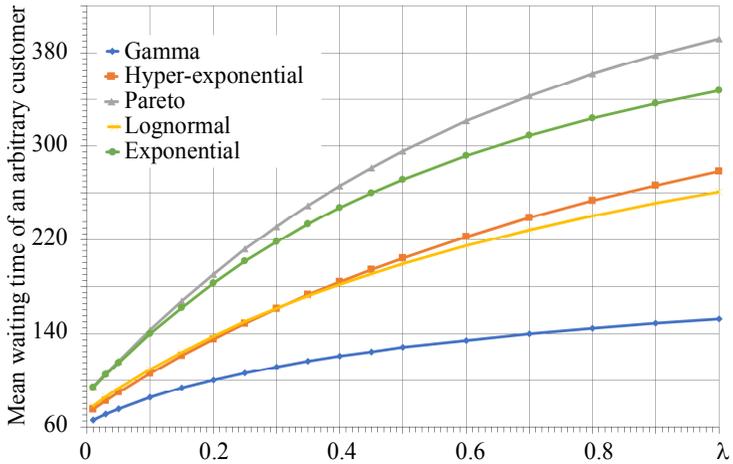


Fig. 5. Mean waiting time of an arbitrary customer vs arrival intensity using various distributions, $\sigma = 0.001$, $\tau = 0.002$

Conclusions

We studied the development of performance measures like the mean number of customers in the orbit or the mean waiting time of an arbitrary customer in a retrial queueing system of type M/G/1 with a non-reliable server and impatient customers in the orbit. Simulation has been carried out, the obtained results demonstrate that the number of customers in the orbit corresponds to the normal distribution in the case of every applied distribution. It is also displayed how the different distributions affect the performance measures despite the equality of mean value and variance when the squared coefficient of variation is more than one.

REFERENCES

1. *Danilyuk E.Yu., Moiseeva S.P., Sztrik J.* Asymptotic analysis of retrial queueing system M/M/1 with impatient customers, collisions and unreliable serv // J. Sib. Fed. Univ. Math. Phys. 2020. V. 13. No. 2. P. 218–230.
2. *Fiems D., Phung-Duc T.* Light-traffic analysis of random access systems without collisions // Ann. Oper. Res. 2019. V. 277. P. 311–327.
3. *Kim J., Kim B.* A survey of retrial queueing systems // Ann. Oper. Res. 2016. V. 247. P. 3–36.
4. *Tóth Á., Sztrik J.* Simulation of finite-source retrial queueing systems with collisions, non-reliable server and impatient customers in the orbit // Proceedings of 11th International Conference on Applied Informatics. 2020. V. 2650. P. 408–419.
5. *Jouini Q., Koole G., Roubos A.* Performance indicators for call centers with impatient customers // IIE Transactions. 2013. V. 45. No. 3. P. 341–354.
6. *Kumar S., Sharma S.* Transient performance analysis of a single server queueing model with retention of reneuing customers // Yugoslav Journal of Operations Research. 2018. V. 28. No. 3. P. 315–331.
7. *Kim C.S., Dudin S., Dudin A., Samouylov K.* Analysis of a semi-open queueing network with a state dependent marked Markovian arrival process, customers retrials and impatience // Mathematics. 2019. V. 7. P. 715–734.
8. *Francini A., Neri. F.* A comparison of methodologies for the stationary analysis of data gathered in the simulation of telecommunication networks // Proceedings of MASCOTS '96 – 4th International Workshop on Modeling, Analysis and Simulation of Computer and Telecommunication Systems. 1996. P. 116–122.
9. *Chen E., Kelton D.* A procedure for generating batch-means confidence intervals for simulation: Checking independence and normality // Simulation. 1996. V. 83. P. 683–694.
10. *Sztrik J., Tóth Á., Pintér Á., Bács Z.* Simulation of finite-source retrial queues with two-way communications to the orbit // Dudin A., Nazarov A., Moiseev A. (eds) Information Technologies and Mathematical Modelling. Queueing Theory and Applications. ITMM 2019. Communications in Computer and Information Science 2019. V. 1109. P. 270–284.

problems can arise like a power outage, human error, or other failures. Various factors have effect on the transmission rate of the wireless channel in a wireless communication scenario and these are apt to suffer transmission failure, interruptions throughout transferring the packets. Investigating retrial queueing systems with random server breakdowns and repairs has great importance as the operation of non-reliable systems modifies system characteristics and performance measures. In this paper, we assume that in the case of a failure of the server, the request generation from the source continues, and calls go to orbit.

The novelty of this investigation is to carry out sensitivity analysis using different distributions of impatient calls on performance measures. Different Figures help to understand the special features of the system. The paper is a continuation of the works [4, 6].

Mathematical model

A retrial queueing system of type M/M/1//N is considered with a non-reliable server and impatient customers as Fig. 1 illustrates. In the finite-source, N customers reside and each of them can generate calls towards the server with rate λ/N so the inter-request time is exponential with parameter λ/N . A customer cannot generate a new call until the previous call returns to the source. Every incoming customer has an random impatience time which determines how much time the customer spends in the orbit without fulfilling its service requirement. Exceeding this time results that the customer no longer waits for the service unit and departs without being served properly.

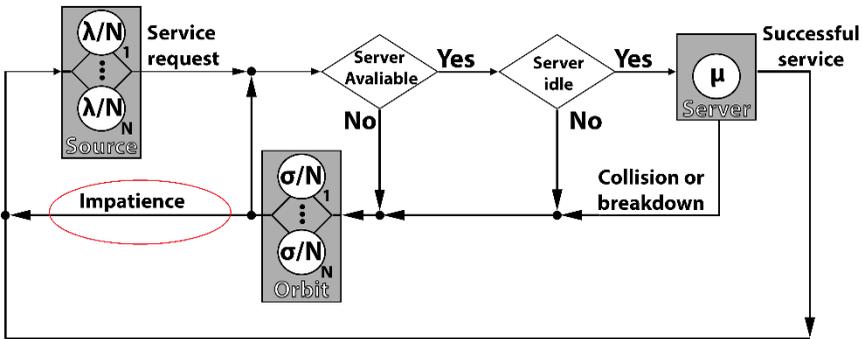


Fig. 1. Operational scheme of the system

This random time follows gamma, hypo-exponential, hyper-exponential, Pareto, and lognormal distribution with different parameters but with the same mean value. In absence of a waiting queue if an incoming customer finds the server in an idle state its service starts immediately. The service times of the customers are exponentially distributed with parameter μ . After its successful service customers return to the source. Encountering the service unit in a busy state an arriving (primary or repeated) customer involves into collision with customer under service and they both move into the orbit. After an exponentially distributed time with parameter σ/N customers located in the orbit make another attempt to get into the service.

The server is not reliable so from time to time it is supposed to break down. The lifetime is assumed to be an exponentially distributed random variable with a different parameter depending on whether the server is idle or busy, that is γ_0 and γ_1 , respectively. In other words, the probability of failure of the server is $\gamma_0 h + o(h)$, $\gamma_1 h + o(h)$ in any time interval $(t, t+h)$, respectively. The repair process starts immediately upon the breakdown that also follows an exponential distribution with parameter γ_2 . If server failure takes place during the service of a customer then it is transferred to the orbit. The source, service, retrial, impatience, life, and repair times are supposed to be independent of each other.

We aim to examine how the different distributions of impatient calls have an effect on the performance measure when the mean and variance are equal, respectively. The parameters are chosen in such a way that the squared coefficient of variation would be either greater than one or less than one.

For comparison hyper-exponential, hypo-exponential, gamma, lognormal and Pareto distributions are utilized besides the case when the mean value is constant. Several Figures are generated to illustrate the effect of the type of distribution on some important performance measures.

Simulation examples

In this section, we give some simulation results with the following input parameters. The mean of the impatient time is 500 and the squared coefficient of variation is greater than one.

N	λ/N	γ_0	γ_1	γ_2	σ/N	μ
100	0.01	0.1	0.1	1	0.01	1

Figure 2 shows the comparison of steady-state distribution of number of customers. This distribution can be approximated by a normal distribution, as it was noticed in similar systems, see for example in [5]. However, this Figure displays the difference between the applied distributions. Although the shape of the curves is almost the same the average number of customers in the system varies a little bit especially in the case of Pareto distribution and when the average impatience of calls is constant the mean is more compared to the others.

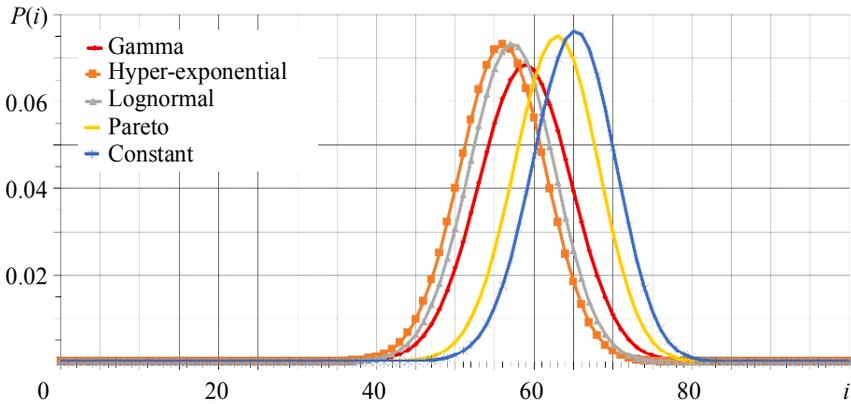


Fig. 2. Comparison of steady-state distributions

The mean response time of a successfully served customer, an impatient customer, and an arbitrary customer are shown in Figures 3 – 5, respectively in the function of arrival intensity. Of course, the mean response time of an arbitrary customer can be obtained with the help of the law of total expectation. The corresponding probabilities can be seen in Fig. 6. Interestingly, differences can be observed even though the first two moments are equal, especially in the case of the gamma distribution. Results clearly illustrate the effect of various distributions. We draw your attention to Fig. 4, where the mean waiting time of an impatient customer is much less than the average of the impatient time distribution that could have been expected.

Despite the increasing arrival intensity, the maximum property characteristic of a finite-source retrial queueing system occurs under suitable parameter settings.

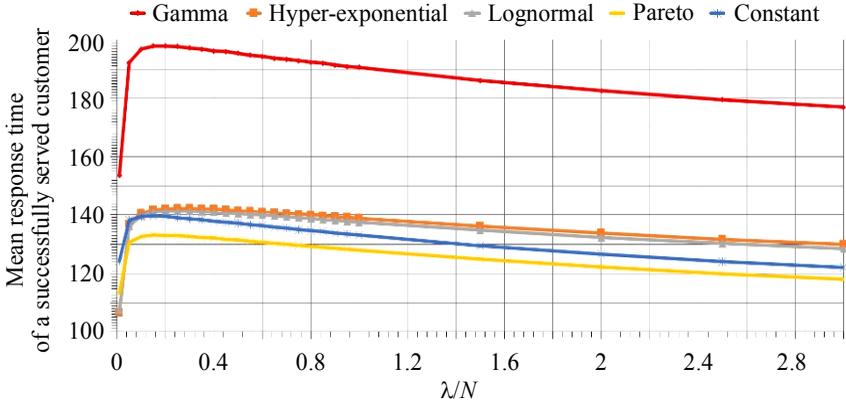


Fig. 3. Mean response time vs arrival intensity using various distributions

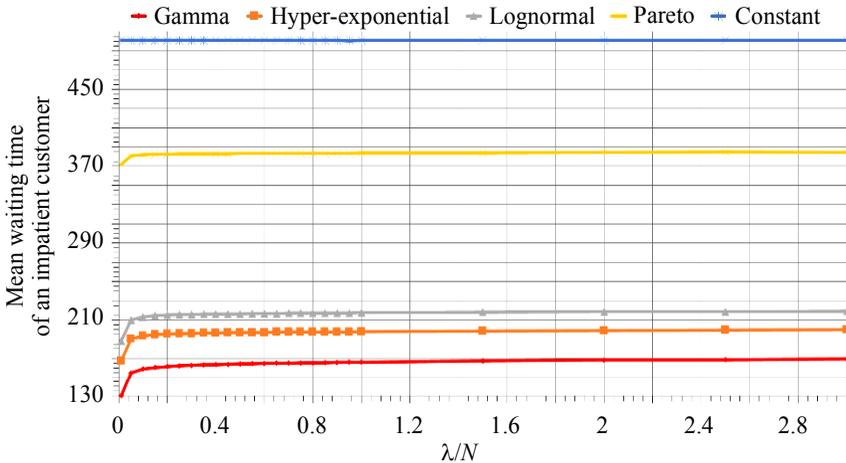


Fig. 4. Mean response time vs arrival intensity using various distributions

Figure 6 demonstrates how the probability of abandonment of a customer changes with the increment of the arrival intensity. By probability of abandonment, we mean the probability that a customer leaves the system without getting its full-service requirement (through the orbit). After a slow increase of the value of this performance measure, it stagnates which is true for every used distribution of impatience of calls but they differ significantly from each other. At gamma distribution, the tendency of leaving the system

earlier is much higher than the others especially compared to at constant mean of impatience of calls. Here the disparity is much higher among the applied distributions compared to the previous figures.

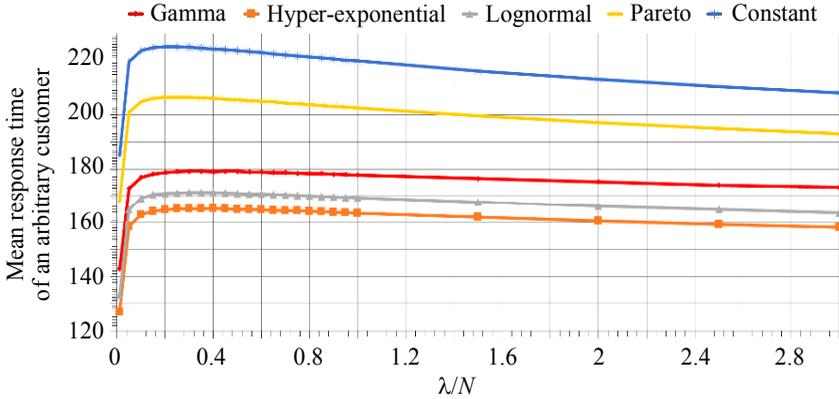


Fig. 5. Mean response time vs arrival intensity using various distributions

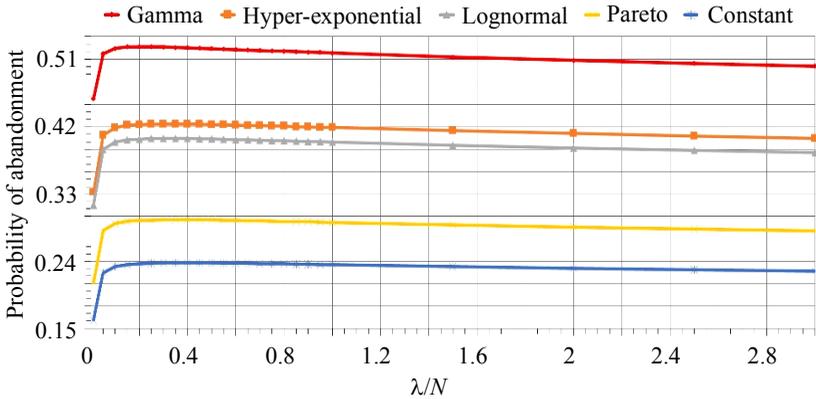


Fig. 6. Probability of abandonment of a customer vs arrival intensity using various distributions

The case when the squared coefficient of variation is less than one hopefully will be an investigation in the extended paper.

Conclusions

In this paper, a finite-source retrial queueing system was presented with a non-reliable server, collisions, and impatient customers in the orbit. The obtained results fully demonstrated how essential is the selection of distribution of impatient calls because it has a great influence on the system characteristics despite the fact that the mean and the variance are the same. Figures in connection with the probability of abandonment clearly assure this phenomenon. Results evidently indicated the distinction was noticeable and significant among the performance measures having the same mean and variance of different distributions when the squared coefficient of variation is greater than one. Soon, we would like to investigate the waiting time distribution of this system with help of asymptotic methods as N tends to infinity and that is the reason why we used rates λ/N and σ/N .

REFERENCES

1. *Artalejo J.R. Pla V.* On the impact of customer balking, impatience and retrials in telecommunication systems // *Computers & Mathematics with Applications*. 2009. V. 57. No. 2. P. 217–229.
2. *Haight. F.A.* Queueing with reneging // *Metrika*. 1959. V. 2. No. 1. P. 186–197.
3. *Kim J.S.* Retrial queueing system with collision and impatience // *Communications of the Korean Mathematical Society*. 2010. V. 25. No. 4. P. 647–653.
4. *Kuki A. Bérczes T. Tóth Á. Sztrik J.* Numerical analysis of finite source Markov retrial system with non-reliable server, collision, and impatient customers // *Annales Mathematicae et Informaticae*. 2020. V. 51. P. 53–63.
5. *Nazarov A., Sztrik J., Kvach A.* A survey of recent results in finite-source retrial queues with collisions // *Information Technologies and Mathematical Modelling. Queueing Theory and Applications*. Springer, 2018. V. 912. P. 1–15.
6. *Tóth Á., Sztrik J.* Simulation of finite-source retrial queueing systems with collisions, non-reliable server, and impatient customers in the orbit // *Proceedings of 11th International Conference on Applied Informatics, Eger, Hungary, CEUR Workshop Proceedings (CEUR-WS.org)*. 2020. V. 2650. P. 408–419.

Исследование двумерного выходящего потока марковской модели узла обработки запросов с повторными обращениями и вызываемыми заявками*

Алексей Благинин, Иван Лапатин, Анатолий Назаров

*Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

В работе предлагается рассмотреть двумерный выходящий поток системы массового обслуживания [1] с повторными вызовами [2] и вызываемыми заявками [3]. Такую систему можно интерпретировать как узел сети связи со случайным множественным доступом, который в свободное от обработки запросов время может запросить самодиагностику или другую процедуру, которая будет продолжаться случайное время. Также данная система подходит для моделирования узла связи с разными типами заявок. Заявки одного типа не теряются и в любом случае получают обслуживание, а заявки другого типа получают обслуживания только при свободном ресурсе.

Отдельные узлы образуют модель сети связи, в которой выходящий поток одного узла является входящим для другого. В случае разнотипных заявок после обслуживания на определенном узле заявки разных типов уходят по своим маршрутам. Поэтому результаты исследования выходящих потоков систем массового обслуживания широко применимы для проектирования реальных систем передачи данных и анализа сложных процессов, состоящих из нескольких этапов.

В данной работе для исследования системы используется метод асимптотического анализа для нахождения вида предельного двумерного распределения числа обслуженных заявок входящего потока и числа обслуженных вызываемых заявок за некоторое время t при условии большой задержки заявок на орбите [4].

Математическая модель

Рассмотрим систему массового обслуживания с одним обслуживающим прибором, на вход которого поступает простейший поток заявок с интенсивностью λ . Заявка входящего потока, поступая в систему и

* Работа выполнена при финансовой поддержке РФФИ в рамках научного проекта № 18-01-00277.

обнаруживая прибор свободным, занимает его, где обслуживается экспоненциально распределенное случайное время с параметром μ_1 . Если же при поступлении в систему заявка входящего потока обнаруживает прибор занятым, она мгновенно уходит на орбиту и осуществляет там случайную задержку в течение экспоненциально распределенного времени с параметром σ , после чего вновь обращается к прибору. В свободное от обслуживания заявок из входящего потока время прибор сам вызывает заявки другого типа с интенсивностью α и обслуживает их в течение экспоненциально-распределенного времени с параметром μ_2 .

Введем следующие обозначения: $i(t)$ – количество заявок на орбите в момент времени t , $k(t)$ – состояние прибора в момент времени t : 0 – прибор свободен, 1 – прибор занят обслуживанием заявки с входящего потока, 2 – прибор занят обслуживанием вызываемой заявки; $m_1(t)$ – количество обслуженных заявок из входящего потока к моменту времени t , $m_2(t)$ – количество обслуженных заявок, вызванных прибором, к моменту времени t .

Система дифференциальных уравнений Колмогорова

Рассмотрим четырехмерный марковский процесс:

$$\{k(t), i(t), m_1(t), m_2(t)\}.$$

Для вероятностей $P\{k(t) = k, i(t) = i, m_1(t) = m_1, m_2(t) = m_2\} = P_k(i, m_1, m_2, t)$ составим систему дифференциальных уравнений Колмогорова

$$\begin{aligned} \frac{\partial P_0(i, m_1, m_2, t)}{\partial t} &= -(\lambda + i\sigma + \alpha)P_0(i, m_1, m_2, t) + \\ &+ P_1(i, m_1 - 1, m_2, t)\mu_1 + P_2(i, m_1, m_2 - 1, t)\mu_2, \\ \frac{\partial P_1(i, m_1, m_2, t)}{\partial t} &= -(\lambda + \mu_1)P_1(i, m_1, m_2, t) + (i + 1)\sigma P_0(i + 1, m_1, m_2, t) + \\ &+ \lambda P_0(i, m_1, m_2, t) + \lambda P_1(i - 1, m_1, m_2, t), \end{aligned} \quad (1)$$

$$\frac{\partial P_2(i, m_1, m_2, t)}{\partial t} = -(\lambda + \mu_2)P_2(i, m_1, m_2, t) + \lambda P_2(i - 1, m_1, m_2, t) + \alpha P_0(i, m_1, m_2, t).$$

Решить данную систему аналитически не получится, так как это система бесконечного числа дифференциальных конечно-разностных уравнений с переменными коэффициентами.

Для того чтобы перейти к конечному числу уравнений, введем частичные характеристические функции, обозначив $j = \sqrt{-1}$,

$$H_k(u, u_1, u_2, t) = \sum_{i=0}^{\infty} \sum_{m_1=0}^{\infty} \sum_{m_2=0}^{\infty} e^{ju_i} e^{ju_1 m_1} e^{ju_2 m_2} P_k(i, m_1, m_2, t).$$

Тогда перепишем систему (1) в виде

$$\begin{aligned} \frac{\partial H_0(u, u_1, u_2, t)}{\partial t} &= -(\lambda + \alpha)H_0(u, u_1, u_2, t) + j\sigma \frac{\partial H_0(u, u_1, u_2, t)}{\partial u} + \\ &+ \mu_1 e^{ju} H_1(u, u_1, u_2, t) + \mu_2 e^{ju_2} H_2(u, u_1, u_2, t), \\ \frac{\partial H_1(u, u_1, u_2, t)}{\partial t} &= -(\lambda + \mu_1)H_1(u, u_1, u_2, t) - \sigma j e^{-ju} \frac{\partial H_0(u, u_1, u_2, t)}{\partial u} + \\ &+ \lambda H_0(u, u_1, u_2, t) + \lambda e^{ju} H_1(u, u_1, u_2, t), \end{aligned} \quad (2)$$

$$\frac{\partial H_2(u, u_1, u_2, t)}{\partial t} = -(\lambda + \mu_2)H_2(u, u_1, u_2, t) + \lambda e^{ju} H_2(u, u_1, u_2, t) + \alpha H_0(u, u_1, u_2, t).$$

Асимптотический анализ

Полученную систему дифференциальных уравнений в частных производных (2) будем решать методом асимптотического анализа в предельном условии большой задержки заявок на орбите ($\sigma \rightarrow 0$).

Обозначим $\varepsilon = \sigma, u = \varepsilon w, F_k(w, u_1, u_2, t, \varepsilon) = H_k(u, u_1, u_2, t)$, тогда система запишется в виде

$$\begin{aligned} \frac{\partial F_0(w, u_1, u_2, t, \varepsilon)}{\partial t} &= -(\lambda + \alpha)F_0(w, u_1, u_2, t, \varepsilon) + j \frac{\partial F_0(w, u_1, u_2, t, \varepsilon)}{\partial w} + \\ &+ \mu_1 e^{ju} F_1(w, u_1, u_2, t, \varepsilon) + \mu_2 e^{ju_2} F_2(w, u_1, u_2, t, \varepsilon), \\ \frac{\partial F_1(w, u_1, u_2, t, \varepsilon)}{\partial t} &= -(\lambda + \mu_1)F_1(w, u_1, u_2, t, \varepsilon) - j e^{-j\varepsilon w} \frac{\partial F_0(w, u_1, u_2, t, \varepsilon)}{\partial w} + \\ &+ \lambda F_0(w, u_1, u_2, t, \varepsilon) + \lambda e^{j\varepsilon w} F_1(w, u_1, u_2, t, \varepsilon), \\ \frac{\partial F_2(w, u_1, u_2, t, \varepsilon)}{\partial t} &= -(\lambda + \mu_2)F_2(w, u_1, u_2, t, \varepsilon) + \lambda e^{j\varepsilon w} F_2(w, u_1, u_2, t, \varepsilon) + \\ &+ \alpha F_0(w, u_1, u_2, t, \varepsilon). \end{aligned} \quad (3)$$

Заметим, что, используя условие согласованности для многомерных распределений, характеристическая функция процессов $m_1(t)$ и $m_2(t)$ выражается следующим образом через введенные функции

$$M\{\exp(ju_1 m_1(t)) \exp(ju_2 m_2(t))\} = \sum_{k=0}^2 H_k(0, u_1, u_2, t) = \sum_{k=0}^2 F_k(0, u_1, u_2, t, \varepsilon).$$

Теорема. Асимптотическое приближение двумерной характеристической функции числа обслуженных заявок входящего потока и числа обслуженных вызываемых заявок за некоторое время t имеет вид

$$\begin{aligned} \mathbf{F}(u_1, u_2, t) &= \lim_{\sigma \rightarrow 0} M\{\exp(ju_1 m_1(t)) \exp(ju_2 m_2(t))\} = \\ &= \lim_{\varepsilon \rightarrow 0} \sum_{k=0}^2 F_k(0, u_1, u_2, t, \varepsilon) = \mathbf{R} \cdot \exp\{\mathbf{G}(u_1, u_2)t\} \cdot \mathbf{E}, \end{aligned} \quad (4)$$

где

$$\mathbf{G}(u_1, u_2) = \begin{pmatrix} -(\lambda + \alpha + \kappa) & \mu_1 e^{ju_1} & \mu_2 e^{ju_2} \\ \kappa + \lambda & -\mu_1 & 0 \\ \alpha & 0 & -\mu_2 \end{pmatrix}^T,$$

вектор-строка $\mathbf{R} = \{R_0, R_1, R_2\}$ – стационарное распределение вероятностей состояния прибора [5],

$$\mathbf{R} = \left\{ \frac{\mu_2(\mu_1 - \lambda)}{\mu_1(\mu_2 + \alpha)}, \frac{\lambda}{\mu_1}, \frac{\alpha(\mu_1 - \lambda)}{\mu_1(\mu_2 + \alpha)} \right\},$$

κ – нормированное среднее значение числа заявок на орбите [5]

$$\kappa = \frac{\lambda(\lambda\mu_2 + \alpha\mu_1)}{\mu_2(\mu_1 - \lambda)},$$

а \mathbf{E} – единичный вектор-столбец соответствующей размерности.

Данная формула позволяет найти асимптотическое приближение характеристической функции числа заявок из входящего потока, обслуженных системой за некоторое время t .

Переход к распределению вероятности числа обслуженных заявок

Характеристическая функция (4) полностью описывает процессы $m_1(t)$ и $m_2(t)$, однако для проведения вычислений необходимо получение из нее распределения вероятности.

Для обращения функции применим обратное преобразование Фурье для дискретных случайных величин:

$$P(m_1, m_2, t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{-i u_1 m_1} e^{-i u_2 m_2} \mathbf{F}(u_1, u_2, t) du_1 du_2.$$

Полученное распределение характеризует вероятность обслуживания m_1 заявок из входящего потока и m_2 заявок, вызванных прибором, к моменту времени t в рассматриваемой системе.

Коэффициент корреляции

Полученное асимптотическое приближение характеристической функции (4) позволяет нам подробнее изучить выходящие потоки рассматриваемой системы, а именно – найти корреляционную зависимость случайных процессов $m_1(t)$ и $m_2(t)$.

Рассмотрим нахождение коэффициента корреляции, который будет зависеть от параметра t :

$$r(t) = \frac{\text{cov}(m_1(t), m_2(t))}{\sqrt{D(m_1(t))} \sqrt{D(m_2(t))}}.$$

Воспользуемся свойством характеристической функции о существовании ее n -й производной, соответствующей n -му начальному моменту случайной величины. Тогда ковариация и дисперсия будут вычисляться следующим образом:

$$\begin{aligned} \text{cov}(t) &= M_{u_1 u_2}(t) - M_{u_1}(t) M_{u_2}(t) = \\ &= \frac{1}{j^2} \frac{\partial^2}{\partial u_1 \partial u_2} \mathbf{F}(u_1, u_2, t) \Big|_{u_2=0}^{u_1=0} - \frac{1}{j^2} \frac{\partial}{\partial u_1} \mathbf{F}(u_1, u_2, t) \Big|_{u_2=0}^{u_1=0} \frac{\partial}{\partial u_2} \mathbf{F}(u_1, u_2, t) \Big|_{u_2=0}^{u_1=0}, \\ D_{u_1}(t) &= M_{u_1}^2(t) - (M_{u_1}(t))^2 = \\ &= \frac{1}{j^2} \frac{\partial^2}{\partial u_1^2} \mathbf{F}(u_1, u_2, t) \Big|_{u_2=0}^{u_1=0} - \left(\frac{1}{j} \frac{\partial}{\partial u_1} \mathbf{F}(u_1, u_2, t) \Big|_{u_2=0}^{u_1=0} \right)^2, \\ D_{u_2}(t) &= M_{u_2}^2(t) - (M_{u_2}(t))^2 = \\ &= \frac{1}{j^2} \frac{\partial^2}{\partial u_2^2} \mathbf{F}(u_1, u_2, t) \Big|_{u_2=0}^{u_1=0} - \left(\frac{1}{j} \frac{\partial}{\partial u_2} \mathbf{F}(u_1, u_2, t) \Big|_{u_2=0}^{u_1=0} \right)^2. \end{aligned}$$

Полученные формулы позволяют численно исследовать поведение системы при разных параметрах.

Численный пример

Приведем результаты численного примера, показывающие, как изменяется корреляционная зависимость случайных процессов $m_1(t)$ и $m_2(t)$ при разных параметрах системы:

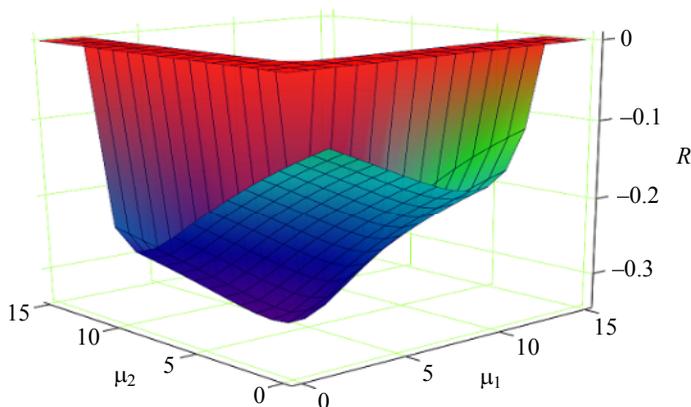


Рис 1. Изменение коэффициента корреляции процессов $m_1(t)$ и $m_2(t)$ в зависимости от параметров μ_1 и μ_2

Параметры системы: $\lambda = 1, \alpha = 3, t = 150, \mu_1 \in [0, 15], \mu_2 \in [0, 15]$. При заданных параметрах системы видно, что наибольшее абсолютное значение коэффициента корреляции процессов $m_1(t)$ и $m_2(t)$ достигается при $\mu_1 = 4, \mu_2 = 5$.

Заключение

Таким образом, в работе получена формула для нахождения асимптотического приближения двумерной характеристической функции числа заявок из входящего потока, закончивших обслуживание в марковской системе массового обслуживания с повторными вызовами и вызываемыми заявками в условии большой задержки заявок на орбите. Был проведен численный эксперимент, показывающий корреляционную зависимость случайных процессов $m_1(t)$ и $m_2(t)$. Также, было показано, что с помощью обратного преобразования Фурье возможно вычисление значений распределения вероятностей по найденному асимптотическому приближению характеристической функции процессов $m_1(t)$ и $m_2(t)$.

ЛИТЕРАТУРА

1. *Kendall D.G.* Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain // *The Annals of Mathematical Statistics*. 1953. С. 338–354.
2. *Artalejo J.R., Gomez-Corral A.* Retrial Queueing Systems: A Computational Approach. Berlin; Heidelberg: Springer, 2008.
3. *Kulkarni V.G.* On queueing systems by retrials // *Journal of Applied Probability*. 1983. V. 20. No. 2. P. 380–389
4. *Nawel Gharbi, Claude Dutheillet.* An algorithmic approach for analysis of finite-source retrial systems with unreliable servers // *Computers & Mathematics with Applications*. V. 62. Iss. 6. 2011. P. 2535–2546
5. *Nazarov A., Paul S., Gudkova I.* Asymptotic analysis of markovian retrial queue with two-way communication under low rate of retrial condition // *Proceedings 31st European Conference on Modelling and Simulation, ECMS 2017*. Budapest, 2017. P. 687–693.
6. *Burke P.J.* The output process of a stationary M/M/s queueing system // *The Annals of Mathematical Statistics*. 1968. T. 39. № 4. С. 1144–1152.

СМО вида $GI^{(k)}/GI/\infty$ с групповым обслуживанием

Анна Бояркина, Светлана Моисеева,
Ирина Туренова, Алексей Шкуркин

*Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

В данной статье рассматривается бесконечнолинейная система массового обслуживания с групповым рекуррентным входящим потоком, для исследования которой применяются метод динамического просеивания и асимптотический метод. Данная работа является обобщением результатов [1, 2] на случай неординарного рекуррентного входящего потока.

Математическая модель

Рассмотрим СМО вида $GI^{(k)}/GI/\infty$ с неординарным рекуррентным входящим потоком и групповым обслуживанием. Под групповым обслуживанием будем понимать следующее. Поступившая на обслуживание случайного размера группа заявок после обслуживания в том же составе покидает систему. Рассматриваемая система массового обслуживания имеет неограниченное число обслуживающих приборов, то есть является системой с моментальным обслуживанием (рис. 1).

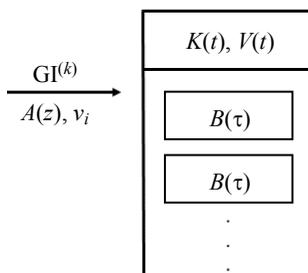


Рис. 1. СМО с неординарным рекуррентным входящим потоком и групповым обслуживанием

Для входящего рекуррентного потока моментов прихода групп определим случайную величину τ , как интервал между наступлением последовательных событий, все τ независимы и одинаково распределены. Запишем τ в виде $\tau = \frac{\xi}{N}$, где ξ – некоторая неотрицательная случайная величина с функцией распределения $A(z)$, величина $N > 0$, $N \rightarrow \infty$ – параметр растущей интенсивности потока. Тогда для функции распределения длин интервалов τ имеем

$$P\{\tau < z\} = P\left\{\frac{\xi}{N} < z\right\} = P\{\xi < Nz\} = A(Nz).$$

Для положительной случайной величины, определяемой функцией распределения вероятностей $A(z)$ существуют конечные математическое ожидание и дисперсия, то есть

$$a = \frac{1}{\lambda} = \int_0^{\infty} (1 - A(z)) dz, \quad \sigma^2 = \int_0^{\infty} (u - a)^2 dA(z). \quad (1)$$

Неординарность поступающих требований определяется дискретной случайной величиной ξ , заданной рядом распределения $P\{\xi = i\} = v_i$. Входящая группа заявок размерности i занимает любые свободные приборы, где происходит их обслуживание в течение случайного времени $\tau > 0$ с функцией распределения вероятностей $B(x) = P\{\tau < x\}$. По окончании обслуживания вся группа освобождает все приборы одновременно.

Ставится задача исследования числа занятых приборов в системе. Сложность аналитического исследования числа заключается в том, обязательное условие, что все заявки группы покидают систему одновременно. Следовательно, возникает необходимость отслеживать также и число поступивших групп.

Для решения задачи предлагается использовать авторский метод динамического просеивания [3, 4] в применении к ресурсным системам [5]. Указанный подход в сочетании с асимптотическими методами в дальнейшем позволит исследовать более сложные системы с коррелированными входящими потоками.

Метод динамического просеивания

Так как особенность системы предполагает уход заявок группами того же состава, то необходимо учитывать количество находящихся в системе групп, и общее число занятых дискретных ресурсов (приборов). Обозначим $V(t) = \sum_{i=1}^{K(t)} v_i$ – общее количество занятых приборов в системе в момент времени и пусть $K(t)$ – число групп в системе в момент времени $t, t \geq t_0$. Будем считать, что в начальный момент времени система была пуста.

Очевидно, что процесс $\{K(t), V(t)\}$ не будет являться марковским, так как время обслуживания есть положительная случайная величина, имеющая произвольное распределение и мы не знаем, сколько осталось времени до окончания обслуживания той или иной группы. Поэтому будет использоваться метод динамического просеивания [3], который позволяет учитывать только те группы, которые не закончили к рассматриваемому моменту времени своё обслуживание в системе.

В применении к рассматриваемой системе метод заключается в следующем. Фиксируется момент времени $T \geq t_0$. Вводятся динамические вероятности: $S(t, T) = 1 - B(T - t)$ – вероятность того, что группа размера v , поступившая в систему в момент времени $t < T$, к моменту времени T не закончит обслуживание и будет занимать v приборов, и противоположная вероятность $1 - S(T, t)$ для тех пачек, которые покинули. Обозначим $N(t)$ – число групп событий просеянного потока, а $W(t)$ – общее число занимаемых просеянными заявками дискретных ресурсов, наступивших на промежутке $[t_0, T)$, в просеянном потоке. Известно, что в момент времени T законы распределения вероятностей двумерных случайных величин $\{K(t), V(t)\}$ и $\{N(t), W(t)\}$ совпадают [4].

Для построения системы дифференциальных уравнений Колмогорова воспользуемся методом введения дополнительной компоненты. Для марковизации двумерного случайного процесса $\{N(t), W(t)\}$ добавим компоненту $z(t)$ – остаточное время от момента t до момента наступления следующего события в исходном рекуррентном потоке. Таким образом, трехмерный случайный процесс $\{z(t), N(t), W(t)\}$ является марковским, следовательно, для его распределения вероятностей

$P\{z(t) < z, N(t) = n, W(t) = w\} = P(z, n, w, t)$, для $z > 0$, запишем систему дифференциальных уравнений Колмогорова для $z > 0, w, n = 0, 1, 2, \dots$:

$$\frac{1}{N} \frac{\partial P(z, n, w, t)}{\partial t} = \frac{\partial P(z, n, w, t)}{\partial z} + \frac{\partial P(0, n, w, t)}{\partial z} (A(z) - 1) + S(t) A(z) \left[\sum_{k=0}^w \frac{\partial P(0, n-1, w-k, t)}{\partial z} v_k - \frac{\partial P(0, n, w, t)}{\partial z} \right], \quad (2)$$

с начальным условием

$$P(z, n, w, t_0) = \begin{cases} R(z), & n = w = 0, \\ 0, & \text{иначе,} \end{cases}$$

где $R(z)$ – стационарное распределение вероятностей значений случайного процесса $z(t)$, определяемое выражением

$$R(z) = \lambda \int_0^z (1 - A(x)) dx, \quad \text{где } \lambda = \left[\int_0^{\infty} (1 - A(x)) dx \right]^{-1}.$$

Определим функцию вида

$$h(z, u, v, t) = \sum_{n=0}^{\infty} e^{jun} \sum_{w=0}^{\infty} e^{jvw} P(z, n, w, t), \quad z > 0, \quad (3)$$

которую будем называть частичной характеристической функцией.

Очевидно, что

$$h(\infty, u, v, t) = h(u, v, t) = M \left\{ e^{juN(t)} e^{jvW(t)} \right\} = \sum_{n=0}^{\infty} e^{jun} \sum_{w=0}^{\infty} e^{jvw} P(n, w, t).$$

Также характеристическую функцию числа событий в пачке, будем обозначать

$$V^*(v) = M \left\{ e^{jv\xi} \right\} = \sum_{k=0}^{\infty} e^{jvk} v_k.$$

Учитывая (3), уравнение (2) запишем в виде

$$\frac{1}{N} \frac{\partial h(z, u, v, t)}{\partial t} = \frac{\partial h(z, u, v, t)}{\partial z} + \frac{\partial h(0, u, v, t)}{\partial z} (A(z) - 1) + \frac{\partial h(0, u, v, t)}{\partial z} S(t) A(z) (e^{ju} V^*(v) - 1), \quad (4)$$

с начальным условием

$$h(z, u, v, t_0) = R(z). \quad (5)$$

Асимптотический анализ

Далее для нахождения асимптотического решения уравнения (4) с начальными условиями (5) воспользуемся методом асимптотического анализа при условии растущей интенсивности входящего потока [3, 4]. В работе доказаны теоремы.

Теорема 1. Асимптотическая характеристическая функция первого порядка трехмерного процесса $\{z(t), N(t), W(t)\}$ для системы $GI^{(K)}/GI/\infty$ при условии растущей интенсивности входящего потока имеет вид

$$h(z, u, v, t) \approx R(z) \exp \left\{ N\lambda [ju + jva] \int_{t_0}^t S(\tau) d\tau \right\},$$

где $a = M\{\xi\} = \sum_{i=0}^{\infty} i v_i$ – математическое ожидание размера группы.

Теорема 2. Асимптотическое при условии растущей интенсивности входящего потока совместное стационарное распределение вероятностей двумерного процесса числа групп, находящихся на обслуживании суммарного числа занятых приборов в системе $GI^{(K)}/GI/\infty$ является двумерным гауссовским распределением вероятностей с параметрами:

- вектором математических ожиданий

$$\mathbf{a} = [N\lambda b \quad N\lambda ab],$$

- ковариационной матрицей

$$\mathbf{K} = \begin{bmatrix} N(\lambda b + \kappa\beta) & aN(\lambda b + \kappa\beta) \\ aN(\lambda b + \kappa\beta) & N(\lambda a_2 b + \kappa a^2 \beta) \end{bmatrix},$$

где $b = \int_0^{\infty} (1 - B(\tau)) d\tau$ – математическое ожидание времени обслуживания

группы; $\beta = \int_0^{\infty} (1 - B(\tau))^2 d\tau$ – математическое ожидание минимума случайных величин времени обслуживания группы заявок; λ – интен-

сивность входящего рекуррентного потока моментов поступления групп заявок; $\kappa = \lambda^3 \left(\sigma^2 - \frac{1}{\lambda^2} \right)$.

Выводы

Таким образом, построена математическая модель системы массового обслуживания с рекуррентным неординарным входящим потоком и групповым обслуживанием типа $GI^{(K)} / GI / \infty$, отличие которой заключается в том, что поступаемые группы покидают систему в том же составе. Построена модификация метода асимптотического анализа для исследования общего числа занимаемых приборов в условии растущей интенсивности входящего потока. Показано, что исследуемый процесс является асимптотически нормальным.

ЛИТЕРАТУРА

1. *Tikhonenko O.M.* Generalized Erlang problem for queueing systems with bounded total size // *Probl. Peredachi Inf.* 2005. V. 41. No. 3. P. 64–75; *Problems Inform. Transmission.* 2005. V. 41. No. 3. P. 243–253.
2. *Tikhonenko O.M., Kempa W.* Queueing system with processor sharing and limited memory under control of the AQM mechanism // *Avtomat. i Telemekh.* 2015. No. 10. P. 90–105; *Autom. Remote Control.* 2015. V. 76. No. 10. 1784–1796.
3. *Моисеев А.Н., Назаров А.А.* Бесконечнолинейные системы и сети массового обслуживания. Томск: Изд-во НТЛ, 2015. 240 с.
4. *Lisovskaya E., Moiseeva S., Pagano M., Potatueva V.* Study of the MMPP/GI/ ∞ queueing system with random customers' capacities // *Информ. и её примен.* 2017. V. 11. No. 4. P. 109–117.
5. *Bushkova T., Pavlova E., Rozhkova S., Moiseeva S., Pagano M.* Resource queueing system MMPP(2,v)|GI2| ∞ with parallel service of multiple paired customers // *CCIS.* 2019. V. 1109. P. 136–149.
6. *Galileyskaya A.A., Lisovskaya E.Yu., Moiseeva S.P., Danilyuk E.Yu.* Asymptotic stationary probability distribution of total amount of physics experimental data // *Russ. Phys. J.* 2020. V. 62. No. 10. P. 1779–1788.

Асимптотический анализ ресурсной гетерогенной СМО (ММРР+2М)^(v)/М/∞ *

Татьяна Бушкова, Анастасия Галилейская
Екатерина Лисовская, Светлана Моисеева

*Национальный исследовательский
Томский государственный университет, Томск, Россия*

В теории массового обслуживания приборы/линии, а также места для ожидания (буфер, очередь) играют роль необходимых для обслуживания дискретных ресурсов. При рассмотрении ресурсных систем массового обслуживания (РСМО) заявкам могут потребоваться различные дополнительные ресурсы. Это может быть некоторый случайный объем ресурса, занимаемого на время ожидания начала обслуживания, либо на время обслуживания, либо на все время пребывания заявки в системе. Например, таким образом может моделироваться объем памяти некоторого устройства или отдельной системы. Интерес к РСМО объясняется возможностью их применения для моделирования достаточно широкого спектра технических устройств и, в целом, информационно-вычислительных систем [1, 2].

В статье рассматривается ресурсная гетерогенная система массового обслуживания с неограниченным числом обслуживающих приборов, на вход которой поступают три потока заявок: два пуассоновских с параметрами $\lambda^{(1)}$ и $\lambda^{(2)}$ и марковски модулированный пуассоновский поток (ММРР-поток) сдвоенных заявок двух типов. Заявки каждого типа обслуживаются в течение случайного времени на первом и втором блоке соответственно. Для получения вероятностных характеристик будем использовать метод многомерного динамического просеивания и метод асимптотического анализа в двух условиях: эквивалентно растущей интенсивности входящих потоков и эквивалентно растущего времени обслуживания в блоках системы.

* Исследование выполнено при финансовой поддержке РФФИ и Томской области в рамках научного проекта № 19-41-703002.

Математическая модель

Рассмотрим РСМО с двумя блоками обслуживающих приборов, на вход которой поступают три потока: два пуассоновских с параметрами $\lambda^{(1)}$ и $\lambda^{(2)}$ и марковски модулированный пуассоновский поток (ММРР-поток), управляемый цепью Маркова $k(t)$ с конечным числом состояний, $k(t) = 1, 2, \dots, K$, заданной матрицей инфинитезимальных характеристик $\mathbf{Q} = \|q_{ij}\|$, $i, j = 1, 2, \dots, K$, и матрицей условных интенсивностей $\Lambda = \text{diag } \lambda_k$, $\lambda_k \geq 0$ [3]. Заявки простейшего 1-потока обслуживаются в течение случайного времени $\xi_1 \geq 0$ с функцией распределения $B_1(\tau) = P\{\xi_1 < \tau\}$ на первом блоке и требуют случайное количество ресурса $v_1 \geq 0$ с функцией распределения $G_1(y) = P\{v_1 < y\}$. Аналогично для простейшего 2-потока: случайное время обслуживания $\xi_2 \geq 0$ с функцией распределения $B_2(\tau) = P\{\xi_2 < \tau\}$, случайное количество ресурса $v_2 \geq 0$ с функцией распределения $G_2(y) = P\{v_2 < y\}$. Заявки ММРР-потока расщепляются на две копии, каждая из которых обслуживается на соответствующем блоке. Когда обслуживание заканчивается, требование покидает систему, освобождая прибор и все занимаемые ресурсы. Количество занимаемых ресурсов и время обслуживания не зависят друг от друга.

Обозначим $V_i(t)$ – суммарный объем занятого ресурса на i -м блоке в момент времени t , $i = 1, 2$.

Ставится задача исследования случайного двумерного процесса $\mathbf{V}(t) = \{V_1(t), V_2(t)\}$. Данный процесс не является марковским, поэтому для его исследования применим модификацию метода многомерного динамического просеивания [4].

Метод динамического просеивания

Зададим функции $S_i(t) \in [0, 1]$, $i = 1, 2$, определяющие динамические вероятности просеивания. Зафиксируем произвольный момент времени T в будущем. Полагаем, что заявка входящего потока, поступившая в i -й блок системы в момент времени $t < T$, с вероятностью $S_i(t) = 1 - B_i(T - t)$ будет находиться на обслуживании в момент времени T , а значит, и сформирует событие просеянного потока, а с вероятностью $1 - S_i(t)$ – закончит обслуживание до момента времени T и ос-

вободит занимаемый ресурс, следовательно, в просеянном потоке событие сформировано не будет.

Обозначим $W_i(t)$ – суммарный объем занятого ресурса i -го типа просеянными требованиями на i -ю ось.

В [3] показано, что распределение вероятностей числа занятых приборов совпадает с распределением вероятностей числа просеянных заявок в момент времени T . Аналогично устанавливается связь между построенным процессом $\mathbf{W}(t) = \{W_1(t), W_2(t)\}$ и процессом $\mathbf{V}(t) = \{V_1(t), V_2(t)\}$. Законы распределения вероятностей значений случайных процессов $\mathbf{W}(t)$ и $\mathbf{V}(t)$ в момент времени $t = T$ совпадают:

$$P\{\mathbf{V}(T) < \mathbf{x}\} = P\{\mathbf{W}(T) < \mathbf{x}\}, \quad (1)$$

для любых $\mathbf{x} = \{x_1, x_2\}$. Равенство (1) является основной формулой метода динамического просеивания для рассматриваемой системы.

Обозначим распределение вероятностей этого процесса:

$$P(k, w_1, w_2, t) = P\{k(t) = k, W_1(t) < w_1, W_2(t) < w_2\}.$$

Для этого распределения составим Δt -методом прямую систему дифференциальных уравнений Колмогорова:

$$\begin{aligned} \frac{\partial P(k, w_1, w_2, t)}{\partial t} = & (\lambda_k + \lambda^{(1)})S_1(t)P(k, w_1, w_2, t) + \\ & + (\lambda_k + \lambda^{(2)})S_2(t)P(k, w_1, w_2, t) - (\lambda_k + \lambda^{(1)})S_1(t) \int_0^{w_1} P(k, w_1 - y_1, w_2, t) dG_1(y) - \\ & - (\lambda_k + \lambda^{(2)})S_2(t) \int_0^{w_2} P(k, w_1, w_2 - y, t) dG_2(y) + \lambda_k S_1(t)S_2(t) [P(k, w_1, w_2, t) - \\ & - \int_0^{w_1} P(k, w_1 - y_1, w_2, t) dG_1(y) - \int_0^{w_2} P(k, w_1, w_2 - y, t) dG_2(y) + \\ & + \int_0^{w_1} \int_0^{w_2} P(k, w_1 - y_1, w_2 - y_2, t) dG_1(y_1) dG_2(y_2)] + \sum_v q_{vk} P(v, w_1, w_2, t) \end{aligned}$$

для $k = 1, \dots, K$; $w_1 > 0, w_2 > 0$.

Начальное условие для решения $P(k, w_1, w_2, t)$ в момент времени t_0 определим в виде

$$P(k, w_1, w_2, t_0) = \begin{cases} r(k), w_1 = w_2 = 0, \\ 0, \text{ иначе,} \end{cases}$$

где $r(k)$ – стационарное распределение вероятностей состояний цепи Маркова $k(t)$.

Введем частичные характеристические функции вида

$$h(k, v_1, v_2, t) = \int_0^\infty e^{jv_1 w_1} \int_0^\infty e^{jv_2 w_2} P(k, dw_1, dw_2, t),$$

где $j = \sqrt{-1}$ – мнимая единица.

Тогда можем записать следующие дифференциальные уравнения:

$$\begin{aligned} \frac{\partial h(k, v_1, v_2, t)}{\partial t} = & h(k, v_1, v_2, t) \left[(\lambda_k + \lambda^{(1)}) S_1(t) (G_1^*(v_1) - 1) + \right. \\ & \left. + (\lambda_k + \lambda^{(2)}) S_2(t) (G_2^*(v_2) - 1) + \lambda_k S_1(t) S_2(t) (1 - G_1^*(v_1)) (1 - G_2^*(v_2)) \right] + \\ & + \sum_v q_{vk} h(v, v_1, v_2, t), \end{aligned}$$

для $k = 1 \dots K$, где

$$G_i^*(v_i) = \int_0^\infty e^{jv_i y} dG_i(y).$$

Перепишем эту систему в матричном виде:

$$\begin{aligned} \frac{\partial \mathbf{h}(v_1, v_2, t)}{\partial t} = & \mathbf{h}(v_1, v_2, t) \left[(\mathbf{\Lambda} + \mathbf{\Lambda}^{(1)}) S_1(t) (G_1^*(v_1) - 1) + \right. \\ & + (\mathbf{\Lambda} + \mathbf{\Lambda}^{(2)}) S_2(t) (G_2^*(v_2) - 1) + \\ & \left. + \mathbf{\Lambda} S_1(t) S_2(t) (1 - G_1^*(v_1)) (1 - G_2^*(v_2)) + \mathbf{Q} \right], \end{aligned} \quad (2)$$

с начальным условием:

$$\mathbf{h}(v_1, v_2, t_0) = \mathbf{r}. \quad (3)$$

Здесь

$$\mathbf{h}(v_1, v_2, t) = [h(1, v_1, v_2, t), \dots, h(K, v_1, v_2, t)],$$

$\mathbf{r} = [r(1), \dots, r(K)]$ – вектор стационарного распределения вероятностей состояний управляющей цепи Маркова $k(t)$, удовлетворяющий системе:

$$\begin{cases} \mathbf{r}\mathbf{Q} = \mathbf{0}, \\ \mathbf{r}\mathbf{e} = 1, \end{cases}$$

\mathbf{e} – единичный вектор-столбец,

$$\mathbf{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdot & 0 \\ 0 & \lambda_2 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \lambda_K \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} q_{11} & q_{12} & \cdot & q_{1K} \\ q_{21} & q_{22} & \cdot & q_{2K} \\ \cdot & \cdot & \cdot & \cdot \\ q_{K1} & q_{K2} & \cdot & q_{KK} \end{bmatrix},$$

$$\mathbf{\Lambda}^{(1)} = \begin{bmatrix} \lambda^{(1)} & 0 & \cdot & 0 \\ 0 & \lambda^{(1)} & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \lambda^{(1)} \end{bmatrix}_{[K \times K]} = \lambda^{(1)} \mathbf{I}, \quad \mathbf{I} = \begin{bmatrix} 1 & 0 & \cdot & 0 \\ 0 & 1 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & 1 \end{bmatrix},$$

$$\mathbf{\Lambda}^{(2)} = \begin{bmatrix} \lambda^{(2)} & 0 & \cdot & 0 \\ 0 & \lambda^{(2)} & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \lambda^{(2)} \end{bmatrix}_{[K \times K]} = \lambda^{(2)} \mathbf{I}.$$

Таким образом, поставленная задача сводится к решению дифференциального матричного уравнения (2) – (3). Следует отметить, что построенная система матричных уравнений имеет структуру, аналогичную похожим задачам без дополнительных потоков, например в статьях [4–6], что позволяет сделать вывод о том, что поставленную задачу можно решать методом асимптотического анализа в двух предельных условиях: условия эквивалентно растущего времени обслуживания и условия эквивалентно растущей интенсивности входящих потоков.

Для этого были разработаны и реализованы модификации некоторых этапов метода асимптотического анализа, что позволило сформулировать и доказать теоремы об асимптотической гауссовости исследуемых процессов и получить асимптотические вероятностные характеристики.

Заключение

Таким образом, построена математическая модель поликомпонентной РСМО типа $(\text{MMPP} + 2M)^{(v)}/M/\infty$. Предложена модификация метода асимптотического анализа для исследования рассматриваемых процессов в двух новых условиях: эквивалентно растущей интенсивности входящих потоков и эквивалентно растущего времени обслуживания в блоках системы. Аналитически получен вид асимптотической характеристической функции двумерного процесса – суммарных объемов занимаемых ресурсов в системе. Найдены асимптотические числовые характеристики указанных процессов: первого и второго моментов, дисперсии и коэффициент корреляции. Сделаны выводы об области применимости.

ЛИТЕРАТУРА

1. *Tikhonenko O.M.* Generalized Erlang problem for queueing systems with bounded total size // *Probl. Peredachi Inf.* 2005. V. 41. No. 3. P. 64–75; *Problems Inform. Transmission.* 2005. V. 41. No. 3. P. 243–253.
2. *Tikhonenko O.M., Kempa W.* Queueing system with processor sharing and limited memory under control of the AQM mechanism // *Avtomat. i Telemekh.* 2015. No. 10. P. 90–105; *Autom. Remote Control.* 2015. V. 76. No. 10. 1784–1796.
3. *Моисеев А.Н., Назаров А.А.* Бесконечнолинейные системы и сети массового обслуживания. Томск: Изд-во НТЛ, 2015. 240 с.
4. *Bushkova T., Pavlova E., Rozhkova S., Moiseeva S., Pagano M.* Resource queueing system $\text{MMPP}(2,v)|G|2|\infty$ with parallel service of multiple paired customers // *CCIS.* 2019. V. 1109. P. 136–149.
5. *Lisovskaya E., Moiseeva S., Pagano M., Potatueva V.* Study of the $\text{MMPP}/GI/\infty$ queueing system with random customers' capacities // *Информ. и её примен.* 2017. V. 11. No. 4. P. 109–117.
6. *Galileyskaya A.A., Lisovskaya E.Yu., Moiseeva S.P., Danilyuk E.Yu.* Asymptotic stationary probability distribution of total amount of physics experimental data // *Russ. Phys. J.* 2020. V. 62. No. 10. P. 1779–1788.

Аналитический метод анализа случайных процессов с непрерывным временем и дискретными состояниями при времязависимых вероятностях переходов*

Константин Вытовтов, Елизавета Барабанова,
Владимир Вишнеvский

Институт проблем управления им. В.А. Трапезникова, Москва, Россия

Развитие систем массового обслуживания предполагает изучение различного рода марковских случайных процессов и разработку методов их исследования [1–5], в частности решение уравнений Колмогорова. Отметим, что на сегодняшний день отсутствуют аналитические методы решения уравнений Колмогорова с времязависящими коэффициентами [3–5].

В данной работе впервые представлен точный аналитический метод решения системы уравнений Колмогорова третьего порядка и M -го порядка с произвольными кусочно-постоянными коэффициентами.

Постановка задачи

В данной работе, прежде всего, рассмотрен неоднородный марковский процесс с тремя дискретными состояниями и непрерывным временем (рис. 1). Предполагается, что вероятности переходов из одного состояния в другое $\lambda_{ij}(t)$ зависят

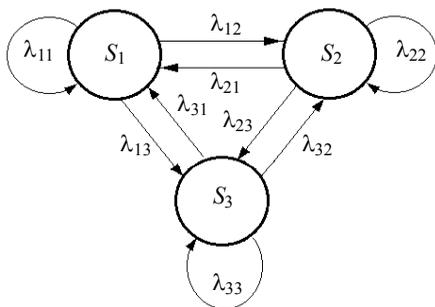


Рис. 1. Граф состояний

от времени и описываются произвольными кусочно-постоянными функциями (например, рис. 2). Далее полученный результат обобщается на случай марковского процесса с произвольным числом M состояний и произвольными кусочно-постоянными вероятностями переходов $\lambda_{ij}(t)$.

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-29-06043.

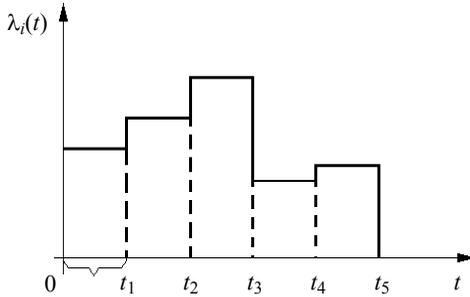


Рис. 2. График зависимости потоков от времени

Система уравнений Колмогорова для случая с тремя состояниями (рис. 1) имеет вид

$$\begin{aligned} \frac{dp_1}{dt} &= -(\lambda_{12} + \lambda_{13})p_1 + \lambda_{21}p_2 + \lambda_{31}p_3, \\ \frac{dp_2}{dt} &= \lambda_{12}p_1 - (\lambda_{21} + \lambda_{23})p_2 + \lambda_{32}p_3, \\ \frac{dp_3}{dt} &= \lambda_{13}p_1 + \lambda_{23}p_2 - (\lambda_{31} + \lambda_{32})p_3. \end{aligned} \quad (1)$$

Здесь p_i – вероятность нахождения процесса в состоянии S_i . Прежде всего, найдем аналитическое решение системы (1) для случая постоянных потоков. Действительно, в пределах некоторого интервала времени потоки являются пуассоновскими, а уравнение (1) имеет постоянные коэффициенты. В этом случае характеристическое уравнение этой системы записывается как

$$\begin{aligned} \gamma[\gamma^2 + (\lambda_{12} + \lambda_{13} + \lambda_{21} + \lambda_{23} + \lambda_{31} + \lambda_{32})\gamma + \lambda_{12}\lambda_{23} + \lambda_{12}\lambda_{31} + \lambda_{12}\lambda_{32} + \\ + \lambda_{13}\lambda_{21} + \lambda_{13}\lambda_{23} + \lambda_{13}\lambda_{32} + \lambda_{21}\lambda_{31} + \lambda_{21}\lambda_{32} + \lambda_{23}\lambda_{31}] = 0. \end{aligned} \quad (2)$$

С учетом (2) решение для вероятности p_1 нахождения системы в состоянии S_1 может быть получено в виде

$$p_1 = A + Be^{\gamma_1 t} + Ce^{\gamma_2 t}. \quad (3)$$

Из первых двух уравнений системы (1) найдем вероятности нахождения системы в состояниях S_2 и S_3 :

$$p_2 = \xi_0 A + \xi_1 Be^{\gamma_1 t} + \xi_2 Ce^{\gamma_2 t}; \quad p_3 = \zeta_0 A + \zeta_1 Be^{\gamma_1 t} + \zeta_2 Ce^{\gamma_2 t}, \quad (4)$$

где
$$\xi_1 = \frac{\gamma_i \lambda_{32} + \lambda_{12} \lambda_{31} + \lambda_{12} \lambda_{32} + \lambda_{13} \lambda_{32}}{\gamma_i \lambda_{31} + \lambda_{12} \lambda_{31} + \lambda_{12} \lambda_{32} + \lambda_{23} \lambda_{31}}, \quad (5)$$

$$\zeta_1 = \frac{\gamma_i^2 + (\lambda_{12} + \lambda_{13} + \lambda_{21} + \lambda_{23})\gamma_i + \lambda_{12}\lambda_{23} + \lambda_{13}\lambda_{21} + \lambda_{13}\lambda_{23}}{\gamma_i \lambda_{31} + \lambda_{12} \lambda_{31} + \lambda_{21} \lambda_{32} + \lambda_{23} \lambda_{31}}. \quad (6)$$

Далее, нашей основной задачей является нахождение аналитического решения системы (1) для случая, когда все вероятности переходов являются произвольными кусочно-постоянными функциями, поскольку такое решение на сегодняшний день отсутствует. Кроме того, метод обобщается для случая M состояний случайного процесса и кусочно-постоянных вероятностей переходов.

Фундаментальная матрица процесса с постоянными вероятностями переходов

В этом разделе найдем фундаментальную матрицу решений системы (1) на i -м интервале с постоянными коэффициентами. В соответствии с методикой 3×3 матрица фундаментальных решений может быть получена в виде

$$\mathbf{M}_i = \mathbf{A}_{0i} + \mathbf{A}_{1i} \exp(\gamma_1^{(i)} t_i) + \mathbf{A}_{2i} \exp(\gamma_2^{(i)} t_i), \quad (7)$$

где
$$\mathbf{A}_{ji} = \frac{1}{\Delta} \begin{vmatrix} \Delta_{j1} & \Delta_{j2} & \Delta_{j3} \\ \xi_j \Delta_{j1} & \xi_j \Delta_{j2} & \xi_j \Delta_{j3} \\ \zeta_j \Delta_{j1} & \zeta_j \Delta_{j2} & \zeta_j \Delta_{j3} \end{vmatrix}. \quad (8)$$

Здесь $\Delta = \xi_0(\zeta_1 - \zeta_2) + \xi_1(\zeta_2 - \xi_0) + \xi_2(\zeta_0 - \zeta_1)$; $\Delta_{11} = \xi_1 \zeta_2 - \xi_2 \zeta_1$;

$$\Delta_{21} = \xi_2 \zeta_0 - \xi_0 \zeta_2; \Delta_{31} = \xi_0 \zeta_1 - \xi_1 \zeta_0; \Delta_{12} = \zeta_1 - \zeta_2; \Delta_{22} = \zeta_2 - \zeta_0;$$

$$\Delta_{32} = \zeta_0 - \zeta_1; \Delta_{13} = \xi_2 - \xi_1; \Delta_{23} = \xi_0 - \xi_2; \Delta_{33} = \xi_1 - \xi_0,$$

i – номер интервала.

Обобщая результат для случая с M состояниями процесса, получим

$$\mathbf{M}_i = \sum_{j=0}^M \mathbf{A}_{ji} \exp(\gamma_j^{(i)} t_i). \quad (9)$$

Таким образом, матрица фундаментальных решений (7) определяет вероятности $p_1(t)$, $p_2(t)$ и $p_3(t)$ в любой момент времени при заданных начальных вероятностях состояний процесса и постоянными вероятностями переходов. Аналогично, матрица (9) определяет состояние случайного процесса с M возможными состояниями и постоянными вероятностями переходов.

Фундаментальная матрица процесса со скачкообразным изменением вероятностей переходов

Далее рассмотрим процесс (рис. 1) с зависящими от времени вероятностями переходов, описываемыми произвольными кусочно-постоянными функциями (рис. 2) и найдем аналитическое решение системы уравнений Колмогорова (1) для рассматриваемого случая. На интервалах с постоянными вероятностями переходов вероятности состояний процесса в произвольный момент времени описываются матрицей фундаментальных решений (7) при заданных начальных условиях в начале каждого интервала. А фундаментальная матрица для произвольного числа интервалов в соответствии с методом находится как

$$\mathbf{M} = \prod_{i=N}^1 \mathbf{M}_i(t_i - t_{i-1}). \quad (10)$$

Учитывая выражение (7) и проводя алгебраические преобразования, фундаментальная матрица (10) для произвольного количества скачков вероятностей переходов и трех состояний случайного процесса (рис. 1) принимает вид

$$\mathbf{M} = \sum_{j_1=1}^3 \dots \sum_{j_N=1}^3 \exp \left[\sum_{i=1}^N (\ln A_{ij_i} + \gamma_{ij_i} \Delta t_i) \right]. \quad (11)$$

Обобщая данный результат для случая M состояний процесса и N скачков вероятности переходов, получим

$$\mathbf{M} = \sum_{\substack{j_1=1 \\ \dots \\ j_N=1}}^M \exp \left[\sum_{i=1}^N (\ln A_{ij_i} + \gamma_{ij_i} \Delta t_i) \right]. \quad (12)$$

Таким образом, нами представлен аналитический метод нахождения фундаментальной матрицы, описывающей поведение рассматриваемого случайного процесса с M дискретными состояниями и произвольными кусочно-постоянными вероятностями переходов.

Численные расчеты

В данном разделе проведены численные расчеты, подтверждающие правильность полученных аналитических результатов. Прежде всего, расчеты проведены для случая двух интервалов с постоянными вероятностями переходов (рис. 3, *а*). Скачок величин этих вероятностей происходит при $t_1 = 2$. Величины вероятностей переходов до и после скачка: для первого интервала $\lambda_{12} = 0.1$; $\lambda_{13} = 0.9$; $\lambda_{21} = 0.2$; $\lambda_{23} = 0.8$; $\lambda_{31} = 0.4$; $\lambda_{32} = 0.6$; для второго интервала $\lambda_{12} = 0.7$; $\lambda_{13} = 0.3$; $\lambda_{21} = 0.65$; $\lambda_{23} = 0.35$; $\lambda_{31} = 0.45$; $\lambda_{32} = 0.55$. Кроме того, представлены результаты расчетов поведения системы для трех интервалов с постоянными значениями переходных вероятностей (рис. 3, *б*). Скачки происходят при $t_1 = 2$; $t_1 = 5$. Параметры первых двух интервалов соответствуют предыдущему случаю, а для третьего интервала $\lambda_{12} = 0.3$; $\lambda_{13} = 0.7$; $\lambda_{21} = 0.1$; $\lambda_{23} = 0.9$; $\lambda_{31} = 0.6$; $\lambda_{32} = 0.4$. В обоих случаях выбраны начальные условия $\{p_1(0); p_2(0); p_3(0)\}^T = \{0.9; 0.1; 0\}^T$.

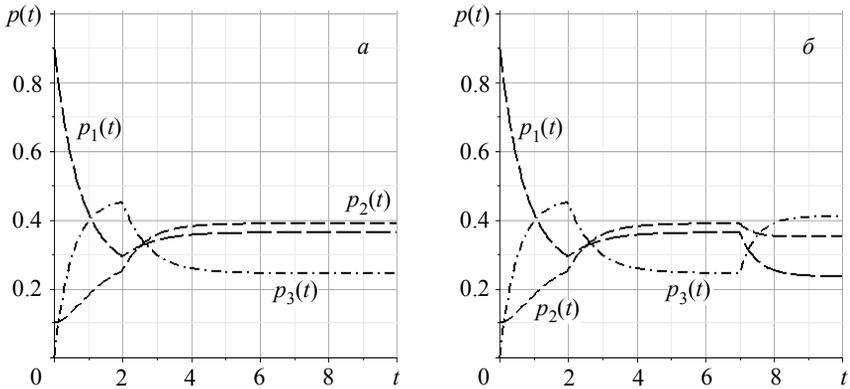


Рис. 3. Зависимость вероятности состояний системы от времени

Результаты проведенных расчетов, прежде всего, показывают корректность представленного метода. Действительно, как в переходном, так и в стационарном режимах сумма вероятностей состояний равна единице. Для случая одного скачка стационарный режим устанавливается при $t > 6$, для случая двух скачков стационарный режим устанавли-

вается при $t > 9$. Более того, процесс не испытывает скачков вероятностей на границе интервалов, хотя производные этих вероятностей имеют скачки на этих границах.

Выводы

В данной работе рассмотрен неоднородный марковский процесс с тремя дискретными состояниями и непрерывным временем (рис. 1). Предполагается, что все вероятности переходов из одного состояния в другое являются произвольными кусочно-постоянными функциями времени (рис. 2). Кроме того, результаты обобщены для случая с M состояниями процесса. В работе впервые представлен точный аналитический метод решения системы дифференциальных уравнений Колмогорова третьего (11) и M -го (12) порядка с произвольными кусочно-постоянными коэффициентами. Также проведены численные расчеты, подтверждающие правильность полученных аналитических результатов.

ЛИТЕРАТУРА

1. *Vishnevsky V., Dudin A.N., Klimenok V.I.* Theory of Queueing Systems with Correlated Flows. Springer International Publishing, 2020. 460 p.
2. *Гайдамака Ю.В., Сопин Э.С., Гудкова И.А., Андреев С.Д., Шоргин С.Я., Самуйлов К.Е.* Модели и методы анализа и расчета показателей эффективности беспроводных гетерогенных сетей. М.: ФИЦ ИУ РАН, 2018. 71 с.
3. *Бондрова О.В., Головкин Н.И., Жук Т.А.* Вывод уравнений типа Колмогорова – Чепмена с интегральным оператором // Дальневосточный математический журнал. 2017. Т. 17. № 2. С. 135–146.
4. *Бондрова О.В., Крылова Д.С., Головкин Н.И., Жук Т.А.* Вывод уравнений для систем массового обслуживания с бесконечным накопителем и скачкообразной интенсивностью входного потока // Вестник ВГУ. Серия: Физика. Математика. 2015. № 4. С. 89–100.
5. *Nezhel'skaya L., Pagano M., Sidorova E.* Distribution parameters estimation in recurrent synchronous generalized doubly stochastic flow of the second order // International Conference on Distributed Computer and Communication Networks DCCN 2019: Distributed Computer and Communication Networks. P. 276–288.

Об использовании четырехфазных систем массового обслуживания для описания работы грузовых и сортировочных железнодорожных станций*

Максим Жарков¹, Михаил Пavidис²

*¹ Институт динамики систем и теории управления
имени В.М. Матросова СО РАН, г. Иркутск, Россия*

*² Иркутский государственный университет путей сообщения,
г. Иркутск, Россия*

Одним из ключевых элементов инфраструктуры железнодорожного транспорта являются грузовые и сортировочные железнодорожные станции (ГСЖС). От качества их работы зависят объемы перевозимого груза (товаров), устойчивость перевозочного процесса и эффективность работы железнодорожной сети в целом [1]. Эти объекты обеспечивают переработку поступающих вагонопотоков, погрузку / разгрузку грузов (для грузовых станций) и формирование новых поездов в соответствии с общесетевым и дорожным планами. Поэтому повышение производительности и обеспечение их бесперебойной работы является актуальной задачей.

Ранее авторами был предложен подход к моделированию функционирования ГСЖС на основе трехфазных систем массового обслуживания (СМО) [2]. В этом случае каждая фаза описывает работу одной функциональной подсистемы: парка прибытия с сортировочной горкой, сортировочного парка, парка отправления (для сортировочной станции) или грузового двора (для грузовой). Это позволяет обобщенно отобразить устройство и функционирование станции.

В ходе дальнейших исследований выяснилось, что парк прибытия (ПП) и сортировочная горка (СГ) являются отдельными функциональными элементами. Так, во-первых, в них выполняются различные операции с поездами: в ПП происходит принятие поезда и отцепка от него локомотива, затем на СГ полученный состав разбивается на группы вагонов по направлениям. Во-вторых, в ПП есть свои обслуживающие устройства – электровозы надвига, а СГ имеет свою перерабатываю-

* Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта 18-07-00604, РФФИ и Правительства Иркутской области в рамках проекта 20-47-383002.

щую способность. В-третьих, в зависимости от структуры станции ПП и СГ, могут находиться на значительном удалении друг от друга. Таким образом, логичным видится описание парка прибытия с сортировочной горкой в виде двух обособленных фаз.

В этой связи предлагается использовать четырехфазные СМО для моделирования работы ГСЖС. Это позволит описать четырехэтапное обслуживание потока вагонов на станции, а также определить пропускную способность каждой из подсистем в отдельности.

Математическая модель

Математическую модель функционирования ГСЖС будем строить на основе четырехфазной СМО специального вида, в которой каждая фаза соответствует подсистеме выбранного объекта и обладает следующими характеристиками [3]:

- в фазе имеется n обслуживающих устройств (каналов);
- допустимо групповое обслуживание заявок в каждом канале;
- на каждой фазе присутствует очередь, ее максимальная длина m конечна;
- длина очереди, распределения времени обслуживания группы заявок в канале и ее размера различны для каждой фазы;
- заявки из очереди выбираются согласно FIFO (первый вошел, первый вышел);
- каналы первых трех фаз временно блокируются при отсутствии свободных мест для принятия заявок в очереди последующих фаз.

Для описания входящего вагонопотока, обладающего сложной структурой, применяется модель ВМАР-потока [4, 5]. Благодаря этому возможно учесть наличие нескольких подпотоков заявок с различными параметрами, включая законы распределения размеров групп заявок и интенсивности их поступления. Прибывшие в систему группы заявок принимаются согласно дисциплине полного отказа (Complete rejection): если не хватает места хотя бы для одной заявки, то теряется вся группа заявок [5].

В терминах теории массового обслуживания [3] модель работы ГСЖС запишется следующим образом:

$$\text{ВМАР}/G^{X1}/n_1/m_1 \rightarrow */G^{X2}/n_2/m_2 \rightarrow */G^{X3}/n_3/m_3 \rightarrow */G^{X4}/n_4/m_4,$$

где G – произвольный закон распределения времени обслуживания заявок в канале; X_i – распределение размеров групп заявок, выбранных на

обслуживание в канал фазы i ; m_i – длина очереди, n_i – количество каналов на фазе i .

Для параметрической идентификации модели конкретного объекта в виде многофазной СМО необходимо, во-первых, определить параметры *ВМАР*-потока, во-вторых, для каждой фазы – установить число мест в очереди, количество каналов, распределения времени обслуживания и размеров обслуживаемых групп заявок.

Полученная четырехфазная СМО оказалась сложна для аналитического исследования, поэтому авторами применяется имитационное моделирование [6] для определения ее показателей эффективности [3]: вероятности обслуживания заявки и группы заявок, среднего времени пребывания заявки в системе, средней длины очереди для каждой фазы, времени блокировки каналов и других характеристик (в соответствии с целью исследования).

Имитационная модель и вычислительный эксперимент

Имитационная модель работы многофазной СМО основана на дискретно-событийном подходе моделирования и реализована в виде программного модуля [7]. Он предназначен для численного нахождения вероятностей состояний и на их основе определения показателей эффективности выбранной СМО, а также проведения многовариантных сценарных расчетов.

Объектом исследования была выбрана сортировочная станция И-С, расположенная на ВСЖД в Иркутской области. Она является двухсистемной с последовательным расположением парков. Обе системы практически идентичны, поэтому рассмотрим работу нечетной, в структуре которой имеются парк прибытия, сортировочные парк и горка, парк отправления. В систему поступает поездопоток с различными параметрами для дневной и ночной смен.

В терминах теории массового обслуживания модель нечетной системы станции И-С имеет вид

$$\text{ВМАР}/G^B/2/720 \rightarrow * /G^B/1/0 \rightarrow * /G^B/2/3500 \rightarrow * /G^B/2/750,$$

где B – биномиальное распределение.

Более подробная информация об объекте исследования и результаты численного анализа ее модели будут представлены в докладе.

Выводы

В работе представлено развитие ранее разработанного подхода к математическому моделированию работы грузовых и сортировочных железнодорожных станций за счет использования четырехфазных СМО с ВМАР-потоками. Их применение позволяет описать четыре функциональные подсистемы, которые являются этапами обслуживания вагонного потока в ГСЖС. На основе предложенного подхода разработана обобщенная модель работы железнодорожной станции, преимуществом которой является то, что она может быть адаптирована за минимальное время под конкретную систему. Использование четырехфазных СМО значительно увеличило сложность получаемых моделей, однако современные средства имитационного моделирования позволяют эффективно исследовать подобные объекты.

ЛИТЕРАТУРА

1. Правдин Н.В., Вакуленко С.П., Головнич А.К. и др. Проектирование инфраструктуры железнодорожного транспорта (станции, железнодорожные и транспортные узлы). М.: ФГБОУ «Учебно-методический центр по образованию на железнодорожном транспорте», 2012. 1086 с.
2. Лемперт А.А., Павидис М.М., Жарков М.Л. Моделирование зарубежных сортировочных станций на основе теории массового обслуживания // Вестник УрГУПС. 2019. № 3(43). С. 24–32.
3. Medhi J. Stochastic models in queueing theory. London: Academic Press, 2003. 482 p.
4. Lucantoni D.M. New results on single server queue with a batch Markovian arrival process // Stochastic Models. 1991. V. 7. P. 1–46.
5. Дудин А.Н., Клименок В.И. Системы массового обслуживания с коррелированными потоками. Минск: БГУ, 2000. 176 с.
6. Кельтон В., Лоу А. Имитационное моделирование. СПб.: Питер, 2004. 846 с.
7. Бычков И.В., Жарков М.Л., Казаков А.Л. Интеллектуальная технология моделирования объектов транспортно-логистической инфраструктуры на основе теории массового обслуживания // Труды Восьмой Межд. конф. «Системный анализ и информационные технологии» (САИТ – 2019), г. Иркутск, 8–14 июля 2019 г. М.: ФИЦ ИУ РАН, 2019. С. 541–548.

Метод бесконечных разметок в системах с неизвестным временем обслуживания поступающих заявок

Владимир Задорожный, Татьяна Захаренкова

Омский государственный технический университет, г. Омск, Россия

В статье исследуются возможности практического применения метода бесконечных разметок (МБР) [1, 2], позволяющего эффективно обеспечить конечное среднее время ожидания (конечную среднюю длину очереди) в системах с бесконечной дисперсией времени обслуживания. Если время x обслуживания поступающей заявки известно, то ей присваивается абсолютный приоритет, определяемый соответствующим интервалом бесконечной разметки диапазона возможных значений x . Чем дальше находится интервал разметки от начала диапазона возможных значений x , тем ниже соответствующий этому интервалу приоритет. МБР может использоваться для решения различных классов задач, возникающих при обслуживании фрактального трафика современных компьютерных сетей. Такой трафик описывается степенными распределениями времени обслуживания заявок, характеризуемыми конечным математическим ожиданием и бесконечной дисперсией [3]. Формула Полячека – Хинчина для такого фрактального трафика определяет при любом коэффициенте загрузки ρ системы, меньшем единицы, бесконечное среднее время ожидания [4] (бесконечную среднюю длину очереди). Рассматриваются следующие случаи:

- когда x известно и длина очереди определяется не только числом заявок в ней, но и их длиной (как при передаче файлов в интернете),
- важный для практики случай, когда длина очереди определяется только числом заявок в ней, но время обслуживания поступающих заявок неизвестно. Для этого случая предлагается версия МБР, основанная на динамически настраиваемых приоритетах. Устанавливается, что в этом случае МБР также позволяет эффективно обеспечивать конечное среднее время ожидания заявок.

Случай, когда фактическое время x обслуживания каждой поступающей заявки известно и длина очереди определяется как число заявок в ней, исследован в [1, 2] – это основной случай применения МБР.

Математическая модель

Время x обслуживания заявок распределено по закону Парето:

$$F(t) = P(x \leq t) = 1 - \left(\frac{K}{t}\right)^\alpha, \quad \alpha > 0, \quad t \geq K,$$

где $K > 0$ – минимальное значение случайной величины (с.в.) и, одновременно, масштабный параметр распределения; α – параметр формы.

Первые два момента с.в. x выражаются формулами

$$M(x) = \frac{\alpha K}{\alpha - 1}, \quad (\alpha > 1),$$

$$M(x^2) = \frac{\alpha K^2}{(\alpha - 1)^2 (\alpha - 2)}, \quad (\alpha > 2),$$

$$D(x) = \frac{\alpha K^2}{(\alpha - 1)^2 (\alpha - 2)}, \quad (\alpha > 2).$$

При $\alpha < 1$ все моменты с.в. x бесконечны.

При $1 < \alpha \leq 2$ имеем $M(x) < \infty$, $D(x) = \infty$. Этот случай в моделировании фрактального трафика наиболее востребован. В системе M/Pa/1/∞ (с временем x , принадлежащим распределению Парето) при $1 < \alpha \leq 2$ в соответствии с формулой Полячека – Хинчина [4] средняя длина L очереди бесконечна:

$$L = \frac{\lambda^2 M(x^2)}{2(1 - \rho)} = \infty, \quad (1)$$

где коэффициент загрузки $\rho = \lambda M(x) < 1$, λ – интенсивность входящего пуассоновского потока заявок.

В соответствии с формулой Литтла среднее время ожидания $W = L/\lambda$. Следовательно, при $L = \infty$ также и $W = \infty$.

Введение в систему абсолютных приоритетов в соответствии с МБР определяется разбиением диапазона $K \leq t < \infty$ возможных значений времени x бесконечной последовательностью точек

$$\{t_k\} = t_0, t_1, \dots, t_k, \dots$$

на интервалы $[t_0, t_1), [t_1, t_2), \dots, [t_{k-1}, t_k), \dots$,

где $t_0 = K$ (рис. 1).

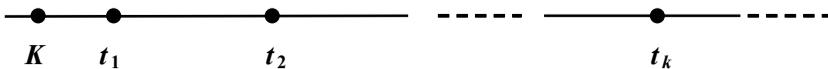


Рис. 1. Бесконечная разметка лиапазона возможных значений x

Если у входящей заявки время x принадлежит интервалу $[t_0, t_1)$, то ей присваивается абсолютный приоритет P_1 ; если $x \in [t_1, t_2)$, то заявке присваивается приоритет $P_1 - 1$; если $x \in [t_2, t_3)$, то заявке присваивается приоритет $P_1 - 2$ и т.д.

При таком назначении абсолютных приоритетов средняя длина L очереди и среднее время W ожидания становятся конечными [1, 2].

Далее будет использоваться экспоненциальная разметка, в которой

$$t_0 = K, \quad t_k = K + ce^{ak}, \quad (k = 1, 2, \dots). \quad (2)$$

Значения параметров a, c оптимизируются по критерию $W \rightarrow \min$. При этом для расчета W используются известные формулы расчета систем с абсолютными приоритетами [4].

Влияние МБР на очередь, определяемую с учетом длины заявок

Если время обслуживания заявки пропорционально ее длине – как при передаче файла – то длину очереди можно рассчитывать двумя способами. Первым способом, традиционным, длина l очереди определяется как число заявок в ней. В этом случае применение МБР к системе с бесконечной дисперсией времени обслуживания приводит к тому, что средняя длина L очереди становится конечной. Вторым способом длина очереди определяется как сумма длин заявок в ней. Чтобы отличать такую длину очереди от традиционной, будем называть ее продолжительностью очереди, обозначать через ω , а среднюю продолжительность очереди обозначать через Ω . При этом, полагая, что время обслуживания заявки пропорционально ее длине с коэффициентом пропорциональности единица, получаем

$$\omega = \sum_{i=1}^l x_i, \quad (3)$$

где x_i – время обслуживания заявки, находящейся в очереди.

Определим тестовую систему М/Ра/1/ ∞ как систему с параметрами $\lambda = 1/6, K = 1, \alpha = 1.5$. В этой системе $M(x) = \frac{K\alpha}{\alpha - 1} = 3, D(x) = \infty$, коэффициент загрузки $\rho = 0.5$. Средняя длина L очереди (1) бесконечна.

Оптимизированные значения параметров a , c экспоненциальной разметки (2) составляют $a \approx 1$, $c \approx 0.5$. Из $L = \infty$ вытекает $\Omega = \infty$.

В табл. 1 приведены результаты имитационного моделирования тестовой системы М/Ра/1/∞ (без применения МБР) при разных длинах прогона модели. Длина прогона определяется моделируемым периодом времени, на котором в систему поступает в среднем N заявок. Имитационные оценки \hat{L} , $\hat{\Omega}$ и \hat{W} показателей L , Ω и W с ростом длины прогона сходятся к своим стационарным значениям – к бесконечностям.

Таблица 1

Изменение оценок \hat{L} , $\hat{\Omega}$ и \hat{W} с ростом N в модели беспriorитетной системы

Длина прогона N	\hat{L}	$\hat{\Omega}$	\hat{W}
10^4	1.093	2.773	3.737
10^5	3.025	7.884	15.222
10^6	64.486	161.834	383.375
10^7	129.524	332.182	773.538

В табл. 2 приведены результаты имитационного моделирования тестовой системы М/Ра/1/∞ с применением МБР. Оценки для L и для W стабилизируются вблизи конечных стационарных значений этих показателей, близких, соответственно, к 0.22 и 1.3. В то же время оценка средней продолжительности Ω очереди с ростом длины прогона растет, что свидетельствует о том, что $\Omega = \infty$. Учитывая, что среднее число $L \approx 0.22$ слагаемых в формуле (3) невелико, стационарное среднее $\Omega = \infty$ обуславливается тем, что в очередь попадают заявки со все большим временем обслуживания. Это вполне согласуется с физическим смыслом МБР, дающего возможность «коротким» заявкам за счет высокого абсолютного приоритета обслуживаться без задержек.

Таблица 2

Изменение оценок \hat{L} , $\hat{\Omega}$ и \hat{W} с ростом N в модели, использующей МБР

Длина прогона N	\hat{L}	$\hat{\Omega}$	\hat{W}
10^4	0.133	3747	0.903
10^5	0.176	45688	1.070
10^6	0.233	514686	1.391
10^7	0.222	5055579	1.329

Применение МБР при неизвестном времени обслуживания

Другой важный случай применения МБР связан с длиной очереди, определяемой лишь числом заявок в ней (как при передаче математических задач для решения на сайт wolframalpha.com). В этом случае время обслуживания заявки зависит только от ее особенностей (от сложности передаваемой математической задачи). При этом, как правило, время обслуживания поступающих заявок (время решения задачи) неизвестно.

Для рассматриваемого случая можно предложить такую версию МБР, в которой приоритеты заявок настраиваются в ходе их обслуживания. Всякой поступающей в систему заявке вначале присваивается высший приоритет P_1 . Если заявка обслуживается за время, принадлежащее интервалу $[t_0, t_1)$, то заявка покидает систему обслуженной и в итоге получается, что высший приоритет ей был присвоен правильно. Если заявка обслуживается дольше, т.е. если время x достигает величины t_1 , то приоритет заявки уменьшается на единицу. Далее, если ее обслуживание завершается за время, не превышающее t_2 , то заявка покидает систему обслуженной и в итоге получается, что последний приоритет ей был присвоен правильно. Если время обслуживания достигнет величины t_3 , то приоритет заявки уменьшается еще на единицу, и т.д.

В табл. 3 представлены результаты имитационного моделирования тестовой системы с версией МБР с динамической настройкой приоритетов. Средняя длина очереди в этой версии МБР составляет приблизительно 0.53, среднее время ожидания составляет около 3.14 единиц времени. При известном времени обслуживания МБР обеспечивает значения этих показателей, составляющие 0.22 и 1.3 соответственно (см. табл. 2), которые, естественно, несколько лучше. В то же время при неизвестном времени обслуживания МБР за счет динамической настройки приоритетов обеспечивает принципиальное улучшение этих показателей по сравнению с беспriorитетной системой (см. табл. 1), где они бесконечны.

Таблица 3

Характеристики МБР при неизвестном времени обслуживания заявок

Длина прогона N	\hat{L}	\hat{W}
10^4	0.366	2.337
10^5	0.419	2.534
10^6	0.556	3.325
10^7	0.525	3.144

Относительно высокая эффективность алгоритма настройки приоритетов при неизвестном времени обслуживания обусловлена применением экспоненциальной разметки (2). Поскольку длина интервалов разметки быстро растет с ростом номера интервала, время обслуживания заявок в начальных интервалах (с неадекватно высоким приоритетом) занимает относительно малую долю от всего времени их обслуживания.

Заключение

Выполненные в статье исследования позволяют сформулировать проблему сохраняющейся при использовании МБР бесконечной средней продолжительности очереди.

Там, где продолжительность очереди равна ее длине, МБР может эффективно использоваться для кардинального сокращения очереди и времени ожидания даже при неизвестном времени обслуживания заявок. Экспоненциальная разметка обеспечивает малое влияние этапов динамической настройки приоритетов заявок на эффективность МБР.

ЛИТЕРАТУРА

1. *Zadorozhnyi V.N., Zakharenkova T.R.* Methods to reduce loss probability in systems with infinite service time dispersion // *Communications in Computer and Information Science*. 2019. V. 1109. P. 296–311.
2. *Zadorozhnyi V.N., Zakharenkova T.R., Pagano M.* Queue normalization methods in systems GI/GI/1/m with infinite variance of service time // *Journal of Physics: Conference Series*. 2020. P. 1–11.
3. *Leland W.E. et al.* On the self-similar nature of Ethernet traffic // *ACM SIGCOMM'93*. San Fransisco, 1993. P. 183–193.
4. *Kleinrock L.* Queueing Systems. Vol. II. Computer Applications. New York: Wiley Interscience, 1976. 576 p.

Применение метода бесконечных разметок к сетям с коммутацией пакетов

Владимир Задорожный¹, Микеле Пагано²,
Татьяна Захаренкова¹

¹ Омский государственный технический университет, г. Омск, Россия

² Пизанский университет, г. Пиза, Италия

Рассматривается возможность практического применения в сетях с фрактальным трафиком [1] метода, предложенного авторами в [2, 3] для снижения средних длин очередей в системах с бесконечной дисперсией времени обслуживания. Метод основан на введении абсолютных приоритетов, назначаемых в соответствии с бесконечными разметками полуоси времени обслуживания. Из формулы Полячека – Хинчина следует, что в системах M/GI/1 с бесконечной дисперсией времени обслуживания средняя длина очереди бесконечна. В [4] это детально показано для времени обслуживания, распределенного по закону Парето. В [2] доказано, что введение абсолютных приоритетов по методу бесконечных разметок делает среднюю длину очереди в системе конечной.

Однако непосредственное применение метода бесконечных разметок (МБР) к компьютерным сетям сталкивается с двумя проблемами. Во-первых, передача пакетов, на которые разбиваются файлы, не прерывается, поэтому использовать здесь дисциплину абсолютных приоритетов не удастся. Кроме того, почти все пакеты, на которые разбиваются передаваемые файлы, имеют одинаковые относительно небольшие размеры, поэтому дисперсия времени передачи пакетов не бесконечна.

Тем не менее МБР можно применять и здесь, если в качестве заявок рассматривать не пакеты сообщений, а файлы, разбиваемые на пакеты. В этом случае абсолютные приоритеты файлов достаточно точно реализуются относительными приоритетами пакетов. Это подтверждается имитационными экспериментами, выполняемыми с одноканальной, многоканальной и сетевой моделями. Применение МБР к сетям с фрактальным трафиком с бесконечной дисперсией времени обслуживания позволяет принципиально сократить среднее время передачи файлов (с бесконечного на конечное). Но платой за это является некоторое увеличение среднего числа пакетов в системе. Далее исследуются эти особенности применения МБР к компьютерным сетям.

Математическая модель

Базовой математической моделью передачи файлов по каналу является система M/Pa/1/∞ с временем x обслуживания заявок, распределенным по закону Парето:

$$F(t) = P(x \leq t) = 1 - \left(\frac{K}{t}\right)^\alpha, \quad \alpha > 0, \quad t \geq K, \quad (1)$$

где $K > 0$ – минимальное значение случайной величины (с.в.) и, одновременно, масштабный параметр распределения; α – параметр формы, который при моделировании фрактального трафика принадлежит диапазону $1 < \alpha \leq 2$. При таких α имеем $M(x) < \infty$, $D(x) = \infty$.

В этой системе M/Pa/1/∞ при коэффициенте загрузки $\rho < 1$ имеет место стационарный режим: существует стационарное распределение времени w ожидания. Но это стационарное распределение характеризуется бесконечным средним $W = \infty$ [4].

Введение в систему абсолютных приоритетов в соответствии с МБР определяется разбиением диапазона $K \leq t < \infty$ возможных значений времени x бесконечной последовательностью точек

$$\{t_k\} = t_0, t_1, \dots, t_k, \dots \quad (2)$$

на промежутки

$$[t_0, t_1), [t_1, t_2), \dots, [t_{k-1}, t_k), \dots,$$

где $t_0 = K$. Если у входящей заявки время x принадлежит k -му интервалу – интервалу $[t_{k-1}, t_k)$, то ей присваивается абсолютный приоритет $P_k = P_1 + 1 - k$, где P_1 – приоритет, соответствующий первому интервалу – интервалу $[t_0, t_1)$. Появление с ростом k отрицательных приоритетов не меняет логики их использования.

Из разметок, рассмотренных в [2], наиболее эффективной является экспоненциальная разметка, в которой

$$t_0 = K, \quad t_k = K + ce^{ak}, \quad (k = 1, 2, \dots), \quad (3)$$

определяемая двумя параметрами $a > 0$, $c > 0$. Значения a , c легко оптимизируются в электронных таблицах по критерию $W \rightarrow \min$. Для расчета W используются известные формулы расчета систем с абсолютными приоритетами [2, 5].

Возможность применения МБР к сетям с коммутацией пакетов основана на том, что относительные приоритеты передаваемых файлов, разбиваемых на пакеты, функционируют практически так же, как абсолютные приоритеты. На рис. 1 показано, как происходит временное прекращение передачи пакетов некоторого файла (изображенных не закрашенными прямоугольниками в очереди к каналу) при поступлении в очередь пакетов более приоритетного файла (черные прямоугольники).

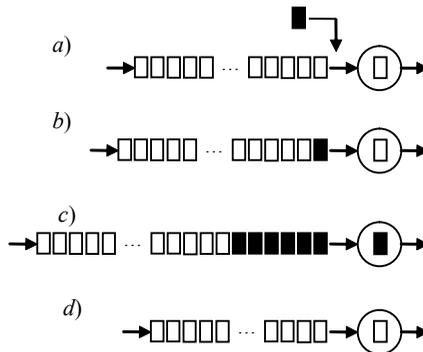


Рис. 1. Передача приоритетного файла, разбитого на пакеты (черные прямоугольники), при дисциплине относительных приоритетов

Рис. 1, *a, b* иллюстрирует поступление приоритетного «черного» пакета и ожидание им окончания передачи «белого» пакета, уже занимающего канал. Сразу после окончания передачи «белого» пакета начинает передаваться «черный», и в это время за ним в начало очереди поступают остальные «черные» пакеты приоритетного файла (рис. 1, *c*). Когда передача приоритетного файла заканчивается, возобновляется передача «белых» пакетов менее приоритетного файла (рис. 1, *d*). И, поскольку передаваемые файлы состоят из нескольких пакетов, то задержка начала обслуживания приоритетного файла, составляющая долю времени передачи одного пакета, относительно невелика, что приводит к небольшому отличию процесса передачи файлов с относительными приоритетами от процесса передачи файлов с абсолютными приоритетами.

Возможность применения МБР к передаче файлов далее обосновывается посредством имитационного моделирования.

Имитационное моделирование типового фрагмента сети

Моделируемый фрагмент сети представлен на рис. 2. Потоки файлов, поступающие на входы 1, 2 и 3, пуассоновские. Время x передачи файлов по каналу (определяемое размерами файлов) распределено по закону Парето с параметрами $K = 1$, $\alpha = 1.5$. Следовательно, $M(x) = 3$, $D(x) = \infty$.

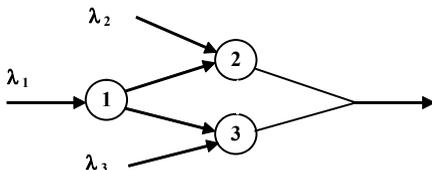


Рис. 2. Моделируемый фрагмент сети (символы очередей к каналам опущены)

Среднее время между поступлениями файлов на входы 1, 2 и 3 составляет 6, 12 и 12 единиц соответственно. Файлам присваиваются приоритеты в соответствии с МБР, с тем отличием от базовой версии МБР, что эти приоритеты не абсолютные, а относительные. При назначении приоритетов используется экспоненциальная разметка (3). В результате оптимизации ее параметров a , c установлено, что на всех входах приблизительно оптимальными являются одни и те же значения $a = 1$, $c = 0.5$.

Перед передачей файлов они разбиваются на пакеты единичной длительности, которые наследуют приоритеты файлов. Пакеты могут перемещаться разными маршрутами. Пакет из узла 1, например, поступает в узел 2, если очередь в нем короче, чем в узле 3, иначе пакет поступает в узел 3. На выходе из сети пакеты вновь собираются в файлы. В момент завершения сборки файла он считается переданным.

Результаты моделирования, полученные при прогоне приблизительно 10 млн файлов на первом входе (и, соответственно, примерно по 5 млн файлов на втором и третьем входах) представлены в табл. 1.

Два варианта сети отличаются лишь тем, что в одном из них используется МБР. Соответственно оценки коэффициентов загрузки ρ_1 , ρ_2 и ρ_3 каналов 1, 2 и 3 для сравниваемых вариантов сети получились приблизительно одинаковыми. Однако оценки среднего времени T_1 , T_2 и T_3

передачи с учетом времени ожидания файлов, поступающих на входы 1, 2 и 3 существенно различаются. Для варианта сети с МБР эти оценки сходятся к конечным значениям, показанным в последней строке табл. 1. Для беспriorитетного варианта сети соответствующие оценки не сходятся к конечным значениям и растут с ростом времени моделирования. Тем самым имитационное моделирование подтверждает предположение, что относительные приоритеты файлов в сети с коммутацией пакетов функционируют приблизительно так же, как и абсолютные приоритеты, что позволяет использовать МБР и что тем самым при бесконечных дисперсиях «чистого» времени x передачи файлов мы можем обеспечивать конечное среднее время их ожидания.

Таблица 1

Результаты прогона модели в течение 60 млн единиц модельного времени

Вариант сети	ρ_1	ρ_2	ρ_3	T_1	T_2	T_3
Без приоритетов	0.502	0.364	0.634	798.4	23.5	50.5
Сеть с МБР	0.503	0.362	0.636	5.82	3.70	3.92

Применение МБР не приводит к существенному снижению средних длин L_1 , L_2 и L_3 очередей пакетов перед каналами 1, 2, и 3 соответственно, хотя и не ухудшает эти показатели (табл. 2). Показатели L_1 , L_2 и L_3 со временем растут, поскольку их стационарные значения бесконечны. В рассматриваемой модели с бесконечными буферами это не имеет значения, но на практике приведет к проблеме снижения вероятностей потерь. Результаты, представленные в табл. 2, показывают, что скорость роста очередей чувствительна к коэффициентам загрузки каналов.

Таблица 2

Средние длины очередей пакетов при модельном времени 60 млн единиц

Вариант сети	ρ_1	ρ_2	ρ_3	L_1	L_2	L_3
Без приоритетов	0.752	0.635	0.862	2591.5	84.45	111.2
Без приоритетов	0.502	0.364	0.634	876.14	10.85	49.59
Без приоритетов	0.251	0.153	0.346	282.63	10.84	19.10
Сеть с МБР	0.754	0.632	0.864	2262.8	61.25	105.8
Сеть с МБР	0.503	0.362	0.636	806.53	19.49	44.26
Сеть с МБР	0.251	0.152	0.347	266.78	13.06	16.68

Для беспriorитетной сети и сети с МБР в табл. 2 приведены по три строки показателей, полученные соответственно при снижении быстродействия каналов по сравнению с сетью, рассмотренной в табл. 1, в полтора раза (первые строки двух вариантов сети), при таком же их быстродействии (вторые строки) и при их вдвое увеличенном быстродействии (третьи строки). Как видим, показатели L_i , а следовательно, и вероятности потерь (при конечных буферах) могут значительно снижаться за счет повышения производительности сети. При коэффициентах загрузки каналов порядка 0.25 за время передачи 10 млн файлов средние длины очередей пакетов еще далеки от значений, угрожающих потерями пакетов.

Заключение

Анализ логики функционирования относительных приоритетов в сетях с коммутацией пакетов сообщений позволяет предположить, что процессы передачи файлов с относительными приоритетами имеют свойства, близкие к свойствам процессов передачи файлов с абсолютными приоритетами и что этим обеспечивается возможность применения метода бесконечных разметок. Имитационное моделирование типового фрагмента сети подтверждает справедливость сделанного предположения. Таким образом, в условиях фрактального трафика, при бесконечной дисперсии чистого времени передачи файлов мы можем снижать бесконечное среднее время ожидания файлов до конечной величины. За счет этого столь же существенно снижается и полное среднее время передачи файлов.

ЛИТЕРАТУРА

1. *Leland W.E. et al.* On the self-similar nature of Ethernet traffic // ACM SIGCOMM'93. San Francisco, 1993. P. 183–193.
2. *Zadorozhnyi V.N., Zakharenkova T.R.* Methods to reduce loss probability in systems with infinite service time dispersion // Communications in Computer and Information Science, 2019. V. 1109. P. 296–311.
3. *Zadorozhnyi V.N., Zakharenkova T.R., Pagano M.* Queue normalization methods in systems GI/GI/1/m with infinite variance of service time // Journal of Physics: Conference Series. 2020. P. 1–11.
4. *Likhanov N., Tsybakov B., Georganas N.* Analysis of an ATM buffer with self-similar («fractal») input traffic // IEEE INFOCOM'95, 1995. V. 3. P. 985–992.
5. *Kleinrock L.* Queueing Systems. Vol. II. Computer Applications. New York: Wiley Interscience, 1976. 576 p.

Метод решения стационарных уравнений для процесса приоритетного обслуживания с разделением времени в случайной среде

Андрей Зорин, Ксения Сизова

*Нижегородский государственный университет им. Н. И. Лобачевского,
г. Нижний Новгород, Россия*

В работах [1–5] рассматривалась задача об обслуживании конфликтных ординарных или неординарных пуассоновских потоков в классе алгоритмов с разделением времени. Класс алгоритмов управления состоял в переключении прибора между очередями по информации о длинах очередей. Основным результатом этих работ – установление оптимальных свойств алгоритма обслуживания с относительным приоритетом в задаче минимизации среднего времени пребывания всех требований в системе в единицу времени или за рабочий акт обслуживающего устройства. Ключевой момент доказательства в упомянутых работах состоял в получении явных выражений для стационарных вероятностей состояний прибора, которые оказывались независимыми от выбранного алгоритма управления.

В работах [6, 7] рассматривалось обобщение задач из [3–5]. Входные потоки системы формировались в случайной среде с двумя состояниями. Явные выражения для стационарных вероятностей состояний прибора были получены только в частном предположении, что интенсивности потоков при каждом состоянии среды одинаковы. В общем случае найти эти вероятности аналитически не удастся и, следовательно, остается открытым вопрос о зависимости этих вероятностей от алгоритма управления.

В настоящей работе приведена методика решения стационарных уравнений при обслуживании с относительными приоритетами с использованием средств компьютерной алгебры.

Описание системы обслуживания

В систему обслуживания поступают m потоков $\Pi_1, \Pi_2, \dots, \Pi_m$. Требования потока Π_j помещаются в накопитель O_j с неограниченной емкостью, $j = 1, 2, \dots, m$. Положим $n = m + 1$. После обслуживания требование из очереди O_j перенаправляется на повторное обслуживание в

очередь O_r с вероятностью $p_{j,r}$, а с вероятностью $p_{j,n} = 1 - \sum_{r=1}^m p_{j,r}$ покидает систему. После каждого акта обслуживания прибор проводит переналадки и функции управления потоками. Длительность акта обслуживания требования из очереди O_j задана функцией распределения $B_j(t)$, $B_j(+0) = 0$, а длительность последующего акта переналадки и управления – функцией распределения $\bar{B}_j(t)$, $\bar{B}_j(+0) = 0$. Если по окончании акта переналадки и управления очереди пусты, прибор переходит в режим ожидания поступления требований и на обслуживание выбирается первое поступившее. Если после акта переналадки длины очередей заданы ненулевым вектором $x = (x_1, x_2, \dots, x_m)$, то немедленно начинается обслуживание заявки из очереди с номером $j = h(x)$. Здесь $h(\cdot)$ – некоторое заданное отображение целочисленной неотрицательной решетки $X = \{0, 1, \dots\} \times \dots \times \{0, 1, \dots\}$ размерности m на множество $\{1, 2, \dots, n\}$, такое, что $h(x) = j$ влечет $x_j > 0$ при $j = 1, 2, \dots, m$. Прообразом точки n является только нулевой вектор $\mathbf{0} = (0, 0, \dots, 0) \in X$.

Поступление первичных требований в систему по каждому из потоков определяется внешней случайной средой с двумя состояниями $e^{(1)}$ и $e^{(2)}$. Состояния среды могут меняться только в моменты окончания актов обслуживания и актов переналадок и управления. Вероятность перехода внешней среды из состояния $e^{(k)}$ в состояние $e^{(l)}$ равна $a_{k,l}$, где $k, l \in \{1, 2\}$. На промежутках постоянства состояний случайной среды потоки $\Pi_1, \Pi_2, \dots, \Pi_m$ условно независимы. Требования потока Π_j при состоянии среды $e^{(k)}$ поступают группами, группа содержит b требований с вероятностью $p_j(b; k)$, $b = 1, 2, \dots$. В то же время, группы образуют поток Пуассона с интенсивностью $\lambda_j^{(k)}$.

Пусть $\tau_0 = 0$ и τ_i есть либо момент окончания обслуживания требования, либо момент окончания переналадки, причем $\tau_{i+1} > \tau_i$. Введем используемые далее случайные величины и элементы. Пусть $\chi_i \in \{\chi^{(1)}, \chi^{(2)}\}$ – состояние внешней случайной среды на промежутке времени $(\tau_i, \tau_{i+1}]$. Пусть случайный элемент $\Gamma_i \in \{\Gamma^{(1)}, \Gamma^{(2)}, \dots, \Gamma^{(n)}\}$ при $i = 1, 2, \dots$ описывает состояние обслуживающего устройства на промежутке $(\tau_{i-1}, \tau_i]$. При этом равенство $\Gamma_i = \Gamma^{(s)}$ при $s = 1, 2, \dots, m$ имеет место, если происходит или только что завершился акт обслуживания очереди O_s , а при $s = n$ – происходит или только что завершился акт переналадки. Элемент Γ_0 со значениями из Γ задает состояние обслуживающего устройства в момент τ_0 . Через $\kappa_{j,i}$ обозначим длину очереди O_j в момент τ_i , $\kappa_i = (\kappa_{1,i}, \kappa_{2,i}, \dots, \kappa_{m,i})$, $i = 0, 1, \dots$. В сделанных выше пред-

положениях случайная последовательность $\{(\Gamma_i, \kappa_i, \chi_i); i = 0, 1, \dots\}$ при заданном начальном векторе $(\Gamma_0, \kappa_0, \chi_0)$ является марковской цепью, существенные состояния которой образуют один класс с двумя циклическими подклассами.

Обозначим через $Q(j, x, k), j = 1, 2, \dots, n, x \in X, k = 1, 2$, стационарную вероятность состояния $(\Gamma^{(j)}, x, e^{(k)})$. Для целочисленного вектора $x \in X$ и вектора $v = (v_1, v_2, \dots, v_m) \in \mathbf{C}^m$ введем краткую запись $v^x = v_1^{x_1} \times \dots \times v_2^{x_2} \times \dots \times v_m^{x_m}$, в частности определим $\mathbf{0}^0 = 1$. Введем производящие функции

$$f_j^{(k)}(z) = \sum_{b=1}^{\infty} p_j(b; k) z^k, \quad z \in \mathbf{C},$$

$$R_j(v) = v_j^{-1} \left(p_{j,n} + \sum_{r=1}^m p_{j,r} v_r \right), \quad v \in \mathbf{C}^m,$$

$$q_j^{(k)}(v) = \int_0^{\infty} \prod_{r=1}^m \exp\{\lambda_r^{(k)} t (f_r^{(k)}(v_r) - 1)\} dB_j(t),$$

$$\bar{q}_j^{(k)}(v) = \int_0^{\infty} \prod_{r=1}^m \exp\{\lambda_r^{(k)} t (f_r^{(k)}(v_r) - 1)\} d\bar{B}_j(t)$$

$$\Psi(v, s, k) = \sum_{x \in X} Q(s, x, k) v^x, \quad \Phi(v, j, k) = \sum_{x \in X_j} Q(n, x, k) v^x$$

и обозначим $\lambda_+^{(k)} = \lambda_1^{(k)} + \lambda_2^{(k)} + \dots + \lambda_m^{(k)}$ для $k = 1, 2$. Ранее в [6, 7] были найдены уравнения

$$\Psi(v, j, l) = \sum_{k=1}^2 a_{k,l} q_j^{(k)} R_j(v) \left(\Phi(v, j, k) + \frac{\lambda_j^{(k)}}{\lambda_+^{(k)}} f_j^{(k)}(v_j) Q(m+1, \mathbf{0}, k) \right), \quad j = \overline{1, m}, \quad (1)$$

$$\Psi(v, n, k) = \sum_{k=1}^2 a_{k,l} \sum_{j=1}^m \bar{q}_j^{(k)}(v) \Psi(v, j, k), \quad (2)$$

справедливые, по меньшей мере, в поликруге $|v_j| \leq 1, j = 1, 2, \dots, m$. Подставив в уравнения (1), (2) $v = \mathbf{1} = (1, 1, \dots, 1) \in X$, получим, с учетом условия нормировки: $\Psi(\mathbf{1}, n, k) = a_{1,k} (2(a_{1,2} + a_{2,1}))^{-1}, l \neq k$.

Метод решения стационарных уравнений для приоритетного алгоритма

Пусть $h(x) = \min\{j: x_j \neq 0\}$, что соответствует приоритетному алгоритму, причем очередь O_1 имеет наивысший приоритет, O_2 – следующая по приоритету, и т.д. Воспользуемся оператором $\mathbf{0}^j v = (0, \dots, 0, v_j, v_{j+1}, \dots, v_m)$, $v \in \mathbf{C}^m$. Тогда имеет место равенство

$$\Phi(v, j, k) = \Psi(\mathbf{0}^j v, n, k) - \Psi(\mathbf{0}^{j+1} v, n, k).$$

Тогда задача сводится к отысканию функций $\Psi(v, n, k)$, $k = 1, 2$, аналитических в области $|v_1| < 1$, $|v_2| < 1$ и удовлетворяющих уравнениям

$$\Psi(v, n, k) = \sum_{l=1}^2 a_{l,j} \sum_{j=1}^m \bar{q}_j^{(l)}(v) \sum_{l=1}^2 a_{l,l} q_j^{(l)}(v) R_j(v) \times \left(\Psi(\mathbf{0}^j v, n, l) - \Psi(\mathbf{0}^{j+1} v, n, l) + \Psi(\mathbf{0}, n, l) f_j^{(l)}(v_j) \frac{\lambda_j^{(l)}}{\lambda_+^{(l)}} \right). \quad (3)$$

Задачу можно решить, последовательно подбирая аналитические функции $v_s = v_{s,j}(v_{j+1}, v_{j+2}, \dots, v_m)$, $s = 1, 2, \dots, j-1$, чтобы исключить $\Psi(v, n, k)$, $\Psi(\mathbf{0}^2 v, n, k)$, \dots , $\Psi(\mathbf{0}^{j-1} v, n, k)$, $k = 1, 2$ из равенства (3).

Например, для исключения $\Psi(v, n, 1)$ и $\Psi(v, n, 2)$, обозначим через $Q_1(v)$ матрицу

$$\begin{pmatrix} v_1 - \sum_{l=1}^2 a_{l,1} a_{l,1} \bar{q}_1^{(l)}(v) q_1^{(l)}(v) v_1 R_1(v) & - \sum_{l=1}^2 a_{2,l} a_{l,1} \bar{q}_1^{(l)}(v) q_1^{(2)}(v) v_1 R_1(v) \\ - \sum_{l=1}^2 a_{l,1} a_{l,2} \bar{q}_1^{(l)}(v) q_1^{(l)}(v) v_1 R_1(v) & v_1 - \sum_{l=1}^2 a_{2,l} a_{l,2} \bar{q}_1^{(l)}(v) q_1^{(2)}(v) v_1 R_1(v) \end{pmatrix}$$

и рассмотрим уравнение $\det Q(v) = 0$.

Утверждение 1. При $|v_2| < 1, \dots, |v_m| < 1$ уравнение $\det Q(v) = 0$ имеет в круге $|v_1| \leq 1$ с учетом кратности ровно два нуля.

Доказательство утверждения 1 проводится на основании известной теоремы Руше. Однако доказать, что частная производная $\partial / \partial v_1 (\det Q(v))$ не обращается в ноль при указанных v_2, \dots, v_m и тем самым выделить регулярные ветви многозначной аналитической функции $v_1(v_2, \dots, v_m)$ не удается.

В данной работе используются численные методы теории функций комплексной переменной [8, 9] и система компьютерной алгебры [10] для численного решения уравнений (3) для двух потоков ($m = 2$). При

этом на каждом шаге алгоритма решения численно верифицируется выполнение соответствующих теорем анализа. Например, для проверки отсутствия у неявно заданных функций $v_1 = v_1(v_2)$ особых точек в области $|v_2| < 1$ используется формула для числа решений системы нелинейных уравнений $\det Q(v) = 0$, $\partial / \partial v_1 \det Q(v_1, v_2) = 0$, использующая многомерный аналог логарифмического вычета Коши [11].

Применение комбинации символьных вычислений и численных методов позволяет прояснить вопрос о зависимости стационарных вероятностей некоторых состояний рассматриваемой системы от назначенных приоритетов для очередей.

ЛИТЕРАТУРА

1. *Климов Г. П.* Системы обслуживания с разделением времени. I // Теория вероятностей и ее применения. 1974. Т. XIX. Вып. 3. С. 558–576.
2. *Рыков В.В., Кутаев М.Ю.* Система обслуживания с ветвящимися потоками вторичных требований // Автоматика и телемеханика. 1980. № 9. С. 52–61.
3. *Федоткин М.А.* Оптимальное управление конфликтными потоками и маркированные точечные процессы с выделенной дискретной компонентой. I // Литовский математический сборник. 1988. Т. 28. № 4. С. 783–794.
4. *Федоткин М.А.* Оптимальное управление конфликтными потоками и маркированные точечные процессы с выделенной дискретной компонентой. II // Литовский математический сборник. 1989. Т. 29. № 1. С. 148–159.
5. *Fedotkin M.A., Vysotsky A.A.* Bartlett flow control in time-sharing systems // Twelfth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes. Booklet of abstracts. Prague, 1994. С. 110–122.
6. *Зорин А.В., Федоткин М.А.* Оптимизация управления дважды стохастическими неординарными потоками в системах с разделением времени // Автоматика и телемеханика. 2005. № 7. С. 102–111.
7. *Fedotkin M.A., Zorine A.V.* Optimization of control of conflict flows with repeated service // Journal of Mathematical Sciences. 2013. V. 191. No. 4. P. 492–505.
8. *Delves L.M., Lyness J.N.* A numerical method for locating the zeros of an analytic function // Mathematics of Computation. 1967. V. 21. No. 100. P. 543–560.
9. *Kravanja P., Van Barel M.* Computing the Zeros of Analytic Functions. Springer, 2000. 116 p.
10. *Maxima*, a Computer Algebra System. Version 5.44.0 (2020). URL: <http://maxima.sourceforge.net/>
11. *Айзенберг Л.А., Южаков А.П.* Интегральные представления и вычеты в многомерном комплексном анализе. Новосибирск: Наука, 1979. 368 с.

Стационарные характеристики системы массового обслуживания с повторными вызовами и поиском на орбите

**Валентина Клименок, Александр Дудин,
Иван Ванькович**

Белорусский государственный университет, г. Минск, Беларусь

В данной статье рассматривается однолинейная экспоненциальная система массового обслуживания, которая отличается от классической системы $M/M/1$ с повторными вызовами наличием поиска запросов на орбите. В такой системе в момент окончания обслуживания включается механизм поиска, во время которого освободившийся прибор ищет запрос на орбите, чтобы принять его на обслуживание. Таким образом, когда прибор свободен, борьба за его занятие происходит в условиях конкуренции между первичными запросами, запросами, поступающими на прибор с орбиты в результате повторной попытки и в результате успешного поиска. Актуальность исследования такой системы следует из того, что возможность поиска на орбите может сократить время простоя прибора.

Поиск запросов на орбите был введен в статье [2], где рассмотрена система $M/G/1$ с повторными вызовами. Статья [3] значительно обобщает результаты [2]: в ней вводится время поиска, зависящее от числа повторных вызовов на орбите, рассматривается групповой марковский поток (BMAP – Batch Markovian arrival process) и два типа обслуживания: для первичных запросов и запросов с орбиты. Система с поиском запросов в буфере (а не на орбите) рассматривалась в [1], где после окончания обслуживания прибор начинает поиск запроса в буфере, а интенсивность экспоненциально распределенного времени поиска пропорциональна числу запросов в системе. Частный случай системы [3] с групповым пуассоновским потоком был рассмотрен в работе [4]. Система с групповым пуассоновским потоком, повторными вызовами и двумя типами поиска, в которой число запросов, взятых на обслуживание, зависит от числа запросов на орбите, рассматривалась в статье [5].

В процитированных статьях [2–5] рассмотрены системы с повторными вызовами со сложными процессами поступления и(или) обслуживания и(или) поиска, что делает невозможным получение простых

явных формул для стационарного распределения этих систем. Для нахождения этих распределений используются достаточно сложные матричные алгоритмы, при помощи которых удастся вычислить векторы стационарных вероятностей. Вместе с тем, представляется интересным проанализировать более простую экспоненциальную систему с повторными вызовами и поиском на орбите и получить ее распределение и характеристики производительности в явном, легко трактуемом виде. Анализ такой системы со стационарным пуассоновским потоком, экспоненциальными распределениями времен обслуживания, поиска и интервалов между повторными попытками проводится в данной статье.

Рассматривается однолинейная система массового обслуживания с повторными вызовами и бесконечной орбитой. Запросы поступают в систему в стационарном пуассоновском потоке с параметром λ . Если в момент поступления запроса обслуживающий прибор занят, то запрос идёт на орбиту, откуда повторяет попытки попасть на обслуживание через случайное время, распределённое по экспоненциальному закону с параметром ν . В момент окончания обслуживания включается механизм поиска запроса на орбите. Время поиска распределено по экспоненциальному закону с параметром $\gamma > 0$. В результате прибор оказывается занятым либо первичным запросом, либо повторным запросом в результате повторной попытки, либо повторным запросом в результате поиска на орбите. Времена обслуживания любых запросов распределены по экспоненциальному закону с параметром μ .

Стационарное распределение. Характеристики производительности

Процесс функционирования системы описывается двумерной цепью Маркова $\xi_t = \{i_t, n_t\}$, $t \geq 0$, где в момент времени t

- i_t – число заявок в системе;
- $n_t = \begin{cases} 0, & \text{если прибор свободен,} \\ 1, & \text{если прибор занят.} \end{cases}$

Цепь Маркова ξ_t является регулярной неприводимой и ее эргодическое распределение совпадает с единственным стационарным. Введем обозначения для стационарных вероятностей состояний цепи:

$$\alpha_i = \lim_{t \rightarrow \infty} P\{i_t = i, n_t = 0\}, \quad \kappa_i = \lim_{t \rightarrow \infty} P\{i_t = i, n_t = 1\} \quad i \geq 0. \quad (1)$$

Интуитивно понятно и нетрудно доказать, что условие существования пределов (1) совпадает с условием существования стационарного

распределения системы $M/M/1$ с повторными вызовами и имеет вид неравенства $\lambda/\mu < 1$. В дальнейшем будем предполагать, что это условие выполнено.

Теорема 1. Стационарные вероятности цепи Маркова $\xi_t, t \geq 0$, имеют следующий вид:

$$\alpha_i = \alpha_0 \rho^i \prod_{l=0}^{i-1} \frac{l+a}{l+(b+1)}, \quad i \geq 0; \quad (2)$$

$$\kappa_i = \alpha_0 \frac{\lambda + \gamma}{\mu} \rho^i \prod_{l=0}^{i-1} \frac{l+(a+1)}{l+(b+1)}, \quad (3)$$

где
$$\rho = \frac{\lambda}{\mu}, \quad a = \frac{\lambda + \gamma}{\mu}, \quad b = \frac{\gamma}{\mu};$$

$$\alpha_0 = \left[F(a, 1, b+1, \rho) + \frac{\lambda + \gamma}{\mu} F(a+1, 1, b+1, \rho) \right]^{-1}, \quad (4)$$

$F(\varphi, \phi, \chi, z)$ – гипергеометрическая функция с параметрами φ, ϕ, χ, z .

Доказательство. Используя диаграмму интенсивностей переходов, получим следующую систему уравнений равновесия для стационарных вероятностей:

$$\alpha_{i+1}((i+1)v + \gamma) = \lambda \kappa_i, \quad \alpha_i(iv + \gamma + \lambda) = \mu \kappa_i, \quad i \geq 0. \quad (5)$$

Уравнения (5) позволяют выразить все вероятности α_i через α_0 . С учетом обозначений, введенных в теореме, эти вероятности будут иметь вид (2). Далее, подставив выражения (2) во вторую подсистему в (5), получим следующую цепочку преобразований:

$$\begin{aligned} \kappa_i &= \alpha_i \frac{(iv + \lambda + \gamma)}{\mu} = \alpha_0 \frac{v}{\mu} \rho^i \prod_{l=0}^{i-1} \frac{l+a}{l+(b+1)} (a+i) = \\ &= \alpha_0 \frac{v}{\mu} \rho^i \prod_{l=0}^{i-1} \frac{l+(a+1)}{l+(b+1)} = \alpha_0 \frac{\lambda + \gamma}{\mu} \rho^i \prod_{l=0}^{i-1} \frac{l+(a+1)}{l+(b+1)}, \end{aligned}$$

из которой следуют формулы (3) для вероятностей κ_i .

Неизвестную вероятность α_0 найдем из условия нормировки

$$\sum_{i=0}^{\infty} \alpha_i + \sum_{i=0}^{\infty} \kappa_i = 1. \quad (6)$$

Для первой бесконечной суммы в (6) справедлива следующая цепочка соотношений:

$$\begin{aligned} \sum_{i=0}^{\infty} \alpha_i &= \alpha_0 \left(1 + \sum_{i=1}^{\infty} \prod_{l=0}^{i-1} \frac{l+a}{l+b+1} \rho^i \right) = \\ &= \alpha_0 \left(1 + \sum_{i=1}^{\infty} \prod_{l=0}^{i-1} \frac{(l+a)(l+1)}{(l+1)(l+b+1)} \rho^i \right) = \alpha_0 F(a, 1, b+1, \rho). \end{aligned} \quad (7)$$

Используя формулу (3), для второй бесконечной суммы в (6) получим следующее соотношение:

$$\begin{aligned} \sum_{i=0}^{\infty} \kappa_i &= \alpha_0 \frac{\lambda + \gamma}{\mu} \left(1 + \sum_{i=1}^{\infty} \rho^i \prod_{l=0}^{i-1} \frac{l+(a+1)}{l+(b+1)} \right) = \\ &= \alpha_0 \frac{\lambda + \gamma}{\mu} F(a+1, 1, b+1, \rho). \end{aligned} \quad (8)$$

Подставляя (7) и (8) в (6), получим формулу (4) для α_0 . Вычислив стационарные вероятности α_i и κ_i , можем вычислить ряд характеристик производительности системы. Некоторые из них приведены ниже.

- Вероятность того, что прибор занят

$$P_{busy} = \alpha_0 \frac{\lambda + \gamma}{\mu} F(a+1, 1, b+1, \rho).$$

- Вероятность того, что поступивший первичный запрос сразу пойдет на обслуживание $P_{imm} = \alpha_0 F(1, a, b+1, \rho)$.
- Среднее число заявок на орбите, когда прибор свободен

$$L_0 = \alpha_0 (aF(1, a+1, b+1, \rho) - aF(1, a, b+1, \rho)).$$

- Среднее число заявок на орбите, когда прибор занят

$$L_1 = \alpha_0 \frac{\lambda + \gamma}{\mu} (a+1)F(1, a+2, b+1, \rho) - F(1, a+1, b+1, \rho)$$

- Среднее число заявок на орбите $L = L_0 + L_1$.

Численный эксперимент

Как было сказано выше, использование механизма поиска на орбите позволяет в некоторых случаях существенно снизить время простоя прибора в ситуациях, когда на орбите есть запросы. В экспериментах, представленных ниже, исследуется зависимость P_{busy} и P_{imm} от интен-

сивности входного потока λ при фиксированных значениях μ и ν и различных величинах параметра поиска γ . Для экспериментов в качестве входных данных возьмем следующие величины: λ изменяется в интервале $[1, 9]$, $\mu = 10$, $\nu = 5$, $\gamma = 10, 20, 30$. На рис. 1 и 2 приведены графики зависимостей вероятностей P_{busy} и P_{imm} от интенсивности входного потока λ при фиксированных значениях μ и ν и различных величинах параметра поиска γ .

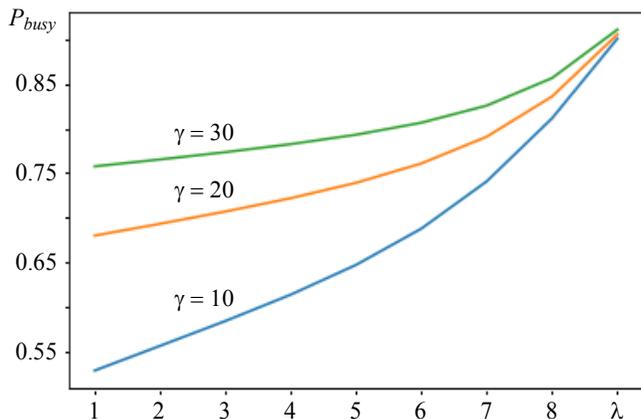


Рис. 1. Зависимость P_{busy} от λ при различных значениях γ

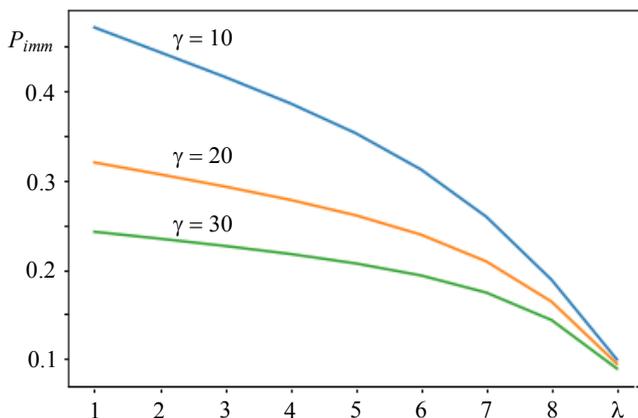


Рис. 2. Зависимость P_{imm} от λ при различных значениях γ

Из приведенных графиков следует очевидный факт, что при росте параметра λ до уровня, где нарушается стационарный режим в системе, вероятность того, что прибор занят, стремится к 1, а вероятность того, что поступивший первичный запрос пойдёт на обслуживание, не посещая орбиту, стремится к 0. Более интересно проследить зависимость этих вероятностей от параметра поиска γ . Из рисунков видно, что увеличение этого параметра может значительно сократить время простоя прибора. Если ввести критерий качества функционирования системы, включающий стоимость пребывания запроса на орбите и стоимость поиска, то может ставиться задача нахождения оптимального параметра поиска.

Заключение

В статье получены аналитические формулы для стационарного распределения экспоненциальной системы с повторными вызовами и поиском на орбите и для ее основных вероятностных характеристик. Проведенные численные эксперименты показали, что сокращение времени поиска уменьшает время простоя обслуживающего прибора. Полученные результаты могут использоваться при разработке реальных стохастических систем с повторными вызовами.

ЛИТЕРАТУРА

1. *Neuts M.F., Ramalhoto M.F.* A service model in which the server is required to search for customers // *Journal of Applied Probability*. 1984. V. 21. P. 157–166.
2. *Artalejo J.R., Joshua V.C., Krishnamoorthy A.* An M/G/1 retrial queue with orbital search by the server // *Advances in Stochastic Modeling*. New Jersey: Notable Publications Inc., 2002. P. 41–54.
3. *Dudin A.N., Krishnamoorthy A., Joshua V.C., Tsarankov G.V.* Analysis of the BMAP/G/1 retrial system with search of customers from the orbit // *European Journal of Operational Research*. 2004. V. 157. P. 169–179.
4. *Deepak T.G., Dudin A.N., Joshua V.C., Krishnamoorthy A.* On an M(X)/G/1 retrial system with two types of search of customers from the orbit // *Stochastic Analysis and Applications*. 2013. V. 31. P. 92–107.
5. *Deepak T.G.* On a Retrial queueing model with single/batch service and search of customers from the orbit // *TOP*. 2015. V. 23. P. 493–520.
6. *Dudin A.N., Deepak T.G., Varghese C. J., Krishnamoorthy A., Vishnevsky V.M.* On a BMAP/G/1 retrial system with two types of search of customers from the orbit // *Communications in Computer and Information Science*. V. 800. P. 1–12.

Анализ ожидаемого дохода в открытой сети с ограниченным числом заявок и обходами ими систем обслуживания

Дмитрий Копать, Михаил Маталыцкий

*Гродненский государственный университет имени Янки Купалы,
г. Гродно, Беларусь*

Проведено исследование открытой марковской сети массового обслуживания (СеМО) с ограниченным числом заявок, обходами ими систем обслуживания и доходами. Параметры обслуживания сети постоянны, маршрут заявок определяется в соответствии с матрицей вероятностей переходов. Состояние сети описывается цепью Маркова (ЦМ) с непрерывным временем и конечным числом состояний. С помощью асимптотического анализа при большом числе заявок найден полный ожидаемый доход в такой сети.

Введение. Описание сети

СеМО с обходами узлов заявками введена в работе [1]. В ней показано, что такая модель включает возможность обхода систем за счет ограничений на количество заявок или на предполагаемое время ожидания. Найден стационарные вероятности состояний сети в форме произведения. В переходном режиме сеть с обходами была исследована методом многомерных производящих функций в работе [2]. Выражения для ожидаемых доходов такой сети, когда доходы от переходов являются случайными величинами (СВ) были получены в статье [3]. Средние характеристики для сети с обходами с помощью асимптотического анализа были найдены авторами [4]. Применение сети с обходами связана с возможностью клиента, прибывшего в сервисный центр информационной сети, не присоединяться к очереди по тем или иным причинам, а перейти в другой сервисный центр. В данной статье с помощью асимптотического анализа найден ожидаемый доход сети с обходами систем обслуживания (СМО) и доходами.

Рассмотрим открытую экспоненциальную СеМО с однотипными заявками, состоящую из n однолинейных СМО. Состояние сети в момент времени t описывается вектором размерности $n + 1$: $\vec{k}(t) = (\vec{k}, t) =$

$= (k_1, k_2, \dots, k_n, t)$, который образует ЦМ с непрерывным временем и счетным числом состояний, где k_i – число заявок в i -й СМО, $i = \overline{1, n}$.

В i -ю систему из внешней среды поступает простейший поток заявок с интенсивностью λ_{0i} , $i = \overline{1, n}$. Все потоки заявок, которые поступают в сеть, независимы. Длительности обслуживания заявок в i -й СМО распределены по показательному закону с параметром μ_i , $i = \overline{1, n}$.

Заявка, направленная в i -ю СМО извне или из другой системы, когда сеть находится в состоянии k , с вероятностью $f^{(i)}(k_i)$ присоединяется к очереди, а с дополнительной вероятностью $1 - f^{(i)}(k_i)$ не присоединяется к очереди, считаясь мгновенно обслуженной (т. е. обходит СМО).

Заявка, обслуженная в СМО S_i , с вероятностью p_{ij} направляется в СМО S_j и с вероятностью $\sum_{j=0}^n p_{ij} = 1$ уходит из сети (в СМО S_0), $i, j = \overline{1, n}$. Введём следующие условные вероятности.

Пусть $\phi_i(\vec{k})$ – условная вероятность того, что заявка, поступающая в i -ю СМО, когда сеть находится в состоянии \vec{k} , не будет обслужена ни одной из СМО и не изменит состояние сети;

$\psi_{ij}(\vec{k})$ – условная вероятность того, что заявка, поступающая в i -ю СМО, когда сеть находится в состоянии \vec{k} , впервые получит обслуживание в j -й СМО, $j = \overline{1, n}$;

$\alpha_i(\vec{k})$ – условная вероятность того, что заявка, обслуженная в i -й СМО, завершена и когда сеть находится в состоянии \vec{k} , не будет больше обслужена ни в одной из СМО и уйдёт из сети;

$\beta_{ij}(\vec{k})$ – условная вероятность того, что заявка, обслуживание которой в i -й СМО завершено, когда сеть находится в состоянии \vec{k} , впервые после этого получит обслуживание в j -й СМО, $i, j = \overline{1, n}$.

На основании формулы полной вероятности получаем [1], что

$$\phi_i(\vec{k}) = (1 - f^{(i)}(\vec{k})) \left(p_{i0} + \sum_{j=1}^n p_{ij} \phi_j(\vec{k}) \right), \quad i = \overline{1, n}; \quad (1)$$

$$\psi_{ij}(\vec{k}) = f^{(i)}(\vec{k})\delta_{ij} + (1 - f^{(i)}(\vec{k}))\sum_{l=1}^n p_{il}\psi_{lj}(\vec{k}), \quad i, j = \overline{1, n}; \quad (2)$$

$$\alpha_i(\vec{k}) = p_{i0} + \sum_{j=1}^n [p_{ij}\phi_j(\vec{k} - I_i)], \beta_{ij}(\vec{k}) = \sum_{l=1}^n p_{il}\psi_{lj}(\vec{k} - I_i), \quad i, j = \overline{1, n}, \quad (3)$$

где I_i – нулевой вектор размерности n , за исключением компоненты с номером i , которая равна 1, $u(x)$ – единичная функция Хевисайда. В результате перехода заявок из одной системы в другую доход первой системы убывает, а второй возрастает на эту же величину.

Обозначим через $V(\vec{k}, t)$ – полный ожидаемый доход, который получит сеть за время t , если в начальный момент времени она находится в состоянии \vec{k} . Очевидно, что $V(\vec{k}, t) = \sum_{i=0}^n V_i(\vec{k}, t)$, где $V_i(\vec{k}, t)$ – ожидаемый доход, который получает система S_i за время t , если в начальный момент времени сеть находится в состоянии \vec{k} .

Уравнение в частных производных для плотности распределения ожидаемого дохода сети

Аналогично как в [6] можно показать справедливость леммы.

Лемма. Ожидаемый доход сети удовлетворяет следующей системе разностно-дифференциальных уравнений (РДУ):

$$\begin{aligned} \frac{dV(\vec{k}, t)}{dt} = & R - \left[\lambda_{0i} (1 - \phi_i(\vec{k})u(k_i)) + \mu_i u(k_i) (1 - \beta_{ii}(\vec{k})) + \lambda_{0i} u(k_i) \right] \times \\ & \times V(\vec{k}, t) + \sum_{i=1}^n \mu_i \alpha_i(\vec{k} - I_i) u(k_i) (-R_{i0} + V(\vec{k} - I_i, t)) + \\ & + \sum_{i,j=1}^n \left[\lambda_{0j} \psi_{ji}(\vec{k} + I_i) u(k_i) (r_{0i} + V(\vec{k} + I_i, t)) + \right. \\ & \left. + \mu_j \beta_{ji}(\vec{k} - I_j + I_i) u(k_j) (R_{ij} + V(\vec{k} - I_i + I_j, t)) \right]. \end{aligned} \quad (4)$$

Решить систему РДУ (4) аналитически при большом числе n не представляется возможным. Исследование этой системы будем проводить, используя методику, описанную в [5]. Перейдем к относительным переменным и будем исследовать вектор $\vec{\xi}(t) = (k_1(t)K^{-1}, k_2(t)K^{-1},$

..., $k_n(t)K^{-1}$), где K – величина, ограничивающая общее число заявок в сети. Возможные значения вектора $\vec{\xi}(t)$ принадлежат ограниченному замкнутому множеству $G = \left\{ \vec{x} = (x_1, x_2, \dots, x_n) : x_i \geq 0, \sum_{i=1}^n x_i \leq 1 \right\}$, в котором они располагаются в узлах n -мерной решетки на расстоянии $\varepsilon = K^{-1}$ друг от друга. При увеличении значений K «плотность заполнения» множества G возможными значениями рассматриваемого вектора $\vec{\xi}(t)$ увеличивается, и становится возможным считать, что он имеет непрерывное распределение в области G . При этих предположениях можно считать, что полный ожидаемый доход сети непрерывно изменяется в зависимости от начального состояния (\vec{x}, t) .

Поэтому можем ввести в рассмотрение функцию плотности распределения (концентрации) ожидаемого дохода $\rho(\vec{x}, t)$ в области G , которую определим как следующий предел:

$$\rho(\vec{x}, t) = \lim_{\varepsilon \rightarrow 0} V(x_1 \leq \xi_1 \leq x_1 + \varepsilon, \dots, x_n \leq \xi_n \leq x_n + \varepsilon) \varepsilon^{-n}. \quad (5)$$

Из (5) получим, что для $V(\vec{k}, t)$ справедлива следующая аппроксимация при $K \rightarrow \infty$:

$$V(\vec{k}, t) = V(\vec{x}K, t) = \varepsilon^n \rho(\vec{x}, t) \quad \text{или} \quad \rho(\vec{x}, t) = K^n V(\vec{x}K, t). \quad (6)$$

Введём обозначения:

$$K^n R = r, \quad K^n r_{0i} = r_{0i}^{(1)}, \quad K^n R_{i0} = R_{i0}^{(1)}, \quad K^n R_{ij} = R_{ij}^{(1)}. \quad (7)$$

Из уравнения (4) с учетом (5) – (7), следует, что для плотности распределения дохода $\rho(\vec{x}, t)$ уравнение может быть представлено в виде

$$\begin{aligned} \frac{d\rho(\vec{x}, t)}{dt} = & r + K \sum_{i=1}^n \left\{ \mu_i \alpha_i (\vec{x} - I_i) u(x_i) [\rho(\vec{x} - e_i, t) - \rho(\vec{x}, t)] - \mu_i \times \right. \\ & \times \alpha_i (\vec{x} - e_i) u(x_i) R_{i0}^{(1)} \left. \right\} + K \sum_{i,j=1}^n \left[\lambda_{0j} \psi_{ji} (\vec{x} + e_j) u(x_j) [\rho(\vec{x} + e_j, t) - \rho(\vec{x}, t)] + \right. \\ & + \mu_j \beta_{ji} (\vec{x} - e_j + e_i) u(x_j) [\rho(\vec{x} - e_i + e_j, t) - \rho(\vec{x}, t)] - \\ & \left. + [\lambda_{0j} \psi_{ji} (\vec{x} + e_j) u(x_j) R_{0i}^{(1)} + \mu_j \beta_{ji} (\vec{x} - e_j + e_i) u(x_j) R_{ij}^{(1)}] \right]. \end{aligned}$$

Теорема. Плотность распределения дохода сети с точностью до $O(\varepsilon^2)$ удовлетворяет следующему ДУ в частных производных:

$$\begin{aligned} \frac{\partial \rho(\bar{x}, t)}{\partial t} = & - \sum_{i=1}^n A_i(\bar{x}, t) \frac{\partial \rho(\bar{x}, t)}{\partial x_i} - \sum_{i=1}^n \mu_i \alpha_i (\bar{x} - e_i) u(x_i) R_{i0}^{(1)} + \\ & + \sum_{i,j=1}^n \left[\lambda_{0j} \psi_{ji}(\bar{x} + e_i) u(x_i) r_{0i}^{(1)} + \mu_j \beta_{ji}(\bar{x} - e_j + e_i) u(x_j) R_{ij}^{(1)} \right], \end{aligned} \quad (9)$$

где $A_i(\bar{x}, t) = (\mu_i \alpha_i (\bar{x} - e_i) - \lambda_{0j} \psi_{ji}(\bar{x} + e_i)) u(x_i) + \mu_j \beta_{ji}^*(\bar{x} - e_j + e_i) u(x_j)$.

Нахождение ожидаемого дохода сети

Интегрируя обе части (9) по \bar{x} в области G и разделив обе части уравнения на объем области G , равный $m(G) = (n!)^{-1}$ [6], получим

$$\begin{aligned} (m(G))^{-1} \iint_G \dots \int \frac{\partial \rho(\bar{x}, t)}{\partial t} d\bar{x} = & - (m(G))^{-1} \iint_G \dots \int \sum_{i=1}^n A_i(\bar{x}, t) \frac{\partial \rho(\bar{x}, t)}{\partial x_i} d\bar{x} + \\ & + (m(G))^{-1} \iint_G \dots \int \left\{ \sum_{i=1}^n \mu_i \alpha_i (\bar{x} - e_i) u(x_i) R_{i0}^{(1)} + \right. \\ & \left. + \sum_{i,j=1}^n \left[\lambda_{0j} \psi_{ji}(\bar{x} + e_i) u(x_i) r_{0i}^{(1)} + \mu_j \beta_{ji}(\bar{x} - e_j + e_i) u(x_j) R_{ij}^{(1)} \right] d\bar{x} \right\}. \end{aligned}$$

Будем считать, что в левой части этого равенства допустима переменная порядка интегрирования и дифференцирования (пусть в замкнутой области G функция $\rho(\bar{x}, t)$ является непрерывной):

$$(m(G))^{-1} \iint_G \dots \int \frac{\partial \rho(\bar{x}, t)}{\partial t} d\bar{x} = (m(G))^{-1} \frac{\partial}{\partial t} \iint_G \dots \int \rho(\bar{x}, t) d\bar{x} = \frac{d}{dt} \bar{v}_G(t),$$

где $\bar{v}_G(t)$ – среднее по \bar{x} значение дохода при условии изменения начального состояния (\bar{x}, t) в области G . Тогда:

$$\begin{aligned} \frac{d}{dt} \bar{v}_G(t) = & \sum_{i=1}^n \frac{\partial A_i(x, t)}{\partial x_i} \bar{v}_G(t) + (m(G))^{-1} \iint_G \dots \int \left\{ \sum_{i=1}^n \mu_i \alpha_i (\bar{x} - e_i) u(x_i) R_{i0}^{(1)} + \right. \\ & \left. + \sum_{i,j=1}^n \left[\lambda_{0j} \psi_{ji}(\bar{x} + e_i) u(x_i) r_{0i}^{(1)} + \mu_j \beta_{ji}(\bar{x} - e_j + e_i) u(x_j) R_{ij}^{(1)} \right] \right\} d\bar{x}. \end{aligned}$$

Его решение имеет вид

$$\begin{aligned} \overline{v}_G(t) = & e^{\int_0^t \sum_{i=1}^n \frac{\partial A_i(\bar{x}, z)}{\partial x_i} dz} \left(\overline{v}_G(0) - \frac{1}{m(G)} \int_0^t \iiint_G \dots \int \left\{ \sum_{i=1}^n \mu_i \alpha_i (\bar{x} - e_i) u(x_i) R_{i0}^{(1)} + \right. \right. \\ & \left. \left. + \sum_{i,j=1}^n \left[\lambda_{0j} \Psi_{ji} (\bar{x} + e_i) u(x_i) r_{0i}^{(1)} + \mu_j \beta_{ji} (\bar{x} - e_j + e_i) u(x_j) R_{ij}^{(1)} \right] d\bar{x} d\tau \right. \right) \end{aligned}$$

Заключение

Проведен асимптотический анализ ожидаемого дохода для сети с ограниченным числом заявок и обходами ими систем обслуживания. Получены выражения для среднего по вектору относительных переменных дохода сети.

ЛИТЕРАТУРА

1. *Малинковский Ю.В.* Сети массового обслуживания с обходами узлов заявками // Автоматика и телемеханика. 1991. № 2. С. 102–110.
2. *Матальцкий М.А., Науменко В.В.* Анализ сети с обходами систем обслуживания разнотипными заявками // Вестник ГрГУ. Сер. 2. Математика, физика, информатика, вычислительная техника и управление. 2013. № 1. С. 152–159.
3. *Matalytski M., Naumenko V.* Finding incomes of HM-network with one-type messages bypass of systems // Scientific Research of the Institute of Mathematics and Computer Science Czestochowa University of Technology. 2012. V. 11. No. 3. P. 111–123.
4. *Галицкая-Петровская А.О., Матальцкий М.А.* Асимптотический анализ открытой сети обслуживания с ограниченным числом разнотипных заявок и обходами // Вестник ГрГУ. Сер. 2. 2019. Т. 9. № 1. С. 128–139.
5. *Медведев Г.А.* Замкнутые системы массового обслуживания и их оптимизация // Известия АН СССР. Техническая кибернетика. 1978. № 6. С. 199–203.
6. *Копать Д.Я., Матальцкий М.А.* Анализ ожидаемого дохода в открытой марковской сети обслуживания с ограниченным числом заявок и случайным временем их ожидания в очередях // Известия Гомельского государственного университета имени Ф. Скорины. 2019. № 6. С.137–143.

Исследование СМО вида ММРР|М|N с обратной связью методом асимптотически диффузионного анализа

Анатолий Назаров, Екатерина Павлова

*Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

Представлены результаты исследования СМО с N обслуживающими приборами, с обратной связью и орбитой. Заявки могут поступать на орбиту в неограниченном количестве. Входящий поток является марковским модулированным пуассоновским (ММРР). Методом асимптотически диффузионного анализа находятся распределения вероятностей числа занятых приборов в системе и числа заявок на орбите.

Введение

Под обратной связью понимают повторное обращение заявки к обслуживающему прибору. Самым простым примером систем с обратной связью являются сети коммуникации, в ходе использования которых повреждение данных приводит к повторному отправлению данных адресату.

В [1] представлены результаты исследований СМО с мгновенной обратной связью, а в [2] изучены СМО с отсроченной обратной связью. Также ряд работ [3–5] посвящен исследованию систем с двумя типами обратной связи.

Математическая модель и постановка задачи

Рассмотрим систему массового обслуживания с N обслуживающими устройствами и обратной связью. На вход системы поступает ММРР-поток заявок, заданный диагональной матрицей условных интенсивностей $\Lambda = [\lambda_k]$, $k = 1, 2, \dots, K$, матрицей инфинитезимальных характеристик $\mathbf{Q} = [q_{ij}]$, $i, j = 1, 2, \dots, N$, управляющей потоком цепи Маркова $k(t) = 1, 2, \dots, K$.

Заявка, поступая в систему, занимает один из свободных приборов и обслуживается в течение случайного времени, распределенного экспоненциально с параметром μ . Если, при поступлении в систему, заявка обнаружит все приборы занятыми, она мгновенно отправляется на ор-

биту, где осуществляет задержку в течение случайного времени, экспоненциально распределенного с параметром σ .

В момент завершения обслуживания заявка может покинуть систему с вероятностью r_0 ; осуществляя мгновенную обратную связь, отправиться на повторное обслуживание с вероятностью r_1 ; осуществляя отсроченную обратную связь, отправиться на орбиту с вероятностью r_2 .

Обозначим $k(t) = k$ – состояние цепи Маркова, управляющей ММРР-поток в момент времени t , $k = 1, 2, \dots, K$, $n(t)$ – число занятых приборов в системе в момент времени t , $n = 0, 1, \dots, N$, $i(t)$ – число заявок на орбите в момент времени t .

Ставится задача получения трехмерного стационарного распределения вероятностей $P(k, n, i) = P\{k(t) = k, n(t) = n, i(t) = i\}$.

Система уравнений Колмогорова

Рассмотрим марковский процесс $\{k(t), n(t), i(t)\}$, для его нестационарного распределения вероятностей $P\{k(t) = k, n(t) = n, i(t) = i\} = P(k, n, i, t)$ запишем систему дифференциальных уравнений Колмогорова

$$\begin{aligned} \frac{\partial P(k, n, i, t)}{\partial t} = & -(\lambda_k + (1 - r_1)n\mu + i\sigma)P(k, n, i, t) + \lambda_k P(k, n - 1, i, t) + \\ & + (i + 1)\sigma P(k, n - 1, i + 1, t) + (n + 1)r_0\mu P(k, n + 1, i, t) + \\ & + (n + 1)r_2\mu P(k, n + 1, i - 1, t) + \sum_v q_{vk} P(v, n, i, t), 0 \leq n \leq N - 1, \\ \frac{\partial P(k, N, i, t)}{\partial t} = & -(\lambda_k + (1 - r_1)N\mu)P(k, N, i, t) + \lambda_k P(k, N - 1, i, t) + \\ & + \lambda_k P(k, N, i - 1, t) + (i + 1)\sigma P(k, N - 1, i + 1, t) + \sum_v q_{vk} P(v, N, i, t). \end{aligned} \quad (1)$$

Введем частичные характеристические функции вида

$$H(k, n, u, t) = \sum_{i=0}^{\infty} e^{jui} P(k, n, i, t),$$

где $j = \sqrt{-1}$ мнимая единица. Тогда можем записать систему (1) для характеристических функций

$$\begin{aligned}
\frac{\partial H(k, n, u, t)}{\partial t} &= -(\lambda - k + n\mu(1 - r_1))H(k, n, u, t) + \lambda_k H(k, n - 1, u, t) + \\
&+ j\sigma \frac{\partial H(k, n, u, t)}{\partial u} + (n + 1)\mu(r_0 + r_2 e^{ju})H(k, n + 1, u, t) - \\
&- j\sigma e^{-ju} \frac{\partial H(k, n - 1, u, t)}{\partial u} + \sum_v q_{vk} H(v, n, u, t), 0 \leq n \leq N - 1, \\
\frac{\partial H(k, N, u, t)}{\partial t} &= (\lambda_k (e^{ju} - 1) - N\mu(1 - r_1))H(k, N, u, t) + \\
&+ \lambda_k H(k, N - 1, u, t) - j\sigma e^{-ju} \frac{\partial H(k, N - 1, u, t)}{\partial u} + \sum_v q_{vk} H(v, N, u, t). \quad (2)
\end{aligned}$$

Обозначим вектор-строки

$$\mathbf{H}(n, u, t) = \{H(1, n, u, t), \dots, H(K, n, u, t)\},$$

$$\mathbf{H}(u, t) = \{\mathbf{H}(0, u, t), \mathbf{H}(1, u, t), \dots, \mathbf{H}(N, u, t)\}$$

и перепишем (2) в матричном виде с учетом введенных обозначений

$$\frac{\partial \mathbf{H}(u, t)}{\partial t} = \mathbf{H}(u, t)(\mathbf{A} + e^{ju}\mathbf{B}) + j\sigma \frac{\partial \mathbf{H}(u, t)}{\partial u} (\mathbf{I}_0 - e^{-ju}\mathbf{I}_1).$$

Здесь \mathbf{I} – единичная матрица размерности $K \times K$,

$$\mathbf{A} = \begin{bmatrix} \mathbf{Q} - \mathbf{\Lambda} & \mathbf{\Lambda} & 0 & \dots & 0 \\ \mu r_0 \mathbf{I} & \mathbf{Q} - \mathbf{\Lambda} - \mu \mathbf{I}(r_0 + r_2) & \mathbf{\Lambda} & \dots & 0 \\ 0 & 2\mu \mathbf{I} r_0 & \mathbf{Q} - \mathbf{\Lambda} - 2\mu \mathbf{I}(r_0 + r_2) & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \mathbf{Q} - \mathbf{\Lambda} - N\mu \mathbf{I}(r_0 + r_2) \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ \mu \mathbf{I} r_2 & 0 & \dots & 0 & 0 \\ 0 & 2\mu \mathbf{I} r_2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & N\mu \mathbf{I} r_2 & \mathbf{\Lambda} \end{bmatrix},$$

$$\mathbf{I}_0 = \begin{bmatrix} \mathbf{I} & 0 & \dots & 0 & 0 \\ 0 & \mathbf{I} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \mathbf{I} & 0 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}, \quad \mathbf{I}_1 = \begin{bmatrix} 0 & \mathbf{I} & 0 & \dots & 0 & 0 \\ 0 & 0 & \mathbf{I} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & \mathbf{I} \\ 0 & 0 & 0 & \dots & 0 & 0 \end{bmatrix}.$$

Домножая матричное уравнение на единичный вектор-столбец \mathbf{e} , принимая во внимание $(\mathbf{A} + \mathbf{B})\mathbf{e} = 0$ и $(\mathbf{I}_0 - \mathbf{I}_1)\mathbf{e} = 0$, получим скалярное уравнение и запишем систему

$$\begin{aligned} \frac{\partial \mathbf{H}(u, t)}{\partial t} &= \mathbf{H}(u, t)(\mathbf{A} + e^{ju}\mathbf{B}) + j\sigma \frac{\partial \mathbf{H}(u, t)}{\partial u}(\mathbf{I}_0 - e^{-ju}\mathbf{I}_1), \\ \frac{\partial \mathbf{H}(u, t)}{\partial t} \mathbf{e} &= [\mathbf{H}(u, t)\mathbf{B}\mathbf{e} + j\sigma e^{-ju} \frac{\partial \mathbf{H}(u, t)}{\partial u} \mathbf{I}_0 \mathbf{e}] (e^{ju} - 1). \end{aligned} \quad (3)$$

Будем искать решение задачи (3) методом асимптотически диффузионного анализа в условии большой задержки заявки на орбите, то есть $\sigma \rightarrow 0$.

Асимптотически диффузионный анализ

Обозначим $\sigma = \varepsilon$ и выполним следующие замены в (3):

$$u = \varepsilon w, \tau = t\varepsilon, \mathbf{H}(u, t) = \mathbf{F}(w, \tau, \varepsilon).$$

С учетом замен перепишем (3)

$$\begin{aligned} \varepsilon \frac{\partial \mathbf{F}(w, \tau, \varepsilon)}{\partial \tau} &= \mathbf{F}(w, \tau, \varepsilon)(\mathbf{A} + e^{j\varepsilon w}\mathbf{B}) + j \frac{\partial \mathbf{F}(w, \tau, \varepsilon)}{\partial w}(\mathbf{I}_0 - e^{-j\varepsilon w}\mathbf{I}_1), \\ \varepsilon \frac{\partial \mathbf{F}(w, \tau, \varepsilon)}{\partial \tau} \mathbf{e} &= [\mathbf{F}(w, \tau, \varepsilon)\mathbf{B}\mathbf{e} + j e^{-j\varepsilon w} \frac{\partial \mathbf{F}(w, \tau, \varepsilon)}{\partial w} \mathbf{I}_0 \mathbf{e}] (e^{j\varepsilon w} - 1). \end{aligned}$$

Теорема 1. В системе массового обслуживания с обратной связью в предельном условии $\sigma \rightarrow 0$ предельная характеристическая функция нормированного числа $i(t)$ заявок на орбите имеет вид

$$\lim_{\sigma \rightarrow 0} M \{e^{jw\sigma i(t)}\} = e^{jw\chi(\tau)},$$

где функция $\chi(\tau)$ является решением дифференциального уравнения $\chi'(\tau) = a(\chi)$, а функция $a(\chi)$ определяется равенством

$$a(x) = \mathbf{R}(\mathbf{B} - x(\tau)\mathbf{I}_0)\mathbf{e}. \quad (4)$$

Здесь вектор-строка $\mathbf{R} = \{R(0), R(1), \dots, R(N)\}$ является решением системы

$$\mathbf{R}\{(\mathbf{A} + \mathbf{B}) - x(\tau)(\mathbf{I}_0 - \mathbf{I}_1)\} = 0,$$

$$\mathbf{R}\mathbf{e} = 1.$$

Теорема 2. Предельная плотность распределения вероятностей нормированного числа заявок на орбите в рассматриваемой системе с обратной связью имеет вид

$$\pi(z) = \frac{C}{b(z)} \exp\left\{\frac{2}{\sigma} \int_0^z \frac{a(x)}{b(x)} dx\right\}, \quad (5)$$

где C – нормирующая константа, функция $a(x)$ определяется выражением (4), а $b(x)$ – равенством

$$b(x) = a(x) + 2\mathbf{g}[\mathbf{B} - x\mathbf{I}_0]\mathbf{e} + 2\mathbf{R}x\mathbf{I}_0\mathbf{e}, \quad (6)$$

а вектор \mathbf{g} является решением системы уравнений

$$\mathbf{g}(\mathbf{A} + \mathbf{B} + x(\mathbf{I}_1 - \mathbf{I}_0)) = a(x)\mathbf{R} + \mathbf{R}(x\mathbf{I}_1 - \mathbf{B}),$$

$$\mathbf{g}\mathbf{e} = 0.$$

Рассмотрим выражение (5), подставляя в него набор аргументов $k\sigma$, где $k = 0, 1, 2, \dots$, получим набор чисел

$$\pi(k\sigma) = \frac{C}{b(k\sigma)} \exp\left\{\frac{2}{\sigma} \int_0^{k\sigma} \frac{a(x)}{b(x)} dx\right\},$$

к которому применим условие нормировки, получим дискретное распределение вероятностей

$$P(k) = \frac{\pi(k\sigma)}{\sum_{k=0}^{\infty} \pi(k\sigma)}.$$

Таким образом, получена аппроксимация $P(k)$ дискретного распределения числа заявок на орбите в рассматриваемой системе с обратной связью.

Заклучение

При предельном условии большой задержки ($\sigma \rightarrow 0$) заявок на орбите для СМО вида $MPP|M|N$ с обратной связью получены асимптотические распределения вероятностей числа занятых приборов в системе и нормированного числа заявок на орбите.

ЛИТЕРАТУРА

1. *Takacs L.* A single-server queue with feedback // Bell System Technical Journal. 1963. V. 42. P. 505–519.
2. *Lee H.W., Seo D.W.* Design of a production system with feedback buffer // Queueing Systems. 1997. V. 26. Iss. 1. P. 187–198.
3. *Melikov A.Z., Ponomarenko L.A., Rustamov A.M.* Methods for analysis of queuing models with instantaneous and delayed feedbacks // Communications in Computer and Information Sciences. 2015. V. 564. P. 185–199.
4. *Koroliuk V.S., Melikov A.Z., Ponomarenko L.A., Rustamov A.M.* Methods for analysis of multi-channel queuing models with instantaneous and delayed feedbacks // Cybernetics and System Analysis. 2016. V. 52. Issue 1. P. 58–70.
5. *Melikov A.Z., Ponomarenko L.A., Rustamov A.M.* Hierarchical space merging algorithm to analysis of open tandem queuing networks // Cybernetics and System Analysis. 2016. V. 52. Iss. 6. P. 867–877.

Исследование системы с обратной связью, рекуррентным обслуживанием и неординарным пуассоновским входящим потоком

Анатолий Назаров¹, Светлана Рожкова^{1,2},
Екатерина Титаренко^{1,2}

¹ *Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

² *Национальный исследовательский
Томский политехнический университет, г. Томск, Россия*

В настоящее время для обеспечения надежности инфокоммуникационных систем используют модель RQ-системы с обратными связями, которая предполагает повторное облуживание заявок, уже получивших облуживание [1].

Данное исследование посвящено системам массового облуживания (СМО) с мгновенными и отложенными обратными связями, неординарным пуассоновским входящим потоком, рекуррентным облуживанием и является продолжением исследования [2], в котором рассматривалось экспоненциальное облуживание. При исследовании систем с произвольным облуживанием используются дополнительные переменные, такие, как истекшее и остаточное время [3].

В работе при построении системы уравнений Колмогорова используется истекшее время. В этом случае рассматриваемая система допускает аналитическое решение, в отличие от метода с использованием остаточного времени, при котором необходимо применять асимптотические способы решения. Получено стационарное распределение вероятностей числа заявок на орбите.

Описание модели и постановка задачи

Рассмотрим СМО $M^n/GI/1$ с повторными вызовами (рис. 1). На вход поступает пуассоновский неординарный поток заявок с параметром λ и заданными вероятностями q_v появления v заявок в группе ($v > 1$, $q_0 = 0$, $\sum_{v=1}^{\infty} q_v = 1$). Если облуживающий прибор свободен, то одна заявка поступает на облуживание, остальные попадают на орбиту. Заявки, поступившие в момент, когда прибор занят, также попадают на орбиту.

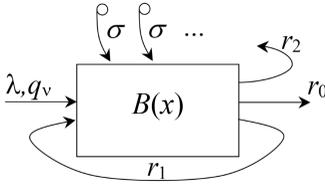


Рис. 1. Модель СМО

Длительность обслуживания заявки является случайной величиной с произвольной функцией распределения $B(x)$. Заявка, обслуживание которой завершено, покидает систему с вероятностью r_0 , повторно поступает на обслуживание с вероятностью r_1 или переходит на орбиту с вероятностью r_2 , таким образом $r_0 + r_1 + r_2 = 1$. На

орбите заявки ожидают повторного обслуживания в течение времени, распределенного по экспоненциальному закону с параметром σ , после чего повторяют попытку занять прибор. В случае неудачной попытки заявки остаются на орбите.

Обозначим $i(t)$ – число заявок на орбите в момент времени t . Процесс $n(t)$ определяет состояние прибора: $n(t) = 0$, если прибор свободен, $n(t) = 1$, если прибор занят. Требуется найти стационарное распределение числа заявок на орбите с учетом состояния прибора.

Уравнения Колмогорова с использованием истекшего времени обслуживания

Случайный процесс $\{i(t)\}$, а также процесс $\{i(t), n(t)\}$ не являются марковскими, поэтому введем дополнительную переменную $y(t)$. Процесс $y(t)$ представляет собой истекшее время обслуживания, т.е. длину интервала времени от момента начала обслуживания заявки до момента t . Если прибор свободен, то процесс $y(t)$ не определяется. Тогда марковским является случайный процесс с переменным числом компонент $\{i(t), n(t), y(t)\}$. Этот процесс состоит из двух компонент $i(t), n(t)$, если прибор свободен, и из трех компонент $i(t), n(t), y(t)$, если прибор занят.

Обозначим вероятности числа заявок на орбите в момент времени t

$$P_0(i, t) = P\{i(t) = i, n(t) = 0\},$$

$$P_1(i, y, t) = \partial P\{i(t) = i, n(t) = 1, y(t) < y\} / \partial y.$$

Рассмотрим интенсивность обслуживания $\mu(y) = B'(y) / [1 - B(y)]$, тогда

$$\mu(y)\Delta t = \frac{B'(y)\Delta t}{1 - B(y)} = \frac{B(y + \Delta t) - B(y)}{1 - B(y)} = P\{y < \tau < y + \Delta t / \tau > y\}$$

– условная вероятность того, что время обслуживания τ закончится за $y + \Delta t$ при условии, что оно продолжается дольше y .

Запишем вероятности в момент времени $t + \Delta t$:

$$P_0(i, t + \Delta t) = \int_0^{\infty} r_0 \mu(y) \Delta t P_1(i, y, t) dy + \int_0^{\infty} r_2 \mu(y) \Delta t P_1(i-1, y, t) dy + \\ + (1 - \lambda \Delta t)(1 - i \sigma \Delta t) P_0(i, t) + o(\Delta t);$$

$$P_1(i, y + \Delta t, t + \Delta t) = (1 - \lambda \Delta t)(1 - \mu(y) \Delta t) P_1(i, y, t) + \\ + \sum_{v=1}^i \lambda q_v \Delta t P_1(i-v, y, t) + o(\Delta t).$$

Если истекшее время мало, то

$$\int_0^{\Delta t} P_1(i, x, t + \Delta t) dx = \int_0^{\infty} r_1 \mu(y) \Delta t P_1(i, y, t) dy + \sum_{v=1}^{i+1} \lambda q_v \Delta t P_0(i-v+1, t) + \\ + (i+1) \sigma \Delta t P_0(i+1, t) + o(\Delta t).$$

После преобразований получим систему уравнений Колмогорова

$$\frac{\partial P_0(i, t)}{\partial t} = -(\lambda + i \sigma) P_0(i, t) + \int_0^{\infty} r_0 \mu(y) P_1(i, y, t) dy + \int_0^{\infty} r_2 \mu(y) P_1(i-1, y, t) dy; \\ \frac{\partial P_1(i, y, t)}{\partial t} + \frac{\partial P_1(i, y, t)}{\partial y} = -(\lambda + \mu(y)) P_1(i, y, t) + \sum_{v=1}^i \lambda q_v P_1(i-v, y, t)$$

с краевым условием

$$P_1(i, 0, t) = (i+1) \sigma P_0(i+1, t) + \int_0^{\infty} r_1 \mu(y) P_1(i, y, t) dy + \sum_{v=1}^{i+1} \lambda q_v P_0(i-v+1, t).$$

Перепишем систему с краевым условием для стационарного режима

$$-(\lambda + i \sigma) P_0(i) + \int_0^{\infty} \mu(y) [r_0 P_1(i, y) + r_2 P_1(i-1, y)] dy = 0; \quad (1) \\ \frac{\partial P_1(i, y)}{\partial y} + [\lambda + \mu(y)] P_1(i, y) - \sum_{v=1}^i \lambda q_v P_1(i-v, y) = 0;$$

$$P_1(i, 0) = (i+1) \sigma P_0(i+1) + \int_0^{\infty} r_1 \mu(y) P_1(i, y) dy + \sum_{v=1}^{i+1} \lambda q_v P_0(i-v+1). \quad (2)$$

Введем частичные характеристические функции числа заявок на орбите $H_0(u) = \sum_{i=0}^{\infty} e^{ju i} P_0(i)$, $H_1(u, y) = \sum_{i=0}^{\infty} e^{ju i} P_1(i, y)$ и характери-

ческую функцию числа заявок в группе $Q(u) = \sum_{v=1}^{\infty} e^{juv} q_v$, где $j = \sqrt{-1}$. Преобразуя (1) и (2), получим систему уравнений для функций $H_0(u)$ и $H_1(u, y)$

$$-\lambda H_0(u) + j\sigma \frac{\partial H_0(u)}{\partial u} + (r_0 + r_2 e^{ju}) \int_0^{\infty} \mu(y) H_1(u, y) dy = 0; \quad (3)$$

$$\frac{\partial H_1(u, y)}{\partial y} - [\lambda Q(u) - \lambda - \mu(y)] H_1(u, y) = 0$$

с краевым условием

$$H_1(u, 0) = -j\sigma e^{-ju} \frac{\partial H_0(u)}{\partial u} + \int_0^{\infty} r_1 \mu(y) H_1(u, y) dy + \lambda e^{-ju} Q(u) H_0(u). \quad (4)$$

Полная характеристическая функция числа заявок на орбите имеет вид $H(u) = H_0(u) + H_1(u)$, где $H_1(u) = \int_0^{\infty} H_1(u, y) dy$.

Систему (3) с условием (4) можно решить аналитически. Для этого найдем $H_1(u, y)$ из второго уравнения системы

$$H_1(u, y) = H_1(u, 0) \cdot (1 - B(y)) e^{\lambda(Q(u)-1)y} \quad (5)$$

и перепишем (4) в виде

$$h(u) = \lambda e^{-ju} Q(u) H_0(u) - j\sigma e^{-ju} H_0'(u) + r_1 h(u) \int_0^{\infty} e^{\lambda(Q(u)-1)y} B'(y) dy, \quad (6)$$

где $h(u) = H_1(u, 0)$.

Обозначим $B^*(u) = \int_0^{\infty} e^{\lambda(Q(u)-1)y} dB(y)$ – преобразование Лапласа – Стильтеса распределения $B(x)$, тогда из (6)

$$h(u) = \lambda \frac{e^{-ju}}{1 - r_1 B^*(u)} Q(u) H_0(u) - j\sigma \frac{e^{-ju}}{1 - r_1 B^*(u)} H_0'(u). \quad (7)$$

Подставим (5) в первое уравнение (3), получим уравнение

$$-\lambda H_0(u) + j\sigma H_0'(u) + (r_0 + r_2 e^{ju}) h(u) B^*(u) = 0,$$

которое с учетом (7) перепишем в виде

$$j\sigma \left[1 - \frac{(r_0 e^{-ju} + r_2) B^*(u)}{1 - r_1 B^*(u)} \right] H_0'(u) = \lambda \left[1 - \frac{(r_0 e^{-ju} + r_2) B^*(u)}{1 - r_1 B^*(u)} \right] Q(u) H_0(u).$$

Решением этого уравнения является функция

$$H_0(u) = R \cdot \exp \left\{ -\frac{\lambda j}{\sigma} \cdot \int_0^u \frac{1-F(x)Q(x)}{1-F(x)} dx \right\},$$

где $F(x) = (r_0 e^{-jx} + r_2) B^*(x) / (1 - r_1 B^*(x))$, а константу интегрирования R найдем из условия нормировки $H(0) = 1$.

Окончательно, получим

$$H(u) = R \cdot \exp \left\{ -\frac{\lambda j}{\sigma} \cdot \int_0^u \frac{1-F(x)Q(x)}{1-F(x)} dx \right\} \times \\ \times \left[1 - \frac{\lambda e^{-ju}}{1-r_1 B^*(u)} \cdot \frac{1-Q(u)}{1-F(u)} \int_0^\infty (1-B(y)) e^{\lambda(Q(u)-1)y} dy \right].$$

Для нахождения распределения вероятностей $P(i)$ числа заявок i на орбите применим обратное преобразование Фурье

$$P(i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ju i} H(u) du.$$

Заключение

Исследована система массового обслуживания с обратной связью, рекуррентным обслуживанием и неординарным входящим потоком. Для марковизации случайного процесса числа заявок на орбите введена дополнительная переменная – истекшее время обслуживания. Решение уравнений Колмогорова позволило получить аналитическое выражение для характеристической функции, а также стационарное распределение вероятностей числа заявок на орбите.

ЛИТЕРАТУРА

1. Кудрявцева Е.Н., Росляков А.В. Применение теории сетевого исчисления к исследованию систем массового обслуживания с обратной связью // Т-Comm: Телекоммуникации и транспорт. 2015. № 1. С. 17–21.
2. Назаров А.А., Рожкова С.В., Титаренко Е.Ю. Исследование М/М/1 системы с повторными вызовами, обратной связью и неординарным пуассоновским входящим потоком // Математика, ее приложения и математическое образование (МПМО'20): материалы VII Междунар. конф. Улан-Удэ: Изд-во ВСГУТУ, 2020. С. 160–162.
3. Моисеева Е.А. Исследование RQ-системы M|GI|1 в допредельной ситуации // Математическое и программное обеспечение информационных, технических и экономических систем: материалы II Всероссийской молодежной научной конф. Сер. «Труды ТГУ. Серия физико-математическая». 2013. С. 116–121.

Ресурсная RQ-система с коллизиями

Анна Полховская, Ольга Бобкова, Светлана Моисеева

*Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

В период высоких технологий телекоммуникационная отрасль переживает существенные изменения: модернизация сетей связи, стремительное увеличение числа пользователей, а также увеличение ассортимента услуг, предоставляемых потребителям. Важнейшей задачей операторов сетей связи становится улучшение качества создаваемого ими продукта, сетевые операторы нуждаются в эффективных инструментах оценки производительности, которые учитывают важнейшие особенности современных сотовых сетей.

Последнее десятилетие в области сетей и систем связи ознаменовалось экспоненциальным ростом как числа абонентских устройств, так и общего объема трафика, передаваемого на беспроводном участке доступа в сетях связи общего пользования.

Значительную часть новых устройств составляют носимые устройства, которые в целом генерируют большое количество информации, создавая исключительно высокую нагрузку на абонентском участке доступа, вызывая перегрузки на отдельных участках сети, что ведет к ухудшению качества связи. СМО с заявками случайного объема позволяют решать задачи проектирования информационных систем, объектом преобразования в которых является информация, поступающая порциями в виде дискретных или непрерывных сообщений. Сообщения или заявки обладают различным информационным объемом, который представляет собой случайную величину.

Задача исследования систем массового обслуживания (СМО), в которых каждая поступающая в систему заявка наряду со случайной длиной имеет случайный объем, играет важную роль при моделировании работы самых разнообразных технических устройств, в частности современных информационно-вычислительных и телекоммуникационных систем.

Данная работа направлена на выработку подхода к исследованию ресурсных RQ-систем с целью определения требуемого количества ресурса на орбите при известных интенсивностях запросов на установление соединения и обслуживания.

В [1–3] представлены результаты исследований ресурсных СМО с неограниченным числом обслуживающих приборов, а в [4, 5] изучены СМО с коллизиями.

Математическая модель и постановка задачи

Рассмотрим RQ-систему, на вход которой поступает пуассоновский с параметром λ поток заявок. Предполагаем, что каждое требование характеризуется некоторым объемом $v_k > 0$. Объемы различных требований являются независимыми случайными величинами (дискретными или непрерывными) с функцией распределения $G(y) = P\{v < y\}$.

Заявка случайного объема, заставшая прибор свободным, занимает его и начинает обслуживание, которое заканчивается успешно, если во время ее обслуживания другие заявки не поступали. Продолжительность обслуживания не зависит от объема заявки и имеет экспоненциальную функцию распределения с параметром μ .

Если прибор занят, то поступившая и обслуживаемая заявки вступают в конфликт. Они считаются искаженными и обе вместе со своими объемами переходят на орбиту, в которой находятся случайное время, после чего вновь обращаются к прибору с повторной попыткой его захвата. Случайная задержка, которую осуществляет заявка на орбите в случае возникновения конфликта, экспоненциально распределена с параметром σ .

Если прибор свободен, то заявка из ИПВ занимает его для обслуживания. Если прибор занят, то вновь возникает конфликт, и обе заявки вновь переходят в ИПВ.

Пусть $i(t)$ – число заявок на орбите в момент времени t .

Обозначим $V(t) = \sum_{k=0}^{i(t)} v_k$ – суммарный объем заявок, находящихся на орбите в момент времени t .

$l(t)$ будет определять состояние прибора следующим образом:

$$l(t) = \begin{cases} 0, & \text{если прибор свободен,} \\ 1, & \text{если прибор занят.} \end{cases}$$

Задача заключается в определении основных характеристик суммарного объема заявок на орбите.

Метод производящих функций

Введем характеристическую функцию для суммарного объема находящихся в системе требований

$$\begin{aligned}
 H(u) &= M \left\{ e^{juV(t)} \right\} = M \left\{ \exp \left\{ ju \sum_{k=0}^{i(t)} v_k \right\} \right\} = \\
 &= \sum_{i=0}^{\infty} M \left\{ \exp \left\{ ju \sum_{k=0}^i v_k \right\} \mid i(t) = i \right\} P \{ i(t) = i \} = \\
 &= \sum_{i=0}^{\infty} M \left\{ \exp \{ ju (v_1 + v_2 + \dots + v_i) \} \right\} P(i) = \\
 &= \sum_{i=0}^{\infty} M \left\{ e^{juv_1} \cdot e^{juv_2} \cdot \dots \cdot e^{juv_i} \right\} P \{ i(t) = i \} = \sum_{i=0}^{\infty} \left(M \left\{ e^{juv_1} \right\} \right)^i P \{ i(t) = i \}.
 \end{aligned}$$

Обозначим
$$\varphi(u) = M \left\{ e^{juv} \right\} = \int_0^{\infty} e^{juv} dG(y),$$

тогда
$$H(u) = \sum_{i=0}^{\infty} (\varphi(u))^i P(i) = F(\varphi(u)).$$

Таким образом, характеристическая функция суммарного объема требований имеет вид производящей функции числа занятых приборов с аргументом $\varphi(u)$.

В силу того, что входящий простейший поток и экспоненциальное время обслуживания обладают свойством отсутствия последействия, то случайный процесс $\{l(t), i(t)\}$ изменения во времени состояний описанной RQ-системы является двумерной цепью Маркова с непрерывным временем. Поэтому рассматриваемая система массового обслуживания с конфликтами заявок будет являться марковской.

Нетрудно показать, что стационарное распределение вероятностей $P(l, i)$ исследуемой RQ-системы удовлетворяет системе уравнений

$$\begin{cases}
 -(\lambda + i\sigma)P(0, i) + \mu P(1, i) + \lambda P(1, i - 2) + (i - 1)\sigma P(1, i - 1) = 0, \\
 -(\lambda + \mu + i\sigma)P(1, i) + \lambda P(0, i) + (i + 1)\sigma P(0, i + 1) = 0.
 \end{cases}$$

Теорема 1. Частичные производящие функции

$$F_k(z) = \sum_i z^i P(l, i), \quad k = \{0, 1\},$$

числа заявок в ИПВ определяются формулами

$$F_0(u) = \frac{\mu - \lambda z}{\mu} \cdot \left(\frac{\mu - 2\lambda z}{\mu - 2\lambda} \right)^{\frac{2(\lambda + \sigma) + \mu}{4\sigma}} \exp \left\{ -\frac{\lambda(z-1)}{2\sigma} \right\},$$

$$F_1(u) = \frac{\lambda}{\mu} \cdot \left(\frac{\mu - 2\lambda z}{\mu - 2\lambda} \right)^{\frac{2(\lambda + \sigma) + \mu}{4\sigma}} \exp \left\{ -\frac{\lambda(z-1)}{2\sigma} \right\}.$$

Следовательно, характеристическая функция суммарного объема будет иметь вид

$$H(u) = F(\varphi(u)) = F_0(\varphi(u)) + F_1(\varphi(u)) =$$

$$= \frac{\mu - \lambda(\varphi(u) - 1)}{\mu} \cdot \left(\frac{2\lambda\varphi(u) - \mu}{2\lambda - \mu} \right)^{\frac{2(\lambda + \sigma) + \mu}{4\sigma}} \exp \left\{ -\frac{\lambda(\varphi(u) - 1)}{2\sigma} \right\}.$$

Заключение

Рассмотрена ресурсная марковская RQ-система с коллизиями. Однако, выполнение аналогичного допредельного исследования более сложных систем, выходящих за рамки классических моделей, не представляется возможным. Как правило, в таких случаях применяют метод асимптотического анализа при различных предельных условиях.

ЛИТЕРАТУРА

1. *Bushkova T., Pavlova E., Rozhkova S., Moiseeva S., Pagano M.* Resource queueing system MMPP(2,v)|GI2| ∞ with parallel service of multiple paired customers // CCIS. 2019. V. 1109. P. 136–149.
2. *Lisovskaya E., Moiseeva S., Pagano M., Potatueva V.* Study of the MMPP/GI/ ∞ queueing system with random customers' capacities // Информ. и её примен. 2017. V. 11. No. 4. P. 109–117.
3. *Galileyskaya A.A., Lisovskaya E.Yu., Moiseeva S.P., Daniilyuk E.Yu.* Asymptotic stationary probability distribution of total amount of physics experimental data // Russ. Phys. J. 2020. V. 62. No. 10. P. 1779–1788.
4. *Назаров А.А., Судыко Е.А.* Исследование марковской RQ-системы с конфликтами заявок и простейшим входящим потоком // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 3(12). С. 97–106.
5. *Назаров А.А., Судыко Е.А.* Условия существования стационарного режима в немарковских RQ-системах с конфликтами заявок // Известия Томского политехнического университета. Управление, вычислительная техника и информатика. 2011. Т. 318. № 5. С. 166–168.

Прямая коррекция ошибок на внутрисегментном уровне транспортного протокола

Павел Приступа, Павел Михеев, Сергей Сущенко

*Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

Важнейшим показателем качества взаимодействия сетевых приложений является пропускная способность транспортных соединений. Данная операционная характеристика в значительной мере определяется транспортным протоколом и его параметрами – шириной окна и длительностью тайм-аута [1]. Моделирование абонентского соединения и анализ его потенциальных возможностей выполнялся в [2–11] и др. работах. Математическое моделирование классического транспортного протокола с решающей обратной связью выполнено в [2–8], имитационное исследование быстродействия протокола проведено в [9, 10]. Но результаты получены при существенных ограничениях на параметры протокола и факторы, определяющие быстродействие транспортного соединения. В современных транспортных протоколах для снижения объема повторно передаваемого трафика получают широкое распространение технологии прямой коррекции ошибок [12–14]. Примером реализации такой технологии является протокол QUIC кампании Google [12]. Математическая модель транспортного протокола с технологией прямой коррекции ошибок в межсегментном пространстве рассмотрена в [14]. В данной работе рассмотрена математическая модель транспортного соединения в фазе информационного переноса с механизмом прямой коррекцией ошибок в виде цепи Маркова с дискретным временем. На основе модели предложен анализ условий, обеспечивающих преимущество в быстродействии транспортного протокола с технологией прямой коррекции ошибок в пространстве каждого сегмента перед классическим транспортным протоколом.

Рассмотрим процесс переноса данных между абонентами транспортного протокола, основанного на алгоритме с решающей обратной связью [1] и функционирующего в режиме селективного или группового отказа. Полагаем, что подтверждения получателя о корректности приема данных переносятся в сегментах встречного потока. Считаем,

что управляющий транспортный протокол имеет дополнительную логику внутрисегментной прямой коррекции ошибок. При этом источник перед отправкой каждого сегмента делит его на $A \geq 1$ фрагментов равного размера, к ним добавляет $B - A$, $B \geq A$ избыточных фрагментов той же длины и передает в транспортное соединение расширенный сегмент. Каждый фрагмент снабжается служебной информацией, позволяющей обнаружить в нем возможные ошибки и восстановить из B фрагментов исходный сегмент в точке приема. Искажение до $B - A$ произвольных фрагментов расширенного сегмента позволяет на стороне получателя восстановить исходный сегмент и не инициировать повторные передачи. Пусть f_f и f_r – достоверность передачи фрагмента вдоль транспортного соединения от источника до адресата и обратно соответственно. Тогда вероятность доставки исходного сегмента адресату с учетом возможности его восстановления получателем с помощью механизма прямой коррекции ошибок задается соотношением

$$\Psi_f = \sum_{i=A}^B C_i^B f_f^i (1 - f_f)^{B-i}. \text{ Данные параметрические зависимости от } A \text{ и}$$

B определяют математическую модель различных методов прямой коррекции ошибок. Считаем, что участки переприема вдоль тракта передачи данных имеют одинаковое быстроедействие, а длительность цикла передачи расширенного сегмента в отдельном звене составляет t . Управление потоком данных реализуется механизмом скользящего окна [1] размера $W \geq 1$. Процесс информационного переноса расширенных сегментов в транспортном соединении может быть описан Марковским процессом с дискретным временем, кратным длительности такта t , в силу того, что время между получениями подтверждений имеет геометрическое распределение [2]. Цепь Маркова задает активность отправителя и изменение размера очереди расширенных сегментов в источнике, ожидающих подтверждения. Область возможных состояний цепи Маркова определяется длительностью тайм-аута ожидания подтверждения S , выраженной в количестве циклов продолжительности t . Сумма длин прямого и обратного трактов может быть интерпретирована как круговая задержка D , выраженная в длительностях t . Состояниям цепи Маркова $i = \overline{0, W}$ соответствует размер очереди переданных, но не подтвержденных сегментов в источнике потока, а состояниям

$i = \overline{W+1, S-1}$ – время, в течение которого отправитель не активен и ожидает получения подтверждения о корректности приема переданной последовательности из W сегментов. Из нулевого состояния в $D-1$ -е источник продвигается с каждым тактом t с вероятностью детерминированного события. В состояниях $i \geq D-1$ после истечения очередного дискретного цикла t к отправителю начинают прибывать подтверждения и, в зависимости от результатов доставки, отправитель передает новые сегменты (при положительном подтверждении), либо повторно – искаженные. Завершение цикла пребывания в состоянии $D-1$ соответствует времени доведения первого сегмента до адресата и получения на него подтверждения. Дальнейший рост номера состояния происходит с вероятностью искажения подтверждения $1 - \Psi_r$ в обратном тракте. В состояниях $i \geq D-1$ в режиме селективного отказа получение подтверждения порождает переход в $D-1$ -е состояние при $W \geq D$ или в состояние $D+W-2-i$ при $W \leq D$. В режиме группового отказа для исходных состояний $i \geq D-1$ возврат в состояния $D-1$ (при $W \geq D$) или $D+W-2-i$ (при $W \leq D$) происходит при получении подтверждения только в случае успешной доставки адресату дошедших к данному моменту до получателя $i - D + 1$ расширенных сегментов, в противном случае следует возврат в нулевое состояние, поскольку очередь переданных, но не подтвержденных сегментов в этот момент обнуляется. В силу того, что в состояниях $i \geq W$ источник приостанавливает отправку сегментов, получение подтверждений в состояниях $i = \overline{W, D+W-3}$ приводит к переходу в состояния $D+W-2-i$, а из состояний $i = \overline{D+W-2, S-2}$ – в нулевое. Это справедливо для селективного отказа, а в режиме группового отказа указанные изменения состояний реализуются при поступлении положительных подтверждений. В состоянии $S-1$ истекает тайм-аут ожидания подтверждения от получателя о корректности принятых сегментов и происходит безусловный переход в нулевое состояние во всех режимах отказа.

Важнейшей операционной характеристикой протокола является его пропускная способность, определяемая параметрами тракта передачи данных, накладными расходами и особенностями протокольных процедур управления передачей [1, 2]. Нормированное быстродействие транспортного соединения определяется средним числом доставленных

получателю неискаженных сегментов за среднее время между двумя последовательными поступлениями подтверждений. Поскольку время между приходами подтверждений распределено по геометрическому закону, то пропускная способность определится зависимостями [2]

$$Z_s(W, S, D, A, B) = \Psi_r \left\{ \sum_{i=D-1}^{D+W-2} (i-D+2) \Psi_f P_i + W \Psi_f \sum_{i=D+W-1}^{S-1} P_i \right\}$$

и
$$Z_g(W, S, D, A, B) = \Psi_f \Psi_r \left\{ \sum_{i=D-1}^{D+W-2} \frac{1 - \Psi_f^{i-D+2}}{1 - \Psi_f} P_i + \frac{1 - \Psi_f^W}{1 - \Psi_f} \sum_{i=D+W-1}^{S-1} P_i \right\}$$

для селективной и групповой процедур отказа соответственно.

Эффективное применение технологии прямой коррекции ошибок предполагает поиск коэффициента фрагментации сегмента A и числа дополнительных избыточных фрагментов $B-A$ для коррекции ошибок, обеспечивающих максимальное быстродействие транспортного соединения с заданными характеристиками и протокольными параметрами. Наличие избыточных фрагментов в передаваемой последовательности увеличивает вероятность доставки получателю сегмента, однако это достигается за счет роста накладных расходов в виде времени переноса дополнительных данных. В связи с этим возникает задача поиска в многомерном признаковом пространстве области значений характеристик транспортного соединения, параметров транспортного протокола и механизма прямой коррекции ошибок, обеспечивающей превосходство управляющей процедуры с прямой коррекцией ошибок перед классической протокольной процедурой с решающей обратной связью без использования коррекции ошибок. Определим выигрыш в быстродействии от применения механизма ПКО по сравнению с классической протокольной процедурой с решающей обратной связью в виде

$$\Delta(A, B) = Z(W, S, D, A, B) - Z(W, S, D, 1, 1).$$

В общем случае сравнительный анализ удастся провести только численно. В ряде случаев за счет снижения размерности признакового пространства область положительных значений выигрыша удастся найти в простом аналитическом виде. В режиме селективного отказа, полагая $W \rightarrow \infty$ либо абсолютно надежным обратный тракт передачи данных ($f_r = 1$), выигрыш приобретает относительно простой аналитический

вид. Для параметров прямой коррекции ошибок $B = A + 1, A \geq 1$ область положительных значений выигрыша существует для $A \geq 2$ на интервале $f_f \in (0, 1 - 1/A^2)$. Максимальное значение выигрыша достигается для $f_f = 1 - 1/A$. С ростом параметра прямой коррекции ошибок A положение максимума выигрыша дрейфует вправо по координате f_f , а максимальный выигрыш растет. Для технологических параметров механизма прямой коррекции ошибок $B = A + 2, A \geq 1$ положительный выигрыш достигается на области значений достоверности доставки фрагментов $f_f \in (0, (A - 1)(A + 2)/A(A + 1))$, $A \geq 2$, а максимум – в точке $f_f = 1 - \sqrt{2/A(A + 1)}$. В режиме группового повтора для абсолютно надежного обратного тракта передачи данных ($f_r = 1$) положительные значения выигрыша при $A = 1, B = 2$ достигаются для достоверности передачи фрагмента в прямом тракте транспортного соединения на интервалах $f_f \in (0, D - 2/D - 1), W \geq D, D > 2$ и $f_f \in (0, 2 - D/W - 1), 1 + D/2 < W < D$. Для нагруженного транспортного соединения ($W \geq D$) с ростом длительности круговой задержки D область положительных значений выигрыша расширяется, максимум растет по кривой с насыщением и его положение сдвигается в область большей достоверности доставки фрагментов f_f . Зависимость выигрыша от параметра A при $B = A + 1, A \geq 1$ имеет экстремальный характер. Для недогруженного транспортного соединения ($1 + D/2 < W < D$) выигрыш с расширением окна от значения $W = 1 + D/2$ до $W = D - 1$ область положительных значений выигрыша по координате f_f расширяется, а максимум выигрыша растет и дрейфует вправо. Такой же эффект имеет место с увеличением параметра прямой коррекции ошибок A для $B = A + 1$. При выполнении условия $W = 1 + D/2$ рост длительности круговой задержки D приводит к сужению по координате f_f области положительных значений выигрыша и снижению преимуществ технологии прямой коррекции ошибок.

ЛИТЕРАТУРА

1. *Fall K., Stevens R.* TCP/IP Illustrated. Volume 1: The Protocols (2nd Edition). Addison-Wesley Professional Computing Series, 2012. 1017 p.
2. *Kokshenev V.V., Mikheev P.A., Sushchhenko S.P.* Comparative analysis of the performance of selective and group repeat transmission models in a transport protocol // *Automation and Remote Control.* 2017. V. 78. No 2. P. 247–261.
3. *Kokshenev V., Mikheev P., Suschenko S., Tkachyov R.* Analysis of the throughput in selective mode of transport protocol // *CCIS.* 2016. V. 678. P. 168–181.
4. *Bogushevsky D., Mikheev P., Pristupa P., Suschenko S.* The time-out length influence on the available bandwidth of the selective failure mode of transport protocol in the load data transmission path // *CCIS.* 2018. V. 919. P. 120–131.
5. *Kassa D.F.* Analytic Models of TCP Performance: PhD Thesis. University of Stellenbosch, 2005. 199 p.
6. *Giordano S., Pagano M., Russo F., Secchi R.* Modeling TCP startup performance // *Journal of Mathematical Sciences.* 2014. V. 200. Iss. 4. P. 424–431.
7. *Kravets O.Ya.* Mathematical modeling of parameterized TCP protocol // *Automation and Remote Control.* 2013. V. 74. No. 7. P. 1218–1224.
8. *Arvidsson A., Krzesinski A.* A model of a TCP link // *Proceedings of the 15th International Teletraffic Congress Specialist Seminar.* 2002.
9. *Olsen Y.* Stochastic modeling and simulation of the TCP protocol // *Uppsala Dissertations in mathematics* 28. 2003. 94 p.
10. *Mikheev P., Pristupa P., Suschenko S.* Performance of transport connection with selective failure mode when competing for throughput of data transmission path // *CCIS.* 2019. V. 1141. P. 89–103.
11. *Nikitinskiy M.A., Chalyy D.Ju.* Performance analysis of trickles and TCP transport protocols under high-load network conditions // *Automatic Control and Computer Sciences.* 2013. V. 47. No. 7. P. 359–365.
12. *Shalin R., Kesavaraja D.* Multimedia data transmission through TCP/IP using hash based FEC with AUTO-XOR scheme // *ICTACT Journal on Communication Technology.* 2012. V. 03. Iss. 03. P. 604–609.
13. *Herrero R.* Modeling and comparative analysis of forward error correction in the context of multipath redundancy // *Telecommunication Systems. Modelling, Analysis, Design and Management.* 2017. V. 65(4). P. 783–794.
14. *Mikheev P., Suschenko S., Tkachev R.* Estimation of high-speed performance of the transport protocol with the mechanism of forward error correction // *Communications in Computer and Information Science.* 2017. V. 700. P. 259–268.

Математическое моделирование и исследование приоритетной управляющей системы с непостоянной интенсивностью обслуживания требований*

Екатерина Пройдакова, Виктория Санникова

ННГУ им. Н.И. Лобачевского, г. Нижний Новгород, Россия

Рассматривается система обслуживания и управления m независимыми и конфликтными потоками $\Pi_1, \Pi_2, \dots, \Pi_m$ требований. Конфликтность потоков означает, что их обслуживание должно происходить в непересекающиеся промежутки времени. Входные потоки делятся на три типа: Π_1 – приоритетный поток малой интенсивности, Π_2, \dots, Π_{m-1} – потоки средней интенсивности и без преимуществ по обслуживанию, Π_m – интенсивный поток без преимуществ по обслуживанию. Приоритет потока Π_1 заключается в том, что если поступила хотя бы одна заявка по первому направлению, то она должна быть обслужена как можно быстрее, но не прерывая уже проводящееся обслуживание других требований. Входные потоки считаем простейшими (пуассоновскими) с интенсивностями $\lambda_1, \lambda_2, \dots, \lambda_m$ соответственно, причем $\lambda_1 \ll \lambda_m$. У каждого из m потоков есть основной этап обслуживания и этап переналадки. Во время основного этапа поток Π_j , $j = \overline{2, m}$, обслуживается с *непостоянной интенсивностью*. Введено одно дополнительное состояние для обслуживания наиболее интенсивного потока Π_m . Таким образом, обслуживающее устройство имеет $(2m+1)$ состояния $\Gamma^{(1)}, \Gamma^{(2)}, \dots, \Gamma^{(2m+1)}$. В состоянии $\Gamma^{(1)}$ обслуживается поток Π_1 с постоянной интенсивностью μ_1 . В состоянии $\Gamma^{(2j)}$ не обслуживается ни один из потоков. В состоянии $\Gamma^{(2j-1)}$ обслуживается поток Π_j с непостоянной интенсивностью $\mu_j(t)$. В состоянии $\Gamma^{(2j)}$ не обслуживается ни один из потоков. В состоянии $\Gamma^{(2m+1)}$ обслуживается только поток Π_m , но уже с постоянной интенсивностью μ''_j . Заявки, пришедшие в систему, могут образовывать неограниченные очереди. Длительности состояний $\Gamma^{(1)}, \Gamma^{(2)}, \dots, \Gamma^{(2m+1)}$ известны и равны $T_1, T_2, \dots, T_{2m+1}$ единиц времени соответственно.

* Работа выполнена при финансовой поддержке РФФИ (проект № 18 – 413-520005).

Построение математической модели

При построении математической модели исследуемой системы обслуживания применялся кибернетический подход [1]. Система наблюдалась в дискретные моменты времени $\tau_i, i = 1, 2, \dots$, переключений состояний обслуживающего устройства или на промежутках $[\tau_i, \tau_{i+1})$. При построении математической модели непостоянная интенсивность обслуживания $\mu_j(t)$ аппроксимировалась с помощью кусочно-постоянной функции с одной точкой разрыва, что не повлияло на методику исследования и общность результатов. Каждое состояние $\Gamma^{(2j-1)}$ обслуживающего устройства разбивалось на два виртуальных состояния $\Gamma^{(3j-3)}$ и $\Gamma^{(3j-2)}$, длительностями T_{3j-2} и T_{3j-1} соответственно, $T_{3j-2} + T_{3j-1} = T_{2j-1}$. Таким образом, исходное множество состояний обслуживающего устройства преобразовалось в $\Gamma = \{\Gamma^{(1)}, \Gamma^{(2)}, \dots, \Gamma^{(3j-3)}, \Gamma^{(3j-2)}, \dots, \Gamma^{(3m)}\}$ с известными длительностями $T_1, T_2, \dots, T_{3j-3}, T_{3j-2}, \dots, T_{3m}$ соответственно.

Определим при $j = \overline{1, m}$ и $i = 0, 1, \dots$ следующие случайные объекты:

1) $\eta_{j,i}$ – число заявок потока Π_j , пришедших за время $[\tau_i, \tau_{i+1})$, $\eta_{j,i} \in X = \{0, 1, 2, \dots\}$;

2) $\xi_{j,i}$ – максимальное число заявок, которое может быть обслужено за $[\tau_i, \tau_{i+1})$ из очереди потока Π_j , $\xi_{j,i} \in \{0, l_{3j}\}$ при $j = \overline{1, m}$; здесь $l_{3j} = \max\{[\mu_j^1 T_{3j-2}], [\mu_j^2 T_{3j-1}]\}$; $\xi_{m,i} \in \{0, l'_{3m}, l_{3m}\}$ при $j = m$, l'_{3m} – максимальное число требований потока Π_{3m} , обслуженное за время работы сигнала $\Gamma^{(3m)}$;

3) Γ_i – состояние устройства во время $[\tau_i, \tau_{i+1})$, $\Gamma_i \in \Gamma$;

4) $\varkappa_{j,i}$ – длина очереди по потоку Π_j в момент времени τ_i , $\varkappa_{j,i} \in X$;

5) $\bar{\xi}_{j,i}$ – число реально обслуженных требований потока Π_j за $[\tau_i, \tau_{i+1})$, причем $\bar{\xi}_{j,i} \in Y_j = \{0, 1, \dots, l_{3j}\}$;

6) $\bar{\xi}_{j,-1}$ – число реально обслуженных заявок потока Π_j за $[0, \tau_0)$, $\bar{\xi}_{j,-1} \in Y_j$;

Переход обслуживающего устройства из состояния Γ_i в состояние Γ_{i+1} осуществлялся согласно следующему соотношению:

$$\Gamma_{i+1} = U(\Gamma_i, \varkappa_{1,i}),$$

$$\text{где } U(\Gamma^{(r)}, w_1) = \begin{cases} \Gamma^{(1)}, & r = 3m - 1, \\ \Gamma^{(r+1)}, & r = \overline{1, 3m - 3}, \\ \Gamma^{(3m-1)}, & r \in \{3m - 2, 3m\}, w_1 > 0, \\ \Gamma^{(3m)}, & r \in \{3m - 2, 3m\}, w_1 = 0. \end{cases}$$

Заявки обслуживаются в соответствии с экстремальной стратегией обслуживания, то есть выполняется

$$\bar{\xi}_{j,i} = \min \{ \varkappa_{j,i} + \eta_{j,i}; \xi_{j,i} \}.$$

Для очереди справедливо соотношение

$$\varkappa_{j,i+1} = \max \{ 0; \varkappa_{j,i} + \eta_{j,i} - \bar{\xi}_{j,i} \}.$$

Состояние всей системы по потоку Π_j на промежутке времени $[\tau_i, \tau_{i+1})$ характеризуется случайным вектором

$$(\Gamma_i, \varkappa_{1,i}, \varkappa_{j,i}, \bar{\xi}_{1,i-1}, \bar{\xi}_{j,i-1}).$$

Пятимерная случайная векторная последовательность

$$\{(\Gamma_i, \varkappa_{1,i}, \varkappa_{j,i}, \bar{\xi}_{1,i-1}, \bar{\xi}_{j,i-1}); i \geq 0\}$$

определяет поведение изучаемой управляющей системы по потокам $\Pi_1, \Pi_j, j = \overline{2, m}$.

Считаем, что в момент времени τ_0 известно распределение начального вектора, то есть известны следующие вероятности:

$$P(\Gamma_0 = \Gamma^{(s)}, \varkappa_{1,0} = x_1, \varkappa_{j,0} = x_j, \bar{\xi}_{1,-1} = y_1, \bar{\xi}_{j,-1} = y_j).$$

Для полученной пятимерной случайной последовательности были доказаны некоторые утверждения.

Лемма 1. При заданном распределении начального вектора

$$(\Gamma_0, \varkappa_{1,0}, \varkappa_{j,0}, \bar{\xi}_{1,-1}, \bar{\xi}_{j,-1})$$

случайная векторная последовательность

$$\{(\Gamma_i, \varkappa_{1,i}, \varkappa_{j,i}, \bar{\xi}_{1,i-1}, \bar{\xi}_{j,i-1}); i \geq 0\}$$

является марковской.

Кроме того, для марковской последовательности

$$\{(\Gamma_i, \varkappa_{1,i}, \varkappa_{j,i}, \bar{\xi}_{1,i-1}, \bar{\xi}_{j,i-1}); i \geq 0\}$$

были выделены несущественные состояния [2], получены рекуррентные соотношения для одномерных распределений. Это фактически позволяет изучать любые конечномерные распределения указанной последовательности.

Имитационное моделирование и численные эксперименты

Чтобы исследовать влияние непостоянной интенсивности обслуживания на характеристики функционирования рассматриваемой управляющей системы, была создана соответствующая программа, являющаяся ее имитационной моделью. При моделировании принимались во внимание найденные аналитически условия существования стационарного режима функционирования [3] для системы с постоянной интенсивностью обслуживания:

$$\lambda_j \Gamma - l_j - l'_j < 0, \quad j = \overline{1, m}, \quad \Gamma = \sum_{r=1}^{2m} \Gamma_r .$$

В начале работы имитационной модели задавались следующие входные параметры:

- 1) λ_j – интенсивность поступления заявок по потоку Π_j , $j = \overline{1, m}$;
- 2) T_{2j-1} – время работы обслуживающих устройств, $j = \overline{1, m}$;
- 3) $\mu_j(t)$ – функция непостоянной интенсивности обслуживания требований потока Π_j , $j = \overline{2, m}$;
- 4) T_{2j} – время переналадки обслуживающего устройства для потока Π_j , $j = \overline{1, m}$;
- 5) t' – длительность состояния $\Gamma^{(2m+1)}$ (петли);
- 6) k – число циклов работы системы;
- 7) t – время обслуживания одной заявки.

Моделирование включало в себя два этапа. На первом этапе определялся момент перехода системы в квазистационарный (близкий к стационарному) режим функционирования. На втором этапе моделировалась работа системы в квазистационарном режиме для нахождения численных оценок следующих характеристик системы:

- 1) \bar{x}_j , $j = \overline{1, m}$ – оценки средних очередей по каждому из потоков;
- 2) \tilde{x} – оценка средневзвешенной очереди, которая рассчитывается по формуле

$$\tilde{x} = \frac{\lambda_1 \bar{x}_1 + \lambda_2 \bar{x}_2 + \lambda_3 \bar{x}_3}{\lambda_1 + \lambda_2 + \lambda_3} .$$

Численные результаты

В качестве примера рассмотрим случай, когда в системе обслуживаются три потока, и вид функции интенсивности обслуживания $\mu_j(t)$, $j = 2, 3$ имеет вид, представленный на рис. 1.

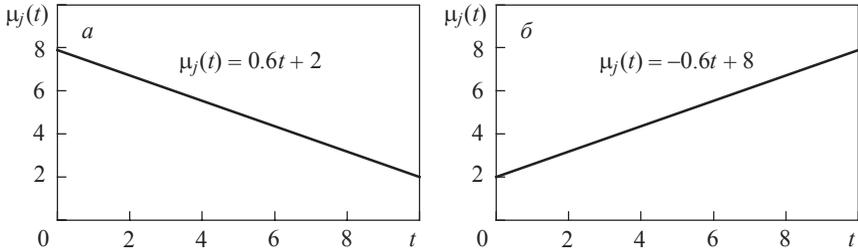


Рис. 1. Вид функции интенсивности обслуживания $\mu_j(t)$, $j = 2, 3$

А именно, $\mu_j(t)$ принимает вид монотонно убывающей (рис. 1, *a*) или монотонно возрастающей (рис. 1, *б*) на основном промежутке обслуживания функции. Отметим, что входные параметры при эксперименте подбирались так, что средняя интенсивность за время обслуживания каждого из потоков была одинаковой и равнялась 5 заявкам в секунду. Входные параметры, зафиксированные по каждому из потоков, измеряются в условных единицах времени: $T_{2j-1} = 10$, $T_{2j} = 5$, $j = \overline{1, m}$, $t' = 5$, $t = 0.1$; число итерацией $k = 100$.

В качестве примера монотонно убывающей функции рассматривалась $\mu_j(t) = -0.6t + 8$, $j = 2, 3$ (рис. 1, *a*), а в качестве монотонно возрастающей $\mu_j(t) = 0.6t + 2$, $j = 2, 3$ (рис. 1, *б*). Для первого приоритетного потока интенсивность поступления заявок и интенсивность их обслуживания зафиксированы и равны $\lambda_1 = 0.01$, $\mu_1(t) = 5$ соответственно. Полученные результаты приведены в таблице.

Результаты численных экспериментов

λ_2 (заяв./ед.вр.)	λ_3 (заяв./ед.вр.)	$\mu_j(t), j = 2, 3$ (заяв./ед.вр.)	\bar{x}_2 (заяв.)	\bar{x}_3 (заяв.)	\tilde{x} (заяв.)
0.3	1	$-0.6t + 8$	89	6	23
0.3	1	$0.6t + 2$	71	6	20
0.5	1.25	$-0.6t + 8$	951	8	262
0.5	1.25	$0.6t + 2$	810	13	228
0.7	1.5	$-0.6t + 8$	1896	17	588
0.7	1.5	$0.6t + 2$	1789	13	553

По данным таблицы можно сделать вывод о том, что вид интенсивности обслуживания дает лучшие показатели (минимальные) для оценки средневзвешенной очереди при возрастающей интенсивности $\mu_j(t)$. Действительно, при смене входных параметров, а именно интенсивности поступления заявок в систему, оценка средней очереди по второму потоку при линейной возрастающей функции всегда меньше, чем при линейной убывающей.

Аналогичное исследование было проведено и для других видов интенсивностей обслуживания.

Заключение

Построена математическая модель приоритетной управляющей системы с непостоянной интенсивностью обслуживания в виде пятимерной случайной векторной марковской последовательности. Данная последовательность изучалась аналитически. Также была создана имитационная модель рассматриваемой управляющей системы, позволяющая исследовать влияние вида функции интенсивности обслуживания на некоторые характеристики функционирования системы, например, на средневзвешенную очередь.

ЛИТЕРАТУРА

1. *Ляпунов А.А., Яблонский С.В.* Теоритические проблемы кибернетики // Проблемы кибернетики. М.: Физматгиз, 1963. С. 5–22.
2. *Пройдакова Е.В., Санникова В.О.* Применение кибернетического подхода для построения математической модели функционирования адаптивной системы обслуживания // Проблемы теоретической кибернетики. Материалы заочного семинара XIX международной конференции / под ред. Ю.И. Журавлева. Казань: Отечество, 2020. 159 с.
3. *Пройдакова Е.В.* Численное исследование циклической и приоритетной систем управления конфликтными потоками требований // Вестник Нижегородского университета им. Н.И. Лобачевского. 2013. № 3(1). С. 199–205.
4. *Пройдакова Е.В.* Исследование приоритетной системы обслуживания и ее выходящих потоков с помощью кибернетического подхода. // Материалы XIII Международной научно-практической конференции имени А.Ф. Терпугова «Информационные технологии и математическое моделирование (ИТММ-2014)», 20–22 ноября 2014 г. Томск: Изд-во Том. ун-та, 2014. Ч. 2. С. 202–207.

Исследование RQ-системы M/M/1 с ненадежным прибором асимптотическим и матричным методами

Светлана Рожкова^{1,2}, Наталья Воронина¹,
Александра Семашко²

¹Национальный исследовательский

Томский политехнический университет, г. Томск, Россия

²Национальный исследовательский

Томский государственный университет, г. Томск, Россия

Математические модели RQ-систем широко используются при анализе процессов в системах производства и обслуживания, телефонных системах, вычислительных комплексах. Особенностью этих моделей является возможность повторной передачи искаженных сообщений. Если при поступлении заявки прибор занят обслуживанием, то эта заявка переходит в источник повторных вызовов (на орбиту) и через некоторое случайное время повторяет попытку получить обслуживание [1].

Нагрузка от потока повторных вызовов, как правило, является не учтенной, что приводит к перегрузке сети, и обслуживающий прибор может выйти из строя. Многочисленные поломки и ограниченные возможности ремонта оказывают значительное влияние на производительность системы [2]. Системы массового обслуживания называются ненадежными, если их приборы могут время от времени выходить из строя и требовать восстановления (ремонта), только после которого они могут возобновить обслуживание запросов как новые. Исследованию систем с ненадежными приборами посвящен ряд работ, обзор которых приведен в [3].

В настоящей работе исследуется одноканальная RQ-система массового обслуживания с ненадежным прибором асимптотическим и матричным методами.

Постановка задачи и описание модели

Рассмотрим однолинейную RQ-систему с ненадежным прибором (рис. 1). На вход поступает простейший поток заявок с интенсивностью λ . Каждая заявка обслуживается случайное время, которое распределено по экспоненциальному закону с параметром μ_1 . Ненадежный прибор

может быть в одном из состояний: свободен, занят обслуживанием или на ремонте. Если прибор свободен и на вход поступает заявка, то прибор сразу же начинает обслуживание поступившей заявки. Если заявка поступает в момент, когда прибор занят обслуживанием другой заявки, тогда поступившая заявка переходит на орбиту и ожидает обслуживания прибором. Время ожидания на орбите распределено по экспоненциальному закону с параметром σ . После задержки на орбите, заявка вновь пробует занять прибор. Если прибор оказался свободным, то заявка обслуживается, иначе остается на орбите.

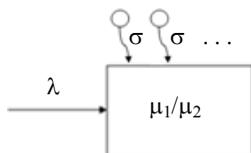


Рис. 1. Модель RQ-системы M/M/1 с ненадежным прибором

Предполагается, что сервер ненадежный. Если сервер простаивает/занят, время бесперебойной работы является случайной величиной, распределенной по экспоненциальному закону с параметром γ_1 или γ_2 , соответственно. Время восстановления прибора после ремонта распределено экспоненциально с параметром μ_2 . Когда сервер на ремонте, все поступающие заявки уходят на орбиту.

Система дифференциальных уравнений Колмогорова

Пусть $i(t)$ – число заявок на орбите в момент времени t , а $k(t)$ определяет состояние прибора:

$$k(t) = \begin{cases} 0, & \text{если прибор свободен,} \\ 1, & \text{если прибор занят,} \\ 2, & \text{если прибор на ремонте.} \end{cases}$$

Требуется найти стационарное распределение вероятностей числа заявок на орбите с учетом состояния прибора

$$P_k(i, t) = P\{k(t) = k, i(t) = i\}, k = \{0, 1, 2\}, i = 0, 1, 2, \dots$$

Для распределения вероятностей $P_k(i, t)$ состояний рассматриваемой RQ-системы составим систему дифференциальных уравнений Колмогорова

$$\begin{aligned} \partial P_0(i, t) / \partial t &= -(\lambda + i\sigma + \gamma_1)P_0(i, t) + \mu_1 P_1(i, t) + \mu_2 P_2(i, t), \\ \partial P_1(i, t) / \partial t &= -(\lambda + \mu_1 + \gamma_2)P_1(i, t) + \lambda P_0(i, t) + \\ &\quad + (i+1)\sigma P_0(i+1, t) + \lambda P_1(i-1, t), \\ \partial P_2(i, t) / \partial t &= -(\lambda + \mu_2)P_2(i, t) + \gamma_1 P_0(i, t) + \gamma_2 P_1(i-1, t) + \lambda P_2(i-1, t). \end{aligned} \quad (1)$$

Обозначим $P_k(i) = \lim_{t \rightarrow \infty} P_k(i, t)$, тогда система (1) в стационарном режиме имеет вид

$$\begin{cases} -(\lambda + i\sigma + \gamma_1)P_0(i) + \mu_1 P_1(i) + \mu_2 P_2(i) = 0, \\ -(\lambda + \mu_1 + \gamma_2)P_1(i) + \lambda P_0(i) + (i+1)\sigma P_0(i+1) + \lambda P_1(i-1) = 0, \\ -(\lambda + \mu_2)P_2(i) + \gamma_1 P_0(i) + \gamma_2 P_1(i-1) + \lambda P_2(i-1) = 0. \end{cases} \quad (2)$$

Так как (2) является системой бесконечной размерности, то решить ее аналитически достаточно сложно. Поэтому рассмотрим численный (матричный) и асимптотический методы решения системы.

Матричный метод

Запишем систему (2) в виде

$$\mathbf{P} \cdot \mathbf{S} = \mathbf{V}. \quad (3)$$

Здесь \mathbf{P} – вектор-строка размерности $3(N+1)$, которая является стационарным распределением вероятностей числа заявок на орбите каждого состояния прибора

$$\mathbf{P} = (P_0(0), \dots, P_0(N), P_1(0), \dots, P_1(N), P_2(0), \dots, P_2(N)),$$

а \mathbf{S} – блочная матрица размерности $3(N+1) \times (3(N+1)+1)$

$$\mathbf{S} = \begin{pmatrix} S_{11} & S_{12} & S_{13} & S_{14} \\ S_{21} & S_{22} & S_{23} & S_{24} \\ S_{31} & S_{32} & S_{33} & S_{34} \end{pmatrix},$$

где

$$\begin{aligned} S_{11} &= \|S_{i,j}^{11}\|_0^N, & S_{12} &= \|S_{i,j}^{12}\|_0^N, & S_{13} &= \|S_{i,j}^{13}\|_0^N, \\ S_{21} &= \|S_{i,j}^{21}\|_0^N, & S_{22} &= \|S_{i,j}^{22}\|_0^N, & S_{23} &= \|S_{i,j}^{23}\|_0^N, \\ S_{31} &= \|S_{i,j}^{31}\|_0^N, & S_{32} &= \|S_{i,j}^{32}\|_0^N, & S_{33} &= \|S_{i,j}^{33}\|_0^N. \end{aligned}$$

Элементы матрицы \mathbf{S} определяются следующим образом:

$$\begin{aligned} S_{i,i}^{11} &= -(\lambda + i\sigma + \gamma_1), \quad S_{i,i}^{12} = \lambda, \quad S_{i+1,i}^{12} = \sigma(i+1), \quad S_{i,i}^{13} = \gamma_1, \\ S_{i,i}^{21} &= \mu_1, \quad S_{i-1,i}^{22} = \lambda, \quad S_{i,i}^{22} = -(\lambda + \mu_1 + \gamma_2), \quad S_{i-1,i}^{23} = \gamma_2, \\ S_{i,i}^{31} &= \mu_2, \quad S_{i,i}^{32} = 0, \quad S_{i,i}^{33} = -(\lambda + \mu_2), \quad S_{i-1,i}^{33} = \lambda. \end{aligned}$$

Блоки S_{14} , S_{24} , S_{34} – единичные векторы-столбцы размерности $(N+1)$, а $\mathbf{B} = \|b_i\|$ – вектор-строка размерности $3(N+1)+1$, которая содержит свободные коэффициенты $b_n = 0$, $n = \overline{0, N-1}$ и $b_N = 1$.

Решение системы (3) находим численно матричным методом, используя математический пакет Mathcad.

При размерности матриц свыше $N = 500$ возникают вычислительные трудности, связанные с обращением матриц больших размерностей.

Поэтому необходимо использование аналитических методов, на работу которых не будет влиять размерность системы, например метод асимптотического анализа.

Асимптотический анализ

Введем частичные характеристические функции

$$H_k(u) = \sum_{i=0}^{\infty} e^{ju} P_k(i), \quad H_k(0) = \sum_{i=0}^{\infty} P_k(i) \triangleq R_k, \quad (4)$$

где $j = \sqrt{-1}$, $k = \{0, 1, 2\}$, R_k – вероятности стационарного состояния процесса $k(t)$. Запишем систему для $H_k(u)$

$$\begin{cases} -(\lambda + \gamma_1)H_0(u) + j\sigma e^{ju} \frac{\partial H_0(u)}{\partial u} + \mu_1 H_1(u) + \mu_2 H_2(u) = 0, \\ -(\lambda + \mu_1 + \gamma_2)H_1(u) + \lambda H_0(u) - j\sigma \frac{\partial H_0(u)}{\partial u} + \lambda e^{ju} H_1(u) = 0, \\ -(\lambda + \mu_2)H_2(u) + \gamma_1 H_0(u) + \gamma_2 e^{ju} H_1(u) + \lambda e^{ju} H_2(u) = 0. \end{cases} \quad (5)$$

Уравнения для характеристической функции (5) решим при асимптотическом условии большой задержки на орбите, т.е. $\sigma \rightarrow 0$. Результат сформулируем в виде теорем.

Теорема 1 (асимптотика первого порядка). Пусть $i(t)$ – число заявок на орбите в RQ-системе с ненадежным прибором, тогда для последовательности характеристических функций выполняется равенство

$$\lim_{\sigma \rightarrow 0} M \{ \exp \{ jw\sigma i(t) \} \} = \exp \{ jwG_1 \},$$

где $G_1 = \lambda(\gamma_2\mu_2 + \mu_1\gamma_1 + \gamma_1\gamma_2 + \mu_2\lambda + \gamma_2\lambda) / (\mu_1\mu_2 - \mu_2\lambda - \gamma_2\lambda)$.

Асимптотика первого порядка характеризует среднее значение числа заявок на орбите. Рассмотрим асимптотику второго порядка, чтобы более детально проанализировать процесс $i(t)$.

Теорема 2 (асимптотика второго порядка). Пусть $i(t)$ – число заявок на орбите RQ-системы с ненадежным прибором, тогда для последовательности характеристических функций выполняется равенство

$$\lim_{\sigma \rightarrow 0} M \{ \exp \{ jw\sqrt{\sigma} (i(t) - G_1/\sigma) \} \} = \exp \left\{ \frac{(jw)^2}{2} G_2 \right\},$$

где $G_2 = [G_1R_0 + (G_1 + \gamma_1 + \lambda)f_0 - \mu_1f_1 - \mu_2f_2] / R_0$.

Из теоремы 2 следует, что асимптотическое распределение вероятностей числа заявок на орбите в RQ-системе является гауссовским с параметрами G_1/σ и G_2/σ , что позволяет для распределения $P(i)$ построить аппроксимацию вида

$$P_{asimpt}(i) = \frac{L(i + 0,5) - L(i - 0,5)}{(1 - L(-0,5))},$$

где $L(x)$ – функция нормального распределения с параметрами G_1/σ и G_2/σ .

Численные результаты

Рассмотрим RQ-систему с параметрами $\mu_1 = 7$, $\mu_2 = 1$, $\gamma_1 = 0.03$, $\gamma_2 = 0.03$, $\lambda = 3$. Сравним распределение вероятностей числа заявок на орбите, полученное матричным методом с результатами асимптотического анализа.

На рис. 2 показано сравнение асимптотического и численного распределений вероятностей числа заявок при $\sigma = 0.1$ и $\sigma = 0.01$.

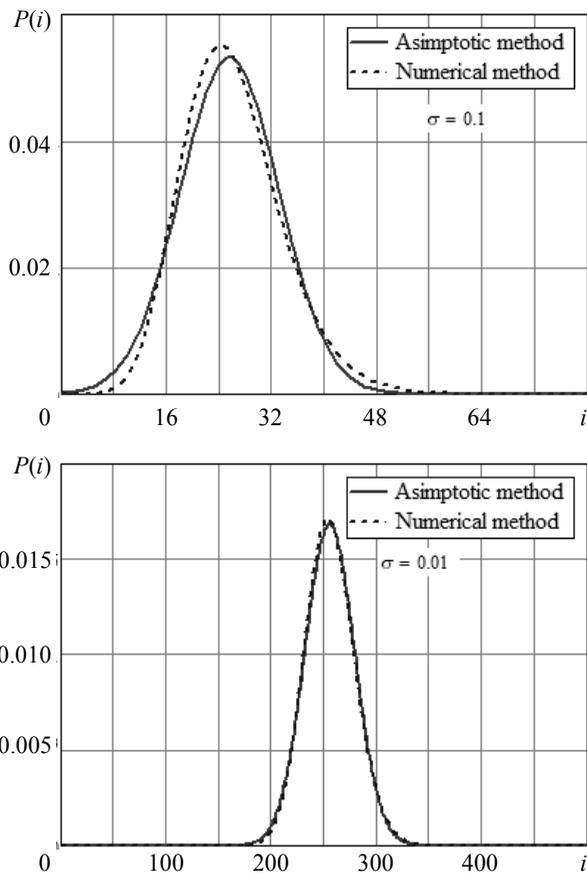


Рис. 2. Асимптотическое и численное распределение вероятностей числа заявок

Чтобы определить точность двух распределений, используем расстояние Колмогорова.

$$\Delta = \max_{0 \leq k \leq N} \left| \sum_{i=0}^k [P_{matrix}(i) - P_{asimp}(i)] \right|,$$

где $P_{matrix}(i)$ – распределение, полученное матричным методом, а $P_{asimp}(i)$ – распределение, полученное асимптотическим методом.

Расстояние Колмогорова

σ	1	0.5	0.1	0.05	0.01
Δ	0.106	0.049	0.028	0.021	0.011

Получаем, что расстояние Колмогорова уменьшается при уменьшении σ .

Заключение

Рассмотрена M/M/1 RQ-система с ненадежным прибором. Доказано, что при условии большой задержки на орбите стационарное распределение вероятностей числа заявок на орбите может быть аппроксимировано гауссовским распределением. Исследована точность аппроксимации с помощью численного матричного метода.

ЛИТЕРАТУРА

1. *Wilkinson R.I.* Theories for toll traffic engineering in the USA // The Bell System Technical Journal. 1956. V. 35. No. 2. P. 421–507.
2. *Gosztory G.* Repeated call attempts and their effect on traffic engineering // Budavox Telecommunication Review. 1976. V. 2. P. 16–26.
3. *Cohen J.W.* Basic problems of telephone traffic and the influence of repeated calls // Philips Telecommunication Review. 1957. V. 18. No. 2. P. 49–100.

Метод анализа замкнутых сетей массового обслуживания с системами типа $M_a/M^{[x,y]}/1^*$

Елена Станкевич, Игорь Тананко

Саратовский национальный исследовательский государственный университет имени Н. Г. Чернышевского, г. Саратов, Россия

В настоящее время достаточно хорошо развиты теория и методы анализа систем и сетей массового обслуживания, в которых каждым прибором в любой момент времени может обслуживаться только одно требование. Однако при решении задач распределения ресурсов, задач анализа производственных систем и процессов, а также тесно связанных с ними транспортных систем, часто требуется учитывать групповые обработку, производство и перемещение сырья, материалов и готовой продукции. Эффективным инструментом для решения этих задач являются системы и сети массового обслуживания с групповым обслуживанием требований [1–4].

Рассматривается замкнутая экспоненциальная сеть массового обслуживания, состоящая из одноприборных систем массового обслуживания. Приборы обслуживают требования группами, нижнее и верхнее ограничения, на размер которых заданы для каждой системы в сети. Результаты данной работы являются обобщением результатов работы [4]. Приведены выражения для основных стационарных характеристик сети массового обслуживания с групповым обслуживанием.

Описание сети массового обслуживания

Рассматривается замкнутая сеть массового обслуживания с непрерывным временем, состоящая из L систем массового обслуживания S_i , $i = 1, \dots, L$, и содержащая H требований одного класса. Маршрутизация требований в сети производится согласно маршрутной матрице $\Theta = (\theta_{ij})$, $i, j = 1, \dots, L$, где θ_{ij} – вероятность перехода требования из системы S_i в систему S_j после завершения обслуживания данного требования в системе S_i . Система S_i , $i = 1, \dots, L$, включает один обслуживающий прибор, одновременно обслуживающий группу, состоящую из g_i , $x_i \leq g_i \leq y_i$, требований, и очередь бесконечной длины. Длительность обслуживания

* Работа выполнена при поддержке Минобрнауки России в рамках выполнения государственного задания (проект № FSRR-2020-0006).

группы требований g_i , в системе S_i является экспоненциально распределенной случайной величиной с параметром μ_i . Выбор на обслуживание из очереди группы, состоящей из g_i требований, осуществляется согласно дисциплине RANDOM. Если в очереди находится меньше, чем x_i требований, то обслуживающий прибор будет простаивать до тех пор, пока в очередь не поступит, по крайней мере, x_i требований. Состояние сети определяется вектором $s = (s_i)$, $i = 1, \dots, L$, где s_i – число требований, находящихся в системе S_i . Множество состояний сети обозначим через X , множество номеров систем – I , множество номеров выходных смежных с S_i систем – V_i .

Переход сети обслуживания из состояния s , $s \in X$, в состояние s' , $s' \in X$, $s \neq s'$, происходит в результате выполнения следующей последовательности действий. В некоторый момент времени в одной из L систем S_i , $i \in I$, завершается обслуживание группы, состоящей из $g_i > 0$ требований. Формируется вектор выходящих требований $d = (d_k)$, $k = 1, \dots, L$, в котором $d_i = g_i$, а остальные элементы $d_k = 0$, $k \neq i$. Множество всех векторов выходящих требований d обозначим через D . Из группы d_i , $i \in I$, требований, выходящих из системы S_i , в соответствии с алгоритмом маршрутизации, определяемым матрицей Θ , формируются направляемые из S_i в S_j подгруппы требований a_j , $j \in V_i$. Размеры групп требований, входящих в системы обслуживания, являются компонентами вектора входящих требований $a = (a_j)$, $j = 1, \dots, L$. Множество всех возможных векторов a обозначим через A . Формируется новое состояние сети $s' = s - d + a$.

Эволюция сети описывается цепью Маркова с непрерывным временем, пространством состояний X и матрицей интенсивностей переходов $Q = (q(s, s'))$, $s, s' \in X$, где

$$q(s, s') = \sum_{\substack{s' \in X, \\ s' = s - d + a}} \sum_{i=1}^L \mu_i \mathbf{1}(s_i \geq x_i) \begin{pmatrix} d_i & & \\ a_1 & \dots & a_L \end{pmatrix} \prod_{j=1}^L \theta_{ij}^{a_j},$$

$$s \in X, d \in D, a \in A.$$

Стационарные характеристики

Стационарное распределение $\pi = (\pi(s))$, $s \in X$, цепи Маркова совпадает со стационарным распределением сети обслуживания и является решением уравнения $\pi Q = 0$ с условием $\sum_{s \in X} \pi(s) = 1$.

Математическое ожидание (м.о.) числа требований в системе S_i вычисляется по формуле

$$\bar{s}_i = \sum_{k=0}^H k \sum_{\substack{s \in X, \\ s_i=k}} \pi(s), \quad i = 1, \dots, L,$$

интенсивность потока требований в систему S_i

$$\lambda_i = \mu_i \sum_{k=x_i}^{y_i} k \left(1 - \sum_{s \in X} \sum_{s_i=0}^{k-1} \pi(s) \right), \quad i = 1, \dots, L,$$

м. о. длительности пребывания требований в системе S_i

$$\bar{u}_i = \frac{\bar{s}_i}{\lambda_i}, \quad i = 1, \dots, L,$$

м. о. длительности простоя прибора в системе S_i

$$\bar{v}_i = \frac{\sum_{k=0}^{x_i-1} (x_i - k) \sum_{\substack{s \in X, \\ s_i=k}} \pi(s)}{\lambda_i \sum_{k=0}^{x_i-1} \sum_{\substack{s \in X, \\ s_i=k}} \pi(s)}, \quad i = 1, \dots, L,$$

м. о. длительности пребывания требований в очереди системы S_i

$$\bar{w}_i = \bar{u}_i - \frac{1}{\mu_i}, \quad i = 1, \dots, L,$$

м. о. числа требований в очереди системы S_i

$$\bar{b}_i = \lambda_i \bar{w}_i, \quad i = 1, \dots, L.$$

Заключение

Предложен метод анализа сети массового обслуживания с непрерывным временем и групповым обслуживанием требований. Разработанный метод может быть использован при решении задач проектирования и анализа производственных систем, в которых необходимо учитывать процессы разделения и слияния различных элементов производственных цепочек, систем распределения ресурсов, систем пассажирских и грузовых перевозок, а также информационных и вычислительных систем с параллельной обработкой и передачей информации.

ЛИТЕРАТУРА

1. *Бояркина А.А., Моисеева С.П.* Исследование бесконечнолинейных СМО с групповым обслуживанием // Труды Томского государственного университета. Т. 304. Серия физико-математическая: Математическое и программное обеспечение информационных, технических и экономических систем: материалы VII Междунар. молодежной науч. конф. Томск, 23–25 мая 2019 г. Томск: Издательский Дом Томского государственного университета, 2019. С. 229–232.
2. *Evequoz C., Tropper C.* Approximate Analysis of Bulk Closed Queueing Networks // INFOR: Information Systems and Operational Research. 1995. V. 33. No. 3. P. 179–204. DOI: 10.1080/03155986.1995.11732280.
3. *Mitici M., Goseling J., Ommereen J.-K., Graaf M., Boucherie R.* On a tandem queue with batch service and its application in wireless sensor networks // Queueing Systems. 2017. V. 87. P. 81–93. DOI: 10.1007/s11134-017-9534-1.
4. *Stankevich E.P., Tananko I.E., Dolgov V.I.* Analysis of closed queueing networks with batch service // Izv. Saratov Univ. (N. S.), Ser. Math. Mech. Inform. 2020. V. 20. Iss. 4. P. 527–533. DOI: 10.18500/1816-9791-2020-20-4-527-533.

Приближенный метод анализа замкнутых сетей массового обслуживания с ненадежными системами и групповым обслуживанием*

Елена Станкевич, Игорь Тананко

Саратовский национальный исследовательский государственный университет имени Н. Г. Чернышевского, г. Саратов, Россия

Исследование сетей массового обслуживания с групповыми переходами, групповым обслуживанием требований и ненадежными приборами в системах является одним из новых направлений развития теории массового обслуживания. Этот класс сетей обслуживания позволяет отображать такие особенности реальных стохастических систем, как групповая обработка объектов и возможность выхода из строя элементов систем. В работах [1, 2] рассмотрены сети массового обслуживания с групповым обслуживанием требований и методы их анализа, в работе [3] – ненадежные сети массового обслуживания с групповыми переходами требований. В данной работе предложен метод анализа замкнутых сетей массового обслуживания с групповым обслуживанием требований, в системах которых обслуживающие приборы могут выходить из строя и восстанавливаться.

Математическая модель

Рассматривается замкнутая сеть массового обслуживания с непрерывным временем, состоящая из L одноприборных систем массового обслуживания S_i , $i = 1, \dots, L$, и содержащая H требований одного класса. Вероятности перехода требований между системами сети определяются маршрутной матрицей $\Theta = (\theta_{ij})$, $i, j = 1, \dots, L$, где θ_{ij} – вероятность того, что требование после завершения обслуживания в системе S_i перейдет в систему S_j . Состояние сети определяется вектором $s = (s_i)$, $i = 1, \dots, L$, где s_i – число требований, находящихся в системе S_i . Множество состояний сети обозначим через X , множество номеров систем – I . Обслуживающие приборы систем сети обслуживают группы требований размера g_i , причем для каждой системы S_i заданы минимальные x_i и

* Работа выполнена при поддержке Минобрнауки России в рамках выполнения государственного задания (проект № FSRR-2020-0006).

максимальные y_i значения для размеров обслуживаемых групп требований, $x_i \leq g_i \leq y_i$. Выбор на обслуживание из очереди группы, состоящей из g_i требований, осуществляется согласно дисциплине RANDOM. Если в очереди находится меньше, чем x_i требований, то обслуживающий прибор будет простаивать до тех пор, пока в очередь не поступит, по крайней мере, x_i требований. Предполагается, что обслуживающий прибор может выходить из строя и восстанавливаться. Если прибор находится в работоспособном состоянии, то обслуживание группы требований производится с интенсивностью μ_i . Если же прибор находится в неработоспособном состоянии, то обслуживание группы прекращается до восстановления прибора. С момента восстановления прибора прерванное обслуживание группы требований начинается заново. Длительности пребывания прибора в системе S_i в работоспособном и неработоспособном состояниях являются экспоненциально распределенными случайными величинами с параметрами α_i и β_i соответственно. Длительность обслуживания группы требований с учетом отказов и восстановлений прибора имеет распределение Кокса второго порядка [4]. Будем рассматривать случай, когда $1/\alpha_i \gg 1/\beta_i$, $i = 1, \dots, L$. Тогда адекватным приближением для случайной величины с распределением Кокса второго порядка является случайная величина, имеющая экспоненциальное распределение, с математическим ожиданием

$$\bar{r}_i = \mu_i^{-1} \frac{\alpha_i + \beta_i}{\beta_i}, \quad i = 1, \dots, L,$$

которую будем использовать в качестве длительности обслуживания требований в системах массового обслуживания.

Переход сети обслуживания из состояния s , $s \in X$, в состояние s' , $s' \in X$, $s \neq s'$, происходит в результате выполнения следующей последовательности действий. В некоторый момент времени в одной из L систем S_i , $i \in I$, завершается обслуживание группы, состоящей из g_i требований. Формируется вектор выходящих требований $d = (d_k)$, $k = 1, \dots, L$, в котором $d_i = g_i$, а остальные элементы $d_k = 0$, $k \neq i$. Множество всех векторов выходящих требований d обозначим через D . Из группы d_i , $i \in I$, требований, выходящих из системы S_i , в соответствии с алгоритмом маршрутизации, определяемым матрицей Θ , формируются направляемые из S_i в S_j подгруппы требований a_j , $j = 1, \dots, L$, $i \neq j$. Размеры групп требований, входящих в системы обслуживания, являются компонентами вектора входящих требований $a = (a_j)$, $j = 1, \dots, L$. Множество всех возможных

векторов a обозначим через A . Формируется новое состояние сети $s' = s - d + a$.

Интенсивности переходов цепи Маркова с непрерывным временем, описывающей эволюцию сети массового обслуживания с ненадежными системами и групповым обслуживанием, имеют следующий вид:

$$q(s, s') = \sum_{\substack{s' \in X, \\ s' = s - d + a}} \sum_{i=1}^L \frac{\mathbf{1}(s_i \geq x_i)}{\bar{r}_i} \begin{pmatrix} d_i & & \\ & \dots & \\ a_1 & \dots & a_L \end{pmatrix} \prod_{j=1}^L \theta_{ij}^{a_j},$$

$$s \in X, d \in D, a \in A.$$

Стационарное распределение $\pi = (\pi(s))$ сети массового обслуживания является решением уравнения $\pi Q = 0$ с условием $\sum_{s \in X} \pi(s) = 1$, где $Q = (q(s, s'))$, $s, s' \in X$.

Заключение

Для замкнутой сети массового обслуживания с групповым обслуживанием, в системах которых приборы могут выходить из строя и восстанавливаться, построен инфинитезимальный оператор. В дальнейшем планируются исследование замкнутых ненадежных сетей массового обслуживания с групповым обслуживанием и точности предложенного метода их анализа. Разработанный метод анализа может быть использован для решения задач проектирования и оптимизации дискретных стохастических систем с сетевой структурой и ненадежными элементами.

ЛИТЕРАТУРА

1. Mitici M., Goseling J., Ommeren J.-K., Graaf M., Boucherie R. On a tandem queue with batch service and its application in wireless sensor networks // *Queueing Systems*. 2017. V. 87. P. 81–93. DOI: 10.1007/s11134-017-9534-1.
2. Stankevich E.P., Tananko I.E., Dolgov V.I. Analysis of closed queueing networks with batch service // *Izv. Saratov Univ. (N. S.), Ser. Math. Mech. Inform.* 2020. V. 20. Iss. 4. P. 527–533. DOI: 10.18500/1816-9791-2020-20-4-527-533.
3. Тананко И.Е., Фокина Н.П. Анализ замкнутых ненадежных сетей массового обслуживания с групповыми переходами требований // *Известия Саратовского университета. Новая серия. Серия: Математика. Механика. Информатика*. 2013. Т. 13. № 2-1. С. 111–117. DOI: 10.18500/1816-9791-2013-13-2-1-111-117.
4. Vinod B., Aliok T. Approximating unreliable queueing networks under the assumption of exponentiality // *J. Opl. Res. Soc.* 1986. V. 37. No. 3. P. 309–316. DOI: 10.1057/jors.1986.49.

Асимптотическая оценка интенсивности сборки пуассоновских потоков

Гурами Цициашвили¹, Анатолий Назаров²,
Александр Моисеев²

¹ *Институт прикладной математики ДВО РАН, г. Владивосток, Россия,
Национальный исследовательский*

² *Томский государственный университет, Томск, Россия*

Рассматривается сборка стационарных пуассоновских потоков событий. Под термином «сборка» понимается такой поток, события которого представляют собой события исходных потоков, «склеенных» по их порядковым номерам. Другими словами, события в сборке наступают в момент, когда наступает последнее из событий с соответствующим порядковым номером среди всех потоков. Подобные модели могут применяться для анализа компьютерных сетей, производственных линий и других систем [1, 2].

В [3] показано, что средняя интенсивность такого потока при растущем времени наблюдения ($t \rightarrow \infty$) стремится к наименьшей из интенсивностей исходных потоков. Однако вычислительные эксперименты, проведенные с помощью аппроксимации пуассоновского распределения с большим параметром нормальным распределением, показали, что имеется возможность улучшения построенных оценок скорости сходимости.

Настоящая работа посвящена получению в некотором смысле наилучшаемых оценок скорости сходимости интенсивности потока сборки. Анализ проблемы показывает, что при решении данной задачи следует сочетать аналитические и численные оценки, постоянно сравнивая их друг с другом. Причем важную роль здесь играет центральная предельная теорема, принимаемая в смысле C -сходимости [4].

Для сокращения объема текста все доказательства и некоторые вспомогательные утверждения в данной статье опущены.

Математическая модель и постановка задачи

Пусть имеется r независимых друг от друга стационарных пуассоновских потоков событий, которые будем называть исходными потоками. Интенсивности этих потоков равны $\lambda_1, \dots, \lambda_r$. Обозначим $t_{k,i}$ – мо-

мент наступления i -го события в k -м исходном потоке $T_k = \{0 \leq t_{k,1} \leq t_{k,2} \leq \dots\}$ ($i = 1, 2, \dots; k = 1, \dots, r$). Сборкой потоков T_1, \dots, T_r (или потоком сборки) называется поток событий

$$A_r = \{0 \leq \max(t_{1,1}, \dots, t_{r,1}) \leq \max(t_{1,2}, \dots, t_{r,2}), \dots\}.$$

Обозначим $n_k(t)$, $t \geq 0$ – число событий, наступивших в k -м исходном потоке до момента t . Тогда число событий $N_r(t)$, наступивших в потоке сборки до момента времени t , можно записать как

$$N_r(t) = \min_{k=1, \dots, r} n_k(t).$$

Следует заметить, что по очевидным причинам рассматриваемый поток сборки не является пуассоновским, что затрудняет его исследование.

Центральная предельная теорема для потока сборки

Пусть существует некоторое число s потоков с одинаковой наименьшей интенсивностью:

$$\lambda = \lambda_1 = \dots = \lambda_s < \lambda_{s+1} \leq \dots \leq \lambda_r.$$

Тогда имеет место следующее утверждение.

Теорема 1. Для любого v : $-\infty < v < \infty$, справедливо предельное соотношение

$$P\left\{\frac{N_r(t) - \lambda t}{\sqrt{\lambda t}} > v\right\} \rightarrow \left[\int_v^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-u^2/2) du\right]^s, \quad t \rightarrow \infty.$$

Замечание 1. Случайный процесс $\frac{n_k(tu) - \lambda tu}{\sqrt{\lambda t}}$ как функция неотрицательной переменной u при $t \rightarrow \infty$ сходится к винеровскому случайному процессу $w_k(u)$, $k = 1, \dots, s$ в смысле C -сходимости [4].

Замечание 2. Пусть $r = s$, тогда случайный процесс $\frac{N_r(tu) - \lambda tu}{\sqrt{\lambda t}}$ при $t \rightarrow \infty$ сходится к винеровскому случайному процессу $\min_{k=1, \dots, r} w_k(u)$ в смысле C -сходимости, где $w_1(u), \dots, w_r(u)$, $u \geq 0$ – независимые винеровские процессы.

Предельные соотношения для интенсивности сборки потоков с одинаковой интенсивностью

Пусть все исходные потоки имеют одинаковую интенсивность (случай $s = r$). Рассмотрим марковский процесс $\{n_1(t), \dots, n_r(t)\}$. Скачок этого процесса в момент времени t из состояния $(n_1, \dots, n_i, \dots, n_r)$, в котором $n_i < \min_{k \neq i} n_k$, в состояние $(n_1, \dots, n_i + 1, \dots, n_r)$ приводит к появлению в этот момент времени события у потока сборки. Следовательно, мгновенная интенсивность потока сборки $\bar{\lambda}(t)$ удовлетворяет равенству

$$\bar{\lambda}(t) = \lambda \sum_{i=1}^r \mathbf{P} \left\{ n_i(t) < \min_{k \neq i} n_k(t) \right\}.$$

Введем следующие обозначения:

$$a = \lambda t, \quad p(k, a) = \frac{e^{-a} a^k}{k!}, \quad k = 0, 1, \dots,$$

$$f(a) = \mathbf{P} \{ n_1(t) = \dots = n_r(t) \} = \sum_{k=0}^{\infty} p^r(k, a).$$

Лемма 1. Выполняется следующее равенство:

$$\bar{\lambda}(t) = \lambda (1 - \mathbf{P} \{ n_1(t) = \dots = n_r(t) \}) = \lambda (1 - f(a)). \quad (1)$$

Аппроксимацию $g(a)$ функции $f(a)$ будем искать в виде

$$g(a) = \int_{-\infty}^{\infty} \left[\frac{1}{\sqrt{2\pi a}} \exp \left(-\frac{(x-a)^2}{2a} \right) \right]^r dx = \frac{1}{\sqrt{r}} (2\pi a)^{(1-r)/2}.$$

Запишем основные результаты исследования.

Теорема 2. Для $\lambda > 0$, $r > 2$ и $\frac{1}{2} < \gamma < \frac{2}{3}$ справедливо следующее предельное выражение:

$$f(a) = g(a) \left(1 + O(a^{3\gamma-2}) \right) \sim g(a), \quad a \rightarrow \infty, \quad (2)$$

и, следовательно, $\lambda(t) \rightarrow \lambda$: $\lambda(t) - \lambda \sim \lambda \frac{(2\pi\lambda t)^{(1-r)/2}}{\sqrt{r}}, \quad t \rightarrow \infty.$

Замечание 3. Приведем результаты численного эксперимента, иллюстрирующего точность полученной аппроксимации (2). Обозначим погрешность $\Delta(a) = \left| \frac{f(a) - g(a)}{f(a)} \right|$. Ее значения для различных возрас-

тающих a при разных значениях числа потоков в сборке представлены в табл. 1. Как видно из таблицы, погрешность $\Delta(a)$ аппроксимации (2) уменьшается с ростом числа событий в исходных потоках $a = \lambda t$ для разного их числа r , что косвенно подтверждает полученные выражения.

Таблица 1

Уменьшение погрешности аппроксимации $\Delta(a)$ при возрастании a для разных r

$r \backslash a$	10	10^2	10^3	10^4	10^5	10^6
2	$6.4 \cdot 10^{-3}$	$6.3 \cdot 10^{-4}$	$6.3 \cdot 10^{-5}$	$6.3 \cdot 10^{-6}$	$6.2 \cdot 10^{-7}$	$6.2 \cdot 10^{-8}$
5	$2.0 \cdot 10^{-2}$	$2.0 \cdot 10^{-3}$	$2.0 \cdot 10^{-4}$	$2.0 \cdot 10^{-5}$	$2.0 \cdot 10^{-6}$	$2.0 \cdot 10^{-7}$
20	$8.3 \cdot 10^{-2}$	$8.3 \cdot 10^{-3}$	$8.3 \cdot 10^{-4}$	$8.3 \cdot 10^{-5}$	$8.3 \cdot 10^{-6}$	$8.2 \cdot 10^{-7}$

Сборка потоков с разной интенсивностью

Рассмотрим случай двух независимых стационарных пуассоновских потоков с разными интенсивностями λ_1 и λ_2 , при этом будем считать, что $\lambda_1 < \lambda_2$. Обозначим

$$d = \lambda_2 t, \quad c = \frac{\lambda_1}{\lambda_2} < 1 \quad (cd = \lambda_1 t).$$

По аналогии с формулой (1) имеем

$$\begin{aligned} \bar{\lambda}(t) &= \lambda_1 P\{n_2(t) > n_1(t)\} + \lambda_2 P\{n_1(t) > n_2(t)\} = \\ &= \lambda_1 - \lambda_1 P\{n_1(t) \geq n_2(t)\} + \lambda_2 P\{n_1(t) > n_2(t)\}, \end{aligned}$$

следовательно,

$$|\bar{\lambda}(t) - \lambda_1| \leq \lambda_2 P\{n_1(t) \geq n_2(t)\},$$

где
$$P\{n_1(t) \geq n_2(t)\} = G(d) = \sum_{k=0}^{\infty} e^{-d} \frac{d^k}{k!} \sum_{i=k}^{\infty} e^{-cd} \frac{(cd)^i}{i!}.$$

Будем называть положительные функции $p(d)$ и $q(d)$ удовлетворяющими соотношению $p(d) \underset{-}{\prec} q(d)$ при $d \rightarrow \infty$, если $\limsup_{d \rightarrow \infty} \frac{p(d)}{q(d)} < \infty$.

Теорема 3. Для любого $c: 0 < c < 1$, справедливо соотношение

$$d^{-1} \exp(-d\alpha(c)) G(d) \leq d^{1/2} \exp(-d\alpha(c)), \quad (3)$$

и, следовательно, $\lambda(t) \rightarrow \lambda: \lambda(t) - \lambda = G(-\lambda_2 t), t \rightarrow \infty$. Здесь множитель

$\alpha(c)$ вычисляется по формуле

$$\alpha(c) = 1 - s^*(c)(1 - \ln s^*(c)), \text{ где } s^*(c) = -\frac{1-c}{\ln c}.$$

Замечание 4. В Замечании 3 мы оценивали вероятность $P\{n_1(t) = \dots = n_r(t)\}$, используя для распределения Пуассона с большим параметром гауссовскую аппроксимацию. Показано, что данная аппроксимация дает результаты близкие к аналитическим. Рассмотрим теперь, как такая аппроксимация работает при оценке вероятности $P\{n_1(t) \geq n_2(t)\}$. Для этого запишем следующие аппроксимации для случайных величин $n_1(t)$ и $n_2(t)$:

$$n_1(t) \approx \sqrt{cd} \xi_1 + cd, \quad n_2(t) \approx \sqrt{d} \xi_2 + d,$$

где ξ_1 и ξ_2 – независимые стандартные нормальные случайные величины. Тогда необходимую гауссовскую аппроксимацию (при больших d) мы можем построить следующим образом:

$$\begin{aligned} P\{n_1(t) \geq n_2(t)\} &\approx P\{\sqrt{cd} \xi_1 + cd \geq \sqrt{d} \xi_2 + d\} = \\ &= P\{\xi_2 \leq \sqrt{c} \xi_1 + \sqrt{d} (c-1)\} = S(d). \end{aligned}$$

Обозначим $h = (c-1)\sqrt{\frac{d}{c+1}}$ и η – стандартная нормальная случайная величина. Так как (ξ_1, ξ_2) – двумерный гауссовский случайный вектор с нулевыми средними и единичной матрицей ковариации, то, применяя известную формулу

$$P\{\eta > R\} \sim \frac{1}{R\sqrt{2\pi}} \exp\left(-\frac{R^2}{2}\right), \quad R \rightarrow \infty,$$

мы можем получить следующее соотношение:

$$\begin{aligned} P\{n_1(t) \geq n_2(t)\} &\approx \frac{1}{h\sqrt{2\pi}} \exp\left(-\frac{h^2}{2}\right) = \frac{\sqrt{c+1}}{\sqrt{2\pi d} (c-1)} \exp\left(-d \frac{(c-1)^2}{2(c+1)}\right) = \\ &= \frac{\sqrt{c+1}}{\sqrt{2\pi d} (c-1)} \exp(-dA(c)) = S(d) \text{ при } d \rightarrow \infty, \end{aligned} \quad (4)$$

где
$$A(c) = \frac{(c-1)^2}{2(c+1)}.$$

Сравним теперь множители $\alpha(c)$ и $A(c)$ в экспонентах формул (3) и (4). Для $c = 5/6$ имеем $\alpha(c) \approx 0.0038$, $A(c) \approx 0.0076$; для $c = 2/3$ – $\alpha(c) \approx 0.0168$, $A(c) \approx 0.0333$. Таким образом, множитель $A(c)$, вычисленный с помощью гауссовской аппроксимации, больше $\alpha(c)$, вычисленного аналитически.

Обозначим погрешность $\delta(d) = \left| \frac{G(d) - S(d)}{G(d)} \right|$ и оценим точность га-

уссовской аппроксимации при $c = 5/6$ и $c = 2/3$. Соответствующие результаты приведены в табл. 2 и 3. Из них видно, что при $c = 5/6$ с ростом d погрешность уменьшается, однако с увеличением d скорость ее убывания сильно падает. При $c = 2/3$ с ростом d функция $\delta(d)$ начинает расти. Таким образом, результаты, представленные в табл. 2 и 3, свидетельствуют о значительно худшем качестве гауссовской аппроксимации, чем для результатов, представленных в табл. 1.

Таблица 2

Изменение погрешности аппроксимации $\delta(d)$ с ростом d для случая $c = 5/6$

d	100	200	500	1000	2000
$\delta(d)$	0.267	0.143	0.051	0.021	0.018

Таблица 3

Изменение погрешности аппроксимации $\delta(d)$ с ростом d для случая $c = 2/3$

d	10	50	100	200	500
$\delta(d)$	0.321	0.059	0.029	0.047	0.192

Замечание 5. Используя полученные результаты, несложно рассмотреть случай сборки потоков с интенсивностями

$$\lambda_1 = \dots = \lambda_s < \lambda_{s+1} \leq \dots \leq \lambda_r$$

и получить неравенство

$$|\bar{\lambda}(t) - \lambda_1| \leq \sum_{i=s+1}^r \lambda_i P\{n_1(t) \geq n_i(t)\}.$$

Заключение

Различные варианты центральной предельной теоремы для сборки получены как в терминах случайных величин, так и в терминах случайных процессов. Получены точные асимптотические формулы для интенсивности сборки идентичных пуассоновских потоков и для случая исходных потоков с разными интенсивностями. Выполнена оценка скорости их сходимости для обоих случаев. Несмотря на кажущуюся простоту рассматриваемой задачи исследования сборки независимых стационарных пуассоновских потоков, результаты исследования показали, что построенная модель достаточно сложна и требует тщательного подхода к ее анализу. При этом в обязательном порядке должен проводиться численный анализ получаемых результатов, так как не во всех случаях построенные аппроксимации позволяют обеспечить приемлемую погрешность.

ЛИТЕРАТУРА

1. Горбунова А.В., Зарядов И.С., Самуйлов К.Е., Сотин Э.С. Обзор систем параллельной обработки заявок. Часть I // Вестник Российского университета дружбы народов. Серия: Математика. Информатика. Физика. 2017. Т. 25. № 4. С. 345–357.
2. Горбунова А.В., Зарядов И.С., Самуйлов К.Е. Обзор систем параллельной обработки заявок. Часть II // Вестник Российского университета дружбы народов. Серия: Математика. Информатика. Физика. 2018. Т. 26. № 1. С. 13–27.
3. Цицаивили Г.Ш., Оситова М.А. Исследование процесса сборки пуассоновских потоков // Вестник ТГУ. Управление, вычислительная техника и информатика. 2019. № 3. С. 51–56.
4. Боровков А.А., Могульский А.А., Саханенко А.И. Пределеные теоремы для случайных процессов // Итоги науки и техники. Серия Современные проблемы математики. Фундаментальные направления. 1995. Т. 82. С. 5–197.

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ПРОГРАММНАЯ ИНЖЕНЕРИЯ

INFORMATION TECHNOLOGIES AND SOFTWARE ENGINEERING

Development of back-end of the service for internationalization of web-applications

Marat Gainutdinov, Aleksey Shkurkin, Anastasia Pichugina

Tomsk State University, Tomsk, Russia

The active process of globalization is ongoing. It is influential in a great variety of human activities such as development of web-applications for businesses, learning and so on. For an application to be used by a large number of users it must be translated into languages that are accessible to them. The growing level of globalization and use of the Internet ensure the relevance of services for localization and internationalization of web applications.

The objective of this article is creation of the server part of a service providing internationalization of web-applications.

Created server part of the service (back-end) is presented as an application programming interface (API) which allows implementation of any user interface and is available to be used by developers in any application by calling necessary API method.

Analysis of the subject field

The key concept of the given work is the difference between internationalization and localization [1]:

Localization is an adaptation of a product and its documentation to the linguistic, cultural and other requirements of the target market; **internationalization**, on the other hand, is design, creating and development of a product in order for it to be localized easily and barrier-free.

A service for internationalization of web-applications is software, in this case a web-application, presenting a convenient tool for specialists doing localization and internationalization.

Internationalization is achieved by a number of technical processes, such as usage of UTF encodings, data formatting and design.

Review of alternatives

There are quite a few similar services, so for a critical review and comparison with the back-end of the service under development (U29N), only those that provide their API were taken. All services have a number of advantages and features apart from the minimal functionality; however, the functionality of U29N is fairly simple and well-documented, which makes the use of it convenient and allows quicker understanding of the mechanism of this service.

Table 1

Comparison of alternatives

	webtranslateit	lingohub	transifex	oneskyapp	u29n
Price	€49	€29	€139	\$0.05 per word	0
Ease of use	-	-	-	+/-	+

1) webtranslateit.com – free rate only for small enclosed projects, minimal rate €49, provides a large but at the same time rather complex functionality [2].

2) lingohub.com – free rate for Open Source projects, minimal rate starts at €29 a month, the convenience of use is similar to the previous alternative [3].

3) transifex.com – free rate for Open Source projects, minimal rate €139, the convenience of use is similar to the previous alternative [4].

4) oneskyapp.com – rate \$0.05 per word, relatively convenient API, has ready solutions for PHP, Python and Ruby [5].

5) U29N – free, fairly simple and well-documented.

Selecting requirements

The following functionality requirements were stated with the help of system analytics. A service must:

- Provide an opportunity to create a language in which a web-application can be translated and also enable editing and deleting of languages.
- Provide an opportunity to select a language by default.
- Provide an opportunity to create, edit and delete the messages that represent phrases in a translated web-application.
- Mark the messages as problematic under certain circumstances.

- Provide an opportunity to add, edit and delete translations of messages.
- Provide an opportunity to find and filter messages and translations in any fields.
- Provide an opportunity for several translators to work simultaneously; therefore, the data showed must always be actual.

The process of internationalization requires two specialists: an **editor** and a **translator**.

The following use case diagram was created after all requirements were considered (Fig. 1):

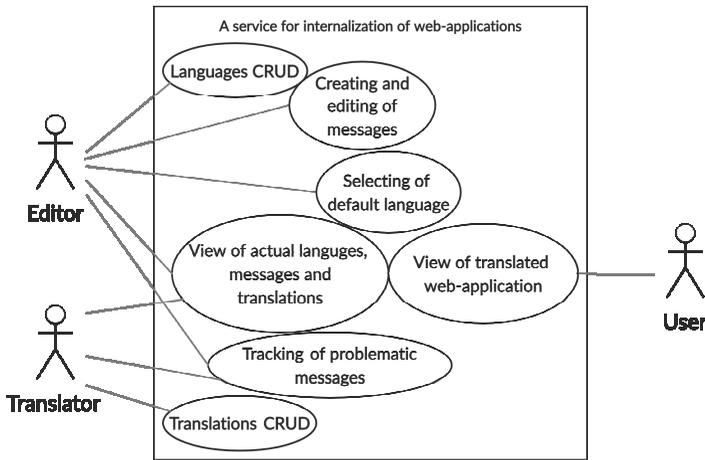


Fig. 1. Use case diagram

The selected functionality requirements are distributed among users in the following way:

Editor: determines what languages the web-application is translated in, prepares the messages that must be translated for the translator. Editor can also modify the translation and evaluate it. Actual messages, translations and languages are defined as correct data.

Translator: while working with the service translates the messages created by the editor.

User: receives the necessary translations, which are placed in the provided spaces of the interface of a web-application.

Domain modelling

A domain model was designed and is presented below (Fig. 2):

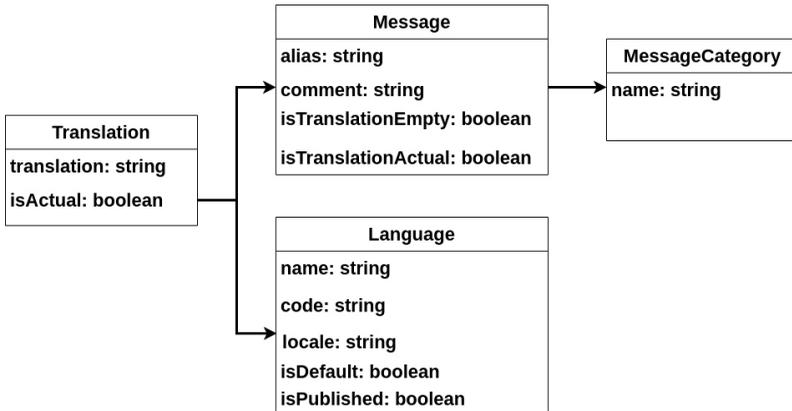


Fig. 2. Domain model

Language – languages that are in the service and in which a web-application is translated.

- name – the name of the language which is reflected in the language selection menu.
- code – two or three letters that represent the language during the intermediate processing of a user in the web-application.
- locale – denotes the regional customizations that can be applied to the language, for example date format and so on.
- isDefault – boolean field which can only be used in one language, which is the main language of the web application.
- isPublished – boolean field which shows, from the editor's perspective, that a translation is ready to be reflected in the user's web-application.

MessageCategory – categories of messages.

- name – name of the category.

Message – messages.

- alias – the content of the translation.
- comment – the description of the context, in which the message is used.
- isTranslationEmpty – boolean field, which shows the absence of a translation in at least one of the languages.
- isTranslationActual – boolean field indicating that there is at least one translation of a message that does not satisfy the editor.

Translation – translation of the message into one of the languages.

- translation – the content of the translation.
- isActual – boolean field, which indicates the correct translation (from the editor’s perspective).

Implementation tools

Framework Yii2 was chosen for implementation of server part. Yii2 is an object-oriented framework in PHP, which implements the MVC paradigm PostgreSQL was used as the DBMS.

Centrifugo was used for the implement of web-sockets. Centrifugo is a separate server that maintains constant connection with users of the service.

Design

The scheme of the architecture of the project, which was implemented with Yii2 framework, was designed and is presented on component diagram [6] below (Fig. 3):

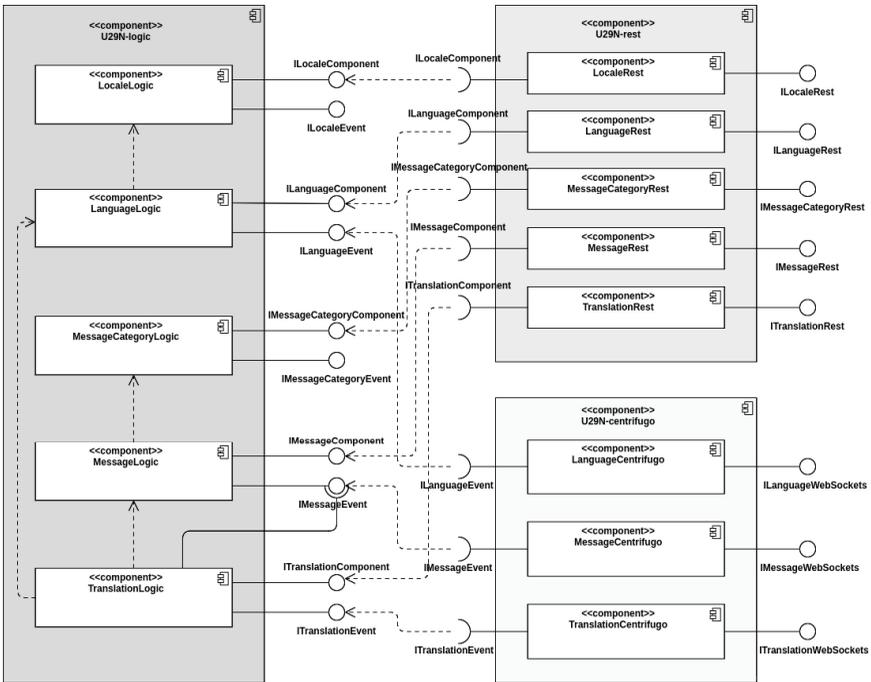


Fig. 3. Component diagram

As shown in Fig. 3, the functionality of the application is divided into three packages:

U29N-logic: contains business-logic and provides it to other packages.

U29N-rest: with the provided business-logic implements RESTful API – a programming interface of an application, implemented with help from the architecture style REST – Representational State Transfer. If a service fulfills the requirements of REST, it is considered RESTful.

U29N-centrifugo: implements the functionality for an asynchronous exchange of data.

Implemented API

The developed application programming interface implements the necessary functionality for entities:

1. **Languages:** CRUD was realized and the change of the main language of the application (default language).

2. **Categories of messages – CRUD.**

3. **Translations – CRUD.**

4. **Messages – CRUD** was realized, the automatic change of marks of problem was realized using the processing of events. It is responsible for presence of problems in a message – `isTranslationEmpty` and `isTranslationNotActual` in case of special events.

Implemented API is available to be used for example in front-end, CMSs, applications developed with other frameworks, etc. by calling necessary API method.

Conclusions

The result of the work is the server part of a service, presented as API and containing all functionality necessary for internationalization of web-applications.

REFERENCES

1. *Localization* and internationalization. URL: <https://www.w3.org/International/question/s/qa-i18n> (Accessed November 14th, 2020)
2. *webtranslateit* – Official site of the presented alternative *webtranslateit*. URL: <https://webtranslateit.com/en> (Accessed November 14th, 2020)
3. *lingohub* – Official site of the presented alternative *lingohub*. URL: <https://lingohub.com> (Accessed November 14th, 2020)
4. *transifex* – Official site of the presented alternative *transifex*. URL: <https://www.transifex.com> (Accessed November 14th, 2020)
5. *oneskyapp* – Official site of the presented alternative *oneskyapp*. URL: <http://www.oneskyapp.com> (Accessed November 14th, 2020)
6. *Fauler M.* UML distilled. 3rd ed. Saint Petersburg: Simvol-Plus, 2004. 192 p.

«IS-A»-отношение, как способ представления взаимосвязи обобщенных и специализированных понятий

Алексей Бабанов, Елена Квач

*Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

«IS-A»-отношения являются важным элементом человеческой умственной деятельности. И на бытовом уровне мы часто используем обобщенные и специализированные понятия: «*Эйнштейн А. есть некоторый ученый*», «*Ученый есть некоторый человек*». Операции обобщения и специализации являются важными мыслительными инструментами, поэтому без них не обойтись во многих областях науки. Для более полной характеристики рассматриваемого вопроса было изучено представление «IS-A»-отношений в логике, искусственном интеллекте, расширенной модели «Сущность – Связь» (Enhanced/Extended Entity-Relationship Model – EER-модель), модели «Объект – Роль» (Object-Role Model – OR-модель), модели «Сущность – Связь – Отображение» (Entity-Relationship-Mapping Model, ERM-модель), ER-модели нотации Баркера (ERB-модели), концепции кластеров Джона М. и Дианы К. Смит и в унифицированном языке моделирования (Unified Modeling Language – UML). Однако, в каждой из представленных областей знаний даны разные определения обобщенным и специализированным понятиям, различны и ограничения целостности, и способы проектирования.

«IS-A»-отношения в различных областях науки

Основополагающие знания об обобщении и ограничении понятий вводятся в логику. По мнению логиков [1, 2], каждый акт обобщения должен быть переходом от вида к некоторому ближайшему роду, а ограничения – от рода к некоторому ближайшему виду. Многие модели (сознательно или не сознательно) используют важный феномен, связанный с рассмотрением «IS-A»-отношений: деление понятий и основание деления. «Цель деления понятия состоит в том, чтобы выделить все возможные виды предметов каждый раз по некоторому определенному основанию» [1]. Заслугой логиков также является рассмотрение условий правильного деления (1 – деление по одному определенному осно-

ванию; 2 – полученные понятия должны быть попарно несовместимы; 3 – члены деления должны исчерпывать объем исходного понятия; 4 – никакой из членов деления не должен быть пустым классом; 5 – деление должно быть непрерывным) и видов деления понятий (дихотомическое деление и деление по видоизменению основания).

В искусственном интеллекте «IS-A»-отношения используются в семантических сетях и являются основным звеном в сети. Связь «IS-A» устанавливает иерархию типов в сети и используется для определения наиболее специфического семантического типа [3]. Семантические сети – одна из немногих областей знаний, где разделяют «IS-A»-отношения между классами (IS-A type of, смысловое значение «класс / класс») и между экземплярами и классами (IS-AN instance of, смысловое значение «объект / класс») [4]. Конечно, подобное деление есть и в логике: отношение включения одного множества в другое и отношение принадлежности элемента множеству. В семантических моделях данных отношение между классом и его объектами всегда предполагалось, однако это явление мало обсуждается специалистами БД.

Главное назначение специализаций в семантических моделях данных – возможность группирования признаков: обобщенные поднимаются на уровень суперкласса, специализированные опускаются до соответствующего подкласса. Если логики обращают внимание, прежде всего, на содержание (признак) понятий деления, то в моделировании данных учитываются преимущественно структурные и ограничительные особенности классов данных. В OR-модели, кроме этого, отмечается, что специализации позволяют делать опциональные роли суперкласса обязательными для подкласса и за счет этого вводить дополнительные ограничения целостности [5]. Взаимоисключающие и исчерпывающие правила для подтипов вводит ERB-модель [6]. Однако этим правилам удовлетворяют лишь полные непересекающиеся специализации. В остальных семантических моделях помимо этого типа специализаций используются также полные пересекающиеся, частичные непересекающиеся и частичные пересекающиеся. Что касается основания деления, то, например, в EER-модели лишь немногие авторы, определяя понятие специализации, уточняют, что подклассы выделяются на основе отличительных характеристик [7] или в соответствии с различными значениями общего атрибута [8]. Однако с практической точки зрения

такое основание характерно лишь для непересекающихся специализаций, что является недостаточным для проектирования всех возможностей предметной области. Лишь ERM-модель [9], на основе опыта логиков, вводит аналогичные понятия: деление, основание деления и виды деления.

Джон М. и Диана К. Смит [10] в своей концепции предлагают сводить иерархии обобщения к отдельным правильным делениям родительских понятий, правда, они их называют кластерами. Стоит отметить, что аналогично с логиками, используется одно основание деления (правильное деление) и непересекаемость подклассов одного кластера (взаимоисключающие родовые классы).

В UML используется термин обобщение – отношение между классом и одной или несколькими его вариациями. Обобщение объединяет классы по их общим свойствам [11]. В [12] дополнительно вводятся 2 правила: 1 – 100% определения суперкласса должно быть применено к подклассу и подкласс должен на 100% соответствовать своему суперклассу; 2 – все элементы множества подкласса должны быть элементами множества его суперкласса. Обратим внимание, что в соответствии со спецификой ООП, помимо атрибутов и отношений подтипы наследуют ещё и методы супертипа. Аналогично основанию деления в логике специалисты, изучающие UML, дополнительно вводят понятие дискриминатор [13] или набор обобщений [12, 13]. Также отмечается, что каждое обобщение должно производиться по одному аспекту (одно основание деления).

Таблица отображает возможные варианты определения «IS-A»-отношений в областях науки, в том числе и различия.

Что касается графической нотации отображения «IS-A»-отношений (рис. 1), то и здесь существуют различия в зависимости от области знаний. К сожалению, даже в непротиворечивой модели представления обобщенных понятий логиков не хватает графического представления с отображением основания деления понятий, что уж говорить о других моделях. Единственной семантической моделью, в графической нотации которой отображается основание деления, является ERM-модель, причина этому – становление модели на основе безупречных канонов логики. Отметим, что в UML можно указывать дискриминатор, но это не является обязательным (а по мнению авторов, должно быть).

«IS-A»-отношение как способ представления обобщенных понятий

Модель	Аналог «IS-A»-отношений	Определение	Ограничения и особенности
Логика	Операции обобщения и ограничения понятий	Обобщение некоторого понятия – операция образования из этого понятия некоторого нового с более широким объемом. Ограничение понятия – операция перехода от некоторого понятия к понятию с меньшим объемом [1].	Правильное деление понятий и основание деления понятий
Искусственный интеллект	«IS-A»-отношение	Это отношение, которое устанавливает иерархию типов в сети и используется для определения наиболее специфического семантического типа [3].	«IS-A»-отношения транзитивны, направленные и не образуют петель. Разделяют отношения между классами и между экземплярами и классами.
Расширенная модель «Сущность – Связь» (EER)	Специализация, отношения «суперкласс-подкласс»	- процесс выделения подклассов в суперклассе, который основан на группировке исключительных характеристик и связей подклассов в соответствии с различными значениями общего атрибута [8]; - процесс выявления различий между экземплярами типа сущностей [14]; - процесс определения набора подклассов для типа сущностей на основе отличительных характеристик [7]; - процесс концептуального уточнения суперкласса путем выделения подклассов [15];	Ограничение пересечения и участия.
Модель «Объект – Роль» (OR)		- процедура, в ходе которой вводятся подклассы более общих типов сущностей, чтобы объявить, что определенные роли зарегистрированы только для этих подклассов [5];	Специализации транзитивны и формируют направленный ациклический граф.

Модель	Аналог «IS-A»-отношений	Определение	Ограничения и особенности
		- граф подтипов, который может возникнуть нисходящим способом, специализируя тип объекта в подтипы [5].	Ограничение непересечения и исчерпывания
ER-модель в нотации Баркера (ERB)		Сущность может быть разделена на два или более взаимоисключающих подтипа, каждый из которых имеет общие атрибуты и/или отношения. [6].	Взаимоисключающее и исчерпывающее правило (полные непересекающиеся специализации).
Кластеры Джона М. и Дианы К. Смит	Обобщение	Абстракция, которая позволяет обобщенно представлять себе некоторый класс индивидуальных объектов как единый именованный объект [10].	Кластеры, одно основание деления и непересекаемость подклассов одного кластера.
Модель «Сущность – Связь – Отображение» (ERM)	Специализация, отношения «суперкласс-подкласс»	Система классов, построенная для ограничений понятия суперкласса, такая, что все понятия подклассов используют в качестве основания выделения этих подклассов одно и то же отображение (предметную функцию).	Основание выделения подкласса и основания деления специализации. Ограничение непересечения и участия.
Унифицированный язык моделирования (UML)	Обобщение, отношение наследования	- отношение между классом и одной или несколькими его вариациями [11]; - связь между сущностью общего характера и более специфичной сущностью [16]; - отношение, при котором подкласс наследует все методы и поля суперкласса и может переопределять наследуемые методы [13]; - процесс, связанный с идентификацией общности между понятиями и определением суперкласса и связанных с ним подклассов [12].	Дискриминатор или набор обобщения. Обобщение транзитивно, не может быть циклов и связь обобщения асимметрична. Правило 100% и правило «IS-A».

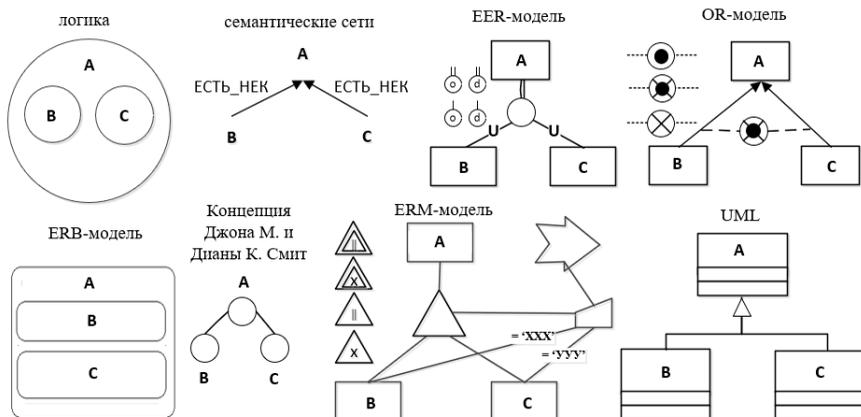


Рис. 1. Графическая нотация (A – суперкласс, B,C – подклассы)

Заключение

Перечисленные области науки внесли серьезный вклад в изучение «IS-A»-отношений, однако можно констатировать различие взглядов всех исследователей на представление обобщенных и специализированных понятий. Даже в моделировании данных нет единого мнения на «IS-A»-структуры, что уж говорить об отличии этих представлений от логических основ. Проанализировав особенности «IS-A»-отношений в различных областях знаний можно сделать вывод о необходимости более строгой формализации обобщений и введения для этого нового базового понятия. Авторами предложено новое понятие «IS-THE»-отношения (отношения «ЕСТЬ-ЭТОТ») в семантической ERM-модели, ориентированного на безупречные логические каноны при сохранении той функциональности этих структур, которая сложилась в моделировании данных [17].

ЛИТЕРАТУРА

1. Войшвилло Е.К., Дегтярев М.Г. Логика как часть теории познания и научной методологии (фундаментальный курс): в 2 кн. М.: Наука, 1994.
2. Бачаров В.А., Маркин В.И. Основы логики. М.: Инфра-М, 1998. 297 с.
3. U.S. National Library of Medicine, The Unified Medical Language System, UMLS Reference Manual. URL: <https://www.ncbi.nlm.nih.gov/books/NBK9676/>
4. Bundy A. Catalogue of Artificial Intelligence Tools. Berlin; Heidelberg; New York; Tokyo: Springer-Verlag, 1984. 150 p.

5. *Halpin T., Morgan T.* Information Modeling and Relational Databases, 2 edition. Morgan Kaufman, 2008. 943 p.
6. *Barker R.* CASE Method: Entity Relationship Modelling. Addison-Wesley, 1990. 240 p.
7. *Elmasi R., Navahe S.B.* Fundamentals of Database Systems, 6th edition. Addison-Wesley, 2011. 1201 p.
8. *Teorey T., Yang D., Fry J.A.* Logical Design Methodology for Relational Databases Using the Extended Entity-Relationship Model // Computing Surveys. 1986. V. 18. No. 2. P. 197–222.
9. *Бабанов А.М.* Семантическая методика проектирования БД и её перспективы, открывающиеся с применением ERM-модели данных // Вестник Том. гос. ун-та. Управление, вычислительная техника и информатика. 2011. № 3(16). С. 58–66
10. *Smith J., Smith D.* Database Abstractions: Aggregation and Generalization // ACM Transactions on Database Systems. 1977. V. 2. No. 2. P. 105 – 133.
11. *Blaha M., Rumbaugh J.* Object-Oriented Modeling and Design with UML. 2nd edition. Pearson, 2004. 496 p.
12. *Larman C.* Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and Iterative Development. 3rd edition. Prentice Hall, 2004. 736 p.
13. *Fowler M., Scott K.* UML Distilled: A Brief Guide to the Standard Object Modeling Language. Addison-Wesley, 1999. 224 p.
14. *Connolly T., Begg C.* Database Systems: A Practical Approach to Design, Implementation, and Management. 4th edition. Addison-Wesley, 2010. 1400 p.
15. *Ferragine V.E., Doorn J.H., Rivero L.C.* Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends. 2009. 1124 p.
16. *Буч Г., Рамбо Д., Якобсон А.* Язык UML. Руководство пользователя. ДМК Пресс, 2007. 496 с.
17. *Бабанов А.М., Квач Е.С.* «IS-THE»-отношения в семантических моделях данных: основные понятия и разновидности // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2016. № 1(34). С. 69–78.

Проектирование базовой архитектуры модуля «Расписание» системы Alterum Med

Людмила Демиденко

*Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

В настоящее время актуальность автоматизации деятельности человека сложно переоценить. Тем более, если это связано с деятельностью медицинского учреждения, где на счету каждая минута. Для упрощения работы с большой базой данных о пациентах, а также для ускорения работы медицинского персонала была создана программа Alterum Med, которая объединяет в себе модули по работе с разными направлениями деятельности медицинского учреждения: регистратура, прием пациента, складской учет, бухгалтерия, пост охраны. Данное программное обеспечение позволяет автоматизировать практически все сферы деятельности больницы в одном продукте.

Одним из самых часто используемых модулей в данной программе является модуль расписания, так как он предоставляет сразу несколько важных функций: просмотр врачами приемов, установление рабочего времени врачам, запись пациента на прием. На данный момент существующий модуль расписания устарел и было принято решение полностью переписать данный модуль с использованием новейших технологий.

Исходя из этого, целью данной работы стало проектирование базовой архитектуры модуля «Расписание» системы Alterum Med.

Для достижения цели необходимо решить следующие задачи:

1. Провести анализ бизнес-процессов предприятия, касающихся модуля расписания.
2. Определить требования к системе.
3. Выбрать платформы реализации модуля приложения.
4. Определить и спроектировать общую структуру модуля.

Анализ бизнес-процесса

Основной бизнес-процесс, который автоматизирует модуль расписания – это запись на прием к врачу. Пациент приходит или звонит в регистратуру. Регистратор берет трубку, узнает у пациента информацию по желаемому времени и специальности врача. В ходе обсуждения они подбирают максимально удобное для пациента время и специалиста. Данные по записи на прием вносятся в программу, время приема считается забронированным, никто больше не может записать другого пациента на это время к данному специалисту.

В том случае, если подходящего времени и специалиста не нашлось, пациент может либо отказаться от записи, либо остаться в резерве. Если по каким-то причинам у интересующего его врача освобождается время, регистратор перезванивает пациенту и предлагает его записать. Диаграмма такого процесса в нотации BPMN представлена на рис. 1 и 2.

Кроме этого, модуль расписания автоматизирует бизнес-процесс назначения врачам рабочего времени, а также позволяет врачам просматривать свое расписание. Их диаграммы здесь не представлены, поскольку эти процессы не являются сложными.

Роли

Как видно из описания основного бизнес-процесса, данный модуль взаимодействует с тремя типами пользователей: регистратор, врач и менеджер. Опишем каждую из ролей подробнее.

Пользователь – лицо, прошедшее авторизацию в системе. Абстрактный тип, остальные роли наследуются от него.

Регистратор – работник медицинского учреждения. Одной из его основных задач является ведение регистрации больных, обратившихся за медицинской помощью. В рамках расписания он может просматривать расписания врачей, создавать, удалять и редактировать приемы, вносить пациентов в резерв.

Менеджер – в рамках модуля расписания основная задача менеджера – создавать рабочие часы врачам.

Врач – может просматривать свое расписание.

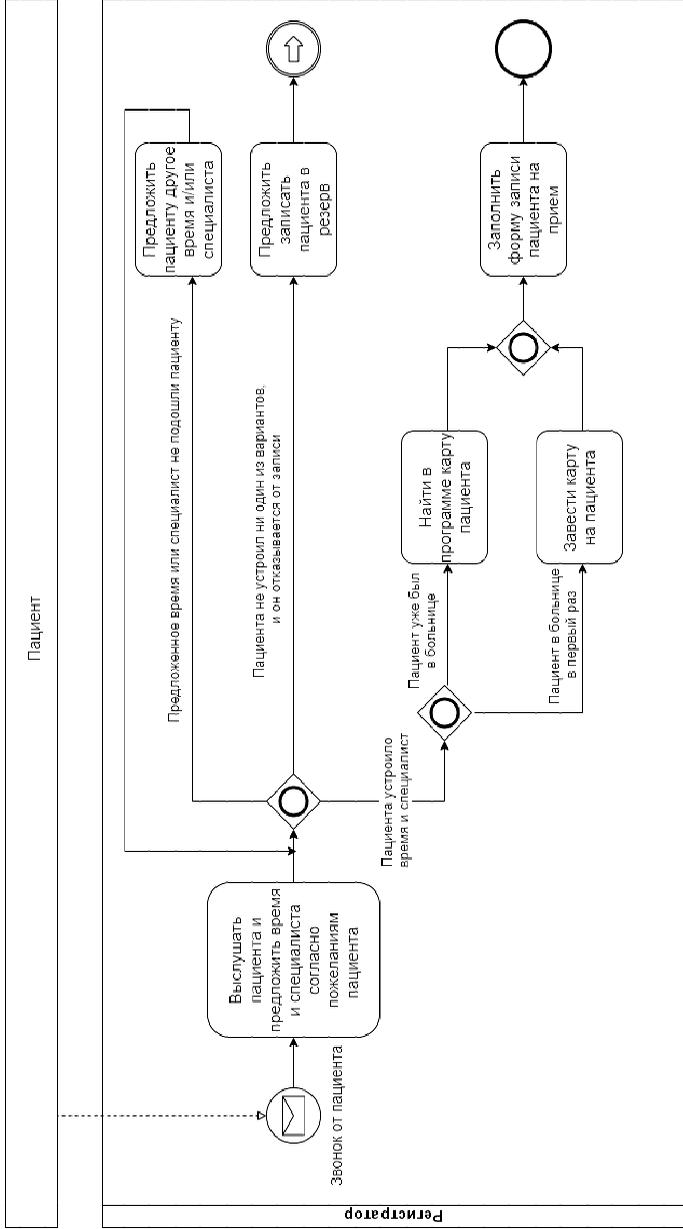


Рис. 1. Модель бизнес-процесса «Запись пациента на прием» (фрагмент)

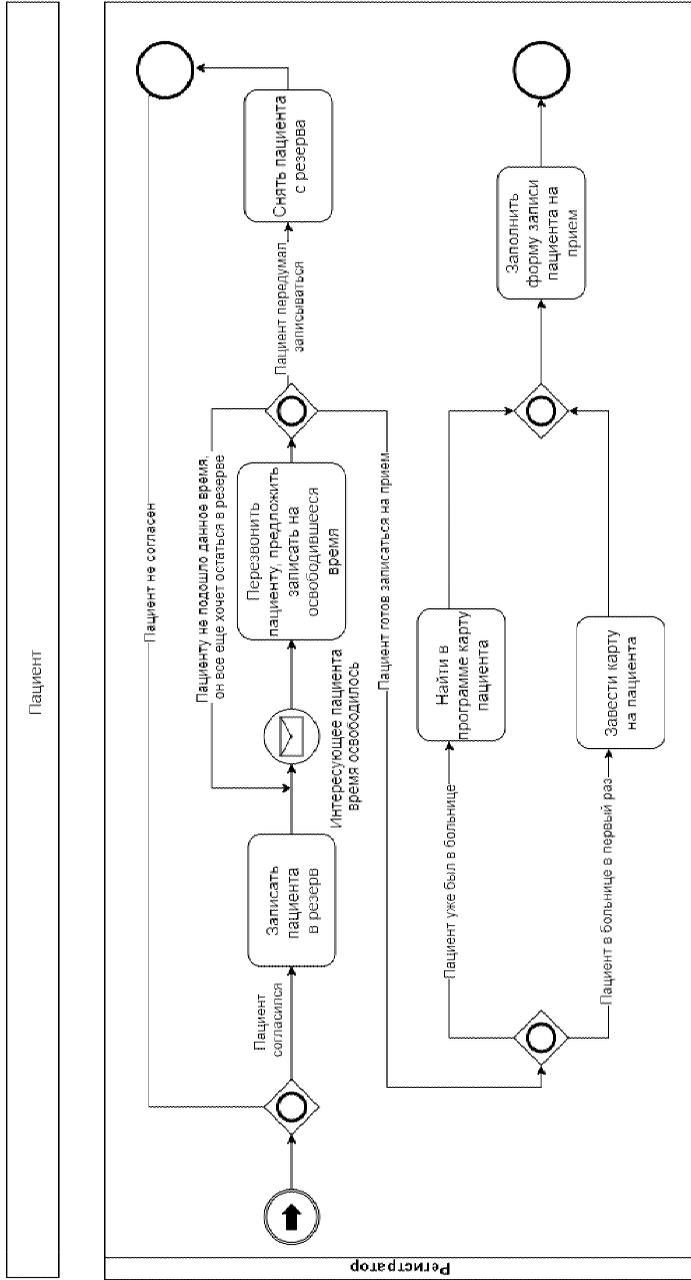


Рис. 2. Модель бизнес-процесса «Запись пациента на прием» (фрагмент)

Варианты использования

Для каждой роли пользователя системы были определены варианты использования, диаграмма которых представлена на рис. 3.

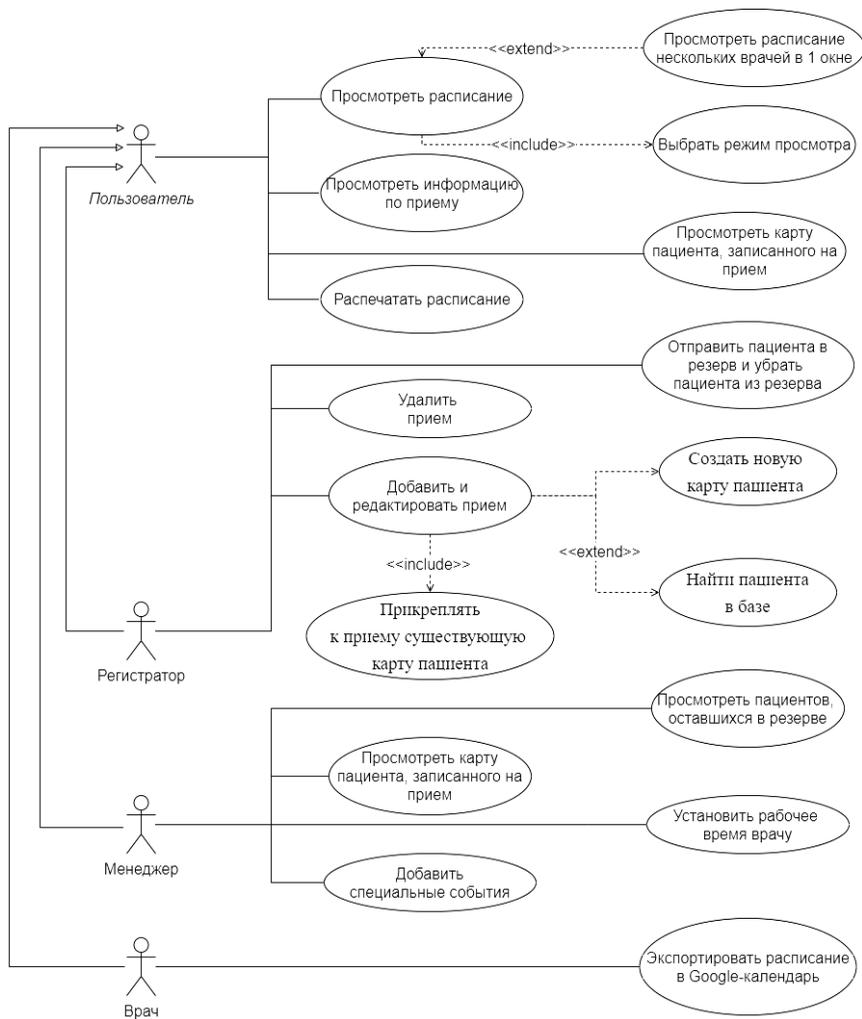


Рис. 3. Варианты использования модуля «Расписание»

Выбор платформы реализации

Выбор большинства платформ был продиктован существующей системой Alterum Med, так как нам необходимо, чтобы разрабатываемая система была реализована как модуль этой программы. Таким образом были выбраны следующие платформы:

- C# [1] в качестве языка программирования.
- Windows Presentation Foundation (WPF) [2] в качестве платформы для реализации интерфейса пользователя.
- Windows Communication Foundation (WCF) [3] как фреймворк для построения распределенных (сервис-ориентированных) систем.
- EntityFramework (EF) [4] для обеспечения доступа к данным.
- Microsoft SQL Server [5] в качестве системы управления базами данных (СУБД).

Основные классы

Диаграмма классов модели модуля расписание представлена на рис. 4.

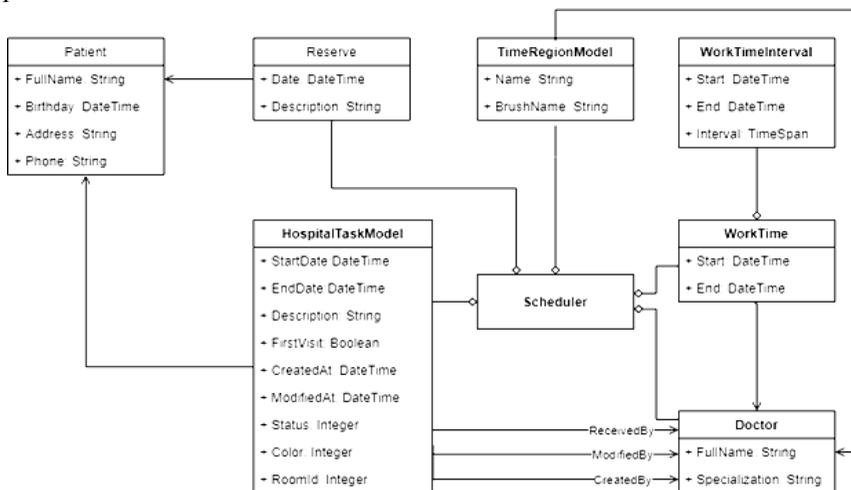


Рис. 4. Основные классы модели приложения Alterum Med

Основным классом, содержащим в себе всю информацию по расписанию, является класс Scheduler. Он содержит коллекции врачей (Doctor), их рабочего времени (WorkTime), приемов (HospitalTaskModel), специальных событий (TimeRegionModel) и резервов (Reserve). Класс WorkTime описывает рабочее время врача. В нем описаны поля Start и End, объекты класса

`DateTime`. Данный класс представляет текущее время, обычно выраженное как дата и время суток. Поля `Start` и `End` означают начало и конец календарного дня, а также коллекция экземпляров класса `WorkTimeInterval`, который описывает конкретные временные интервалы в рамках одного дня (поля `Start` и `End`) и задает длительность приемов в этом временном интервале (поле `Interval`). Поле `Interval` представлено объектом класса `TimeSpan`. Объект этого класса представляет интервал времени, измеряемый как положительное или отрицательное число дней, часов, минут, секунд и долей секунды. Проектирование выполнено именно таким образом, так как у одного врача в рамках одного дня может быть несколько рабочих временных интервалов. Например, первую половину дня врач проводит приемы, у которых одна длительность, а вторую половину – проводит процедуры с другой длительностью.

Класс `HospitaTaskModel` описывает прием врача. В нем записана служебная информация, такая как время создания и время редактирования записи (`CreatedAt` и `ModifiedAt`), пользователь, создавший и отредактировавший запись (`CreatedBy` и `ModifiedBy`), идентификатор записи время начала и конца приема (`StartDate` и `EndDate`), сам пациент (`Patient`), записанный на прием, врач (`Doctor`) и другие.

Заключение

Проведено исследование автоматизируемых бизнес-процессов, выявлены актеры, определены функциональные и нефункциональные требования, расписаны варианты использования и их сценарии для модуля расписания системы `Alterum Med`. Также выполнен анализ архитектуры существующего приложения `Alterum Med`, определено место разрабатываемого модуля в системе, спроектированы основные классы модуля. В настоящее время ведутся работы по реализации модуля и его интеграции в систему.

ЛИТЕРАТУРА

1. *Рихтер Д.* CLR via C#. Программирование на платформе Microsoft.NET Framework 4.5 на языке C#: пер. с англ. Е. Матвеева. М.: Питер, 2019. 896 с.
2. *Петцольд Ч.* Microsoft Windows Presentation Foundation. М.: Питер, 2012. 994 с.
3. *WCF – Краткое руководство / Microsoft.* URL: <https://coderlessons.com/tutorials/microsoft-technologies/uznaite-wcf/wcf-kratkoe>
4. *Документация по Entity Framework / Microsoft.* URL: <https://docs.microsoft.com/ru-ru/ef/>
5. *Ицик Бен-Ган* Microsoft SQL Server 2012. Основы T-SQL. М.: Русская редакция; БХВ-Петербург, 2013. 251 с.

Программная реализация заданий по программированию с многовариантными решениями

Игорь Жуков, Юрий Костюк

*Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

Несмотря на широкое распространение элементов электронного обучения в российской системе образования, остаются востребованными новые решения в области автоматизированного контроля знаний. В частности, в области программирования актуальной задачей является предоставление преподавателям инструментов автоматизированной проверки знаний и практических навыков обучающихся, исключающих ручную проверку программ и автоматическое тестирование методом чёрного ящика. Для решения этой задачи авторами предложена модель представления заданий и алгоритмы контроля [1]. В статье описана программная реализация переработанной и расширенной модели [1], а также алгоритмов контроля.

Краткое описание модели

Модель основана на конструктивно-выборочном методе, при котором обучающийся составляет свой ответ из элементов, являющихся допустимыми смысловыми единицами. В дальнейшем такие единицы будем называть компонентами. Простейшие программы могут быть разбиты на десятки компонент, которые в различных комбинациях образуют верные решения. Модель допускает, что из предложенных компонентов можно составить несколько верных решений, причём в каждом случае некоторые компоненты не входят в верное решение, кроме того, могут быть добавлены компоненты, не входящие ни в одно решение.

Вариант решения может быть верным (правильным) или неверным (неправильным). Все верные варианты решения должны быть описаны конечным множеством шаблонов. Каждый шаблон состоит из элементов. С точки зрения программирования, элемент – это какое-либо действие, состоящее из одной или нескольких строк программного кода. Таким образом, элемент – это последовательность компонентов. В модели выделено три типа элементов: один компонент, один из компонентов, перестановка компонентов. Элемент «один из компонентов» пред-

ставляет собой список компонентов, каждый из которых удовлетворяет верному решению, но любое верное решение содержит только один компонент из этого списка. Элемент «перестановка компонентов» представляет собой список компонентов, любая перестановка которых удовлетворяет верному решению. Каждый элемент может быть отмечен как рубежный или как допускающий отсутствие (опциональный). Отсутствие не помеченного элемента добавляет в итоговую оценку 25% его веса. Если в ответе отсутствует элемент, отмеченный как «допускается отсутствие», при расчёте итогового балла добавляется 50% веса пропущенного элемента.

Приведём описание языка в расширенной форме Бэкуса – Наура (РБНФ), жирным шрифтом выделены метасимволы, относящиеся к РБНФ:

Задание = Шаблон {"|"} Шаблон}

Шаблон = "{" Элемент {";" Элемент} "}"

Элемент = "["?" ((ОдинКомпонент | ОдинИзКомпонентов) "*" | ПерестановкаКомпонентов) "]" ?

ОдинКомпонент = ИдентификаторКомпонента

ЦифрыБезНуля = "1" | "2" | "3" | "4" | "5" | "6" | "7" | "8" | "9"

Цифры = "0" | ЦифрыБезНуля

ИдентификаторКомпонента = ЦифрыБезНуля {Цифры}

ОдинИзКомпонентов = ОдинКомпонент "|" ОдинКомпонент {"|"}
ОдинКомпонент}

ПерестановкаКомпонентов = "(" ОдинКомпонент ";" ОдинКомпонент {";" ОдинКомпонент} ")".

Элементы, содержащие символ "*", являются рубежными. Элементы, заключённые в квадратные скобки, являются опциональными.

Обучающийся составляет решение, используя предложенные компоненты. Этот вариант решения оценивается в баллах. Максимальная оценка принята равной единице. Итоговый балл складывается из оценок за каждый элемент. Если итоговый балл попадает в интервал (0, 1), то решение считается частично правильным. По умолчанию, все элементы имеют одинаковый вес.

Для автоматизированной оценки варианта решения авторами статьи предложен следующий алгоритм.

Соответствие варианта решения (входной последовательности) одному шаблону оценивает автомат, который мы будем называть экспертом. Эксперт считывает по два значения из входной последовательности, если текущий элемент не является «перестановкой компонентов»

или не помечен как рубежный. Для рубежного элемента эксперт считывает только одно значение. Для элемента «перестановка компонентов» эксперт просматривает все значения до следующего рубежного элемента. Эксперт оценивает соответствие считанной последовательности и текущего элемента, а затем выставляет оценку. Верных вариантов решения, как правило, больше одного. Поэтому описание задания содержит более одного шаблона и проверяется несколькими экспертами. На данный момент разработано два алгоритма проверки: наивысшая оценка и выбор эксперта. В алгоритме «наивысшая оценка» ответ обучающегося оценивается всеми экспертами, финальной оценкой считается наибольшая из всех оценок. В алгоритме «выбор эксперта» оценивает компетентность каждого из экспертов относительно конкретно предоставленного ответа обучающегося. Оценка самого компетентного эксперта будет финальной.

Программная реализация

В терминологии Python пакетом называется совокупность модулей, то есть файлов, содержащих программный код на языке Python. В более широком смысле пакет в Python эквивалентен библиотеке классов. Программная реализация модели и алгоритмов контроля выполнена в виде пакета на языке программирования Python. В пакет входят следующие классы: AssessmentResult, Assessor, PatternElement, SingleElement, OneOfElement, PermutationElement, ElementFlag, ElementType, ElementFactory, Expert, ExpertInfo, ExpertOpinion, Task, TaskCreator, TaskCompiler, TaskDecompiler.

Для программной реализации требуется построить компилятор с языка описания заданий в массивы массивов словарей в формате JSON. Работу компилятора принято разделять на две части: анализ (front end) и синтез (back end) [2]. В фазу анализа входит лексический и синтаксический анализ шаблона. В конце фазы анализа шаблон переводится в промежуточное представление. Затем на фазе синтеза промежуточное представление переводится в массив массивов словарей.

Одним из инструментов создания компиляторов является компилятор компиляторов – программа, воспринимающая синтаксическое или семантическое описание языка программирования и генерирующая компилятор для этого языка [3]. Для нашей задачи использование компилятора компиляторов сводится к генерации лексического и синтаксического анализатора. Эта задача была решена с помощью библиотеки

lark (пакета lark-parser). Ключевой для нас особенностью lark является генерация синтаксического анализатора, использующего алгоритм LALR (Look-Ahead Left-to-Right) [4], описанный в [2]. Синтаксический анализатор, генерируемый lark, создаёт аннотированное дерево разбора [4], которое является промежуточным представлением шаблона.

Пример промежуточного представления для шаблона $\{(1;4);10;3*;7|8;11\}$ приведён на рис. 1.

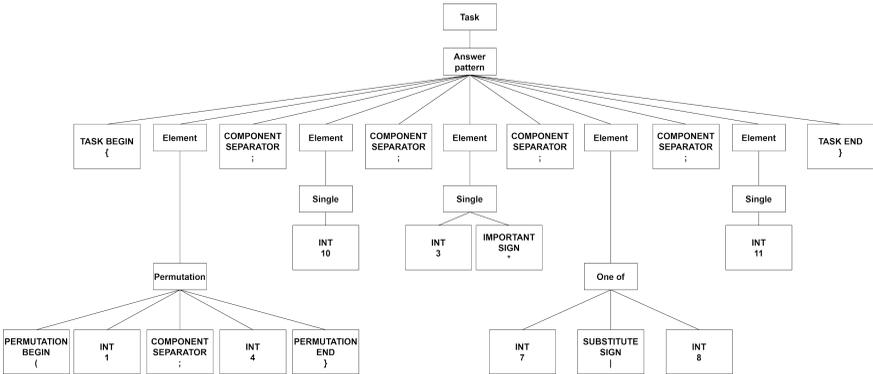


Рис. 1. Пример аннотированного дерева разбора

Для реализации фазы синтеза необходимо разработать алгоритм преобразования аннотированного дерева разбора в массив массивов словарей. В качестве базового алгоритма обхода дерева lark предоставляет класс `Transformer`, построенный по шаблону проектирования «Посетитель», который описывает операцию, выполняемую с каждым объектом структуры. Шаблон «Посетитель» позволяет определить новую операцию над структурой, не изменяя классы объектов, входящих в структуру [5]. Класс `Transformer` описывает только способ обхода самого дерева, поэтому авторами реализован дочерний класс `TreeTransformer`, в котором описан алгоритм преобразования для каждого типа вершин. Сам компилятор реализован в классе `TaskCompiler`.

Помимо компилятора для библиотеки разработан декомпилятор, который реализован в классе `TaskDecompiler`. Задачей декомпилятора является преобразование массива массивов словарей в запись на языке шаблонов. Декомпилятор имеет два режима работы: обычный и педантичный. В педантичном режиме перевод осуществляется с сохранением всех флагов элементов. В обычном режиме часть флагов опускается,

так как ограничения для элемента, стоящего после элемента с типом «перестановка», однозначны.

Для обеспечения удобной работы с результатами компиляции в библиотеке предусмотрен абстрактный базовый класс `PatternElement`, который описывает структуру элемента шаблона и методы, упрощающие работу алгоритма проверки, такие, как проверка соответствия текущего элемента заданной последовательности и индекс элемента в последовательности. От `PatternElement` наследуются три класса для каждого типа элементов `SingleElement`, `PermutationElement` и `OneOfElement`. Для представления заданий в библиотеке реализован класс `Task`, который содержит запись на языке шаблонов, массив массивов словарей и множество всех компонентов, входящих в задание. Перевод шаблона или массива массивов словарей в объект класса `Task` реализован в методе `create_task` класса `TaskCreator`.

Эксперт, проверяющий соответствие ответа шаблону, представлен классом `Expert`. Создание объектов класса `Expert` инкапсулировано в классе `ExpertCreator`. Этот класс имеет два метода: `create_expert` для создания одного эксперта по одному из шаблонов задания и `create_experts` для создания множества экспертов для одного задания. Основным методом класса `Expert` является `do_assessment`, который принимает ответ обучающегося в виде строки, содержащей номера компонентов, разделённых запятой, и возвращает объект класса `ExpertOpinion`. В классе `ExpertOpinion` содержатся поля для исчерпывающего описания оценки представленного ответа экспертом: общую оценку в виде обыкновенной дроби, количество компонентов ответа, не входящих в шаблон, количество пропущенных элементов каждого из типов.

Проверка заданий инкапсулирована в классе `Assessor`. Основным методом класса является `do_assessment`, который принимает на вход объект класса `Task`, ответ обучающегося в виде строки, содержащей номера компонентов, разделённых запятой, и флаг, указывающий алгоритм проверки. Метод возвращает объект класса `AssessmentResult`. Класс `AssessmentResult` описывает результаты проверки: финальную оценку в виде десятичной дроби, значение которой находится в интервале $[0, 1]$, и список объектов класса `ExpertOpinion`. На рис. 2 приведён пример оценки тремя экспертами ответа «4;5;3;7;9» для задания с шаблоном $\{\{1\};2;3^*\}\{(1;4);5;6;3^*;7|8;9\}\{(1;4);10;3^*;7|8;11\}$. Наибольшую оценку ответ получил от второго эксперта, что соответствует представленному решению.

```

Score: 0.25
=====
Expert 1
  Score: 0
  Extra elements: 2
  Missing elements: 1
  Missing optional elements: 1
  Missing important elements: 0
=====
Expert 2
  Score: 1/4
  Extra elements: 0
  Missing elements: 4
  Missing optional elements: 0
  Missing important elements: 1
=====
Expert 3
  Score: 0
  Extra elements: 1
  Missing elements: 3
  Missing optional elements: 0
  Missing important elements: 2
=====

```

Рис. 2. Пример результатов проверки

Заключение

Программная реализация модели представления многовариантных заданий и алгоритмов контроля позволяет создавать инструменты контроля знаний обучающихся в области программирования. Эти инструменты могут применяться, например, при контроле усвоения важных базовых алгоритмов по содержанию созданных обучающимся программ без использования автоматического тестирования методом чёрного ящика.

ЛИТЕРАТУРА

1. Жуков И.А., Костюк Ю.Л. Модель представления заданий с многовариантными ответами для автоматизированного тестирования // Информационные технологии и математическое моделирование (ИТММ-2019): материалы XVIII Междунар. конф. им. А.Ф. Терпугова, 26–30 июня 2019 г. Ч. 1. Томск: Изд-во НТЛ, 2019. С. 31–36.
2. Ахо А., Лам М., Сети Р., Ульман Д. Компиляторы. Принципы, технологии и инструментарий. 2-е изд.: пер. с англ. М.: ООО «И.Д. Вильямс», 2008. 1184 с.
3. Толковый словарь по вычислительным системам / под ред. В. Иллинуорта и др.: пер. с англ. А.К. Белоцкого и др. под ред. Е.К. Масловского. М.: Машиностроение, 1990. 560 с.
4. lark-parser/lark // GitHub [сайт]. URL: <https://github.com/lark-parser/lark> (дата обращения: 20 октября 2020 г.).
5. Гамма Э., Хелм Р., Джонсон Р., Влиссидес Дж. Приемы объектно-ориентированного проектирования. Паттерны проектирования. СПб.: Питер, 2015. 368 с.

О представлении прогресса проекта по разработке программного обеспечения в форме динамической байесовской сети

Денис Змеев, Лидия Иванова, Руфина Рафикова

*Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

Представлен способ описания прогресса проекта по разработке программного обеспечения с использованием динамической Байесовской сети на основе Essence Kernel [1]. Несмотря на существование работ, которые рассматривают применение различных статистических подходов к управлению проектами [2, 3], большинство из них используют только высокоуровневое описание проектного управления, не погружаясь в детали и особенности, критические для отраслей, в которых протекает проект. При этом проекты по разработке программного обеспечения в частности и программная инженерия в целом обладают рядом специфических особенностей, которые делают общие приемы и методы малоприменимыми и неэффективными, что наглядно показано в [4]. Фактически для того, чтобы применять к проектам по разработке программного обеспечения прикладные математические подходы, необходимо получить иной, не общий, способ их представления и фиксации.

Alpha Essence

ALPNa (Abstract Level Project Health – абстрактный уровень здоровья проекта) позволяет оценивать прогресс, который проектная команда достигла в процессе выполнения проекта по разработке программного обеспечения, что достигается за счет нескольких особенностей, которые отличительны для ALPNa-карточек. Рассмотрим их на примере (рис. 1).

Согласно авторам Essence Kernel, в любом проекте по разработке программного обеспечения ровно 7 ALPNa: стейкхолдеры, возможности, требования, программная система, способ работы, команда и работа. Для каждой из этих основных ALPNa определена собственная последовательность состояний. Каждое состояние определяется через набор утверждений, которые либо истинны, либо ложны для текущего проекта. На рис. 1 представлен пример состояния «Узнаны» для ALPNa «Стейкхолдеры» с 3 утверждениями.

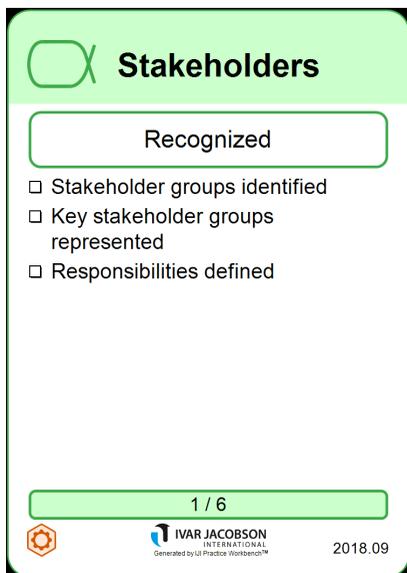


Рис. 1. ALPНа-карта Stakeholders «Recognized» на языке Essence

Применяя ALPНа-карточки, проектная команда может оценивать свое текущее состояние, используя простые логические высказывания и оценивая значения их истинности для каждого момента времени. Для основных 7 ALPНа существует всего 204 различных утверждения, что фактически позволяет представлять состояние проекта битовой строкой из 204 элементов, где каждая строка определяется в зависимости от моментов времени. В результате, такое представление проекта позволяет описывать его прогресс, с учетом специфики, накладываемой программной инженерией, и в результате применять различные прикладные методы для обработки. Существует подход, который называется ALPНа-покером [5], позволяющий команде проекта регулярно оценивать прогресс и планировать следующую итерацию на основе выполненных или невыполненных утверждений, представленных на ALPНа-карточках.

Представление Alpha-покера в форме динамической байесовской сети

Фактически, получив состояние проекта в форме битовой строки в зависимости от измеряемых моментов времени, авторами был выбран прикладной метод анализа при помощи динамической байесовской сети. С одной стороны, такой подход позволяет учитывать зависимость состояния проекта от времени. С другой стороны, применять дополнительные вычислительные алгоритмы для анализа состояния проекта и его прогресса. В рамках данной работы рассмотрим только критические особенности подобного представления.

Первоначально для представления прогресса проекта через ALPНа-покер и динамическую байесовскую сеть необходимо зафиксировать способ представления отдельных утверждений, используемых на ALPНа-карточках. В качестве результата получилась конструкция, представленная на рис. 2.

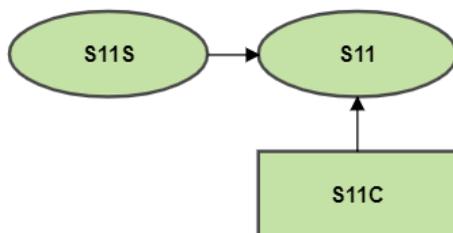


Рис. 2. Представление утверждения № 1 «Stakeholders groups identified» ALPНа -карты Stakeholders «Recognized»

Поскольку динамическая байесовская сеть относится к скрытым цепям Маркова, то было необходимо зафиксировать наблюдаемые и скрытые узлы. В качестве наблюдаемого узла выступает узел S11C, который является представлением утверждения «Stakeholders groups identified» состояния «Recognized» для ALPНа «Stakeholders». Значение истинности S11C мы узнаём напрямую у команды, фиксируя её мнение об этом утверждении для своего проекта. Рассмотрим узел S11, он является скрытым узлом и описывает состояния проекта на самом деле (то есть подлинное значение истинности этого утверждения). Знание значения узла S11 позволяет оценить, с какой вероятностью команда совершает ошибки 1-го рода (команда считает, что это утверждение истинно, но на самом деле это не так, таким образом, команда из-за чрез-

мерной уверенности создает себе риск) или 2-го рода (команда считает, что утверждение ложно, хотя на самом деле это не так, таким образом, команда попадает в паралич аналитики, пытаясь зафиксировать несуществующую информацию для проекта). Так как узел S11 является скрытым, то его значение мы можем только оценить на основании состояний S11C и S11S. Узел же S11S выделен для того, чтобы фиксировать динамический аспект байесовской сети. Во время анализа первого наблюдаемого момента времени в проекте узел S11S будет содержать и определять в себе изначальную функцию распределения вероятности для утверждения S11. Если же состояние проекта оценивается уже не первый раз, то узел S11S будет позволять учитывать влияние предыдущего значения узла S11 на последующее в моментах времени.

Используя описанный выше прием, можно представлять динамическую байесовскую сеть для отдельных состояний ALPНа (пример представлен на рис. 3).

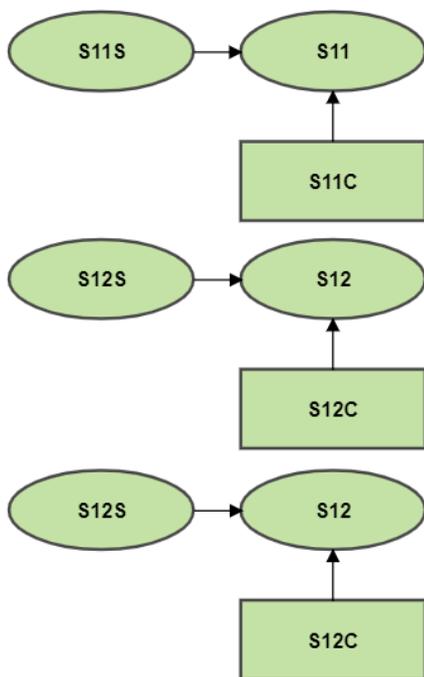


Рис. 3. ALPНа -карта Stakeholders «Recognized» в виде графа

Согласно логике Essence, различные состояния различных ALPНа-карт связаны последовательно через элемент языка, который называется пространством действий. С точки зрения семантики это означает, что утверждения из более ранних состояний ALPНа являются условиями для последующих. Это позволяет построить цепочку последовательных влияний различных утверждений друг на друга (пример на рис. 4).

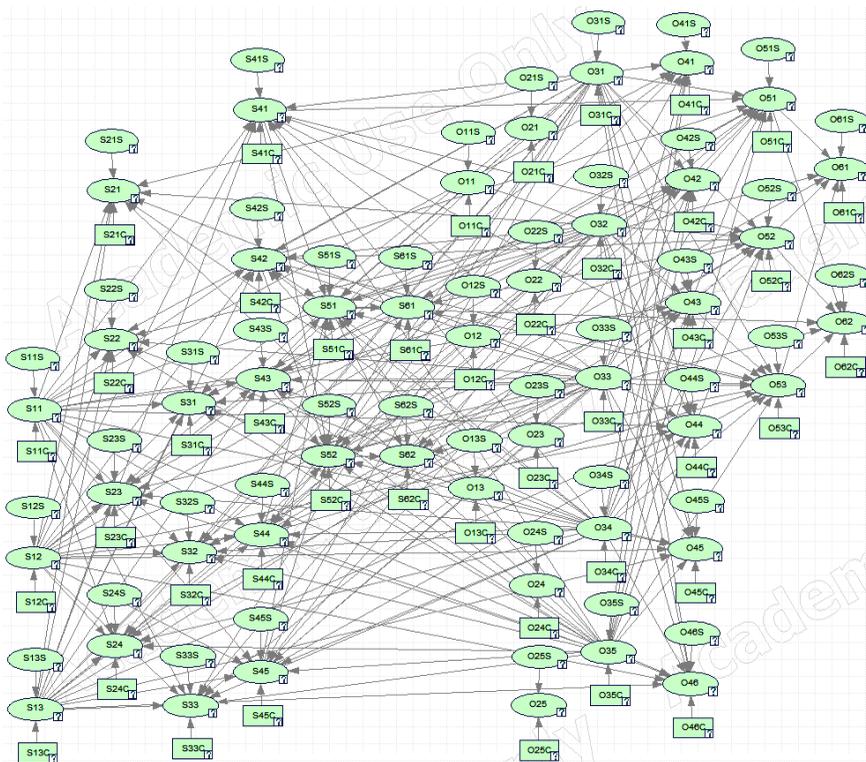


Рис. 4. ALPНа -карты Stakeholders и Opportunity в виде графа со связями зависимостей

После разработки графа для всех ALPНа-карт, следующий шаг – определение значений вероятностей всех переходов в динамической байесовской сети. Для каждой скрытой вершины вероятности всех переходов в неё трансформируются в вероятностный вектор, каждый элемент которого фиксирует вероятность целевого узла принять истинное зна-

чение при условии различных значений истинности вершин, влияющих на целевую. Сформировав динамическую байесовскую сеть для всех ALPНа со всеми состояниями, утверждениями и связями, которые непосредственно следуют из Essence Kernel, была получена следующая сеть (рис. 5):

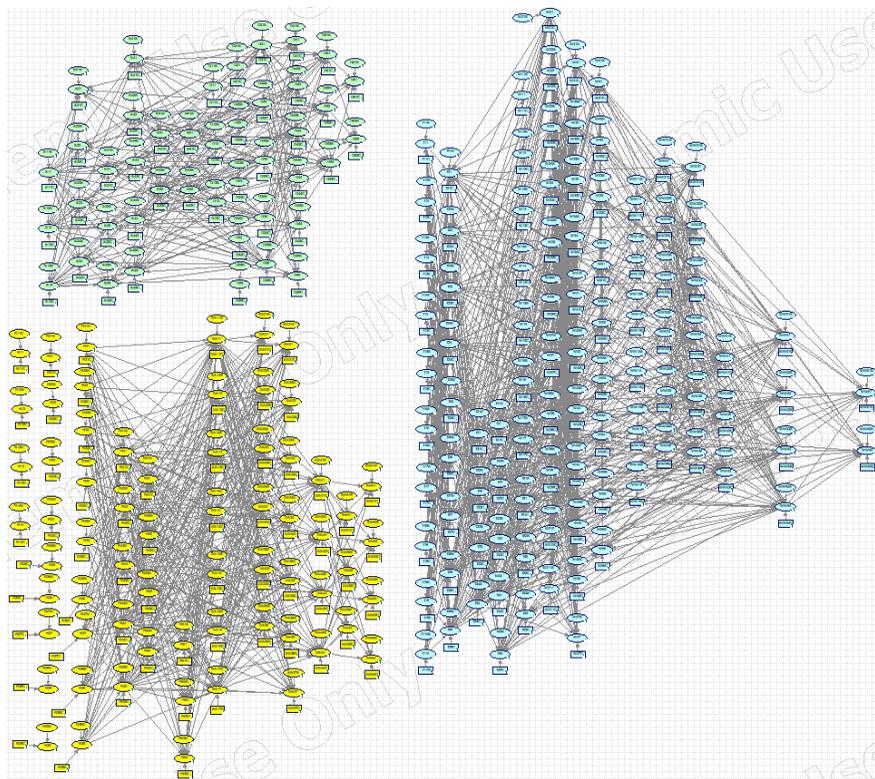


Рис. 5. Представление всех утверждений представленных на ALPНа-картах со всеми связями указанных в Essence Kernel

Подобная сеть, с одной стороны, максимально полна с точки зрения связей зафиксированных в стандарте языка Essence. С другой стороны, объем предполагаемых связей делает применение прикладных методов крайне затруднительным, так как для некоторых утверждений для ALPНа «Work» число влияющих на них вершин составляет 32, что вынуждает проанализировать и составить вектор из 4 294 967 296 (2^{32}) ве-

роятностей, что приводит к очень трудоёмким вычислительным операциям, которые практически мало осуществимы.

Заключение

Использование ALPНа-карт и связанных с ними утверждений позволяет получить более структурированную картину происходящего в проекте. С одной стороны, она является достаточно абстрактной, чтобы на ней можно было производить прикладные расчеты и разрабатывать новый класс методов для управления проектами по программной инженерии. С другой стороны, данные методы должны позволять работать с большим количеством связанных элементов и учитывать внутренние связи между утверждениями. В дальнейших планах авторов реструктурировать полученную байесовскую сеть до объемов, которые позволили бы с ней проводить различные вычисления и эксперименты.

ЛИТЕРАТУРА

1. *Essence – Kernel and Language for Software Engineering Methods Version 1.2*. URL: <http://semat.org/documents/20181/57862/formal-18-10-02.pdf/866c80c0-cdc8-488b-bcf8-0c67cb60b5d7> [дата обращения: 14.11.2020].
2. *Sajad Soltan, Maryam Ashrafi*. Predicting project duration and cost, and selecting the best action plan using statistical methods for earned value management // *Journal of Project Management*. V. 5. 2020. P. 157–166.
3. *Sackey S., Lee D., Kim B.* Duration estimate at completion: Improving earned value management forecasting accuracy // *KSCE Journal of Civil Engineering*. 2020. V. 24. Iss. 3.
4. *National Research Council and National Research Council*. 1996. *Statistical Software Engineering*. National Academy Press, USA.
5. *Progress Poker Flyer*. URL: <https://essence.ivarjacobson.com/publications/brochures/progress-poker-flyer> [дата обращения 01.12.20].

Поддержка настраиваемых типов проектов в системе автоматизации управления Git-репозиториями для использования в процессе обучения

Олег Змеев, Юлия Протасевич, Данила Соколов

*Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

Выполнено проектирование механизма поддержки создания и настройки произвольных типов проектов, которые создаются в рамках предмета, позволяют преподавателю отойти от жестко заданных типов проектов и гибко настраивать проекты под свои нужды. Такое расширение является продолжением развития инструмента по автоматизации использования систем контроля версий в процессе обучения. Основной целью статьи является подробное описание этого механизма: модели типов предметов, модели типов проектов.

Система автоматизации управления Git-репозиториями

Для организации работы на предметах, где студенты пишут код и занимаются разработкой проектов в команде или самостоятельно, был разработан подход [1] для организации взаимодействия преподавателей и студентов с использованием системы контроля версий Git [2] и дополнительной системы. Данная система была разработана с целью адаптировать использование систем контроля версий под образовательные нужды и предназначена для упрощения работы преподавателя с большим количеством репозиториях. Она позволяет автоматизировать управление репозиториями за счет автоматизации рутинных задач, которые выполняют преподаватели во время проведения своих курсов: создание репозиториях, где работают студенты, раздачу и сбор заданий, публикацию материалов и просмотр статистики прогресса учащихся. Работа преподавателя при таком подходе основывается на использовании двух систем – GitLab [3], которая была выбрана в качестве системы хостинга Git-репозиториях и разработанного инструмента, который позволяет управлять репозиториями, необходимыми для проведения предметов.

Типы проектов

Изначально, в разработанном дополнительном инструменте были определены 7 фиксированных типов проектов доступных преподавателям:

- домашняя работа,
- контрольная работа,
- материалы преподавателя,
- материалы студента,
- командная домашняя работа,
- командная контрольная работа,
- материалы команды.

По результатам тестовой эксплуатации системы стало понятно, что данных типов проектов недостаточно для покрытия потребностей всех преподавателей. Например, текущая система не позволяет выборочно включать дополнительные возможности GitLab в рамках репозиториев отдельных проектов. Нет возможности создать пустой репозиторий для студента (без использования шаблона), который был бы доступен преподавателю. Нет возможности предоставить преподавателю права на запись в репозиториях студентов.

Для того чтобы добавить в систему поддержку этих особенностей, а также предусмотреть другие возможные запросы преподавателей, было принято решение создать механизм, позволяющий создавать настраиваемые типы проектов и заменить этим механизмом фиксированные типы.

Модель типов предметов

Настройку типов проектов и других особенностей организации взаимодействия в рамках предмета предлагается выполнять на уровне всего предмета путем определения типа предмета (Subject type), в рамках которого задается следующее (рис. 1):

1. Роли преподавателей и студентов в рамках предмета (Subject role). Роли могут быть назначены участникам при добавлении их на предмет и будут использованы для настройки разграничения прав в системе GitLab (для студентов и преподавателей) и в дополнительном инструменте (только для преподавателей).

2. Группы GitLab создаваемые для предмета (Subject group). В группах преподаватели смогут хранить шаблоны заданий, публиковать материалы и т.д. Для каждой группы определяется ее логическое имя,

формат именования в GitLab, а также кто будет являться членами такой группы и какие права в этой группе иметь.

3. Типы проектов, доступные в рамках предмета и их настройки (Project type).

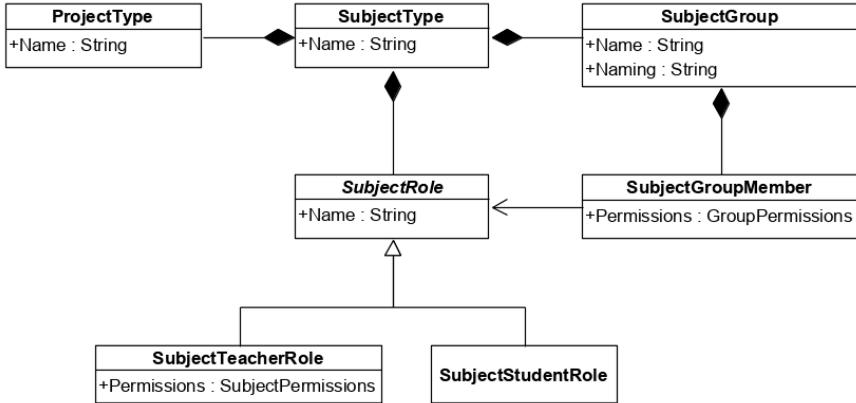


Рис. 1. Тип предмета

В рамках типа проектов определяются (рис. 2):

1. Репозитории (Project repository), которые будут использоваться для данного типа проектов:

- а) Тип репозитория – предметный, студенческий или командный.
- б) Группа, в которой репозиторий будет создан.
- с) Количество репозиториев – один или несколько.
- д) Формат именования – определяет шаблон, на основании которого будет определяться имя репозитория.
- е) Доступные дополнительные возможности GitLab, такие, как Issues, Merge Requests, Wiki, Snippets.
- ф) Роли, имеющие доступ к данному репозиторию, а также их права.

2. Переменные проекта (Project variable), которые могут быть установлены преподавателем. Значения этих переменных могут быть заданы как на уровне всего проекта (например, срок выполнения, который может быть использован для автоматической блокировки репозиториев), так и для каждого студента индивидуально.

3. Действия, которые преподаватели могут совершать в рамках проекта, а также действия, которые совершаются автоматически по заданному расписанию (Project action).

4. Роли, имеющие доступ к проектам данного типа, а также их права (Project member).

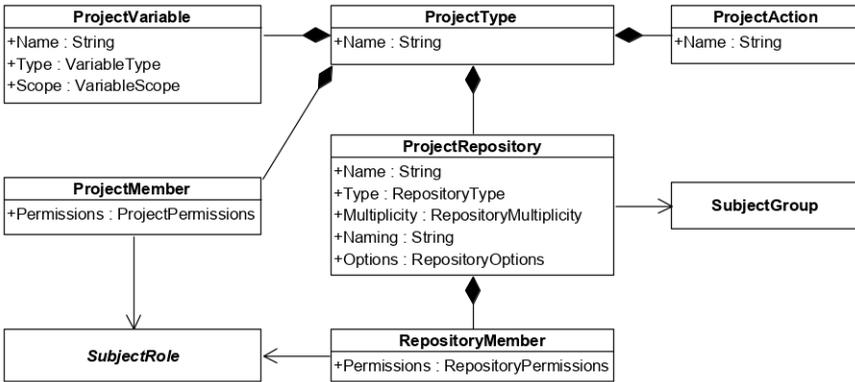


Рис. 2. Тип проекта

Для действия проекта задаются (рис. 3):

1. Последовательность шагов (Action step), которые должны быть выполнены в рамках данного действия, для каждого шага указывается его тип и набор параметров (Action parameter). Типы шагов являются predeterminedными и, как правило, представляют собой различные действия с репозиториями, такие, как создание пустого репозитория, копирование репозитория, копирование коммитов и т.д.

2. Набор условий, которые должны быть выполнены для того, чтобы преподавателю было доступно это действие (Action condition). Для каждого условия определяется его тип и набор параметров. Типы условий, как и типы шагов, также являются predeterminedными.

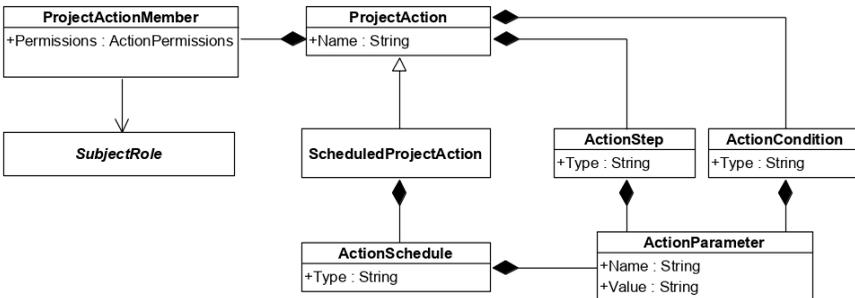


Рис. 3. Действие проекта

3. Расписание выполнения для действий по расписанию (Action schedule).

4. Роли, имеющие доступ к данному действию (Project action member).

Полученный тип предмета может быть выбран при создании нового предмета в системе, в результате чего предмет и все проекты, которые будут созданы в его рамках, будут работать согласно заданным в рамках типа предмета настройкам. Замена типа предмета у существующего предмета, а также внесение изменений в тип предмета, который используется в одном и более предмете, в свою очередь, являются менее тривиальными задачами, поскольку уже могут существовать сущности, зависящие от текущих настроек типа. В частности, наиболее проблемными сущностями являются: группы GitLab, проекты, репозитории проектов. На данном этапе развития системы предлагается ограничить изменение типа предмета, если у него есть зависящие сущности одного из вышеперечисленных типов. В случае необходимости внесения существенных изменений предлагается выполнить клонирование типа и внести изменения в копию.

Заключение

В результате был спроектирован механизм поддержки создания и настройки произвольных типов проектов, создаваемых в рамках предмета. Данная настройка выполняется на уровне типа предмета, позволяет преподавателям отойти от жестко заданных типов проектов и более гибко настраивать систему.

ЛИТЕРАТУРА

1. *Протасевич Ю.А.* Разработка системы для автоматизации управления Git-репозиториями на базе системы GitLab для использования в процессе обучения: выпускная бакалаврская работа по направлению подготовки: 09.03.04 – Программная инженерия. Томск, 2019. URL: <http://vital.lib.tsu.ru/vital/access/manager/Repository/vital:9089>.
2. *Git.* URL: <https://git-scm.com> (retrieval date: 10.11.2020).
3. *The first single application for the entire DevOps lifecycle – GitLab.* URL: <https://about.gitlab.com> (retrieval date: 10.11.2020).

Формальная модель образовательной программы в области компьютерных наук с точки зрения международного стандарта ACM и IEEE

Татьяна Кетова, Евгения Соколова

*Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

На сегодняшний день деятельность Высшей ИТ-школы (HITs) Томского государственного университета направлена на ускоренную профессиональную подготовку студентов в тесном взаимодействии с компаниями цифровой экономики. Кроме этого, образовательная модель школы изначально ориентирована на международный рынок. Это обуславливает необходимость создания образовательной программы с учетом международного опыта, следовательно, и международных стандартов. Также сегодняшнее положение, в эпоху COVID-19, не могло не отразиться на образовании в целом. ИТ-технологии стали занимать не лидирующее, а самое приоритетное направление в мире. Актуальность исследования международных стандартов образования возросла. В настоящее время, в HITs ТГУ разработана модель, которая совмещает в себе не только особенности федерального государственного образовательного стандарта (ФГОС) [2], но и международного стандарта, разработанного ACM [3] и IEEE [4].

В данной статье представлено исследование только Curriculum Guidelines for Undergraduate Programs in Computer Science [1].

Для качественного исследования потребовался глубокий анализ международного стандарта, на основе которого сформирована модель образовательной программы в области компьютерных наук с точки зрения международного стандарта.

Анализ предметной области

Анализ международного стандарта. Система стандартов учебных программ бакалаврской подготовки по дисциплине Computing имеет древовидную структуру [5].

На верхнем уровне располагается методологический документ CC2005. Данный документ – руководства по составлению программ учебных курсов (curricula guidelines или curriculum standarts):

- вычислительная техника (Computer Engineering – CE);
- компьютерные науки (Computer Science – CS);
- информационные системы (Information Systems – IS);
- программная инженерия (Software Engineering – SE);
- системы информационных технологий (Information Technology –IT).

Несмотря на то, что образовательная программа в HITs прежде всего ориентируется на профессиональную подготовку, в качестве ее основы использован стандарт по Computer Science как более подходящий для классического университета.

Для представления объема знаний CS используется типовая (иерархическая) модель. На верхнем уровне содержатся предметные области (areas). Структура знаний содержит 18 предметных областей (рис. 1).



Рис. 1. Предметные области Curriculum Guidelines for Undergraduate Programs in Computer Science

В каждой areas выделяются разделы или модули знаний (units) (рис. 2). Последние, в свою очередь, разбиваются на темы (topics) и результаты обучения (learning outcomes) (рис. 3).

Core Tier-1 должна быть обязательной частью каждого модуля знаний Computer Science, она содержит темы с высоким приоритетом для включения в каждую программу, составляет основу учебной программы.

Темы Core Tier-2, как правило, важны, но в меньшей степени, они не являются обязательными, но их включение в учебную программу рекомендуется.

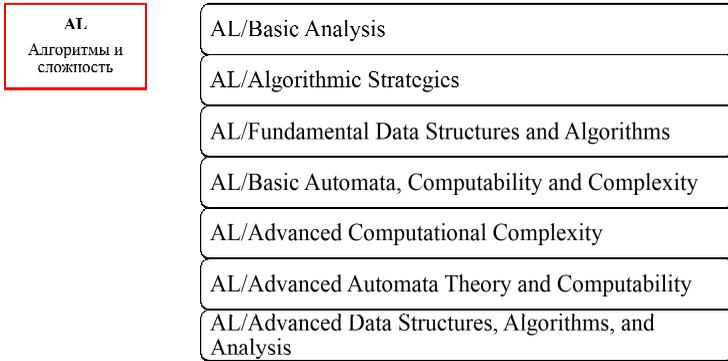


Рис. 2. Модули знаний предметной области AL Curriculum Guidelines for Undergraduate Programs in Computer Science

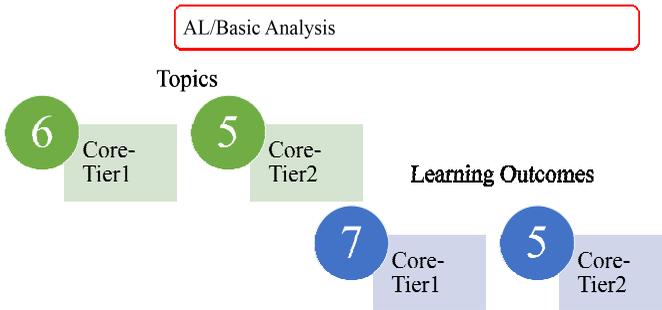


Рис. 3 Topics и Learning Outcomes модуля знания AL/Basic Analysis Curriculum Guidelines for Undergraduate Programs in Computer Science

Одной из центральных в стандарте является концепция ядра или обязательной части объема знаний, которая должна присутствовать во всех учебных программах по данному профилю. Объем почасовой лекционной нагрузки ядра составляет минимум 279.4 лекционных часов, учитывая весь Tier1 + 80% от Tier2 (весь Tier1 + 90% Tier2 – 293.7 лекционных часов, весь Tier1 + весь Tier2 – 308 лекционных часов. «Час» соответствует времени, необходимому для представления материала в традиционном лекционно-ориентированном формате; счетчик часов не включает никаких дополнительных работ, связанных с лекцией (например, самостоятельное изучение, лабораторные занятия и контрольные работы).

В стандарте приведен обновленный весьма обширный список результатов обучения, т.е. ожидаемых характеристик выпускника, состоящий из наборов общих характеристик, когнитивных способностей и навыков, дополнительных профессиональных навыков.

Модель представления Curriculum Guidelines for Undergraduate Programs in Computer Science

Исходя из анализа международного стандарта определена следующая структура:

- Standart – наименование и описание стандарта;
- Characteristics of Graduate – наименование и описание компетенций выпускника бакалавриата;
- Knowledge Area – наименование и описание области знаний;
- Knowledge Unit- наименование и описание модуля знаний;
- KU_Topics – описание топика Knowledge Unit;
- KU_Learning Outcome – описание ожидаемого результата изучения Knowledge Unit;
- Level – номер и описание уровня владения знаниями по Knowledge Unit;
- Duration of Level – длительность изучения уровня Knowledge Unit;
- Level of mastery – наименование и описание уровня мастерства результата изучения.

На рис. 4 представлена диаграмма классов модели предметной области, а именно международного стандарта Compute Science CS2013 [6].

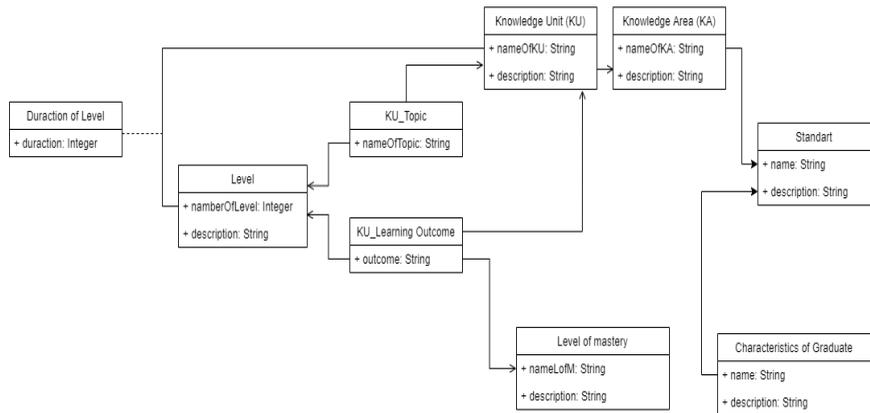


Рис. 4. Модель представления Curriculum Guidelines for Undergraduate Programs in Computer Science

Заключение

На основе анализа международного стандарта сформирована модель представления. Данная модель позволит сформировать дисциплины, соответствующие Федеральному государственному стандарту и международному стандарту одновременно. Она является фундаментом для реализации MVP-системы, представляющей из себя вспомогательный инструмент для составления образовательной программы НИТс ТГУ. Также результаты исследования могут быть полезны для систематизации всех международных стандартов, выявления общей модели представления.

ЛИТЕРАТУРА

1. *Associate-Degree Computing Curricula*. URL: <http://csecc.acm.org/> (дата обращения 01.10.2016).
2. *Федеральный государственный образовательный стандарт высшего образования – бакалавриат по направлению подготовки 09.03.04 Программная инженерия*. Сентябрь 13, 2017.
3. *Association for Computing Machinery*. URL: <https://www.acm.org/> (дата обращения 23.05.2020).
4. *Institute of Electrical and Electronics Engineers*. URL: <https://www.ieee.org/> (дата обращения 31.05.2020).
5. *Computer Science Curricula 2013. Curriculum Guidelines for Undergraduate Degree Programs in Computer Science*. Декабрь 20, 2013
6. *Unified Modeling Language*. URL: <https://www.uml.org/> (дата обращения 01.06.2020).

Сценарии управления беспилотными транспортными средствами в среде «Умного города»

Яна Куликова¹, Дмитрий Качалов²,
Маис Паша Оглы Фархадов²

¹ Волжский государственный технический университет, г. Волгоград, Россия

² Институт проблем управления им. В.А. Трапезникова РАН, г. Москва, Россия

Современное общество очень много внимания уделяет таким важным вопросам, как вопросы безопасности, в том числе на дороге. В качестве подтверждения можно рассмотреть частоту и тщательность проработки и внесения изменений в правила дорожного движения [1]. Одним из наиболее современных способов обеспечения безопасности на дороге является использование беспилотных транспортных средств (далее – БТС) [2].

Для работы со многими вопросами, особенно связанными с безопасностью, очень важно знать нормативно-правовую базу. Так, для вопроса БТС существует Распоряжение Правительства РФ от 25 марта 2020 г. № 724-р «О Концепции обеспечения безопасности дорожного движения с участием беспилотных транспортных средств на автомобильных дорогах общего пользования» (далее – Концепция) [3]. Основная цель Концепции – организация безопасности дорожного движения.

Внедрение БТС

Вопросы внедрения беспилотных транспортных средств носят межведомственный характер, в связи с чем определение ключевых терминов и их использование в нормативных правовых документах и актах технического регулирования имеют важное значение для безопасного развертывания систем беспилотного вождения транспортных средств.

Представим основные термины, связанные с видами автоматизации наземного транспорта в РФ и зарубежом, например: автономный автомобиль (autonomous car); высокоавтоматизированное транспортное средство (highly automated vehicle); беспилотный автомобиль (driverless car); беспилотное транспортное средство (unmanned vehicle); полностью автоматизированное транспортное средство (fully automated vehicle); роботизированный автомобиль (robotic car); самоуправляемое транспортное средство (self-driving vehicle).

Наиболее важным является определение беспилотного транспортного средства – высоко- или полностью автоматизированное транспортное средство, функционирующее в беспилотном режиме, который означает, что во время использования данного режима транспортное средство находится под управлением автоматизированной системы вождения.

Первоначально, разработки, связанные с автономными транспортными средствами были связаны с военными разработками. Тем не менее к первому прототипу БТС можно отнести работу 1961 года студента Стэнфордского университета Джеймса Адамса – самоуправляемую тележку («Стэнфордская тележка»), управление которой осуществлялось путем передачи сигнала по кабелю (рис. 1). Второй прототип Адамс создал радиоуправляемым.

Следующим этапом было усовершенствование тележки при помощи оснащения её системой технического зрения – распознавания линии, по которой тележка должна двигаться.

Создание полностью автономного транспорта принадлежит ученому Эрнсту Дикмансу. В нем впервые были использованы вычислительные механизмы и система имитации движения глаз. Дальнейшее развитие разработок осуществлялось в рамках проекта под названием «Прометей».

Начиная с 2004 года когда прошло первое автосоревнование с участием БТС, беспилотные ТС развивались, появились автопилоты Google, в 2015 – первые серийные БТС Tesla и на данном этапе, все крупные автокомпании ведут разработки собственных прототипов робокаров.

На данный момент существует шесть уровней «беспилотности», где нулевой уровень – полное управление водителем, 5-й уровень – абсолютная беспилотность (полное отсутствие необходимости вмешательства человеком в управление ТС).

Для большего понимания можно привести пример в виде автомобиля, имеющего круиз-контроль и систему экстренного торможения. Та-



Рис. 1. «Стэнфордская тележка»

кой автомобиль имеет первый уровень автономности. Такие автомобили, как Tesla, Mercedes, GM и Volvo, имеют второй уровень.

Некоторые ученые утверждают, что автомобили, имеющие 5-й уровень повсеместно появятся не скоро, но БТС, работающие в ограниченном пространстве (например, по конкретному маршруту), скорее всего, начнут полноценно работать в ближайшие годы.

Умный город и сценарий работы БТС

Предпосылками к организации умной транспортной системе являются главные проблемы дорог больших городов – образование дорожных пробок. Ремонтные работы, ДТП и иные непривычные для людей ситуации на дороге вызывают множество сложностей как для водителей, так и для властей, старающихся решить проблемы. Но для решения много времени уходит на определение причин, формирование альтернативных направлений движения и разгрузки дороги, зачастую – это очень большая потеря времени. «Умный город» создаётся именно для того, чтобы подобные проблемы решать.

В «Умном городе» информация передаётся, преимущественно, от датчиков, которые установлены не только в городе, но и в самих автомобилях. Следовательно, о произошедших ситуациях данные передаются моментально, что позволяет максимально быстро перераспределить движение.

В «Умном городе» движение распределяется не только в зависимости от происшествий на дороге. Например, умные светофоры позволяют эффективно распределять трафик движения в зависимости от времени суток и нагрузки на дорогу. Также, они позволяют обеспечивать безопасность пешеходов, анализируя данные о скоплении групп пешеходов и изменяя фазы работы светофоров в зависимости от ближайшего расположения автомобилей к людям.

В «Умном городе» производится обмен данными между центром управления движением и непосредственно БТС.

В основе конструкции БТС лежит использование программного обеспечения и датчиков.

Датчики могут представлять собой: лидары (дальномер оптического распознавания), радары, камеры, систему глобального позиционирования (GPS, Глонасс), датчики одометрии (информация о движении приводов для оценки перемещения), гиростабилизатор.

Программное обеспечение может использовать машинное зрение и нейросети. Следовательно, некоторые БТС способны передвигаться,

только ориентируясь на определённые метки, другие же способны анализировать пространство без особых меток и специальных линий.

Для обучения автомобилей существуют два подхода: классический и нейросетевой (рис. 2).

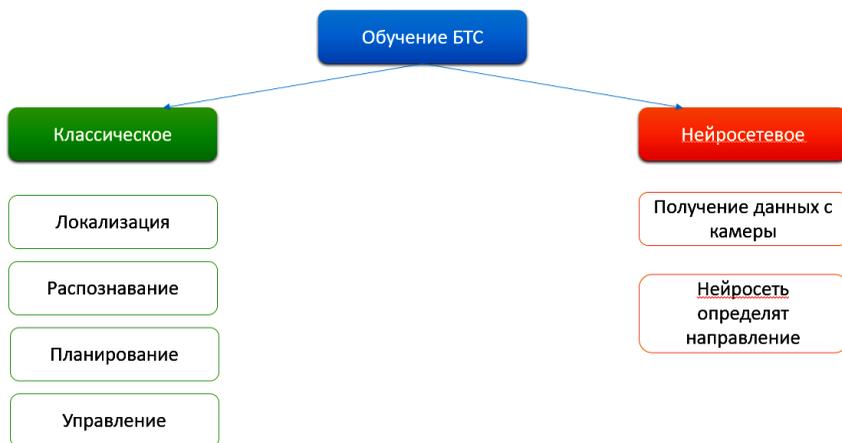


Рис. 2. Подходы обучения БТС

Так как реализация нейросетевого подхода требует обработки больших данных и обширную базу паттернов для обучения ИИ машины, многие БТС работают на основе классического подхода.

Архитектура системы управления БТС в системе «Умного города»

В рамках работы систем управления группами беспилотными транспортными средствами существует множество архитектурных решений, основанных на различных сценариях управления.

Новым подходом построения архитектуры приложения является архитектура, основанная на применении туманных и граничных вычислений.

Задачи пограничных вычислений на уровне БТС:

- Определение препятствий на основе поступающих сигналов с установленных устройств.
- Определение стратегии движения на основе дорожной ситуации и результатов обработки сигналов.
- Взаимодействие с другими БТС.
- Взаимодействие с управляющим узлом сети.
- Взаимодействие с оператором БТС (пассажиры).

На основе обозначенных выше ключевых аспектов задач пограничных вычислений можно предложить архитектуру системы управления (рис. 3).



Рис. 3. Архитектура системы группового управления BTS

Для реализации такой структуры в рамках существующей инфраструктуры, которая только строится или уже существует, хорошо подходит г. Москва (рис. 4).

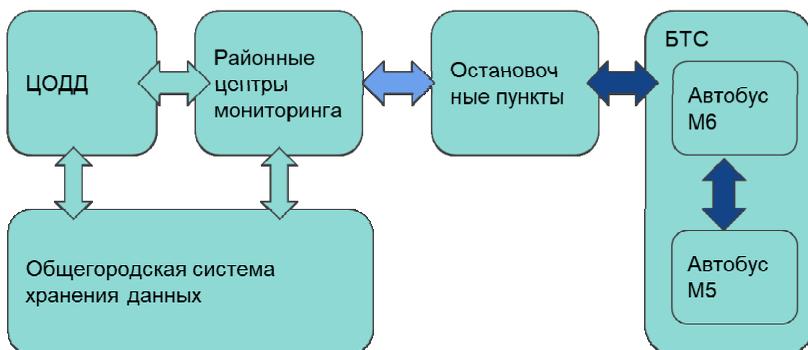


Рис. 4. Пример реализации архитектуры в г.Москва

Таким образом, удаётся обеспечивать работу всей системы, основываясь на ряде основных принципов:

1. Беспилотное транспортное средство принимает решение о корректировке движения на основе данных, собираемых с собственных датчиков и систем а также данных, получаемых от локальных узлов обработки информации.

2. При осуществлении движения БТС обновляет данные, получаемые от локальных узлов и передаёт часть собственных данных (например, данные об авариях и т.д.).

3. Часть вычислений сосредотачивается на локальных узлах, что позволяет избежать избыточной нагрузки на сеть передачи данных и обеспечить целостность системы принятия решений БТС.

4. Часть вычислений передается на территориальные и центральные узлы, что обеспечивает data consistency во всех системе в каждый момент времени, когда БТС запрашивает необходимые данные для обеспечения управления.

5. Туманные и граничные вычисления позволяют быстрее отправлять результаты на БТС.

Вывод

Для создания безопасной системы «Умный город» необходимо тщательно подойти к вопросу сценариев управления БТС. В статье описаны уровни управления, особенности конструкций БТС, а также представлена архитектура системы управления БТС, которая позволяет организовать туманные и граничные вычисления в системе «умного города».

ЛИТЕРАТУРА

1. *Новости* в России и мире // Путин поручил разработать госпрограмму по безопасности дорожного движения до 2030 года. <https://tass.ru/obschestvo/6678424>
2. *Подборка* новостей про робомобили. URL: <http://embeddedsystem.ru/index.php?page=podborka-novostej-pro-robomobili>
3. *Распоряжение* Правительства РФ от 25 марта 2020 г. № 724-р «О Концепции обеспечения безопасности дорожного движения с участием беспилотных транспортных средств на автомобильных дорогах общего пользования». URL: <https://www.garant.ru/products/ipo/prime/doc/73707148/>

Разработка системы автоматизации процессов обращения кассовой техники в банковской организации

Яна Лебедева, Вячеслав Вавилов

*Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

Представлен процесс разработки системы автоматизации процессов обращения кассовой техники в банковской организации. Описаны обязанности системного аналитика, которые были установлены в рамках конкретной команды разработки. Представлены особенности, выявленные при работе с требованиями и Заказчиком. Рассмотрены инструменты, используемые для проектирования системы и представлены результаты работы.

Формирование команды

Как известно, в настоящее время, любой проект разрабатывается командой, в составе которой должны присутствовать специалисты различного профиля, такие как менеджеры, тимлиды, системные и бизнес-аналитики, UI/UX-дизайнеры, архитекторы-проектировщики, front-end и back-end-разработчики, а также QA-инженеры.

Команда разработчиков данной системы состоит из менеджера, системного аналитика, тимлида, трех QA-инженеров и пяти fullstack-разработчиков.

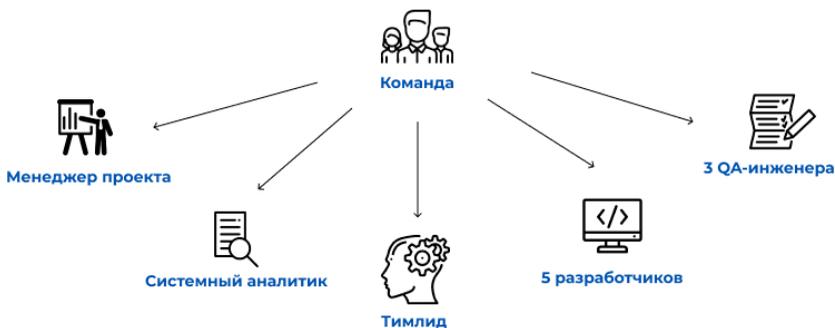


Рис. 1. Состав IT-команды

Каждый член команды имеет свою зону ответственности, например менеджер контролирует соблюдение денежных и временных ограничений проекта, тимлид распределяет нагрузку на команду разработчиков, а QA-инженеры отвечают за качество разрабатываемого продукта. Отлаженная работа всех подразделений команды приведет к высококачественному и надежному продукту, соответствующему заявленным требованиям.

Зоной ответственности системного аналитика является управление требованиями к разрабатываемому продукту. Учитывая, что Заказчик говорит на языке бизнеса и ему, как правило, не знакомы термины программирования, а разработчики в основном мыслят алгоритмами и зачастую не разбираются в предметной области Заказчика, задача аналитика заключается в построении взаимопонимания обеих сторон, являясь переводчиком с одного языка на другой [1].

При работе с требованиями аналитик должен выполнить следующие этапы: этап определения концепции продукта, этап сбора требований, этап анализа требований и этап проектирования системы [2].

На этапе определения концепции продукта вырабатывается единое видение будущей системы. По окончании этого этапа делается вывод о том, будет ли реализовываться проект в будущем.

На этапе сбора требований ведется тесная работа с Заказчиком и заинтересованными пользователями разрабатываемой системы. Цель этапа – выявление бизнес-проблемы, которая может быть решена с помощью системы, а также, непосредственно, сбор требований к функциональности продукта. Качественное выполнение работы на данном этапе напрямую зависит от того, будет ли проект соответствовать ожиданиям Заказчика.

Анализ требований заключается в структурировании заранее собранных требований, которые должны соответствовать таким свойствам, как полнота, корректность, осуществимость, необходимость, недвусмысленность, а также проверяемость [1]. Правильно сгруппированные требования помогут обойтись минимальным количеством функциональности для удовлетворения большего количества целей, что в свою очередь поможет соблюсти временные и финансовые рамки проекта.

Итоги всех предыдущих этапов необходимы для проектирования системы. Результатами этой работы является законченное техническое задание к продукту. На основе этого выполняется подготовка артефактов, понятных как Заказчику, так и членам команды, для последующего согласования. В качестве артефактов могут выступать различные диа-

граммы, прототипы пользовательского интерфейса, а также пользовательские сценарии.

Работа аналитика на этом не заканчивается. На протяжении всего проекта его обязанностями является постановка задач разработчикам, управление изменяющимися требованиями, а также участие в процессе тестирования разрабатываемой системы.

На этапе сдачи проекта Заказчику, аналитик занимается подготовкой проектной документации, такой, как руководство пользователя, руководство системного и прикладного администратора и т.д. В некоторых случаях он проводит демонстрацию системы Заказчику и обучает будущих пользователей.

Учитывая разнообразие методологий управления проектами и внутренних правил ИТ-команд, обязанности аналитика могут не совпадать с указанными выше. Данная статья написана с позиции системного аналитика при работе над крупным долгосрочным проектом по заказной разработке.

Анализ и управление требованиями

Заказчиком системы является банковская организация, проблема которой заключается в неорганизованности и неэффективности процессов деятельности, связанных с учетом и обращением кассовой техники. Система управления специальным банковским оборудованием должна автоматизировать процессы обращения с кассовой техникой, а также обеспечивать единое место хранения учетных данных.

Входными данными для разработки системы является верхнеуровневое техническое задание, которое дает общее представление о функциональности системы, но тем не менее требует дополнительных уточнений и анализ требований. После анализа предметной области и общения с Заказчиком была построена диаграмма классов предметной области, но, ввиду большого объема, она в данной статье не представлена.

Сбор уточняющих требований выполняется в формате интервью, результаты документируются в письменном виде и отправляются второй стороне по каналу связи, установленному в договоре.

Результатом успешного уточнения требований является смена условия доступа пользователя к данным системы с роли пользователя на профиль. Профилем в данном случае называется совокупность роли и уровня подразделения в иерархии филиальной сети банка. Учитывая, что от уровня, на котором находится подразделение зависят практиче-

ски все данные в системе, уточнение этого требования помогло сэкономить много времени на исправлении замечаний от Заказчика.

Дополнительные обсуждения требований к системе зачастую приводят к тому, что пожелания Заказчика выходят за временные и денежные рамки проекта, установленные при заключении договора. В этом случае необходимо собрать все новые требования вместе и провести повторный анализ, учитывая архитектуру и исходные требования к системе.

Проектирование системы

Проектирование системы заключается в построении различных диаграмм, а также в разработке прототипов интерфейса системы.

Для работы с диаграммами используется инструмент визуального моделирования и проектирования Enterprise Architect, который позволяет строить диаграммы, доступные большому количеству пользователей одновременно [3]. Помимо этого, инструмент позволяет автоматически генерировать документы, которые удобно демонстрировать Заказчику при согласовании.

Хранение информации по проекту в Enterprise Architect является огромным плюсом для крупного и долгосрочного проекта, так как помогает вспомнить, почему и как была реализована та или иная функция, а для новых членов команды служит удобным местом изучения системы.

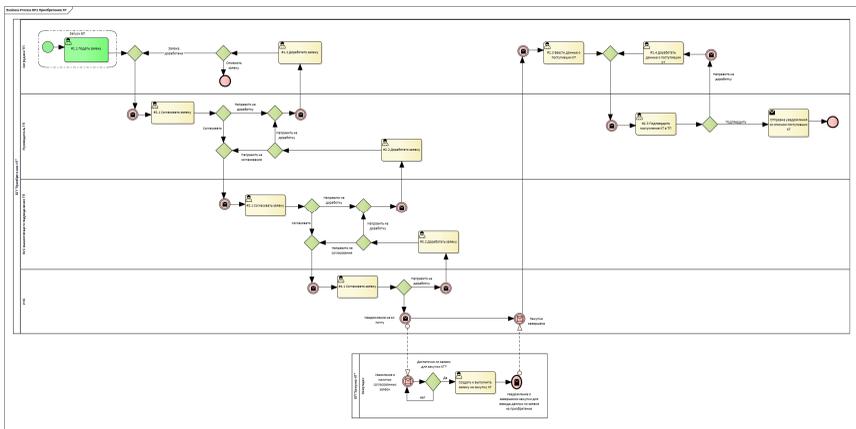


Рис. 2. Пример диаграммы в нотации BPMN 2.0

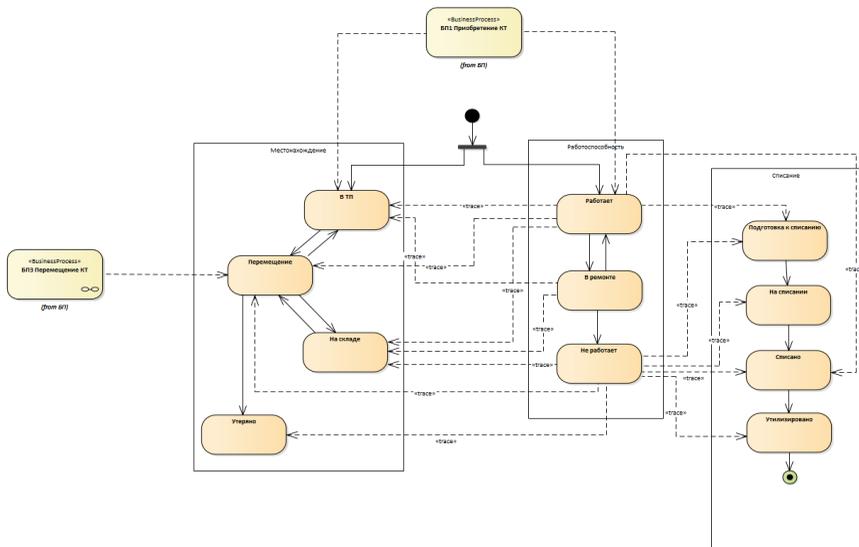


Рис. 3. Пример диаграммы состояний

Для проектирования прототипов интерфейса системы использовался онлайн-сервис Figma. Как и Enterprise Architect, этот инструмент доступен для совместного использования несколькими пользователями.

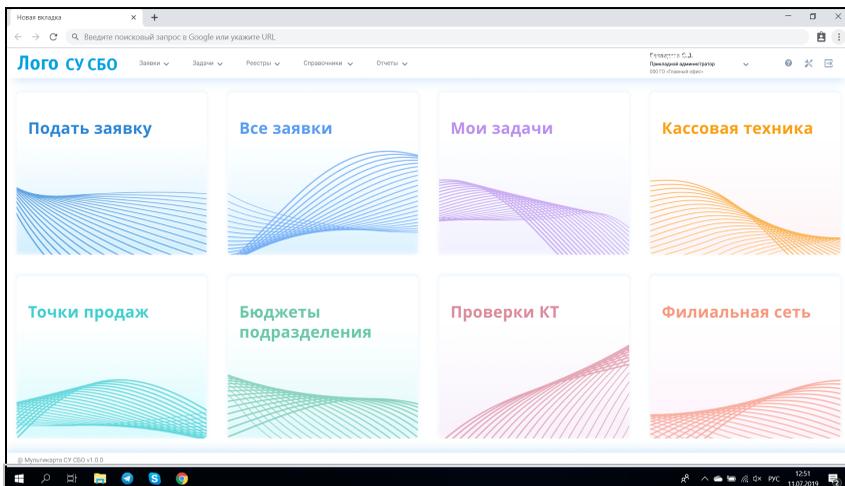


Рис. 4. Прототип главной страницы системы

Разработка прототипов выполнялась для согласования с Заказчиком с целью выявления несоответствий ожиданиям на раннем сроке, то есть до реализации. Для представления основных разделов системы было спроектировано более 100 прототипов, которые были оформлены в виде спецификации.

Заключение

На данный момент выполнено проектирование первого этапа проекта, включающего создание каркаса системы и основных бизнес-процессов. Для разработчиков была выполнена постановка задач, по которым они реализовывали функциональность в соответствии с требованиями. Совместная и своевременная работа помогла команде успешно начать крупномасштабный проект и продолжать разработку в соответствии с установленными сроками. Тем не менее можно отметить сложности, которые впоследствии могут затруднить реализацию системы и на которые следует обратить внимание. К ним относится управление изменяющимися требованиями и поддержание данных, описанных в Enterprise Architect, в актуальном виде.

Подводя итог, стоит отметить важность работы аналитика при разработке продукта, так как именно он должен правильно перевести пожелания бизнеса в техническую плоскость для достижения целей Заказчика и команды.

ЛИТЕРАТУРА

1. *Перерва А., Иванова В.* Коммуникационные способности // Путь аналитика. Практическое руководство IT-специалиста. 2011. Р. 43–45.
2. *Химонин Ю.* Этапы сбора и анализа требований // Сбор и анализ требований к программному обеспечению. 2009. Р. 4–9.

Имитационное моделирование СМО с входящими потоками, параметры которых зависят от состояния системы

Евгений Полин¹, Александр Моисеев¹,
Константин Войтиков²

¹Национальный исследовательский

Томский государственный университет, г. Томск, Россия

²Московский физико-технический институт

(национальный исследовательский университет), г. Москва, Россия

В настоящее время одним из актуальных направлений теории массового обслуживания является исследование моделей с нестационарными входящими потоками, в том числе таких, в которых параметры входящего потока меняются в зависимости от состояния системы, например числа заявок на обслуживании [1, 2]. Основными методами исследования таких моделей остаются приближенные алгоритмы и имитационное моделирование. Разработка приложений для имитационного моделирования таких систем важна, когда используются приближенные методы, так как результаты имитации позволяют оценить погрешности и области применимости приближенных результатов. Но и для тех случаев, когда результат можно получить аналитически, имитационное моделирование не будет лишним, так как сможет подтвердить полученные выводы.

На протяжении ряда лет группой ученых ТГУ разрабатывается программный комплекс для имитационного моделирования систем и сетей массового обслуживания (СМО и СеМО) ODIS [3, 4]. Данный программный комплекс охватывает большой спектр моделей теории массового обслуживания и ежегодно пополняется все новыми библиотеками, однако в нем пока не были реализованы возможности имитационного моделирования СМО с входящими потоками, параметры которых зависят от состояния системы, а такая потребность возникла в связи с последними теоретическими исследованиями в этой области [5, 6].

Для реализации такой возможности в рамках системы ODIS принято решение применить для ее архитектуры на уровне подсистемы объектов, представляющих элементы СМО и СеМО (рис. 1), паттерн Наблюдатель [7]. Как видим, основой иерархии для таких объектов является класс `Element`, представляющий собой абстрактную сущность элемента модели массового обслуживания. Его основные операции `accept(...)` и

processEvent(...) – это единственные предполагаемые точки, где может произойти изменение состояния любого объекта, именно их мы и будем в дальнейшем использовать для отслеживания событий изменения состояния. Более подробную информацию о назначении всех остальных классов можно найти в [8].

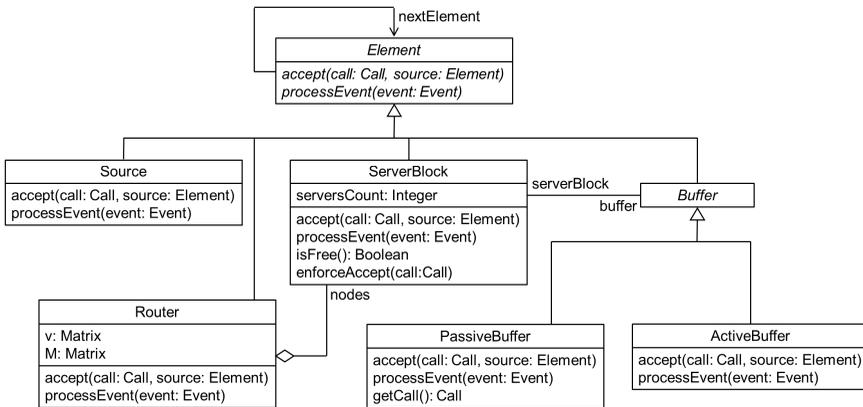


Рис. 1. Иерархия элементов объектной модели СМО и СеМО в системе ODIS

Вообще говоря, для наших целей интересны классы Source и ServerBlock из данной иерархии. Они представляют элементы для моделирования соответственно входящего потока и блока обслуживающих приборов. Однако было принято решение реализовать механизм оповещения об изменении состояния на уровне абстрактного предка Element так, чтобы любой элемент имитационной модели имел возможность оповещать другие об изменении его состояния, а также подписываться на оповещения о подобных изменениях в других элементах. Таким образом, паттерн Наблюдатель был реализован на самом верхнем уровне иерархии элементов имитационной модели массового обслуживания, причем класс Element был назначен и наблюдателем, и объектом наблюдения (рис. 2).

Для реализации механизма оповещения в класс Element добавлена коллекция подписчиков observers, а также неvirtуальные методы addObserver(...) для добавления элемента-подписчика, removeObserver(...) для удаления подписчика, stateChanged() для оповещения всех подписчиков об изменении состояния элемента и виртуальный метод objectIsChanged(...) для обработки оповещения об изменении состояния

у элемента-подписчика. Последний метод в классе `Element` реализован в виде заглушки, но потомки могут переопределить его, чтобы реализовать поведение, связанное с обработкой изменения состояния наблюдаемого объекта.

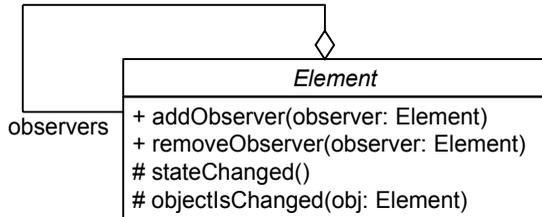


Рис. 2. Реализация паттерна Наблюдатель для обеспечения оповещений об изменениях состояния между элементами модели

Метод `stateChanged()` содержит код, оповещающий всех подписчиков об изменении состояния. Фактически этот метод вызывает в цикле для каждого подписчика его метод `objectIsChanged(...)`, передавая себя в качестве параметра. Вызов данного метода был реализован внутри операций `accept(...)` и `processEvent(...)` класса `Element` (это привело к тому, что данные методы перестали быть абстрактными). Это связано с тем, что в модели каркаса ODIS, как упоминалось выше, эти две операции, по сути, являются единственными возможными точками изменения состояния. Мы не приводим здесь диаграмму последовательностей для метода `stateChanged()`, равно как и для методов `addObserver(...)` и `removeObserver(...)`, ввиду тривиальности этих взаимодействий.

Предложенная реализация паттерна Наблюдатель предполагает использование концепции вытягивания информации: подписчик, получая сообщение об изменении в наблюдаемом объекте, одновременно с этим получает и указатель на изменившийся объект. Это удобно, когда элемент хочет вести наблюдение сразу за несколькими объектами. Затем он обращается к указанному объекту для получения данных, необходимых ему для обработки сообщения. Кроме того, в системе ODIS уже реализована цепочка `nextElement` (см. класс `Element` на рис. 1), по которой также можно добраться до интересующего нас объекта, так как обычно именно этот объект, привязанный к указателю `nextElement`, является источником оповещения об обновлении и/или источником необходимой информации для его обработки. Более того, данный механизм был использован для автоматической привязки наблюдателей к объек-

там при создании объектной структуры всей имитационной модели в классе QueueSimulationModel (см. [8]).

Последнее, что было реализовано для достижения конкретной цели работы – в классе Source был переопределен метод objectIsChanged(...), внутри него реализовано изменение параметров входящего потока в зависимости от состояния привязанного по указателю nextElement блока обслуживания (класс ServerBlock).

Таким образом, с помощью представленных в работе изменений удалось не только реализовать возможность имитационного моделирования входящих потоков СМО, параметры которых зависят от состояния системы, что являлось первоначальной целью работы, но и реализовать универсальный механизм отслеживания изменений в состояниях любых элементов модели, что может оказаться полезным при имитационном моделировании более сложных моделей систем и сетей массового обслуживания.

ЛИТЕРАТУРА

1. *Коротаев И.А., Спивак Л.Р.* Системы массового обслуживания в полумарковской случайной среде // Автомат. и телемех. 1992. Вып. 7. С. 86–92.
2. *Татаишев А.Г.* Система массового обслуживания с переменной интенсивностью входного потока // Автомат. и телемех. 1995. Вып. 12. С. 78–84.
3. *Моисеев А.Н., Синяков М.В.* Разработка объектно-ориентированной модели системы имитационного моделирования процессов массового обслуживания // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 1. С. 89–93.
4. *Moiseev A., Demin A., Dorofeev V., Sorokin V.* Discrete-event approach to simulation of queueing networks // Key Engineering Materials. 2016. V. 685. P. 939–942.
5. *Полин Е.П., Моисеева С.П., Моисеев А.Н.* Исследование бесконечнолинейной СМО с интенсивностью входящего потока, зависящей от состояния системы // Распределенные компьютерные и телекоммуникационные сети: управление, вычисление, связь (DCCN-2020) [Электронный ресурс]: материалы XXIII Международ. науч. конф., 14–18 сент. 2020 г., Москва. С. 638–644.
6. *Полин Е.П., Моисеева С.П., Моисеев А.Н.* Анализ вероятностных характеристик гетерогенной СМО вида $MR(S)/M(S)/\infty$ с параметрами обслуживания, зависящими от состояния вложенной цепи Маркова // Известия Саратовского университета. Новая серия. Серия Математика. Механика. Информатика. 2020. Т. 20. Вып. 3. С. 389–399.
7. *Гамма Э., Хелм Р., Джонсон Р., Влассидес Дж.* Приемы объектно-ориентированного проектирования. Паттерны проектирования. СПб.: Питер, 2010. 368 с.
8. *Моисеев А.Н.* Исследование математических моделей систем и сетей массового обслуживания с высокоинтенсивными непарасоновскими входящими потоками: дис. ... д-ра физ.-мат. наук. Томск, 2016. 333 с.

Обзор исследований по проблеме достижения высокой производительности протокола OPC UA*

Вадим Тренькаев

*Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

Протокол OPC UA [1] предназначен для надёжной и безопасной передачи данных между различными системами в промышленных сетях. При этом OPC UA является усовершенствованной версией «классического» OPC, повсеместно применяемого на современных промышленных объектах. Благодаря своим новым свойствам и продуманной архитектуре OPC UA набирает популярность среди производителей систем автоматизации, активно используется в системах промышленного интернета вещей и коммуникации умных заводов.

При разработке и эксплуатации систем промышленной автоматизации важно знать возможные ограничения производительности протокола OPC UA. Причем в зависимости от области использования предъявляются разные требования к характеристикам протокола OPC UA. Во встроенных системах (промышленных сетях нижнего уровня) скорость передачи данных и минимальная нагрузка на систему являются наиболее важными требованиями. В корпоративных системах (сетях предприятия верхнего уровня) более важна эффективная обработка структурированных данных.

Показателями производительности протокола OPC UA могут служить как скоростные характеристики (время реакции на запрос технологических данных, способность передачи данных в режиме реального времени), так и показатели того, насколько экономно используются ресурсы (процессор, память, сеть) при обеспечении заданного уровня быстродействия.

В данной статье с использованием источников [1–11] представлено современное положение дел в области исследований производительности протокола OPC UA.

* Данное научное исследование выполнено при поддержке Программы повышения конкурентоспособности ТГУ

Краткое описание протокола OPC UA

Клиент-серверная модель. В основе технологии OPC UA лежит клиент-серверная модель. Сервер OPC – программа на устройстве, располагающем данными о технологическом процессе. Сервер предоставляет набор сервисов, включая доступ к актуальным данным, архиву, набору сообщений об авариях и событиях. Сервер ожидает подключения клиентов, после установления подключения принимает от клиента запросы, формирует и отправляет ответы на запросы. Клиент OPC – программа, собирающая данные о технологическом процессе. Чтение данных может быть организовано синхронно/асинхронно, по подписке, т.е. по изменению, запись данных – синхронно/асинхронно. Обмен данными осуществляется путем формирования TCP-пакетов собственного формата, а также с помощью протокола HTTPS или WebSockets. Использование механизма подписки, когда сервер оповещает клиента о наступлении каких-либо событий, позволяет разгрузить сеть от передачи повторяющихся неизменных значений. Сервисы позволяют клиенту подписаться на оповещения от сервера, с помощью которых сервер может сигнализировать о возникновении аварий, событий, изменении значений данных.

Коммуникационный стек. Обмен данными в OPC UA происходит посредством сессий, которые определяются как логическое соединение клиента и сервера. Механизм сессий не зависит от особенностей нижележащего протокола, при этом сессия не прерывается автоматически в случае разрывов нижележащих протоколов. Установление и завершение сессии выполняется с помощью запросов и ответов клиента и сервера. Коммуникационный стек OPC UA содержит три уровня: уровень форматирования, уровень защищенного канала, транспортный уровень. Уровень форматирования представлен тремя возможными форматами сообщений, которыми обмениваются клиент и сервер: UA Binary; UA XML; UA JSON. В начале сессии обмена данными может быть организован защищенный канал связи между клиентом и сервером с поддержкой симметричного шифрования, что обеспечивает конфиденциальность передаваемых данных. При организации защищенного канала используются сертификаты X.509. Транспортный уровень стека OPC UA базируется на протоколе OPC UA Connection Protocol, работающего поверх одного из протоколов: TCP, HTTPS, WebSockets.

Адресное пространство. Сервер OPC связан с объектами наблюдения. Объект представляется как набор переменных, значения которых

можно считать, записать или получить уведомление об их изменении, а также методов, которые клиент может вызывать и получать уведомления после завершения их работы. Совокупность объектов, которые сервер предоставляет клиенту, называется адресным пространством (address space). Адресное пространство представлено клиенту как совокупность узлов (nodes), соединенных связями (references). Для мониторинга состояния узлов из адресного пространства клиентом создаются отслеживаемые узлы (monitored items). Отслеживаемые узлы могут обнаружить изменение данных или возникновение какого-либо события. После чего они генерируют уведомление (notification), которое отправляется клиенту через подписку (subscription).

Параметры режима подписки. Для каждого отслеживаемого узла определяется интервал выборки (sampling interval) – интервал времени проверки сервером источника данных на наличие изменений. Выборка данных может быть быстрее, чем посылка уведомлений клиенту, т.е. меньше интервала опубликования (publish interval). В этом случае сервер буферизирует данные, а потом публикует сразу все накопленные в очереди данные. Размер очереди (queue size) – также настраиваемая величина. Когда данные доставляются клиенту, т.е. происходит публикация (publish), очередь очищается. Каждая очередь имеет политику сброса в случае переполнения, например отбрасывается самое старое значение.

Таким образом, оптимизация работы протокола OPC UA возможна на базе варьирования параметров его коммуникационного стека, а также настройки режима подписки.

Исследования производительности протокола OPC UA

Современные исследования в области выявления оптимальных режимов работы и узких мест протокола OPC UA идут по широкому спектру направлений. Выясняют различные зависимости, например зависимость задержки передачи данных от размера передаваемых данных [2] либо влияние количества клиентов на загруженность серверной машины [3]. При этом исследования могут отличаться предметной областью. Например, проводят тестирование производительности протокола OPC UA в умных электрических сетях [4] или в малоресурсных системах автоматизации [5]. Также исследования отличаются по используемым языкам программирования, библиотекам и средствам разработки (SDK), например в [6] исследуются различные характеристики OPC

UA при реализации на Java. Имеются работы, в которых сравнивают OPC UA с другими протоколами: MQTT [7], CoAP [8].

В данной статье все существующие направления исследований разделены на два класса: анализ скоростных характеристик, выявление требований протокола к аппаратным ресурсам.

Исследование скоростных характеристик OPC UA. Получение технологических данных в рамках OPC UA протокола возможно либо посредством опроса, либо в режиме подписки. В первом случае в качестве скоростной характеристики протокола используют время реакции на запрос данных (round-trip time), т.е. временной интервал между моментом, когда клиент выдал запрос на считывание одной или набора переменных, и моментом, когда соответствующие значения переменных доставлены клиенту. Во втором случае в качестве скоростной характеристикой протокола используют задержку (delay), т.е. временной интервал между моментом, когда новое значение переменной сохранено на сервере, и моментом, когда клиент получает это значение. Стоит отметить, что иногда более важно достичь максимальной общей пропускной способности сети, а не локальных быстрых ответов.

Данные по времени реакции на запрос (без влияния сети) для сообщений *малого* размера в различных статьях варьируются от 0.2 мс [1] до 50 мс [9]. При этом сообщается, что время реакции на запрос растет при увеличении размеров передаваемых сообщений. Например, чтение *большого* блока данных в 2.5 МБ может занимать около 200 мс [10]. Также в [10] экспериментально показано, что время реакции на запрос зависит от количества наблюдаемых переменных и скорости сети, но не от типа данных (булевый, целочисленный, вещественный).

В [11] сообщается, что шифрование, но не подписание, сообщений существенно увеличивает время реакции на запрос. Исследователи [2] утверждают, что при малых размерах сообщений шифрование удваивает время реакции на запрос, операции подписания и шифрования сообщений занимают практически одинаковое время, с увеличением размера передаваемых сообщений влияние операций подписания и шифрования сообщений на скорость передачи данных заметно уменьшается.

Согласно [1], выбор транспортного протокола влияет на время реакции на запрос данных: при использовании протокола SOAP/HTTP с кодировкой XML значительно возрастают временные издержки по сравнению с UA TCP.

В результате проведенного анализа можно выделить следующие факторы, которые существенно влияют на скоростные характеристики

протокола OPC UA: 1) механизмы обновления данных в адресном пространстве сервера OPC UA, в том числе такие параметры подписки как интервал выборки и интервал опубликования; 2) размеры передаваемых данных; 3) настройка режима безопасности. Используемый язык программирования или операционная система мало влияют на производительность протокола OPC UA.

Исследование ресурсоемкости OPC UA. Для обеспечения требуемого уровня качества обслуживания (заданного количество клиентов, скорости передачи данных) необходима оценка требований, который протокол предъявляет к аппаратным ресурсам, что количественно может быть выражено через процент использования ресурсов.

Многие исследования показали, что нагрузка на сервер OPC зависит от количества клиентов, количество подписок, количества наблюдаемых переменных. Так, например, в [10] оценивалось два разных профиля: случай встроенного сервера (500 переменных при 5 подписках) и случай стандартного сервера (37500 переменных при 75 подписках). Загрузка процессора Intel Atom™ Z530 на «встроенном сервере» при интервале выборки в 100 мс составила 2%. Загрузка того же процессора на «стандартном сервере» при интервале выборки в 500 мс составила 40%.

В [3] показано, что загрузка процессора серверной машины почти линейно зависит от количества отслеживаемых клиентом узлов. Это же исследование показывает, что обработка около 1000 отслеживаемых узлов может привести к сильной загруженности процессора сервера, расположенного на средномощной машине, что, в свою очередь, приводит к некорректному обновлению данных.

Кроме того, исследователи выделяют такой комплексный фактор, оказывающий влияние на загруженность процессора сервера, как механизм обработки значений технологических данных на сервере, отвечающий за сбор данных с источников, обработку данных внутри сервера и последующую передачу данных клиентам.

Заключение

Проведен обзор исследований, связанных с задачей достижения высокой производительности протокола OPC UA. В качестве метрик производительности используются скоростные характеристики протокола и различные показатели потребления аппаратных ресурсов. Выделены факторы, существенно влияющие на производительность протокола OPC UA.

ЛИТЕРАТУРА

1. *Mahnke W., Leitner S.H., Damm M.* OPC Unified Architecture. Berlin; Heidelberg: Springer-Verlag, 2009. 339 p.
2. *Cavalieri S., Chiacchio F.* Analysis of OPC UA performances // Computer Standards & Interfaces. 2013. V. 36. Iss. 1. P. 165–177.
3. *Fojcik M., Folkert K.* Introduction to OPC UA performance // Communications in Computer and Information Science. 2012. V. 291. P. 261–270.
4. *Claassen A., Rohjans S., Lehnhoff S.* Application of the OPC UA for the smart grid // Proceedings of 2nd IEEE PES International Conference and Exhibition on Innovative Smart Grid Technologies. 2011. P. 1–8.
5. *Burger A., Koziolok H., Rückert J., Platenius-Mohr M., Stomberg G.* Bottleneck identification and performance modeling of OPC UA communication models // Proceedings of 10th ACM/SPEC International Conference on Performance Engineering. 2019. Article 4. 12 p.
6. *Folkert K., Fojcik M., Cupek R.* Efficiency of OPC UA communication in Java-based implementations // Communications in Computer and Information Science. 2011. V. 160. P. 348–357.
7. *Rocha M.S., Sestito G.S., Dias A.L., Turcato A.C., Brandão D.* Performance comparison between OPC UA and MQTT for data exchange // Workshop on Metrology for Industry 4.0 and IoT. Brescia, 2018. P. 175–179.
8. *Durkop L., Czybik B., Jasperneite J.* Performance evaluation of M2M protocols over cellular networks in a lab environment // 18th International Conference on Intelligence in Next Generation Networks. Paris, 2015. P. 70–75.
9. *Freund M., Martin C., Braune A., Steinkrauss U.* JSUA – An OPC UA JavaScript framework // IEEE 18th Conference on Emerging Technologies & Factory Automation (ETFA). Cagliari, 2013. P. 1–4.
10. *Intel, AscoLab, Unified Automation.* Reducing Product Development Effort for OPC Unified Architecture. White paper, 2009.
11. *Post O., Seppälä J., Koivisto H.* The performance of OPC-UA security model at field device level // Proceedings of the 6th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2009). Milan, 2009. P. 337–341.

**МАТЕМАТИЧЕСКОЕ
И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ
ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ**

**MATHEMATICAL AND COMPUTER MODELLING
OF TECHNOLOGICAL PROCESSES**

Stability analysis and motion of the Kapitza pendulum*

Mary Michel Begre, Warren Kassy Dougg Feussi,
Shakhmurad Kanzitdinov, Sergey Vasilyev

RUDN University, Moscow, Russia

The generalization of Kapitza pendulum is investigated. Lyapunov stability analysis of the motion for this pendulum subjected to excitation of periodic driving forces that act in the vertical and horizontal planes has been studied with a small parameter.

We study the generalization of the Kapitza pendulum in the preposition of the existence of vertical and horizontal periodic oscillations of the suspension point. The Kapitza pendulum is a rigid pendulum in which the suspension point vibrates vertically, up and down. The unique feature of the Kapitza pendulum is that the vibrating suspension can cause it to balance steadily in an inverted position (inverted pendulum). A pendulum with a vibrating point is a classical problem that is actively investigated by methods of perturbation theory. The phenomenon of stabilization of the upper vertical position of the pendulum with fast vertical oscillations of the suspension point was discovered by A. Stephenson [11, 12]. P.L. Kapitza developed a method for analyzing the stability of an inverted pendulum by using the separation of slow and fast movements for such a pendulum [5, 6]. Various aspects of this problem have been discussed in many publications [1, 2, 8–10].

It is well known that time-varying dynamic systems are more difficult to analyze than time-invariant systems [13–15]. It is also known that for linear time-invariant dynamical systems. There are only two types of stability such

* The publication was financially supported by RFBR according to the research project No.18-07-00567.

as Lyapunov stability and asymptotic stability, which are completely determined by the set of eigenvalues of the system matrix. However, for linear systems that change over time, there are more types of stability, such as non-uniformly asymptotic stability, uniformly asymptotic stability, non-uniformly exponential stability, and uniformly exponential stability. In addition, in contrast to linear time-invariant dynamical systems, the stability of linear time-varying systems can not be directly related to the set of eigenvalues of matrices [14], so the analysis of time-varying dynamical systems is more complex [3, 4, 7].

In this paper, we propose a model of a generalized Kapitza pendulum, in which the suspension point moves in the vertical and horizontal planes, and analyze its Lyapunov stability in the presence of a small parameters in the system, which is the length of the suspension. The results of numerical investigation of the motion of the generalized Kapitza pendulum are also presented. The existence of a stable motion of this pendulum is shown.

Mathematical model of the Kapitza pendulum

Let the function $x(t)$ и $y(t)$ are horizontal and vertical Cartesian coordinates of the suspension point, l is the length of the massless pendulum rod and m is a mass of the pendulum. Let $\theta(t)$ be corner between the pendulum rod and the vertical. Position coordinates of a pendulum bob have the following form

$$\begin{aligned} x(t) &= l \sin \theta(t) + a(t) \sin \omega_1 t, \\ y(t) &= l \cos \theta(t) + b(t) \sin \omega_2 t, \end{aligned} \tag{1}$$

where ω_1 and ω_2 are frequencies of harmonic vertical and horizontal forced suspension vibrations, functions $a = a(t)$ и $b = b(t)$ are time-dependent amplitudes of forced oscillations along the axes x and y . If we take a derivative of the time from the functions $x(t)$ and $y(t)$ in (1), then we can write the coordinates of the velocity of a pendulum bob in the form

$$\begin{aligned} v_x(t) &= l \cos \theta(t) \dot{\theta}(t) + a \omega_1 \cos \omega_1 t, \\ v_y(t) &= -l \sin \theta(t) \dot{\theta}(t) + b \cos \omega_2 t, \end{aligned}$$

where functions $v_x(t) = \dot{x}(t)$, $v_y(t) = \dot{y}(t)$ are velocities along each of the axes.

Let functions $K(T)$ and $V(t)$ are kinetic and potential pendulum energies, then we can present them in the form

$$K(t) = 0.5m(\dot{x}^2(t) + \dot{y}^2(t)), V(t) = mg y(t),$$

$$K(t) = 0.5m \left[\left(l \cos \theta(t) \dot{\theta}(t) + a \omega_1 \cos \omega_1 t \right)^2 + \left(-l \sin \theta(t) \dot{\theta}(t) + b \omega_2 \cos \omega_2 t \right)^2 \right],$$

$$V(t) = mg(l \cos \theta(t) + b \sin \omega_2 t),$$

where g is the acceleration of free fall.

The full energy of the system is the sum of kinetic and potential energies

$$E(t) = K(t) + V(t) = 0.5m^{-1} \left(p_x^2(t) + p_y^2(t) \right) + mg y(t),$$

where $p_x(t) = mv_x$ and $p_y(t) = mv_y$ are momentum projection $p(t)$ on the axis x and y .

Lagrangian of such a system has the form

$$L(t) = K(t) - V(t) = 0.5m \left(\dot{x}^2(t) + \dot{y}^2(t) \right) - mgl \cos \theta(t).$$

The Euler – Lagrange equation for pendulum model has the form

$$\frac{d}{dt} \frac{\partial L(t)}{\partial \dot{\theta}} - \frac{\partial L(t)}{\partial \theta} = 0,$$

and we can write the pendulum movement equation in this case

$$l\ddot{\theta}(t) - a(t)\omega_1^2 \sin \omega_1 t \cos \theta + \left(b(t)\omega_2^2 \sin \omega_2 t - g \right) \sin \theta = 0.$$

Let the length of the massless pendulum rod be a small parameter in this equation $l \ll 1$. Let's introduce the generally accepted designation of a small parameter $\varepsilon = l$ for convenience in this case, then the equation has the form

$$\varepsilon \ddot{\theta}(t) - a(t)\omega_1^2 \sin \omega_1 t \cos \theta + \left(b(t)\omega_2^2 \sin \omega_2 t - g \right) \sin \theta = 0,$$

which is singularly perturbed equation, since if we put the value of a small parameter to zero $\varepsilon = 0$, then we can get a degenerate equation in the form

$$\text{tg } \theta = a\omega_1^2 \sin \omega_1 t / \left(b\omega_2^2 \sin \omega_2 t - g \right),$$

which is not a differential equation, but an algebraic equation.

We can rewrite the singularly perturbed equation in the form of a first-order differential equation system with a small parameter ε in the form

$$\begin{aligned} \dot{\theta} &= \varphi(t), \\ \varepsilon \dot{\varphi}(t) &= a\omega_1^2 \sin \omega_1 t \cos \theta(t) - \left(b\omega_2^2 \sin \omega_2 t - g \right) \sin \theta(t), \end{aligned} \quad (2)$$

where $\varphi(t)$ is an auxiliary variable.

We can analyze the Lyapunov stability for the pendulum movement in case when the length of the suspension is a small parameter $l \ll 1$.

Let θ_0 and φ_0 are parameters of the pendulum equilibrium and let $\Delta\theta$ and $\Delta\varphi$ are deviations from the equilibrium state θ_0 and φ_0 . We can describe the movement of the pendulum as follows in this case

$$\theta = \theta_0 + \Delta\theta(t), \quad \varphi = \varphi_0 + \Delta\varphi(t).$$

We can find parameters θ_0 and φ_0 from the solution of such system

$$\begin{aligned} \varphi_0 &= 0, \\ \operatorname{tg} \theta_0 &= a\omega_1^2 \sin \omega_1 t / (b\omega_2^2 \sin \omega_2 t - g). \end{aligned}$$

This equation is exactly the same as the equation for a degenerate case $\varepsilon = 0$.

We can get a system of differential equations to study functions $\Delta\theta(t)$ and $\Delta\varphi(t)$ if we use linearization method

$$\begin{aligned} \Delta\dot{\theta} &= \Delta\varphi(t), \\ \Delta\dot{\varphi} &= -\varepsilon^{-1} \left(\sqrt{(a\omega_1^2 \sin \omega_1 t)^2 + (b\omega_2^2 \sin \omega_2 t - g)^2} \right) \Delta\theta(t). \end{aligned}$$

The characteristic polyn of this system has the form

$$D(p) = p^2 + \varepsilon^{-1} \left(\sqrt{(a\omega_1^2 \sin \omega_1 t)^2 + (b\omega_2^2 \sin \omega_2 t - g)^2} \right).$$

We have the necessary and sufficient condition $D(p) > 0$ for the stability of the generalized movement of the Kapitza pendulum in this case. This stability condition take place when $\varepsilon > 0$ and for any ω_1, ω_2 и $t \geq 0$.

Simulation and numerical examples

The numerical study of the movement of the generalized pendulum of Kapitza is presented in this section. Let functions $a(t)$ и $b(t)$ have the forms like this $a(t) = A \sin v_1 t$, $b(t) = B \sin v_2 t$, where parameters A and B are amplitudes, v_1 and v_2 are the frequencies of harmonic vertical and horizontal forced vibrations of the suspension, then we can rewrite the system of differential equations (2) in the form

$$\begin{aligned} \dot{\theta} &= \varphi(t), \\ \dot{\varphi} &= \varepsilon^{-1} A v_1^2 \sin v_1 t \cos \theta(t) + \varepsilon^{-1} (B v_2^2 \sin v_2 t + g) \sin \theta(t). \end{aligned}$$

We have assumed parameters of the pendulum $l = 0.05$ m, $v_1 = 40$ s⁻¹, $v_2 = 10$ s⁻¹, $g = 10$ m/s² for numerical simulations. The result of the numerical study is presented on the Fig. 1. We can see that for these parameters of the system we can have a steady movement of the pendulum.

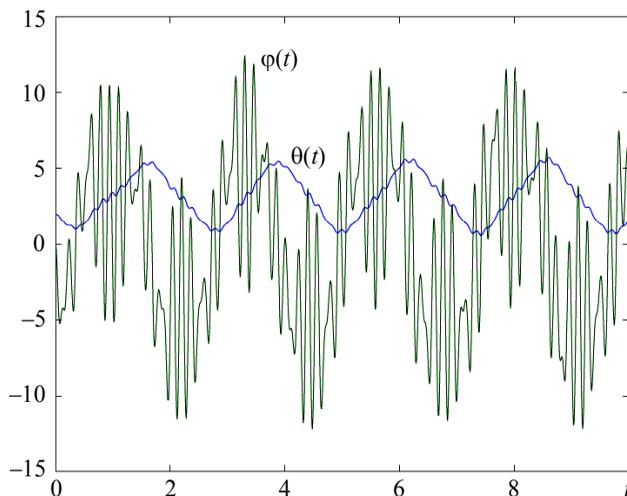


Fig. 1. The numerical study of dynamics pendulum

Conclusions

A generalization of the pendulum Kapitza is considered in this work. We investigate the type of inverted pendulum in the case when the suspension point moves in vertical and horizontal planes. We consider the Lyapunov stability of movement in the presence of a small parameter in the system. The numerical study of its movement is presented and showed the existence of the steady movement of this pendulum. Analysis of the movement of such a pendulum can be used to determine the stability parameters of buildings and structures in an earthquake.

REFERENCES

1. *Agarana M.C., Akinlabi E.T.* Lagrangian–Laplace dynamic mechanical analysis and modeling of inverted pendulum // *Procedia Manufacturing*. 2019. V. 35. P. 711–718.
2. *Bardin B.S., Markeyev A.P.* On the stability of equilibrium of a pendulum with vertical oscillations of its suspension point // *J. Appl. Math. Mech.* 1995. V. 59. P. 879–886.

3. *Du H., Qian C., Yang S., Li S.* Recursive design of finite-time convergent observers for a class of time-varying nonlinear systems // *Automatica*. 2013. V. 49. No. 2. P. 601–609.
4. *Jiang C., Teo K.L., Duan G.* A suboptimal feedback control for nonlinear time-varying systems with continuous inequality constraints // *Automatica*. 2012. V. 48. No. 4. P. 660–665.
5. *Kapitsa P.L.* Dynamic stability of a pendulum with oscillating point of suspension // *Sov. Phys. JETP (in Russian)*. 1951. V. 21. P. 588–597.
6. *Kapitsa P.L.* A pendulum with vibrating point of suspension // *Usp. Phys. Nauk (in Russian)*. 1951. V. 44. P. 7–20.
7. *Mazenc F., Andrieu V., Malisoff M.* Design of continuous–discrete observers for time-varying non-linear systems // *Automatica*. 2015. V. 57. P. 135–144.
8. *Levi M.* Geometry of Kapitsa's potential // *Nonlinearity*. 1988. V. 11. P. 1365–1368.
9. *Neishtadt A.I., Shenga K.* Bifurcations of phase portraits of pendulum with vibrating suspension point // 2016 <https://arxiv.org/abs/1605.09448>
10. *Ovseyevich A.I.* The stability of an inverted pendulum when there are rapid random oscillations of the suspension point // *J. Appl. Math. Mech.* 2006. V. 70. P. 761–768.
11. *Stephenson A.* On induced stability // *Philosophical Magazine Series*. 1908. V. 6. No. 15. P. 233–236.
12. *Stephenson A.* On a new type of dynamical stability // *Mem. Proc. Manch. Lit. Phil. Soc.* 1908. V. 52. No. 8. P. 1–10.
13. *Zhang X., Liu L., Feng G.* Leader–follower consensus of time-varying nonlinear multi-agent systems // *Automatica*. 2015. V. 52. P. 8–14.
14. *Zhou B., Cai G., Duan G.* Stabilization of time-varying linear systems via Lyapunov differential equations // *International Journal of Control*. 2013. V. 86. No. 2. P. 332–347.
15. *Zhou B.* Global stabilization of periodic linear systems by bounded controls with applications to space-craft magnetic attitude control // *Automatica*. 2015. V. 60. P. 145–154.

Kadyshevsky equation numerical analysis with periodic boundary conditions on adaptive grids*

Mohamed Adel Bouatta, Irina Kolosova,
Evgeniya Korshok, Darya Vasilyeva

RUDN University, Moscow, Russia

The boundary value problem for the relativistic Schrodinger equation (Kadyshevsky equation) which is a singularly perturbed differential equation of infinite order is considered. The truncation method is applied to analyze its solution. The numerical methods on an irregular grid is constructed and the solutions is obtained with a quasipotential.

The description of the interaction of relativistic particles within the quasi-potential approach is very widely known in modern physics. This description is based on the so-called covariant single formulation of quantum field theory, in which the theory is considered on the spatial three-dimensional hypersurability in the Minkowski space. This approach focuses on methods of constructing various quasi-potentials, as well as using a quasi-potential approach to describe the characteristics of relativistic particle interactions in quark models, such as the amplitudes of elastic hadron scattering, mass spectra and the width of meson decays, as well as cross-sectional sections of deep non-suprudian scattering of leptons on the adrons [1–5].

In this work, we formulate Sturm-Liouville problem with periodic conditions on the segment for the truncated $2m$ -order Kadyshevsky equation with a small parameter. A modification of the adaptive grids method is proposed for Kadyshevsky equation numerical analysis with periodic boundary conditions. The proposed modification grid is piecewise homogeneous Shishkin-type grids adapted to the small size of the boundary layer.

Mathematical model of bound states of elementary particles and numerical analisys

A method is proposed for constructing asymptotic solutions to the boundary value problem for the Kadyshevsky equation (a singularly

* The publication was financially supported by RFBR according to the research project No. 18-07-00567.

perturbed differential equation of infinite order) [1–5]

$$L\psi(r, l) = \lambda_c \psi(r, l), \quad (1)$$

$$L = H_{inf}^{rad} + V(r), \lambda_c = 2c\sqrt{q^2 + m^2 c^2},$$

$$\begin{aligned} H_{inf}^{rad} &= 2mc^2 \operatorname{ch}\left(\frac{i\hbar}{mc} D\right) + \frac{\hbar^2 l(l+1)}{mr\left(r + \frac{i\hbar}{mc}\right)} \exp\left(\frac{i\hbar}{mc} D\right) = \\ &= \sum_{p=0}^{\infty} \frac{(-1)^p 2mc^2}{(2p)!!} \left(\frac{\hbar}{mc}\right)^{2p} D^{2p} + \frac{\hbar^2 l(l+1)}{mr\left(r + \frac{i\hbar}{mc}\right)} \sum_{p=0}^{\infty} \frac{1}{p!} \left(\frac{i\hbar}{mc}\right)^p D^p, \end{aligned}$$

where $D^p = d^p/dr^p$ is differential operator, m – mass, q – momentum and $l = 0, 1, \dots$ – moment of bounded particles, $V(r)$ – quasipotential; \hbar/mc is a small parameter.

For this equation we can formulate boundary value problem with periodic boundary conditions for eigenfunctions and eigenvalues on segment $[0, r_0]$

$$\begin{aligned} L\psi(r, l) &= \lambda_c \psi(r, l), \\ \psi^i(0, l) &= \psi^i(r_0, l), \quad i = 0, 1, 2, \dots \end{aligned}$$

We can apply truncation procedure to the equation (1) and write it in the form

$$L^{2k} \psi_{2k}(r, l) = \lambda \psi_{2k}(r, l), \frac{1}{2}. \quad (2)$$

$$L^{2k} = H_{inf, 2k}^{rad} + V(r),$$

$$H_{inf, 2k}^{rad} = \sum_{p=0}^k \frac{(-1)^p 2mc^2}{(2p)!!} \left(\frac{\hbar}{mc}\right)^{2p} D^{2p} + \frac{\hbar^2 l(l+1)}{mr\left(r + \frac{i\hbar}{mc}\right)} \sum_{p=0}^k \frac{1}{p!} \left(\frac{i\hbar}{mc}\right)^p D^p,$$

for this equation we can also formulate boundary value problem with periodic boundary conditions for eigenfunctions and eigenvalues on segment $[0, r_0]$

$$\begin{aligned} L^{2k} \psi_{2k}(r, l) &= \lambda \psi_{2k}(r, l), \\ \psi_{2k}^i(0, l) &= \psi_{2k}^i(r_0, l), \quad i = 0, 1, 2, \dots, 2k - 1. \end{aligned}$$

For equation (2) with quasipotential $V(0)$ for $k = 4, 6$ an irregular mesh has constructed on the segment $[0, r_0]$. This grid is modification of piecewise homogeneous Shishkin-type grids adapted to the small size of the boundary layer [6, 8]. It is shown that the numerical solution of the boundary value problem (3) converges uniformly for different values of a small parameter.

Conclusions

We formulate Sturm – Liouville problem with periodic conditions on the segment for the truncated $2m$ -order Kadyshevsky equation with a small parameter. We proposed a modification of the adaptive grids method for the truncated 4 and 6-order Kadyshevsky equation numerical analysis with periodic boundary conditions. The proposed modification grid is piecewise homogeneous Shishkin-type grids adapted to the small size of the boundary layer. This type of grids guarantee uniform convergence of numerical solutions when a small parameter ε changes to zero.

REFERENCES

1. *Kadyshevsky V.G.* // Nucl. Phys. 1968. V. B6. P. 125.
2. *Kadyshevsky V.G., Mateev M.* // Nuovo Cimento. 1967. V. 55A. P. 275.
3. *Kadyshevsky V.G., Mir-Kasimov R.M., Skachkov N.B.* // Nuovo Cimento. 1968. V. 55A. P. 1233.
4. *Амирханов И.В., Жидков Е.П., Жидкова И.Е., Васильев С.А.* // Дифференциальные уравнения. 2001. Т. 37. № 1. С. 83–90.
5. *Амирханов И.В., Жидков Е.П., Жидкова И.Е., Васильев С.А.* // Математическое моделирование. 2003. Т. 15. № 9. С. 3–16.
6. *Andreev V., Savin I.* The uniform convergence with respect to a small parameter of A.A. Samarskii's monotone scheme and its modification // Computational Mathematics and Mathematical Physics. 1995. V. 35. P. 581–591.
7. *Linss T., Roos H.-G., Vulanovic R.* Uniform pointwise convergence on Shishkin-type meshes for quasi-linear convection-diffusion problems // SIAM Journal on Numerical Analysis. 2000. V. 38. No. 3. 2000. P. 897–912.
8. *Vulanovic R., Teofanov L.* A modification of the Shishkin discretization mesh for one-dimensional reaction–diffusion problems // Applied Mathematics and Computation. 2013. V. 220. P. 104–116.

A reaction-diffusion-advection competition model with a free boundary

Jozil Takhirov

Institute of Mathematics, Tashkent, Uzbekistan

In this paper, we study a competitive diffusion quasilinear system with a free boundary. First, we investigate the mathematical questions of the problem. A priori estimates of Schauder type are established, which are necessary for the solvability of the problem. One of two competing species is an invader, which initially exists on a certain sub-interval. The other is initially distributed throughout the area under consideration. Examining the influence of baseline data on the success or failure of the first invasion. We conclude that there is a dichotomy of spread and extinction and give precise criteria for spread and extinction in this model.

Introduction

Today, humanity is faced with increasingly serious environmental and epidemiological problems, such as environmental pollution, invasion of exotic species, the emergence of new infectious diseases and the resumption of existing epidemiological diseases. Mathematical modeling has been used successfully to study many biomedical and epidemiological problems, and in all these contexts non-linear and complex dynamics have been observed.

Over the past twenty years, significant progress has been made in the mathematical modeling of biomedical processes, which has led to the creation of more complex models consisting of systems of nonlinear partial differential equations.

It turns out that the theory developed for coupled systems of parabolic and elliptic equations can be applied to a number of reaction-diffusion models in many areas of applied science. This includes models emerging in biology, ecology, and biomedicine. The bulk of the theory developed in famous articles (see, eg, [1, 2]) is motivated by these model problems.

In the work [3] Pao C.V. and Ruan W. investigate the existence, uniqueness, and asymptotic behavior of solutions for a coupled systems of quasilinear parabolic equations under nonlinear boundary conditions.

Formulation of the problem

In this article discusses the existence and qualitative properties of solutions of a competitive-diffusion system of quasilinear parabolic equations in a domain with a free boundary. The problem of describing the dynamic process of the invasion of a new competitor into the habitat of a local species belongs to Du and Lin [4], who investigated the nonlinear evolution of two species in an unbounded spatial domain.

We focus on the case where the indigenous population undergoes expansion and growth in a limited area to be more realistic, and we investigate the model of diffuse competition described by the following problem:

$$\begin{cases} a(u)u_t - d_1u_{xx} - k_1u_x = u(a_1 - b_1u - c_1v), & t > 0, \quad 0 < x < s(t), \\ b(v)v_t - d_2v_{xx} - k_2v_x = v(a_2 - b_2u - c_2v), & t > 0, \quad 0 < x < l, \\ u(0, x) = u_0(x), \quad 0 \leq x \leq s_0 = s_0 < l, \\ v(0, x) = v_0(x), \quad 0 \leq x \leq l, \\ u_x(t, 0) = v_x(t, 0) = v(t, l) = 0, \quad t \geq 0, \\ u(t, s(t)) = 0, \quad t \geq 0, \\ \dot{s}(t) = -\mu u_x(t, s(t)), t > 0, \end{cases} \quad (1)$$

where $x = s(t)$ is called a free boundary, which is to be determined together with u and v , all the parameters $a_i, b_i, c_i, d_i, k_i (i=1,2)$ and μ are given positive constants.

The coefficients of biocapacity $a(u), b(v)$ and $u_0(x), v_0(x)$ satisfy:

$$\begin{cases} a(u) \geq a_0 > 0, b(v) \geq b_0 > 0 \text{ and Holder continuous in } Q_T, \\ u_0(x) \in C^2([0, s_0]), u'_0(0) = u_0(s_0) = 0 \text{ and } u_0(x) \text{ in } [0, s_0], \\ v_0(x) \in C^2[0, l], v_0(l) = 0, v'_0(0) = 0, v_0(x) \geq 0 \text{ in } [0, l], \end{cases}$$

where $Q_T = \{(t, x) : 0 < t \leq T, 0 < x < l\}$.

This model describes how a new species with population density u invades into the habitat of a native competitor v .

Since the natural biological process is associated with invasion, we consider it necessary to add terms reflecting the transfer and change in concentration to the equations using the convective transfer term and the new «biocapacity» coefficient.

In [4], the authors considered two cases: superior and inferior competitor, when u is the superior competitor, and it was proved that the spread-disappearance dichotomy takes place. Namely, as $t \rightarrow +\infty$, either $h(t) \rightarrow \infty$ and $(u, v) \rightarrow (u^*, 0)$, or $h_\infty < \infty$ and $(u, v) \rightarrow (0, v^*)$. Problem (1) with $a(u) \equiv 1, b \equiv 1$ was studied in [4].

The main contribution of this article is the establishment of the global existence of the classical solution to problem (1) and the study of the behavior of the solution as $t \rightarrow \infty$. A method is proposed for establishing a priori Schauder-type estimates for a new class of free boundary problems for mixed-two-phase equations.

A priori estimates

Local existence and uniqueness results are valid for any quasi-linear parabolic equation, when the given functions have enough smoothness, without any restrictions on the growth type of these functions with respect to u and u_x (see e.g. [5, 6]).

Such conditions are necessary, when the global solvability of the boundary problems is considered. The most challenging aspect of the proof is bounding the spacial gradient of the solution.

For functional spaces and norms, we will employ the notations of [5, 6], and we will also make use of its results.

Lemma 1. Let $(u, v, s(t))$ be a solution of (1) for $t \in [0, T], T > 0$. Then

$$\begin{aligned}
 0 < u(t, x) &\leq \max \left\{ \frac{a_1}{b_1}, \|u_0\|_\infty \right\} = M_1, \quad t > 0, \quad x \in [0, s(t)], \\
 0 < v(t, x) &\leq \max \left\{ \frac{a_2}{c_2}, \|v_0\|_\infty \right\} = M_2, \quad t > 0, \quad x \in [0, l], \\
 0 < \dot{s}(t) &\leq M_3, \quad t > 0,
 \end{aligned} \tag{*}$$

where M_3 depending only data.

The estimate (*) does not guarantee the continuation of the two-phase process till prescribed value of time, that is, $s(t) < l, 0 < t \leq T$. On the other hand, the $s(t) < l$ case is important one from a practical perspective.

For each equation of the system, we formulate the corresponding problem:

$$\begin{cases} u_t = \bar{a}(u)u_{xx} + f_1(u, v, u_x), & D = \{(t, x) : t > 0, 0 < x < s(t)\}, \\ u(0, x) = u_0(x), & 0 \leq x \leq s_0, \\ u_x(t, 0) = 0, & u(t, s(t)) = 0, \quad t > 0, \\ \dot{s}(t) = -\mu u_x(t, s(t)); \end{cases} \quad (\text{A})$$

$$\begin{cases} v_t = \bar{b}(v)v_{xx} + f_2(u, v, v_x), & Q = \{(t, x) : t > 0, 0 < x < l\}, \\ v(0, x) = v_0(x), & 0 \leq x \leq l, \\ v_x(t, 0) = 0, & v(t, l) = 0, \end{cases} \quad (\text{B})$$

where $\bar{a}(u) = \frac{d_1}{a(u)}$, $f_1(\cdot) = \frac{1}{a(u)}[K_1 u_x + u(a_1 - b_1 u - c_1 v)]$,

$$\bar{b}(v) = \frac{d_2}{b(v)}, f_2(\cdot) = \frac{1}{b(v)}[K_2 v_x + v(a_2 - b_2 u - c_2 v)].$$

Theorem 1. Assume that the conditions of Lemma 1 are satisfied and let a continuous in \bar{Q} function v satisfies the conditions of (B). If $v(t, x)$ has derivatives v_{tx}, v_{xx} that are square-integrable in Q , then

$$|v_x(t, x)| \leq M_4(M_2, b_0, d_2, v_0), \quad (t, x) \in \bar{Q},$$

$$|v|_{1+\gamma}^Q \leq M_5(M_4), \quad |v|_{2+\beta}^Q \leq M_6(M_5).$$

Proof. The proof is based on the results of [5]. In the case of (A), using the transformation $\tau = t, y = \frac{x}{s(t)}$, we straighten out the boundary. Then D will be mapped to $\Omega = \{(\tau, y) : 0 < \tau, 0 < y < 1\}$, while for the function $w(\tau, y) = u(\tau, ys(\tau))$ we obtain an equation with bounded coefficients and a right-hand side. Using the results of [5, 6], we establish bounds for $|w_y|, |w|_{1+\gamma}$ in $\bar{\Omega}$. For the highest order derivatives, the bounds are obtained using the linear theory [6].

The existence result

Theorem 2. Suppose that the conditions of Theorem 1 and Lemma 1 are satisfied. Then there exists in $D \cup Q$ a solution $u(t, x) \in C^{2+\alpha}(\bar{D})$, $v(t, x) \in C^{2+\beta}(\bar{Q})$, $s(t) \in C^{1+\gamma}$ ($0 \leq t \leq T$) problems (B).

To prove the solvability of a nonlinear problem, one can use various theorems from the theory of nonlinear equations, remembering that the uniqueness theorem of the classical solution holds for it. We apply the Leray – Schauder principle [6], the established a priori estimates $|\cdot|_{1+\alpha}^Q$ for all possible solutions of nonlinear problems and the solvability theorem in the Hölder classes for linear problems. Moreover, the existence theorems for systems are the same as the theorem for the case of a single equation, since each of the equations can be considered as a linear equation for $u(t, x)$ and $v(t, x)$ with Hölder-continuous coefficients.

Problem (B) is considered simultaneously with a one-parameter family of problems of the same type. The linear problem defines the transformation $w = F(\omega, k)$, $0 \leq k \leq 1$, to which the Leray-Schauder principle is applied. This operator is non-linear and depends on k . Its fixed points with $k = 1$ are solutions to the problem.

Denote by $H^{1+\alpha}$, $\alpha \in (0, 1)$ the Banach space of functions $u(t, x), v(t, x)$ on \bar{Q} with norm $\|\cdot\| = |\cdot|_{1+\alpha}^Q$, which satisfy the corresponding initial and boundary conditions of problem (A) and (B).

For each function $\bar{u}, \bar{v} \in H^{1+\alpha}$ and any number $k \in [0, 1]$ we denote by u^k, v^k solutions to linear problems (A) and (B), the solutions of which exist and are unique; moreover, $u^k, v^k \in C^{2+\alpha}, s(t) \in C^{1+\alpha}$. Moreover, in the region D , as in Theorem 1, we pass to the parabolic equation with Hölder-continuous coefficients in a fixed domain.

Uniform continuity and complete continuity of the transformation operator F with respect to k , uniform estimates for k for solutions, and solvability of linear problems follows from the established a priori estimates of the Hölder norms. The technique of proof is demonstrated in detail, for example, in (Chap. VI, [6]). This completes the proof of the theorem.

On the asymptotic behavior of the solutions

In a competitive environment, each population shrinks at a rate proportional to the size of its competitor's. The problem in ecology is whether two competing species coexist or one species destroys the other. The mathematical question is whether the solution (u, v) of the reaction diffusion system remains positive at $t \rightarrow \infty$, or one (or both) components u, v disappear at $t \rightarrow \infty$. The coexistence of competing components is closely related to the existence of a positive stationary solution and the asymptotic behavior of the time-dependent solution with respect to the stationary solution.

In [3] the authors investigate a coupled system of quasilinear parabolic equations in bounded domain under coupled nonlinear boundary conditions.

To investigate the asymptotic behavior of the solution of the system under consideration, it will be necessary to investigate the corresponding quasilinear elliptic system. Then, using the replacement, the problem for the parabolic system is reduced to an equivalent problem, and the system of equations of the equivalent problem in the case $i = 1, 2$ coincides in structure with the system of equations of our problem (1). This allows us to take advantage of some of the results of the works ([3], Theorem 6.4) and ([2], Ch.12.4, Theorem 4.2).

The main results: The terms of convective transfer and «biological capacity» are introduced, methods of establishing a priori estimates of the Schauder type are proposed, the asymptotic behavior of the solution is studied.

REFERENCES

1. *Murray J. D.* Mathematical biology. II: Spatial models and biomedical applications. Berlin: Springer, 2003.
2. *Pao C.V.* Nonlinear Parabolic and Elliptic Equations. NY: Plenum Press, 1992.
3. *Pao C.V., Ruan W.* Positive solutions of quasilinear parabolic systems with nonlinear boundary conditions // J. Math. Anal. Appl. 2007. V. 333. P. 472–499.
4. *Du Y.H., Lin Z.G.* The diffusive competition model with a free boundary: invasion of a superior or inferior competitor // Discrete Contin. Dyn. Syst., B. 2014. V. 19. P. 3105–3132.
5. *Kruzhkov S.N.* Nonlinear parabolic equations with two independent variables // Transaction of the Moscow Mathematical Society. 1967. V. 16. P. 329–346.
6. *Ladyzhenskaya O.A., Solonnikov V.A., Ural'ceva N.N.* Linear and Quasilinear Equations of Parabolic Type // Amer. Math. Soc., Providence, RI. 1968.

Problem definition for LEO system switching technique development

Sergey Pichugin

Rocket and Space Corporation Energia named by S.P. Korolev, Korolyov, Russia

Related to the latest tendencies of deployment LEOCS (Low Earth Orbit Communication System), having their capabilities almost equal to 5-G terrestrial communications networks, investigations have been launched on performing switching on board of relay satellites [1, 2].

So the task have been put forward to produce an analytical model of SOBS to help developing switching technique for LEOCS to become competitive against terrestrial communications networks in view of technical capabilities, price and time of deployment. The technique is essentially the one for controlling of decision-making and optimization in problem-oriented system of technical objects.

Low Earth orbit communication system description

A sample LEOCS is chosen for investigation, consisting of community of subscribers (subs) and relay satellites (SOBS). The subs are independent in view of their transmit-receive sessions of information exchange, and SOBS are orbiting Earth by circular paths as it is shown in Fig. 1.

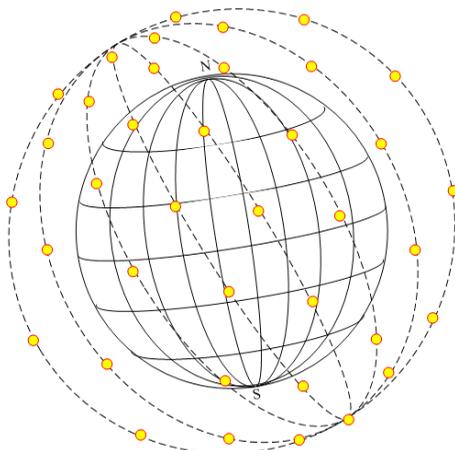


Fig. 1. LEOCS with SOBS on circular Earth orbits

All terrestrial surface is zoned in accordance with number of SOBS [3, 4], these interact with subs by up and down directions, and the messages are addressed to subs in accordance with the zone number of a subscriber. Alternatively to SOBS, the zones are fixed against surface of Earth.

The sample addressing is presented in Fig. 2.

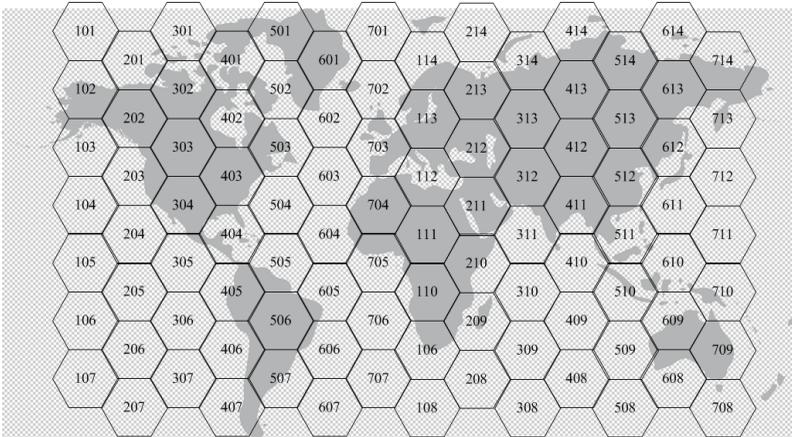


Fig. 2. Addressing (sample) in LEOCS

The model considers a routine operation, when SOBS move synchronously and in orderly manner, from one zone to another, so the order «one satellite per one zone» is maintained. Each SOBS interacts over intersatellite links (ISLs) with his closest neighboring SOBS within his orbit and in two neighboring orbits. The sample interaction of SOBS over ISLs is shown in Fig. 3.

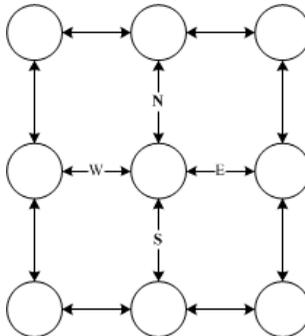


Fig. 3. Sample interaction of SOBS over ISLs

W, N, E, and S symbols correspond to westbound, northbound, eastbound and southbound ISLs of SOBS.

Given mutual movements of subs and SOBS relative to each other and the zones mentioned above, a primary model is developed for control of decision-making and optimization in problem-oriented system of technical objects. Data flows in one separate SOBS as per the model are presented in Fig. 4.

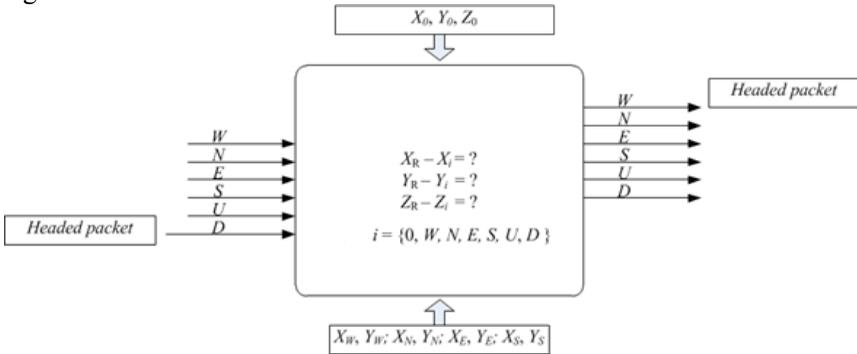


Fig. 4. Data flows sample in one separate SOBS

In accordance with the primary model, headed packet arrives to input D of SOBS switch. The header contains codes X_R , Y_R , Z_R to deliver the packet to recipient subscriber. Codes of own coordinates of SOBS X_0 , Y_0 и Z_0 and its neighboring SOBS $X_W, Y_W; X_N, Y_N; X_E, Y_E; X_S, Y_S$ arrive to the second and to the third inputs of the switch. The switch makes decision to transport the packet from its input to its corresponding output W based on the data received. Symbol «0» means SOBS failure, symbols U and D mean up and down directions.

Mathematical modelling and tools for the investigation

For investigation of the system described above, we plan to use tools of theory of finite machines, tools of queuing theory and tools of probability theory. When the system is investigated by methods of queuing theory, primarily system idle time and refusal probability have to be found. Markov flows are planned to be applied in the model for requests for service arrival and serving to meet the required accuracy, so model M/M/N/m is planned to be used for ISL and for subscriber links investigation [6, 7]. We plan to use tools of queuing theory mainly for investigation of LEOCS and SOBS

regular operation, and tools of probability theory for failure scenarios investigation.

To simplify analysis of the model, data flows in SOBS, as per Fig. 4, are transformed into finite machine pattern [5], as it is presented in Fig. 5.

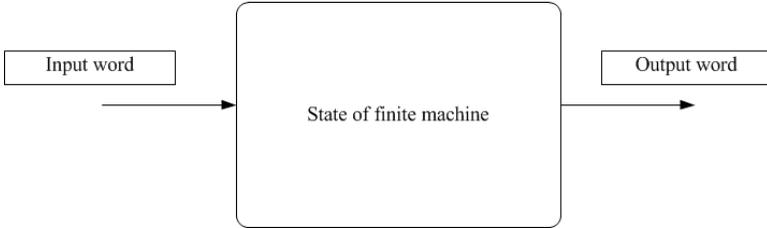


Fig. 5. SOBS presentation as finite machine

It worth mentioning that the finite machine pattern is applicable not only for SOBS investigation, but also for investigation of LEOCS.

Aimed to development of switching technique, analytical presentation of SOBS have to include technical performances of SOBS together with cost estimations of space communications services and timeframes for LEOCS implementation, and all these to be made in view of the end result, that is, based on the requirement to be compared against pre-set values. So, mathematically, the task of the development of the technique providing control of decision-making and optimization in problem-oriented system of technical objects, in accordance with Fig. 5, is formulated as follows:

$$\text{freal}(T, C, F) - \text{fforecast}(T, C, F) = \min, \text{ under restrictions } n = \text{const}, \quad (1)$$

where

freal(T, C, F) – end result function real value;

fforecast(T, C, F) – end result function forecasted value;

T – type of technical objects in problem-oriented system;

C – cost and timeframe of implementation of technical solution providing control of decision-making and optimization in problem-oriented system of technical objects;

F – performances of technical solution providing control of decision-making and optimization in problem-oriented system of technical objects;

n – types of technical objects, (object «Information» is transported from object «subscriber communications terminal» through object «group of SOBS on circular orbits» towards object «subscriber communications terminal»).

Conclusions

Summarizing the above, the task is defined to produce end-result oriented technique, minimizing controlling time of flow of information inside group of SOBS of LEOCS, synchronized by location and movements of orbits and SOBS. The task formulated defines the desired result, the factors and restrictions, so the result of solution of the task permits objective evaluation.

REFERENCES

1. *Pichugin S.B., Nazarov A.A., Krymov V.S.* Routing algorithm in satellite communications network // Book of abstracts of 9-th Byelorussian winter school-seminar on queuing theory. Minsk: Byelorussian State University, 1993. P. 87.
2. *Pichugin S.B.* Subscriber and intersatellite paths in LEO communications networks // «Radio industry (Russia)» Journal. 2019. V. 29. No. 3. P. 48–55.
3. *Satellite Relay «Aksai»* patent. 2097926 Russian Federation: IPC H 04 B 7/185 / Pichugin S.B.; applicant and patent holder Pichugin S.B. – № 94018404/09; priority date 19.04.1994; published 27.11.1997, Bulletin of Patents of Russian Federation № 33. 11 pages: illustrated.
4. *Tsunoda H., Ohta K., Kato N., Nemoto Y.* Supporting IP/LEO satellite networks by handover-independent IP mobility management // IEEE Journal on Selected Areas in Communications. 2004. V. 22. No. 2.
5. *Hill A.* Introduction to the Theory of Finite-State Machines. Moscow: Nauka, 1966. 272 p.
6. *Nazarov A.A., Terpugov A.F.* Queuing Theory. A Manual for Students in Higher Education. Tomsk: NTL Publ., 2010. 228 p.

Об особенностях выбора ортонормированных систем функций в рандомизированных численных проекционных функциональных алгоритмах*

Анжела Абдразакова¹, Татьяна Булгакова¹,
Антон Войтишек²

¹Новосибирский государственный университет, г. Новосибирск, Россия

²Институт вычислительной математики
и математической геофизики СО РАН, г. Новосибирск, Россия

1. Рандомизированные проекционные численные функциональные алгоритмы

Данная работа развивает соображения в области теории и приложений рандомизированных численных (компьютерных) функциональных алгоритмов, представленные в докладах [1–3]. В докладах [1, 2] (а также в книге [4]) рассматриваются рандомизированные численные алгоритмы приближения решения $\varphi^{(1)}(x)$ интегрального уравнения Фредгольма второго рода

$$\varphi^{(1)}(x) = \int k(x, x')\varphi(x') dx' + f(x), \quad (1)$$

используемые при численном решении задач переноса излучения, кинетических задач, краевых задач и многих других.

В книге [5] отмечено, что с методической точки зрения (с учетом того, что решение уравнения (1) записывается в виде ряда Неймана) целесообразно рассмотреть рандомизированные функциональные алгоритмы для приближения функции

$$\varphi^{(2)}(x) = \int g(x, y) dy, \quad (2)$$

т.е. для интеграла, зависящего от параметра. В работе [3] (и во многих других) рассматриваются численные алгоритмы приближения вероятностных плотностей

$$\varphi^{(3)}(x) = f_{\xi}(x), \quad x \in X \subset R^d \quad (3)$$

случайных величин (векторов) $\xi \in X$ при заданной выборке $\{\xi_1, \dots, \xi_n\}$

* Работа выполнена в рамках государственного задания ИВМиМГ СО РАН (проект 0315-2019-0002).

(такие алгоритмы применяются, например, при обработке больших данных).

Для приближения функций $\varphi(x) = \varphi^{(s)}(x)$, $s = 1, 2, 3$, используются *сеточные, проекционные и проекционно-сеточные (ядерные)* рандомизированные алгоритмы. В данной работе мы выделим новые (по сравнению, в частности, с работой [1]) важные свойства **рандомизированных проекционных функциональных алгоритмов**.

Идея этих алгоритмов состоит в приближении функции $\varphi(x)$ отрезком ряда длины M из разложения

$$\varphi(x) = \sum_{i=1}^{\infty} w^{(i)} \chi^{(i)}(x) \approx L^{(M)}\varphi(x) = \sum_{i=1}^M w^{(i)} \chi^{(i)}(x), \quad (4)$$

где $\Xi^{(\infty)} = \{\chi^{(1)}(x), \chi^{(2)}(x), \dots\}$, $\Xi^{(M)} = \{\chi^{(1)}(x), \dots, \chi^{(M)}(x)\}$; (5)

$$(\chi^{(i)}(x), \chi^{(j)}(x)) = \int \chi^{(i)}(x) \chi^{(j)}(x) dx = \delta^{(i,j)} = \{1 \text{ при } i = j; 0 \text{ иначе}\}$$

– заданная ортонормированная система функций. В рандомизированных проекционных алгоритмах коэффициенты

$$W^{(M)} = \{w^{(1)} = (\varphi, \chi^{(1)}), \dots, w^{(M)} = (\varphi, \chi^{(M)})\}$$

представляются в виде математических ожиданий

$$w^{(s,i)} = (\varphi^{(s)}, \chi^{(i)}) = E\zeta^{(s,i)} \quad (6)$$

и приближаются численно по формуле

$$w^{(s,i)} = E\zeta^{(s,i)} \approx \tilde{w}^{(s,i)} = \frac{\zeta_1^{(s,i)} + \dots + \zeta_n^{(s,i)}}{n},$$

где $\zeta_1^{(s,i)}, \dots, \zeta_n^{(s,i)}$ – соответствующие смоделированные на компьютере выборочные значения.

Окончательные приближения функций $\varphi(x) = \varphi^{(s)}(x)$, $s = 1, 2, 3$, имеют вид

$$\varphi^{(s)}(x) \approx L^{(M)}\tilde{\varphi}^{(s)}(x) = \sum_{i=1}^M \tilde{w}^{(i)} \chi^{(i)}(x). \quad (7)$$

Случайная величина $\zeta^{(s,i)}$ (оценитель коэффициента $w^{(s,i)}$) из соотношения (6) для $s = 3$ имеет вид $\zeta^{(3,i)} = \chi^{(i)}(\xi)$, где ξ – случайная величина, плотность (3) которой приближается с помощью алгоритма (7).

Для $s = 2$ имеем

$$\zeta^{(2,i)} = \frac{g(\xi^{(1)}, \xi^{(2)}) \chi^{(i)}(\xi^{(1)})}{f_{(\xi^{(1)}, \xi^{(2)})}(\xi^{(1)}, \xi^{(2)})},$$

где пара $(\xi^{(1)}, \xi^{(2)})$ имеет выбранную совместную плотность распределения $f_{(\xi^{(1)}, \xi^{(2)})}(x, y)$.

Наконец, для $s = 1$ имеем

$$\zeta^{(1,i)} = \sum_{m=0}^N Q^{(m)} \chi^{(i)}(\xi^{(m)}),$$

(см., например, [1]), где

$$Q^{(0)} = \frac{f(\xi^{(0)})}{\pi(\xi^{(0)})}; Q^{(m)} = Q^{(m-1)} \times \frac{k(\xi^{(m-1)}, \xi^{(m)})}{p(\xi^{(m-1)}, \xi^{(m)})},$$

а $\xi^{(0)}, \xi^{(1)}, \dots, \xi^{(N)}$ – прикладная цепь Маркова [6] с выбранными начальной плотностью $\pi(x)$ и переходной функцией $p(x', x)$; N – случайный номер обрыва цепи.

2. Свойства рандомизированных проекционных численных функциональных алгоритмов

В качестве основного недостатка проекционных методов приближения решения $\varphi^{(1)}(x)$ интегрального уравнения Фредгольма второго рода (1) в работе [1] справедливо отмечено отсутствие соображений теории условной оптимизации (см. также раздел 8.7 книги [6]). В частности, здесь для конкретных функций (5) строится своя теория приближений вида (4) со специфическими требованиями к функции $\varphi(x)$.

Отметим также, что в литературе нет примеров успешного (эффективного, экономичного) применения проекционных алгоритмов (7) для приближения функций $\varphi^{(1)}(x)$ и $\varphi^{(2)}(x)$ из соотношений (1) и (2) (исключение составляют лишь наши исследования из работы [7] для тестовой функции $\varphi^{(1)}(x)$ из раздела 4.10 книги [8]; см. также описание исследований уравнения (8) из раздела 3 данной работы). И даже для алгоритма (7) для приближения плотности $\varphi^{(3)}(x)$ из (3), известного очень давно (см., например, [9]), относительно успешные применения появились относительно недавно (см. работы [10–12]).

Главным ограничением для применения рандомизированных проекционных алгоритмов является то обстоятельство, что выбор ортонормированной системы функций (5) должен быть таким, что приближение (4) уже для $M = 1$ должно давать практически точное приближение решения (т. е. требуется как расширенная информация о приближаемой функции $\varphi(x) = \varphi^{(s)}(x)$, $s = 1, 2, 3$, так и наличие ортонормированной системы (5) с близкой к $\varphi(x)$ начальной функцией $\chi^{(1)}(x)$), иначе экстремальная вычислительная неустойчивость проекционных алгоритмов не позволит получить приемлемого приближения функции $\varphi(x)$.

3. Тестовый пример

Особо выделим нашу работу [7], в которой ярко продемонстрированы упомянутые выше свойства численной неустойчивости рандомизированного проекционного алгоритма (7) при неудачном выборе функций (5). Как указано выше, фактически это первая работа по применению проекционного алгоритма (7) для приближения решения $\varphi^{(1)}(x)$ интегрального уравнения Фредгольма второго рода (1). Правда, все эти замечательные достижения продемонстрированы на примере простейшего тестового, практически мало значимого одномерного уравнения из раздела 4.10 книги [8] вида

$$\varphi^{(1)}(x) = q \int_0^x e^{-(x-x')} \varphi^{(1)}(x') dx' + e^{-x}, \quad x > 0, \quad 0 < q < 1, \quad (8)$$

с известным решением $\varphi^{(1)}(x) = e^{-(1-q)x}$.

Очевидную неустойчивость алгоритма (7) для этого уравнения в работе [7] определил выбор в качестве функций (5) начальный отрезок длины M системы функций Эрмита (см., например, раздел 7.3.7 книги [13])

$$\chi^{(i)}(x) = A^{(i-1)} H^{(i)}(x) e^{-x^2/2}, \quad i = 1, \dots, M,$$

где $H^{(m)}(x)$ – многочлен Эрмита, определяемый соотношением

$$H^{(m)}(x) = (-1)^m e^{x^2} \frac{d^m}{dx^m} e^{-x^2},$$

а $A^{(m)} = 1/\sqrt{2^m m! \sqrt{\pi}}$ – соответствующие нормирующие константы.

Здесь функция $\chi^{(1)}(x) = Ae^{-x^2/2}$ достаточно далека от точного решения $\varphi^{(1)}(x)$ уравнения (8).

Проведенные нами объемные расчеты показали, что положение резко меняется (т. е. алгоритм (7) становится точным), если использовать в качестве системы функций (5) начальный отрезок длины M ортонормированной системы функций Лагерра (см., например, [13, разд. 7.3.7]):

$$\chi^{(i)}(x) = L^{(i-1)}(x) e^{-x/2}, \quad i = 1, \dots, M,$$

где $L^{(m)}(x)$ – многочлен Лагерра, определяемый соотношением

$$L^{(m)}(x) = \frac{e^x}{m!} \times \frac{d^m}{dx^m} (e^x x^{-m}).$$
 Здесь функция $\chi^{(1)}(x) = e^{-x/2}$ достаточно

близка к точному решению $\varphi^{(1)}(x)$ уравнения (8).

Заключение

По результатам данной работы можно сделать достаточно радикальный вывод о том, что применение рандомизированных проекционных численных алгоритмов вида (7) для практически значимых задач приближения функций (1) – (3) является нецелесообразным. Главной причиной здесь является крайняя чувствительность проекционных алгоритмов к выбору ортонормированной системы функций (5). Более предпочтительным для приближения функций $\varphi(x) = \varphi^{(s)}(x)$, $s = 1, 2, 3$, является использование ядерных алгоритмов (точнее – многомерных аналогов метода полигона частот) [14–16].

ЛИТЕРАТУРА

1. *Войтишек А.В.* Особенности построения и оптимизации рандомизированных проекционных функциональных алгоритмов // Материалы XVI Международной конференции имени А.Ф. Терпугова «Информационные технологии и математической моделирование» (Казань, 29 сентября – 3 октября 2017 года). Часть 2. Томск: Изд-во НТЛ, 2017. С. 267–275.
2. *Булгакова Т.Е., Войтишек А.В.* Применение рандомизированных функциональных алгоритмов для решения одной задачи с «вычислимым» ядром интегрального уравнения // Материалы XVIII Международной конференции имени А.Ф. Терпугова «Информационные технологии и математической моделирование» (Саратов, 26–30 июня 2019 года). Часть 2. Томск: Изд-во НТЛ, 2019. С. 329–334.

3. Булгакова Т.Е., Войтишек А.В. Об использовании «ядерных» оценок плотностей при рандомизации математических численных моделей // Материалы XVIII Международной конференции имени А.Ф. Терпугова «Информационные технологии и математической моделирование» (Саратов, 26–30 июня 2019 года). Часть 2. Томск: Изд-во НТЛ, 2019. С. 322–328.
4. Войтишек А.В. Основы метода Монте-Карло в алгоритмах и задачах. Часть 6: Вычисление линейных функционалов от решения интегрального уравнения второго рода. Дискретно-стохастические методы решения интегрального уравнения второго рода. Новосибирск: НГУ, 2004.
5. Войтишек А.В. Основы метода Монте-Карло в алгоритмах и задачах. Часть 5: Вычисление многократных интегралов. Аппроксимация интегралов, зависящих от параметра. Новосибирск: НГУ, 1999.
6. Войтишек А.В. Лекции по численным методам Монте-Карло. Новосибирск: ИПЦ НГУ, 2018.
7. Bulgakova T.E., Voytishchek A.V. On numerical stability of randomized projection functional algorithms // Communications in Statistics – Simulation and Computation. 2019. Latest articles. <https://doi.org/10.1080/03610918.2019.1677914>.
8. Михайлов Г.А., Войтишек А.В. Статистическое моделирование. Методы Монте-Карло. М.: Юрайт, 2018.
9. Ченцов Н.Н. Статистические решающие правила и оптимальные выводы. М.: Наука, 1972.
10. Михайлов Г.А., Трачева Н.В., Ухинов С.А. Рандомизированный проекционный метод для оценки угловых распределений поляризованного излучения на основе численного статистического моделирования // Журнал вычислительной математики и математической физики. 2016. Т. 56. № 9. С. 1560–1570.
11. Rogazinsky S.V. Statistical modelling algorithm for solving the nonlinear Boltzmann equation based on the projection method // Russian Journal of Numerical Analysis and Mathematical Modelling. 2017. V. 32. No. 3. P. 197–202.
12. Tracheva N.V., Ukhinov S.A. Numerical statistical study of the angular distribution of the polarized radiation scattered by medium // Russian Journal of Numerical Analysis and Mathematical Modelling. 2017. V. 32. No. 2. P. 135–146.
13. Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. М.: Наука, 1989.
14. Михайлов Г.А. Рандомизированные алгоритмы метода Монте-Карло для задач со случайными параметрами (метод «двойной рандомизации») // Сибирский журнал вычислительной математики. 2019. Т. 22. № 2. С. 187–200.
15. Михайлов Г.А. Улучшение многомерных рандомизированных алгоритмов метода Монте-Карло с «расщеплением» // Журнал вычислительной математики и математической физики. 2019. Т. 59. № 5. С. 822–828.
16. Voytishchek A.V., Bulgakova T.E. Optimization of kernel estimators of probability densities // Jacimovic M., Khachay M., Malkova V., Posypkin M. (eds). Optimization and Applications. OPTIMA 2019. Communications in Computer and Information Science. 2020: Springer, Cham. V. 1145. P. 254–266. https://doi.org/10.1007/978-3-030-38603-0_19.

Построение решений задач оптимального управления динамическими системами в бесконечномерных пространствах с малым параметром*

Даниэль Перес Акоста, Сергей Васильев,
Шахмурад Канзитдинов, Игорь Левичев

Российский университет дружбы народов, г. Москва, Россия

В прикладных задачах оптимального управления очень часто рассматриваются малые параметры, которые характеризуют относительную малость воздействий или факторов [1–5]. Кроме того, имеется большое число работ, в которых решаются задачи оптимального управления в пространствах с бесконечным числом измерений [6–8].

На данный момент в связи с развитием технологии 5G большой интерес представляют исследования кооперативного управления многоагентными динамическими системами, взаимосвязанными между собой с использованием топологии коммуникационной сети. Беспроводные сети связи являются на сегодняшний день незаменимой частью современной экономики и служат различным целям и приложениям для конечных пользователей. Современные телекоммуникационные сети обладают разнородными характеристиками с точки зрения инфраструктуры доступа, протоколов и пропускной способности. Более того, современные телекоммуникационные сети, такие, как сети 5G, в значительной степени децентрализованы как по своему управлению, так и по ресурсам. Конечные пользователи телекоммуникационных сетей разнообразны, и они управляют различными телекоммуникационными устройствами, на которых установлено различное программное обеспечение.

Так как телекоммуникационные сети носят распределенный характер, то пользователи часто имеют мало информации о топологии и характеристиках используемых сетей. Несмотря на это, пользователи могут предъявлять различные требования к пропускной способности используемых сетей. Учитывая указанные характеристики современных сетей и современных пользователей их ресурсов, фундаментальным на сегодняшний день является вопрос о том, как обеспечить эффективное

* Публикация выполнена при финансовой поддержке РФФИ в соответствии с исследовательским проектом No.18-07-00567.

распределение пропускной способности сетей между ее пользователями. При этом есть понимание того, что решение такого вопроса может быть достигнуто с помощью постановки и решения задач оптимального управления, применения распределенных алгоритмов управления, учитывая при этом проблемы устойчивости.

Необходимо отметить, что в задачах оптимального управления каждая система наделяется собственными переменными состояниями и определенной динамикой. Фундаментальная проблема в мультиагентных динамических сетевых системах – это эффективное проектирование распределенных протоколов, что гарантирует динамический баланс трафика и синхронизацию в том смысле, что элементы системы достигают равновесного значения, а все системы должны быть согласованы для достижения синхронного функционирования.

Фундаментальное значение для сетевых кооперативных динамических систем имеет изучение их взаимодействий и коллективного поведения под влиянием информационных потоков в коммуникационных сетях. Такая коммуникационная сеть может быть смоделирована в виде графа с направленными ребрами или связями, соответствующими разрешенным потокам информации между системами. Эти системы моделируются как узлы графа. Информационные потоки в телекоммуникационных сетях имеют место между ближайшими узлами в графе. В предположении, если граф соединен, то данная локально передаваемая информация в конечном счете поступает каждому агенту. В кооперативных системах управления на графах существуют важные взаимодействия между отдельными агентами сети, но топология графа может сильно ограничить возможную производительность любого элемента. В частности, при кооперативном управлении на графах все элементы управления должны быть распределены в том смысле, что закон управления для каждого агента является допустимым только в зависимости от информации, которая имеется у ближайших соседей по графу.

В данной работе предлагается метод построения решений задачи оптимального управления динамическими системами в бесконечномерных пространствах с малым параметром, который играет роль скейлингового параметра системы. Построение управления ведется путем построения решения дополнительной краевой задачи в пространстве состояний и эквивалентной системы функциональных уравнений в пространстве управлений с помощью методов теории сингулярных возмущений возмущений.

Математическая модель

Рассмотрим задачу оптимального управления, которая является линейной по управлению и квадратичной по состоянию:

$$\Phi(u) = \varphi(x(t_1)) + \int_T (\langle a(x(t), t), u(t) \rangle + d(x(t), t)) dt \rightarrow \min_{u \in V}, \quad (1)$$

$$\begin{aligned} \dot{x}(t) &= A(x(t), t)u(t) + b(x(t), t)dt, \\ x(t_0) &= x^0, u(t) \in U, t \in T = [t_0, t_1], \end{aligned} \quad (2)$$

где $x(t) = (x_1(t), \dots, x_n(t), \dots) \in R^\infty$ – бесконечномерный вектор состояния; $u(t) = (u_1(t), \dots, u_m(t), \dots) \in R^\infty$ – бесконечномерный вектор управления; $u(t), t \in T$, – допустимое управление в классе $V = \{u \in L_2(T): u(t) \in U, t \in T\}$ измеримых и интегрируемых с квадратом по Лебегу на множестве T векторнозначных функций со значениями в выпуклом компактном множестве $U \in R^\infty$; $A(x, t)$ – матричная функция, которая является квадратичной по x и непрерывной по t на множестве $R^\infty \times T$; $\varphi(x)$, $d(x, t)$ – скалярные функции; $a(x, t)$, $b(x, t)$ – векторные функции, при этом начальное состояние $x^0 \in R^\infty$, а также промежуток управления $T \in R$ заданы. Предполагается, что для каждого управления $v \in V$ имеется единственное решение $x(t, v)$, $t \in T$, системы (1), (2). Пусть функция Понтрягина

$$\begin{aligned} H(\psi, x, u, t) &= H_0(\psi, x, t) + \langle H_1(\psi, x, t), u \rangle, \\ H_0 &= \langle \psi, b(x, t) \rangle - d(x, t), H_1(\psi, x, t) = A^T(x, t)\psi - a(x, t), \end{aligned}$$

где $\psi \in R^\infty$ – сопряженная переменная. В этом случае сопряженная система имеет вид

$$\dot{\psi}(t) = -H_x(\psi(t), x(t), u(t), t), \quad t \in T. \quad (3)$$

Для допустимого управления $v \in V$ введем $\psi(t, v)$, $t \in T$ – решение системы (3) при $u(t) = v(t)$, $x(t) = x(t, v)$, $\psi(t_1, v) = -\varphi_x(x(t_1, v))$.

Для допустимого управления $u \in V$ определим такое отображение u^α , $\alpha > 0$,

$$\begin{aligned} u^\alpha(\psi, x, t) &= P_u(u(t) + \alpha H_1(\psi, x, t)), \\ \psi &\in R^\infty, x \in R^\infty, t \in T, \end{aligned} \quad (4)$$

где оператор проектирования на множество U в евклидовой норме P_U .

Для управления $u \in V$ принцип максимума с использованием отображения (4) можно записать в виде

$$u(t) = u^\alpha(\psi(t, u), x(t, u), t), t \in T, \alpha > 0. \quad (5)$$

Рассмотрим задачу улучшения заданного допустимого управления $u^0 \in V$: надо найти управление $v \in V$ с дополнительным условием $\Delta_v \Phi(u^0) = \Phi(v) - \Phi(u^0) \leq 0$.

Рассмотрим модифицированную сопряженную систему в задаче (1), (2) в виде

$$\dot{p}(t) = -H_x(p(t), x(t), u(t), t) - \frac{1}{2} H_{xx}(p(t), x(t), u(t), t), t \in T. \quad (6)$$

Для допустимых управлений $u^0, v \in V, p(t, u^0, v), t \in T$, представим решение системы (6) при $u(t) = u^0(t), x(t) = x(t, v), y(t) = x(t, v) - x(t, u^0)$,

$$p(t_1, u^0, v) = -\varphi_x(x(t_1, u^0)) - \frac{1}{2} \varphi_{xx}(x(t_1, u^0))(x(t_1, v) - x(t_1, u^0)),$$

где $p(t, u^0, u^0) = \psi(t, u^0), t \in T$.

Для заданных $u^0 \in V, \alpha > 0$ построим отображение u^α . Для решения задачи нелокального улучшения u^0 достаточно решить вспомогательную краевую задачу вида

$$\dot{x}(t) = f(x(t), u^\alpha(p(t), x(t), t), t), x(t_0) = x^0; \quad (7)$$

$$p(t) = -H_x(p(t), x(t, u^0), u^0(t), t) - \frac{1}{2} H_{xx}(p(t), x(t, u^0), u^0(t), t)(x(t) - x(t, u^0)),$$

$$p(t_1) = \varphi_x(x(t_1, u^0)) - \frac{1}{2} \varphi_{xx}(x(t_1, u^0))(x(t_1) - x(t_1, u^0)). \quad (8)$$

Предположим, что решение $x^\alpha(t), p^\alpha(t), t \in T$, краевой задачи (7), (8) (возможно, что это решение не будет единственным) будет существовать на интервале T , тогда имеется возможность сформировать выходное управление: $v^\alpha(t) = u^\alpha(p^\alpha(t), x^\alpha(t), t), t \in T$, откуда можно получить

$$x^\alpha(t) = x(t, v^\alpha), p^\alpha(t) = p(t, u^\alpha, v^\alpha)$$

и $v^\alpha = u^\alpha(p(t, u^0, v^\alpha), x(t, v^\alpha), t), t \in T. \quad (9)$

Оценка невозрастания целевого функционала на управлении $v^\alpha(t)$, $t \in T$, примет вид

$$\Phi(v^\alpha) - \Phi(u^0) < -\alpha^{-1} \int_T \|v^\alpha(t) - u^0(t)\|^2 dt. \quad (10)$$

С другой стороны, предположим, что $v^\alpha(t)$, $t \in T$, – допустимое управление, удовлетворяющее условию (9), тогда $(x(t, v^\alpha), p(t, u^0, v^\alpha))$, $t \in T$, есть решение краевой задачи улучшения (7), (8). Краевая задача (7), (8) улучшения в пространстве состояний эквивалентна системе уравнений (9) на множестве допустимых управлений, где управление $u^0 \in V$, удовлетворяющее условию (5) принципа максимума, является очевидным решением системы (9).

Это означает, что краевая задача улучшения (7), (8) для управления $u^0 \in V$, удовлетворяющего принципу максимума, всегда имеет решение $(x(t, v^\alpha), \psi(t, u^0))$, $t \in T$, в пространстве состояний.

Рассмотрим случай линейной по состоянию задачи (1), (2). Эту задачу улучшения (7), (8) можно представить в виде двух задач Коши для фазовой и сопряженной переменных $x(t)$, $p(t)$.

Пусть в краевой задаче улучшения (7), (8) имеет место параметр возмущения $\varepsilon \in [0, 1]$:

$$\begin{aligned} \dot{x}(t) &= f(x(t), u^\alpha(p(t), x(t), t), t), x(t_0) = x^0, \\ \dot{p}(t) &= -H_x(p(t), x(t), u^0(t), t) - \\ &- \varepsilon H_{xx}(p(t), x(t), u^0(t), t)(x(t) - x(t, u^0)); \end{aligned} \quad (11)$$

$$p(t_1) = -\varphi_x(x(t_1, u^0)) - \frac{1}{2} \varepsilon \varphi_{xx}(x(t_1, u^0))(x(t_1) - x(t_1, u^0)). \quad (12)$$

В данном случае невозмущенная краевая задача может быть представлена в виде двух независимых задачи Коши, где решением невозмущенной сопряженной системы является $p_0(t) = \psi(t, u^0)$, $t \in T$, а решение $x_0(t)$, $t \in T$, невозмущенной фазовой системы определяется задачей Коши

$$\dot{x}(t) = f(x(t), u^\alpha(p_0(t), x(t), t), t), x(t) = x^0.$$

Пусть $p(t)$, $x(t)$ можно представить в виде следующих рядов: $p(t) = \sum_{k=0}^{\infty} \varepsilon^k p^k(t)$, $x(t) = \sum_{k=0}^{\infty} \varepsilon^k x^k(t)$, тогда, подставляя эти ряды в

(11), (12) и приравнявая члены при одинаковых степенях ε , для решения возмущенной краевой задачи (11), (12) при $\varepsilon \in (0, 1]$ имеет место итерационный процесс

$$\begin{aligned} \dot{x}^{k+1}(t) &= f(x^{k+1}(t), u^\alpha(p^{k+1}(t), x^{k+1}(t), t), t), x^{k+1}(t_0) = x_0, \\ \dot{p}^{k+1}(t) &= -H_x(p^{k+1}(t), x(t, u^0), u^0(t), t) - \\ &- \varepsilon H_{xx}(p^k(t), x(t, u^0), u^0(t), t)(x^k(t) - x(t, u^0)); \end{aligned} \quad (13)$$

$$p^{k+1}(t_1) = -\varphi_x(x(t_1, u^0)) - \frac{1}{2} \varepsilon \varphi_{xx}(x(t_1, u^0))(x^k(t_1) - x(t_1, u^0)), \quad (14)$$

где на каждой итерации решается задача, аналогичная невозмущенной, и задается начальное приближение в виде $(x^0(t), p^0(t))$, $t \in T$.

Заключение

Идея активного управления телекоммуникационными сетями в реальном времени интенсивно развивается в течение нескольких последних лет. Изучение бесконечномерных динамических систем возникает во многих областях науки и техники, динамические отношения между агентами сети, число которых очень большое, получили большое внимание в течение последних лет. Феномены коллективной синхронизации агентов в крупномасштабных сетях будет иметь решающую роль в развитии телекоммуникационных технологий.

Предложен метод построения решений задачи оптимального управления динамическими системами в бесконечномерных пространствах с малым параметром, который играет роль внутреннего параметра системы. Построение управления осуществляется на основе поиска решения вспомогательной краевой задачи в пространстве состояний и эквивалентной системы функциональных уравнений в пространстве управлений с помощью методов теории сингулярных возмущений. Методы теории сингулярных возмущений развивались в связи с необходимостью поиска приближенных решений (в аналитической форме) сложных задач. Как правило, такие задачи выражаются в терминах дифференциальных уравнений, содержащих хотя бы один малый параметр. Существенной отличительной чертой сингулярной задачи возмущения является то, что простое и прямое приближение (основанное на малости параметра) не дает точного решения во всей области этого решения. Это приводит к тому, что в разных частях области допустимы разные

аппроксимации (обычно требующие «масштабирования» переменных по отношению к параметру). Это, в свою очередь, привело к появлению таких важных понятий, как декомпозиция, сшивка и так далее.

ЛИТЕРАТУРА

1. *Дмитриев М.Г., Курина Г.А.* Сингулярные возмущения в задачах управления // Автомат. и телемех. 2006. Вып. 1. С. 3–51
2. *Булдаев А.С.* Проекционные процедуры нелокального улучшения линейно управляемых процессов // Изв. вузов. Матем. 2004. № 1. С. 18–24
3. *Булдаев А.С.* Методы нелокального улучшения в задачах оптимального управления на основе точных формул приращения // Оптимальное управление: Труды XIII Байкальской международной школы-семинара «Методы оптимизации и их приложения», Иркутск, Байкал, 2 – 8 июля 2005 года. Т. 2. С. 111–115.
4. *Danilin A.R., Kovrizhnykh O.O.* Asymptotics of the optimal time in a time-optimal control problem with a small parameter // Proc. Steklov Inst. Math. 2017. V. 297. P. 62–71. doi.org/10.1134/S0081543817050078.
5. *Kalinin A.I., Lavrinovich L.I.* Application of the small parameter method to the singularly perturbed linear-quadratic optimal control problem // Autom. Remote Control. 2016. V. 77. P. 751–763. doi.org/10.1134/S0005117916050015.
6. *Frankowska H., Marchini E.M., Mazzola M.* Necessary optimality conditions for infinite dimensional state constrained control problems // Journal of Differential Equations. 2018. V. 264 (12). P. 7294–7327. doi.org/10.1016/j.jde.2018.02.012.
7. *Xiaodong Xu, Biao Huang, Stevan Dubljevic.* Optimal continuous-time state estimation for linear finite and infinite-dimensional chemical process systems with state constraints // Journal of Process Control. 2015. V. 35. P. 127–142. doi.org/10.1016/j.jprocont.2015.08.013
8. *Plekhanova M.* Optimal Control Existence for Degenerate Infinite Dimensional Systems of Fractional Order // IFAC-PapersOnLine. 2018. V. 51 (32). P. 669–674. doi.org/10.1016/j.ifacol.2018.11.502.

Система численного моделирования одномерных случайных величин NMPUD: формирование банка плотностей, автоматизация математических выкладок и приложения^{*}

Антон Войтишек¹,
Ярослав Постовалов², Данил Черкашин²

¹Институт вычислительной математики
и математической геофизики СО РАН, г. Новосибирск, Россия
²Лицей № 130 города Новосибирска, г. Новосибирск, Россия

Представлена созданная в лаборатории математического моделирования лицея № 130 города Новосибирска система NMPUD (Numerical Modelling of Probabilistic Univariate Distributions) (см. [1] и <http://monte-carlo-system.herokuapp.com/>), используемая как полезное (и даже необходимое) пособие для исследователей, занимающихся созданием и (или) использованием численных (компьютерных) вероятностных моделей для решения актуальных прикладных задач.

1. Основное наполнение банка системы NMPUD

В банк системы NMPUD входят формулы вероятностных плотностей

$$f_{\xi}(u; \lambda), \quad u \in (a, b), \quad (1)$$

(как правило, элементарных – см. раздел 2.6 книги [2] и раздел 2 данной статьи) случайных величин $\xi \in (a, b)$ и соответствующие формулы (реже – алгоритмы) для численного моделирования выборочного значения ξ_0 этой случайной величины вида

$$\xi_0 = \psi_{\xi}(\alpha_0; \lambda). \quad (2)$$

Здесь $\alpha_0 \in U(0,1)$ – стандартное случайное число (см. раздел 2.6 книги [2]), т.е. выборочное значение случайной величины $\alpha \in U(0,1)$, равномерно распределенной в интервале $(0,1)$. Это выборочное значение

^{*} Работа выполнена в рамках государственного задания ИВМиМГ СО РАН (проект 0315-2019-0002).

реализуется на компьютере с помощью специальной подпрограммы (именуемой в языках программирования как *RAND* или *RANDOM*).

В формулах (1) и (2) буквой $\lambda \in \Lambda$, обозначен параметр (набор параметров) распределения. В набор параметров λ могут быть внесены, в том числе, границы интервала распределения (a, b) .

В левой части основной страницы системы NMPUD (рис. 1) демонстрируется (или заносится) формула плотности (1) с соответствующим выбором параметров $\lambda \in \Lambda$; здесь же представлен график этой функции.



Рис. 1. Вид основной страницы системы NMPUD

Моделирующая формула (2) демонстрируется (или заносится) в правой части основной страницы системы NMPUD. С помощью многократного ($n \gg 1$, раз) обращения к формуле (2)

$$\xi_1 = \Psi_{\xi}(\alpha_1; \lambda), \xi_2 = \Psi_{\xi}(\alpha_2; \lambda), \dots, \xi_n = \Psi_{\xi}(\alpha_n; \lambda) \quad (3)$$

строится гистограмма

$$f_{\xi}^{(M,n)}(u) = \frac{V_i}{n \times h}, \quad u \in I_i = [x_i, x_{i+1}); \quad (4)$$

$$x_i = a + (i-1)h; \quad h = \frac{b-a}{M}, \quad i = 1, \dots, M.$$

Здесь v_i – количество выборочных значений (3), попавших в полуинтервал I_i , из (4); этот процесс можно наблюдать поэтапно.

Правильность моделирующей формулы (алгоритма) (2) определяется визуальным совпадением построенной гистограммы (4) с графиком плотности (1). Также вычисляется и показывается с правой стороны основной страницы среднее время обращения к формуле (алгоритму) (2) в наносекундах.

2. Элементарные плотности. Технология последовательных замен

Стандартным методом численного (компьютерного) моделирования выборочного значения ξ_0 одномерной случайной величины $\xi \in (a, b)$ является *метод обратной функции распределения* (см., например, раздел 2.5 книги [2]). Этот метод реализуется по формуле вида (2):

$$\xi_0 = F_{\xi}^{-1}(\alpha_0; \lambda), \quad (5)$$

где $\alpha_0 \in U(0,1)$ – стандартное случайное число, а $F_{\xi}(x) = P\{\xi < x\} = \int_{-\infty}^x f_{\xi}(u; \lambda) du$ – функция распределения случайной величины ξ .

В качестве промежуточной выкладки при выводе формулы (5) можно записать соотношение

$$\int_a^{\xi_0} f_{\xi}(u; \lambda) du = \alpha_0. \quad (6)$$

Здесь учтено то обстоятельство, что при $u \notin (a, b)$ функция $f_{\xi}(u; \lambda)$ из (1) равна нулю.

Плотность (1) называется *элементарной*, если уравнение (6) разрешимо относительно верхнего предела интегрирования ξ_0 в элементарных функциях. Другими словами, решение уравнения (6) дает формулу (2), где $\psi_{\xi}(v; \lambda) = F_{\xi}^{-1}(v; \lambda)$ – относительно простая («программируемая») композиция элементарных функций (см. раздел 2.6. книги [2]).

Для формирования достаточно объемного банка эффективно (экономично) моделируемых элементарных плотностей целесообразно применять следующую *технология последовательных (вложенных) замен*.

ТЕХНОЛОГИЯ 1 (см. раздел 14.2 книги [2]). Пусть

$$f_{\eta}(v; \lambda), \quad v \in (c, d), \quad \lambda \in \Lambda, \quad (7)$$

– плотность случайной величины η , имеющей элементарное распределение в интервале (c, d) , т.е. из соотношения вида (6)

$\int_c^{\eta_0} f_{\eta}(v; \lambda) dv = \alpha_0$ для соответствующего выборочного значения η_0 случайной величины η можно получить формулу вида (2), (5): $\eta_0 = \Psi_{\eta}(\alpha_0; \lambda)$, где $\Psi_{\eta}(w; \lambda)$ – простая композиция элементарных функций.

Рассмотрим взаимно-однозначное преобразование, задаваемое монотонной (например, возрастающей) дифференцируемой функцией $\varphi(u) \uparrow$ и переводящее интервал (a, b) в интервал (c, d) (для возрастающей функции имеем $\varphi(a) = c, \varphi(b) = d$). Полагаем также, что саму функцию $\varphi(u)$, ее производную $\varphi'(u)$ и обратную к ней функцию $\varphi^{-1}(u)$ можно представить в виде простых композиций элементарных функций.

Пусть случайная величина $\xi \in (a, b)$ имеет плотность распределения

$$f_{\xi}(u; \lambda) = f_{\eta}[\varphi(u); \lambda] \times \varphi'(u), \quad u \in (a, b) \quad (8)$$

(для убывающей функции $\varphi(u)$ в последнем выражении следует использовать модуль производной $\varphi'(u)$).

При сделанных выше предположениях можно утверждать, что плотность $f_{\xi}(u; \lambda)$ вида (8) является элементарной, т.е. уравнение (6) разрешимо относительно ξ_0 в элементарных функциях, а конкретнее

$\xi_0 = \varphi^{-1}[\Psi_{\eta}(\alpha_0; \lambda)]$ (для убывающей функции $\varphi(u)$ имеем

$$\xi_0 = \varphi^{-1}[\Psi_{\eta}(\alpha'_0; \lambda)], \quad \alpha'_0 = 1 - \alpha_0).$$

Действительно, записывая уравнение (6) для сконструированной новой плотности (8), имеем

$$\int_a^{\xi_0} f_{\eta}[\varphi(u); \lambda] \varphi'(u) du = \alpha_0 \quad \text{или} \quad \int_{\varphi(a)}^{\varphi(\xi_0)} f_{\eta}(v; \lambda) dv = \alpha_0,$$

$$\text{или } \varphi(\xi_0) = \Psi_{\eta}(\alpha_0; \lambda), \text{ или } \xi_0 = \varphi^{-1}[\Psi_{\eta}(\alpha_0; \lambda)]$$

(для убывающей функции $\varphi(u)$ выкладки аналогичны).

Название *технология последовательных (вложенных) замен* для технологии 1 здесь связано с тем, что полученную плотность (8) можно взять в качестве исходной плотности (7) и осуществить еще одно преобразование типа технологии последовательных (вложенных) замен. С помощью таких вложенных замен можно получать неограниченное число новых элементарных плотностей.

3. Текущие научные исследования, проводимые с помощью компьютерной системы NMPUD

В настоящее время проводятся следующие исследования с помощью компьютерной системы.

3.1. Изучаются вопросы разумного размера банка системы. Здесь уместно отметить, что для применения описанной в разделе 2 данной работы технологии последовательных (вложенных) замен имеется ограничение по количеству замен, так как каждая замена увеличивает трудоемкость моделирующей формулы. Также активно обсуждаются вопросы использования в системе NMPUD кроме формул (2), (5) метода обратной функции распределения экономичных алгоритмов модифицированного метода дискретной суперпозиции (см. раздел 11.2 книги [2]).

3.2. С проблемами, изложенными в разделе 3.1, тесно связаны изучаемые нами, на первый взгляд «технические», вопросы организации выбора параметров распределений, использования математических пакетов для автоматизации выкладок при применении технологии 1 в системе NMPUD (здесь также возникают свои ограничения, требующие изучения и подробного описания).

3.3. С помощью системы NMPUD проводятся разнообразные исследования, позволяющие дать точные количественные описания многочисленных качественных выводов и рекомендаций из книги [2], касающихся сравнения алгоритмов моделирования одномерных случайных величин. В частности, планируется подтвердить выводы из раздела 12 книги [2] и работы [3] о нецелесообразности применения бета- и гамма-распределений при рандомизации численных прикладных стохастических моделей.

Заключение

Описаны принципы и методические трудности формирования банка плотностей для специальной компьютерной системы моделирования вероятностных одномерных распределений NMPUD. Особо отметим, что система NMPUD является направляющим теоретическим и методическим пособием для создания разнообразных актуальных эффективных (экономичных) компьютерных моделей важных для изучения стохастических процессов и явлений.

ЛИТЕРАТУРА

1. *Васильев Т.В., Постовалов Я.С., Черкашин Д.А.* Проект компьютерной системы для выбора и исследования моделируемых вероятностных распределений // *Материалы 58-й Международной студенческой конференции «Студент и научно-технический прогресс. Математика».* Новосибирск: ИПЦ НГУ, 2020. С. 112.
2. *Войтишек А.В.* Лекции по численным методам Монте-Карло. Новосибирск: ИПЦ НГУ, 2018.
3. *Булгакова Т.Е., Войтишек А.В.* Об использовании бета- и гамма-распределений в численных рандомизированных моделях // *Расширенные тезисы X Международной Петрозаводской конференции «Вероятностные методы в дискретной математике» (22–26 мая 2019 года, Петрозаводск, Россия).* Петрозаводск: КарНЦ, 2019. С. 66–68.

Численный анализ на адаптивных сетках многомерного уравнения Фоккера – Планка с малым параметром*

Мохамед Адель Буатта, Сергей Васильев,
Вячеслав Федорченко

Российский университет дружбы народов, г. Москва, Россия

Традиционно при моделировании в области финансов предполагается, что процентные ставки постоянно изменяются, и эта динамика рассматривается большинством исследователей как диффузионный процесс. Однако исследования показали, что динамика процентных ставок имеет случайную скачкообразную динамику [1–4]. Исследователи полагают, что скачки процентных ставок, вероятно, связаны с различными рыночными явлениями, такими, как рыночная неопределенность или шоки на валютных рынках [5–8].

В данной работе представлен численный анализ решений многомерного уравнения Фоккера – Планка с малым параметром, которое позволяет описать диффузионную динамику процентных ставок на фондовых рынках в условиях неопределенностей. Для анализа решений используется численный метод на нерегулярной сетке, представляющей собой модификацию кусочно-однородных сеток типа Шишкина, адаптированных к малым размерам пограничного слоя. Показано, что численное решение сходится равномерно при различных значениях малого параметра.

Математическая модель динамики

Рассмотрим двухфакторную модель скачкообразной диффузии, которую мы используем для оценки производных процентных ставок. Это исследование предполагает, что переменные состояния являются динамикой мгновенной процентной ставки r и волатильности V .

Мы предполагаем, что определено пространство $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathcal{P})$ – полное вероятностное пространство, которое удовлетворяет обычным условиям, в которых $\{\mathcal{F}_t\}_{t \geq 0}$ – фильтрация.

* Публикация выполнена при финансовой поддержке РФФИ в соответствии с исследовательским проектом No.18-07-00567.

Чтобы учесть резкие изменения процентных ставок на рынках, мы считаем, что мгновенная процентная ставка следует за скачкообразным диффузионным случайным процессом, а волатильность – диффузионным процессом. Поэтому факторы модели опишем следующим стохастическим процессом:

$$r(t) = r(0) + \int_0^t \mu_r(r(z), V(z)) dz + \int_0^t V(z) dW_r(z) + \int_0^t c(r(z-), V(z)) dJ_z; \quad (1)$$

$$V(t) = V(0) + \int_0^t \mu_V(r(z), V(z)) dz + \int_0^t \sigma_V(r(z), V(z)) dW_V(z), \quad (2)$$

где $\mu_r \ll 1$ и $\mu_V \ll 1$ – дрейфы (предполагаются в качестве малых параметров), а σ_V – волатильность процесса подразумеваемой волатильности. Кроме того, W_r и W_V – винеровские процессы, а влияние скачков определяется функцией $c(r(z-), V(z))$ и сложным пуассоновским процессом $J(t) = \sum_{i=1}^{N(t)} Y_i$.

В периоды времени $(\tau_i)_{i \geq 1}$, когда могут произойти скачки, размеры скачков Y_1, Y_2, \dots образуют последовательность одинаково распределенных случайных величин с нормальным распределением вероятностей $\mathcal{N}(0, \sigma_Y)$ и $N(t)$ представляют пуассоновский процесс с интенсивностью $\lambda(r, V)$. Мы предполагаем, что W_r, W_V – распределения по размерам скачков – не зависят от N , но стандартные броуновские движения соотносятся с $[W_r, W_V](t) = \rho t$. Мы также предполагаем, что величина скачка и время скачка не связаны с диффузионными частями процессов. Наконец, мы предполагаем, что функции $\mu_r, \mu_V, \sigma_V, \lambda$ и Π удовлетворяют подходящим условиям регулярности. Будем считать, что рынок свободен от арбитража. Кроме того, существует эквивалентная мера мартигалов, Q-мера, известная как мера, не зависящая от риска. Переменные состояния модели (1), (2) при нейтральной оценке риска следующие:

$$r(t) = r(0) + \int_0^t \mu_r^Q(r(z), V(z)) dz + \int_0^t V(z) dW_r^Q(z) + \int_0^t c(r(z), V(z)) d\hat{J}^Q; \quad (3)$$

$$V(t) = V(0) + \int_0^t \mu_V^Q(r(z), V(z)) dz + \int_0^t \sigma_V(r(z), V(z)) dW_V^Q(z), \quad (4)$$

где $\mu_r^Q = \mu_r - V\theta^{W_r}$; $\mu_V^Q = \mu_V - \sigma_V\theta^{W_V}$; W_r^Q и W_V^Q – процессы Винера с Q-мерой и $[W_r^Q, W_V^Q](t) = \rho t$. Рыночная плата за риск, связанный с процессами Винера W_r и W_r , составляет $\theta^{W_r}(r, V)$ и $\theta^{W_V}(r, V)$ соответственно.

Далее введем соотношение $\hat{J}^Q(t) = \sum_{i=1}^{N(t)^Q} Y(i) - \lambda^Q t E_Y^Q[Y_1]$, которое является компенсированным пуассоновским процессом по Q-мере, а интенсивность пуассоновского процесса $N^Q(t)$ равна $\lambda^Q(r, V)$. Также будем полагать, что функция $c(r, V) = 1$ является константой в (1) и (3).

Таким образом, цена облигации с нулевым купоном в момент времени t со сроком погашения $t \leq T$ в соответствии с вышеизложенными допущениями может быть выражена как

$$P(t, r, V; T) = E^Q \left[\exp \left\{ - \int_t^T r(u) du \right\} \mid r(t) = r, V(t) = V \right], \quad (5)$$

кривая доходности в данном случае

$$R(t, r, V; T) = \frac{-\ln(P(t, r, V; T))}{T - t}.$$

Пусть $C(t, r, V, T_2; T_1)$ – цена европейского колл-опциона, который открывается в момент T , на облигации, срок действия которых истекает при T_2 , $T_1 \leq T_2$, а K является ценой исполнения. Европейский облигационный опцион оценивается как ожидаемая дисконтированная выплата в рамках Q-меры

$$C(t, r, V, T_2; T_1) = E^Q \left[e^{-\int_t^{T_1} r(u) du} \max(P(T_1, r(T_1), V(T_1); T_2) - K, 0) \mid r(t) = r, V(t) = V \right].$$

Кроме того, $\tau_1 = T_1 - t$ и $\tau_2 = T_2 - t$ – срок погашения опционного контракта и облигации с нулевым купоном соответственно.

Для анализа решений уравнения Фоккера – Планка в данном исследовании используются численные методы на нерегулярной сетке, которая является модификацией адаптированных к малой величине пограничного слоя кусочно-однородных сеток типа Шишкина. Показывается, что численное решение равномерно сходится при различных значениях малого параметра.

Заключение

В данной работе представлен численный анализ решений на адаптивной неравномерной сетке многомерного уравнения Фоккера – Планка с малым параметром, которое позволяет описать диффузионную динамику процентных ставок на фондовых рынках в условиях неопределенностей. Анализ решений на нерегулярной сетке, представляющей

собой модификацию кусочно-однородных сеток типа Шишкина, адаптированных к малым размерам пограничного слоя, показал, что численное решение сходится равномерно при различных значениях малого параметра.

ЛИТЕРАТУРА

1. Кокс Дж.К., Ингерсолл Дж.Э., Росс С.А. Теория термина «структура процентных ставок» // *Econometrica*. 1985. Т. 53. С. 385–407.
2. Васичек О. Равновесная характеристика термина структура // *J. Financ. Econ.* 1977. Т. 5. С. 177–188.
3. Дас Ч.П., Фореси С. Точные решения для облигаций и цены опционов с систематическим риском скачка // *Rev. Deriv.* 1996. Т. 1. С. 7–24.
4. Йоханнес М. Статистическая и экономическая роль скачков в непрерывных моделях процентных ставок // *Дж. Финанс.* 2004. Т. 59. С. 227–259.
5. Лин Б.Х. Скачки диффузионного процесса процентной ставки: эмпирическое исследование // *J. Bus. Financ.* 1999. Т. 26. С. 967–995.
6. Andreev V., Savin I. The uniform convergence with respect to a small parameter of A.A. Samarskii's monotone scheme and its modification // *Computational Mathematics and Mathematical Physics*. V. 35. 1995. P. 581–591.
7. Lins T., Roos H.-G., Vulanovic R. Uniform pointwise convergence on Shishkin-type meshes for quasi-linear convection-diffusion problems // *SIAM Journal on Numerical Analysis*. 2000. V. 38. No. 3. P. 897–912.
8. Vulanovic R., Teofanov L. A modification of the Shishkin discretization mesh for one-dimensional reaction–diffusion problems // *Applied Mathematics and Computation*. 2013. V. 220. P. 104–116.

Об использовании конечных автоматов при моделировании наноструктур

Никита Беляков, Рустам Бикмурзин, Дмитрий Федченко

Сибирский федеральный университет, г. Красноярск, Россия

Углеродные наноматериалы: графен, фуллерены и нанотрубки используются во многих областях. Речь идет о производстве композитных наноматериалов, медицине и многом другом. Наименее изученными являются нанолуковицы, которые представляют собой сферические частицы, состоящие из вложенных друг в друга монослойных оболочек (см., например, [1, 2]).

Рассматриваются графические представления детерминированных конечных автоматов специального вида. Речь идет о пазлах – головоломках типа *пятнашек*. Мотивацией к изучению подобных автоматов является их простота и тесная связь с симметрической группой S_d , что позволяет в полной мере задействовать алгебраический аппарат (см., например, [3]). Возникающие на этом пути вопросы требуют обращения к идеям неассоциативной геометрии (см. [4]).

В данной статье мы строим графическое представление детерминированного конечного автомата, внешне схожее с луковичной структурой.

Формулировка задачи

Пусть K_d – полный граф, d пронумерованных вершин которого представим подвижными бусинами, передвигающимися по ребрам. Нам ближе реализация полного графа в виде симплекса в пространстве размерности $d-1$. Символом \emptyset мы обозначили пустую позицию, куда может быть сдвинута любая из бусин. На рис. 1 изображен полный граф K_5 .

Пусть $A = (Q, A, \delta, q_0, F)$ – детерминированный конечный автомат с функцией перехода $\delta: Q \times A \rightarrow Q$, а $S = (Q, A, \delta)$ – полуавтомат (см., например, [3, Глава 2]). Рассмотрим полуавтомат S , множество состояний Q которого совпадает с группой перестановок S_d вершин $\{1, 2, \dots, d-1, 0\}$ графа K_d , а входной алфавит $A = \{1, 2, \dots, d-1\}$, причем $\delta: (\pi, j) \rightarrow \sigma$, где перестановка $\sigma \in S_d$ равна π с точностью до перестановки элементов j и 0 , $j = 1, 2, \dots, d-1$.

При $d = 3$ получим граф полуавтомата, который может быть изображен представленным на рис. 2 способом.

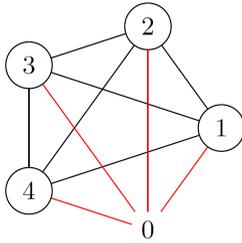


Рис. 1

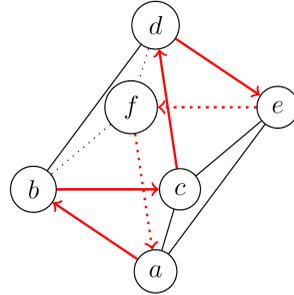


Рис. 2

Здесь a, b, c, d, e и f – элементы симметрической группы S_3 , где $a = (1, 2, 0), b = (1, 0, 2), c = (0, 1, 2), d = (2, 1, 0), e = (2, 0, 1), f = (0, 2, 1)$. Легко видеть, что множество $\{a, c, e\}$ является подгруппой в S_3 . На множестве инволюций (элемент группы называется инволюцией, если его квадрат равен единице группы) $M = \{b, d, f\} \subset S_3$ можно ввести квазигрупповое умножение по правилу $m_i * m_j = \text{proj}(m_i \times m_j)$, см., например, [4], где proj – оператор проектирования на M вдоль гамильтонова цикла, выделенного красным цветом на рис. 2, а $m_i \times m_j$ – произведение в группе. Получим неассоциативную таблицу умножения (таблица). Действительно, различные способы расстановки скобок приводят к различным результатам $(fd)b = d$, а $f(db) = b$. Мы получили квазигруппу третьего порядка. Напомним, что квазигруппой называют произвольную магму с делением. Если говорить очень грубо, то квазигруппы похожи на группы без свойства ассоциативности.

*	b	d	f
b	b	d	f
d	f	b	d
f	d	f	b

Более интересные автоматы получаются, когда мы запрещаем переходы по некоторым ребрам в графе K_d . Рассмотрим граф K_4 без диагональных ребер, которому соответствует пазл размера 2 на 2 (рис. 3). Пазлом мы называем головоломку, в которой одинаковые квадратные костяшки с нанесенными цифрами заключены в квадратную (или прямоугольную) коробку. Самой популярной является игра в 15. Можно заметить, что у множества состояний пазла 2 на 2 есть инвариант: чет-

ность перестановки, выписанной без учета пустой клетки. Игровой конфигурации на рис. 3 соответствует перестановка $(0, 1, 2, 3)$.

Мы видим, что множество Q состояний пазла составляет не всю группу S_4 , а лишь половину (12 элементов). Для визуализации графа автомата мы использовали бесплатную программу Gephi. После экспериментов с настройками притяжения и отталкивания вершин был получен граф (рис. 4). Данный граф называется графом Науру (по аналогии с двенадцатилучевой звездой на флаге государства Науру). Как и октаэдр граф Науру является гамильтоновым.

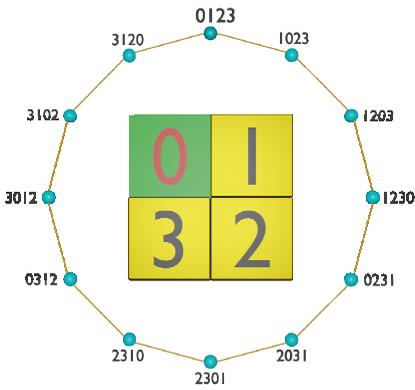


Рис. 3

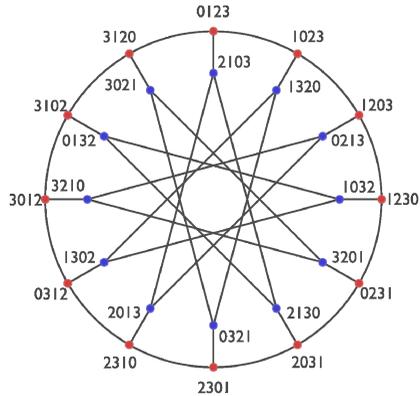


Рис. 4

В конце мы хотели бы обсудить граф конечного автомата, соответствующего пазлу 2 на 3 (рис. 5).

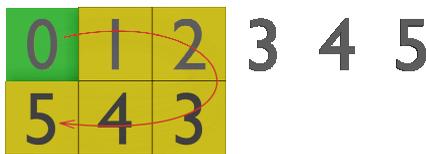


Рис. 5

Как и в предыдущем случае, множество игровых конфигураций пазла составляет лишь половину элементов группы S_6 , которые собираются в 12 колец по 30 элементов в каждом. Визуализация в Gephi дала следующую картинку (рис. 6).

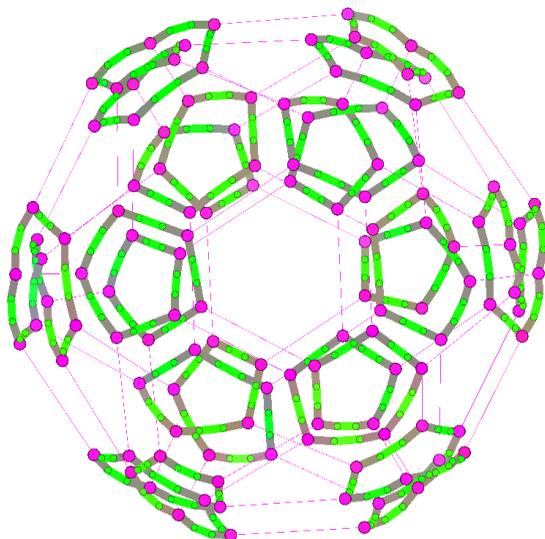


Рис. 6

Заключение

Мы рассмотрели лишь некоторые примеры конечных автоматов специального вида и их графические представления. Хотелось бы особо отметить некоторое внешнее сходство графа пазла 2 на 3 с луковичной углеродной структурой (речь идет о многослойной молекуле фуллерена, состоящей из 360 атомов). Для нас остается открытым вопрос: возможна ли такая структура в реальных условиях при определенных значениях давления и температуры.

ЛИТЕРАТУРА

1. *Рехвиашвили С.Ш., Бухурова М.М.* Устойчивость углеродной нанолуковицы в контакте с графитовой подложкой // Письма в ЖТФ. 2019. Т. 45. № 12.
2. *Баграмов Р.Х. и др.* Луковичные структуры, полученные из фуллеритов C₆₀, C₇₀ и наноалмазов // Известия высших учебных заведений. Химия и химическая технология. 2013. Т. 56. № 5.
3. *Ginzburg A.* Algebraic theory of automata. Academic Press, 2014.
4. *Сабинин Л.В.* О геометрии луп // Математические заметки. 1972. Т. 12. № 5. С. 605–616.

Использование метода фиктивных переменных для исследования пространственной характеристики систем видения через атмосферу

Ирина Гендрина

*Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

Системный подход является универсальным механизмом исследования закономерностей развития различных систем, то есть структур, обладающих определенными свойствами [1]. Его применение для решения задач переноса излучения через атмосферу основывается на том, что структура, состоящая из источника излучения (подстилающей поверхности), среды распространения и приемного устройства, удовлетворяет всем признакам линейной системы [1, 2]. В теории переноса излучения такая структура называется системой видения [3]. Основными задачами при исследовании систем видения является прогнозирование и коррекция формируемых в них изображений. Это обуславливает актуальность исследования закономерностей влияния составляющих частей системы. В последние годы для решения задач, связанных с распространением излучения через атмосферу, широко используются статистические методы: корреляционно-регрессионный анализ, гармонический анализ и др. В нашей работе также сделана попытка применения статистического инструмента – метода фиктивных переменных, распространенного при проведении экономических исследований и позволяющего особенным образом учитывать количественные и качественные факторы.

Характеристики системы видения

Для анализа влияния системы L на произвольный объект, расположенный на подстилающей поверхности, необходимо построить отклик системы L на точечный δ -импульс – функцию размытия точки (ФРТ):

$$L[\delta(x - x_0, y - y_0)] = h(x, y; x_0, y_0).$$

Система называется инвариантной к сдвигу [2–4], если сдвиг входного сигнала приводит к равному сдвигу выходного сигнала:

$$L[\delta(x - x_0, y - y_0)] = h(x - x_0, y - y_0).$$

Произвольная функция $f(x, y)$ может быть представлена как суперпозиция (интеграл) точечных масс

$$f(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x_0, y_0) \delta(x - x_0, y - y_0) dx_0 dy_0 .$$

Поэтому выходной сигнал (изображение произвольного объекта) представляет собой свертку произвольного объекта и ФРТ:

$$g(x, y) = L[f(x, y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x_0, y_0) h(x - x_0, y - y_0) dx_0 dy_0 .$$

Двумерное преобразование Фурье функции размытия точки называется оптической передаточной функцией (ОПФ) системы и представляет собой характеристику системы видения в частотной области:

$$H(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(\xi, \eta) e^{-i(u\xi + v\eta)} d\xi d\eta .$$

Таким образом, для систем, инвариантных к сдвигу, справедливо простое соотношение:

$$G(u, v) = F(u, v)H(u, v) .$$

Здесь $F(u, v)$ – двумерное преобразование Фурье исходного объекта.

ФРТ и ОПФ являются основными характеристиками системы видения. Их используют для решения задач формирования и обработки изображений в широком смысле слова, при построении радиационных моделей и т.д.

В данной работе речь пойдет о ФРТ для различных условий наблюдения.

Результаты статистического моделирования

Функции размытия точки были получены методом Монте-Карло. Их общий вид для различных условий приведен на рис. 1.

На практике применение статистического моделирования нецелесообразно. Его цель заключается, в частности, в том, чтобы на основании результатов моделирования создать быстрые и удобные инженерные методики для расчета соответствующих характеристик. Одним из способов создания таких методик является корреляционно-регрессионный анализ. В своих предыдущих работах мы показали, что построенные регрессии обладают неплохим качеством.

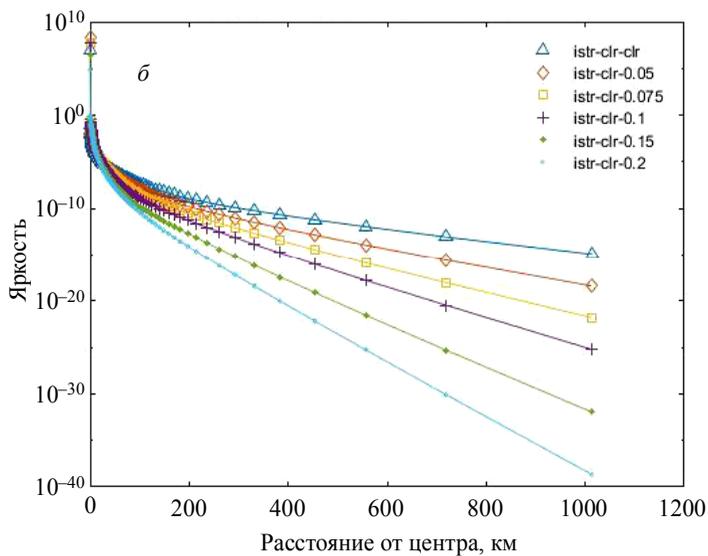
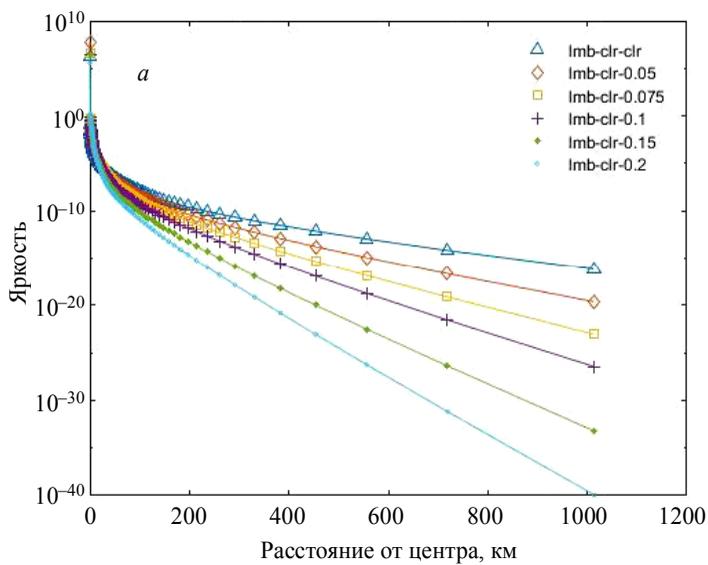


Рис. 1. Функции размытия точки для ламбертовского (а) и изотропного (б) источников

Метод фиктивных переменных

В настоящей работе мы продолжаем использовать для анализа результатов моделирования статистические приемы, одним из которых является аппарат фиктивных переменных. Этот статистический прием эффективен при решении различных экономических задач.

Оценим с его помощью возможность выявления некоторого признака в массивах данных ФРТ.

Будем выявлять такой качественный признак, как тип источника. Моделирование было проведено для двух вариантов: ламбертовского и изотропного. Различие между ними заключается в плотности вероятности распределения начальных направлений.

Введем фиктивную переменную D , принимающую следующие значения:

$$D = \begin{cases} 0, & \text{если источник ламбертовский,} \\ 1, & \text{если источник изотропный.} \end{cases}$$

Будем строить уравнение зависимости яркости y от расстояния x в виде регрессии:

$$\hat{y} = b_0 x^{b_1}.$$

Используем известный прием приведения к линейному виду

$$\ln y = \ln b_0 + b_1 \ln x.$$

Введем зависимость от типа источника:

$$\ln y = \ln b_0 + b_1 \ln x + \gamma D.$$

Будем строить уравнение регрессии в виде

$$y^* = A + Bx^* + \gamma D.$$

Некоторые характеристики регрессии приведены в табл. 1.

Таблица 1

Регрессионная статистика

Множественный R	0.975123
R -квадрат	0.950864
Нормированный R -квадрат	0.950455
Стандартная ошибка	1.184742

В этом случае функция y будет иметь вид

$$\hat{y} = e^A x^{b_1} e^{\gamma D}.$$

Найдем и оценим статистическую значимость коэффициентов полученной регрессии, используя стандартный алгоритм электронных таблиц, реализующий механизм проверки статистических гипотез. Значимость коэффициента γ будет характеризовать значимость качественного фактора.

В табл. 2 приведены значения коэффициентов и статистики, характеризующие значимость.

Таблица 2

	Коэффициенты	t-статистика	Нижние 95%	Верхние 95%
Y-пересечение (A)	-0.165	-0.759	-0.594	0.264
Переменная X (B)	-4.013	-68.066	-4.1293	-3.897
Переменная D (γ)	0.600	3.950	0.301	0.900

t-пороговое	1.970
-------------	-------

На рис. 2 приведены данные сравнения результатов численного моделирования и регрессионных значений.

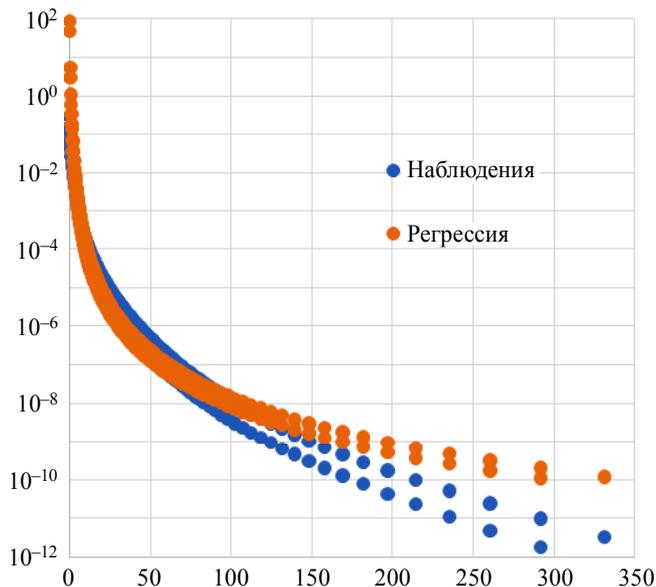


Рис. 2. Результаты сравнения

Заключение

Из приведенных данных следует, что коэффициент при фиктивной переменной является значимым. Следовательно, тип источника – это фактор, который следует учитывать при прогнозировании и коррекции изображений.

ЛИТЕРАТУРА

1. *Тарасенко Ф.П.* Прикладной системный анализ. М.: КНОРУС, 2010. 224 с.
2. *Папулис А.* Теория систем и преобразований в оптике. М.: Мир, 1971. 495 с.
3. *Зуев В.Е., Белов В.В., Веретенников В.В.* Теория систем в оптике дисперсных сред. Томск: Издательство СО РАН, 1997. 402 с.
4. *Белов В.В., Креков Г.М., Макушкина И.Ю.* Изопланарность в системах видения // Оптика атмосферы. 1989. Т. 2. № 10. С. 1011–1018.

Исследование принципов применения моделей многозначной логики в современных приложениях

Антон Есин

Институт проблем управления им. В. А. Трапезникова РАН, г. Москва, Россия

Проведён анализ и представлен обзор нескольких современных приложений, где создание гетерогенных вычислительных систем с небольшими вычислительными блоками на основе трёхзначной логики является математически лучшим решением по сравнению с бинарными моделями.

Известно, что удельная плотность записи информации описывается следующей функцией [7]:

$$Y(a) = \frac{\ln y(a)}{a} = \frac{\ln a}{a},$$

которая достигает максимума при $a = e \approx 2.718$ [7], т.е. математически троичная логика более эффективна, чем двоичная [1, 2].

Исследования и разработка алгоритмов на основе трёхзначной логики вновь актуальны [4], например, в телекоммуникациях [3], в области искусственного интеллекта (ИИ) [1, 3], квантовых вычислениях, медицине, физике [5–7]. Рассмотрим несколько примеров приложений, где построение алгоритмов на основе трёхзначной логики обеспечивает большую эффективность и оказывается предпочтительнее в сравнении с двузначной.

Оценка надёжности технических систем

Многозначная логика позволяет вводит качественные (лингвистические) переменные вместо количественных. Количественные показатели (факторы) дискретизируются путём отображения в некоторую m -интервальную шкалу. Такой подход позволяет соединить в одной модели количественные и качественные показатели, что позволяет провести анализ структурных процессов и оценку влияния факторов на надёжность технических систем. Такой подход позволяет соединить в одной модели количественные и качественные показатели и значительно упростить достаточно сложные вероятностные схемы оценки надёжности [8, 9].

Рассмотрим структурную функцию процесса для оценки качества работы системы Wi-Fi-доступа к сети Internet на подвижных объектах (в

автомобилях, поездах, на объектах городского пассажирского транспорта и т.д.). При построении данной схемы были проанализированы и отобраны следующие ключевые внутренние и внешние параметры, влияющие на рассматриваемый организационный процесс: качество связи (наличие доступа ко внешней сети по одному из трёх каналов связи), качество используемого телекоммуникационного оборудования, стабильность обеспечения электроэнергией. Для построения схемы были отобраны только основные факторы влияния. Список можно существенно расширить за счёт учёта дополнительных параметров, например таких, как: скорость движения объекта, уровень сигнала станций сотовой связи, объём нагрузки на сеть (количество пользователей), ёмкость канала, виды трафика и пр. В этом случае будет возрастать вычислительная сложность структурной функции.

Для уменьшения вычислительной сложности структурной функции переменные группируют по признакам принадлежности к классам: например линии связи (x_{11}), сетевое оборудование (x_{12}), и производят расчет промежуточных переменных на основе экспертной базы правил. Схема построения структурной функции представлена на рис. 1.

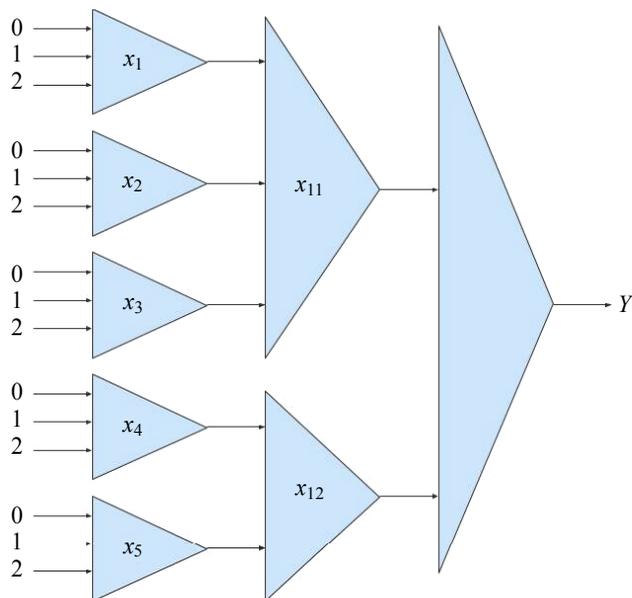


Рис. 1. Схема структурной функции процесса оценки качества работы системы доступа к сети Internet

В табл. 1 приведены переменные и их значения, результаты расчетов промежуточных значений представлены в табл. 2.

Таблица 1

Входные параметры для анализа качества работы системы доступа к сети Internet

Параметр системы	Значение параметра и его описание		
	0	1	2
Канал связи 1 (x_1)	Не доступен	Слабый сигнал	Устойчивое соединение
Канал связи 2 (x_2)	Не доступен	Слабый сигнал	Устойчивое соединение
Канал связи 3 (x_3)	Не доступен	Слабый сигнал	Устойчивое соединение
x_{11}	Не доступен	Слабый сигнал	Устойчивое соединение
Блок питания (x_4)	Не работает	Неустойчивая работа	Работает
Состояние модема (x_5)	Не работает	Неопределено	Работает
x_{12}	Не работает	Неустойчивая работа	Работает
Качество работы системы в целом (Y)	Не работает	Нестабильный доступ к сети	Работает

Таблица 2

Значения выходной функции Y

x_{11}	0	0	0	1	1	1	2	2	2
x_{12}	0	1	2	0	1	2	0	1	2
Y	0	0	0	0	1	2	0	2	2

Временное моделирование процессов

Ещё одним примером реального приложения многозначной логики является временное моделирование сложных систем, которое в системах автоматизированного моделирования часто базируется на принципах трёхзначной логики [7, 5]. Функциональные элементы системы при

построении модели представляются логическими блоками, на входах/выходах которых формируются многозначные сигналы. в докладе будет рассмотрен пример функции для расчёта задержки в сети. Ниже в табл. 3 приведён пример вычислений значений для следующей результирующей функции:

$$f = m_1(m(x_1; m_{01}(x_1; x_1)); A + 2(x_1)).$$

Таблица 3

Таблица значений результирующей функции f

$f(A,B)$	0	1	2	3
0	2	2	2	2
1	1	3	1	3
2	1	2	2	2
3	1	1	1	1

Ниже на рис. 2 приведена логическая схема построения данной функции.

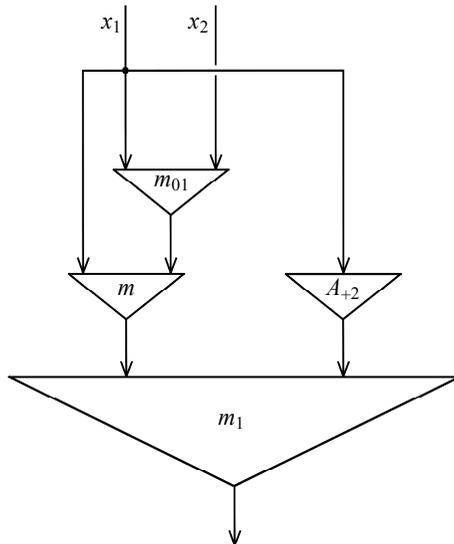


Рис. 2. Пример построения результирующей функции $f(A,B)$

Задачи анализа данных и искусственного интеллекта

В настоящее время инструментарий многозначной логики находит широкое применение в задачах, связанных с анализом данных и построением моделей искусственного интеллекта, например в задачах иерархической кластеризации произвольных наборов данных [1, 3].

Теория многозначной логики над полем комплексных чисел и пороговая функция многозначной логики применяются при построении модели многозначного нейрона. Для дискретнозначного многозначного нейрона отображение ввода / вывода всегда описывается некоторой пороговой функцией многозначной логики. Это применяется при рассмотрении непрерывного многозначного нейрона и построении разделяющей гиперплоскости в n -мерном пространстве, которая определяется функцией активации дискретного многозначного нейрона. Это позволяет решать задачи многоклассовой классификации [1]. Н. Айзенбергом было предложено рассматривать следующую функцию $P(z)$, которая и называется k -значным предикатом:

$$P(z) = \text{CSIN}(z) = \varepsilon_k^j, \quad 2\pi j/k \leq \arg z < 2\pi(j+1)/k, \quad (1)$$

где $z \in \mathbb{C}$. Данная функция, по сути, задаёт деление комплексной плоскости на k равных секторов линиями, проходящими через начало координат, и точки на единичной окружности, соответствующие корням k -й степени из единицы.

Определение 1 [1]. Пороговой функцией k -значной логики (или многозначной пороговой функцией) называется функция

$$f(x_1, \dots, x_n) : T \rightarrow E; \quad T \subseteq E_k^n,$$

если существуют $(n+1)$ комплексные числа w_0, w_1, \dots, w_n , такие, что для любого вектора $(x_1, \dots, x_n) \in T$

$$f(x_1, \dots, x_n) = P(w_0 + w_1 x_1 + \dots + w_n x_n),$$

где $P(z)$ – k -значный предикат, определяемый по формуле (1).

Вектор $W = (w_1, w_2, \dots, w_n)$ – вектор весов для функции f .

Таким образом, введение многозначных функций значительно расширяет набор функций, которые могут быть представлены с использованием $(n+1)$ весов путем добавления многозначных пороговых функций к булевым пороговым функциям, а логическая пороговая функция – это частный случай многозначной пороговой функции. В качестве

простого примера такой функции можно рассмотреть функцию максимума из двух аргументов [1]: $f = \max(x_1, x_2)$.

Данная функция является пороговой функцией (так же как, например, $f = \min(x_1, x_2)$). Многозначные функции над полем комплексных чисел существенно расширяют набор пороговых функций и перспективы в пороговой логике, в нейронных сетях и в решении задач мультиклассовой классификации. Как пример, можно рассмотреть функцию Поста $f = \max(x_1, x_2)$ (пороговая функция трёхзначной логики) [1].

ЛИТЕРАТУРА

1. *Aizenberg I.* Complex-Valued Neural Networks with Multi-Valued Neurons. Studies in Computational Intelligence (Vol. 353). Berlin; Heidelberg: Springer-Verlag, 2011. 273 p. DOI 10.1007/978-3-642-20353-4.
2. *Asadi M., Mosleh M., Haghparast M.* Toward novel designs of reversible ternary 6:2 Compressor using efficient reversible ternary full-adders // The Journal of Supercomputing. 2020. <https://doi.org/10.1007/s11227-020-03485-7>.
3. *Bykovsky Alexey Yu.* Heterogeneous network architecture for integration of AI and quantum optics by means of multiple-valued logic // Quantum Rep. 2020. No. 2. P. 126–165. DOI:10.3390/quantum2010010.
4. *Connelly Jeff.* Ternary Computing Testbed 3-Trit Computer Architecture. PhD thesis. Computer Engineering Department. California Polytechnic State University, 2008. 192 p. URL: <http://xyzyzy.freeshell.org/trinary/CPE%20Report%20-%20Ternary%20Computing%20Testbed%20-%20RC6a.pdf> (дата обращения: 10.10.2020).
5. *Kak Subhash.* On Ternary Coding and Three-Valued Logic. 2018. URL: <https://arxiv.org/abs/1807.06419> (дата обращения: 10.10.2020).
6. *Yi Jin, Huacan He, Yangtian Lü.* Ternary optical computer architecture // Physica Scripta. 2005. T118. DOI: 10.1238/physica.topical.118a00098.
7. *Недвоичная логика.* URL: <https://habr.com/ru/post/160595/> (дата обращения: 10.10.2020).
8. *Kalimulina E.Yu.* Analysis of system reliability with control, dependent failures, and arbitrary repair times // International Journal of System Assurance Engineering. 2017. V. 8. P. 180–188. <https://doi.org/10.1007/s13198-016-0520-5>.
9. *Kalimulina E.Yu.* Rate of Convergence to Stationary Distribution for Unreliable Jackson-Type Queueing Network with Dynamic Routing // Communications in Computer and Information Science. 2017. V. 678. https://doi.org/10.1007/978-3-319-51917-3_23.

Теоретические аспекты построения современных систем управления на базе многозначной логики

Антон Есин

Институт проблем управления им. В. А. Трапезникова РАН, г. Москва, Россия

Создание гетерогенных вычислительных систем с небольшими троячными вычислительными блоками является математически лучшим решением при построении узкоспециализированных вычислительных систем, особенно для задач передачи, обработки и защиты данных, распознавания изображений, квантовые вычисления, в том числе для проведения научных исследований в современных областях физики.

В области телекоммуникаций решения на базе трёхзначной логики фактически могут обеспечить 1.5-кратный рост скорости передачи данных (а в перспективе и экспоненциальный). Для практического решения данной задачи необходимо реализовать схемы из чипов, работа которых основана на трёхзначной логике. Для возможности реализации таких схем должна быть решена принципиально важная задача – **задача полноты классов функций трёхзначной логики**. С практической стороны полнота классов таких функций гарантирует, что на базе произвольного множества чипсетов можно произвести плату с нужной функциональной схемой.

В данной работе рассматриваются **операторы замыкания** на множестве функций трёхзначной логики, являющиеся усилением обычного оператора подстановки. Мы покажем, что задача полноты для этого оператора имеет решение, можно восстановить подрешетку замкнутых классов в общем случае замыкания функций относительно классического оператора суперпозиций, что позволит оптимизировать возможное производство чипов для новых функциональных схем для задач передачи, обработки данных.

Любая из задач построения прикладных систем на базе k -значной логики сводится к определению факторов, влияющих на процесс, и рассмотрению счетного множества P_3 состояний этих факторов (любое счётное число состояний можно аппроксимировать до принципиально трех состояний) [2]: 0, 1, 2.

Для принятия решения необходимо найти значение выходной функции Y . Соответственно мы можем представить выходную функцию Y

как комбинацию предикатов на множестве P_3 [1]: для этого мы будем рассматривать сложные предикаты и суперпозиции этих предикатов на P_3 .

Данные предикаты практически могут быть реализованы как схемы из чипов, работа которых основана на трёхзначной логике.

Для возможности реализации таких схем должна быть решена принципиально важная задача – *задача полноты классов функций трёхзначной логики* [1]. С практической стороны полнота классов таких функций гарантирует, что на базе произвольного множества чипсетов можно произвести плату с нужной функциональной схемой. Для двузначной логики эта задача была также решена Постом, что привело к бурному росту электроники [4].

Классическая теорема Поста описывает пять предполных классов в множестве булевых функций [4].

Для случая трёхзначной логики задача была решена С.В. Яблонским в 1958 г. [1, 2]. Он показал, что для функций трёхзначной логики существует 18 предполных классов. В работах [1, 2] рассматривалось замыкание множества функций относительно оператора подстановки.

К сожалению, для трёхзначной логики было доказано, что в общем случае эта задача не может быть решена [2]. Если в случае двузначной логики решётка замкнутых классов счетная, то в случае трёхзначной логики она экспоненциальная.

Однако можно рассматривать *операторы замыкания* на множестве функций трёхзначной логики, являющиеся усилением обычного оператора подстановки. Решение для этого нового оператора замыкания задач полноты и нахождение структуры решётки замкнутых классов поможет не только восстановить подрешётку замкнутых классов в общем случае замыкания функций относительно классического оператора суперпозиций, но и позволит оптимизировать возможное производство чипов для функциональных схем для решения задач, описанных выше.

Исследование оператора замыкания и описание предполных классов в P_3

Пусть дано множество M функций из P_3 . Через $R^2(M)$ будем обозначать результат замыкания множества функций M относительно операции подстановки и перехода от функции g к эквивалентной функции $f \sim g$, где

$$f \sim g \Leftrightarrow (\forall \bar{x} [(f(\bar{x}) = g(\bar{x})) \vee (f(\bar{x}), g(\bar{x}) \in \{0, 1\})]) \& (|X_f| = |X_g|).$$

Пусть $\alpha^n \in E_3^n$, $\alpha^n = (a_1, \dots, a_n)$, обозначим $O_2(\alpha^n) = \{a_i \mid a_i = 2, i \in \{1, n\}\}$

– порядок двойки набора α .

Определение 2. Для $f \in P_3$, такого, что $|X_f| = n$, считаем $f \in \nabla_2$, если для каждого α^n , такого, что $\alpha^n \in E_3^n$ и $O_2(\alpha^n) > 0$, выполняется $f(\alpha^n) = 2$. Также считаем, что все константы содержатся в ∇_2 .

Можно показать [3], что будет иметь место

Теорема 1. (Полноты) Существует пять R_2 -предполных классов $P_3 : T_2, T_{01}, T_-, \nabla_2, P_3(x)$.

Доказательство получаем из следующих трёх лемм.

Лемма 1. Множество ∇_2 является R_2 -замкнутым.

Доказательство. Очевидно, в силу определения замыкания, что переход к функции, находящейся в отношении эквивалентности, не выводит множество за пределы ∇_2 .

Осталось доказать, что ∇_2 -замкнутое относительно операции подстановки. Пусть $f \in \nabla_2$. Очевидно, что после подстановки констант вместо одной (или нескольких) переменных функции f она не потеряет свойство принадлежности классу ∇_2 . Если же мы вместо всех существенных переменных подставим константы, то получим константу, которая, в свою очередь, также лежит в ∇_2 . Пусть $f \in \nabla_2$, а $g_i, i = 1, \dots, n$, либо переменные, либо $g_i \in \nabla_2$. Возьмем $h = f(g_1, \dots, g_n)$. Без ограничения общности считаем, что h существенно зависит от всех своих переменных, так как если это не так, то берем $h' = h$, полученную отбрасыванием фиктивных переменных. Пусть $|X_h| = k$. Если на вход h подается набор $\gamma_k = (c_1, \dots, c_k)$, и для некоторого i выполняется $c_i = 2$, тогда эта 2-ка попадет на вход к одной из g_i . Но если g_i – переменная, то получаем, что на вход к f подошла 2-ка, а $f \in \nabla_2$, значит, $h(\gamma^k) = 2$. Если же $g_i \in \nabla_2$, то на выходе g_i будем иметь 2, которая также идет на вход f , но $f \in \nabla_2$, значит, $h(\gamma^n) = 2$, следовательно $h \in \nabla_2$.

Лемма 2. Класс $P_3(x)$ является R_2 -предполным.

Доказательство. R_2 -замкнутость класса $P_3(x)$ следует из того, что $[P_3(x)] = P_3(x)$ и по определению из $f \sim g$ вытекает, что $|X_f| = |X_g|$, т.е. расширение не выйдет за пределы $P_3(x)$.

Докажем, что $R_2(P_3(x) \cup f^{P_3(x)}) = P_3$. Пусть $g_1(x) = x + 2$, $g_2(x) = x$. Прямая проверка определений классов из множества предполных классов относительно оператора подстановки показывает, что

$$g_1 \notin M_1^3 \cup M_2^3 \cup M_3^3 \cup T_{01} \cup T_{02} \cup T_{12} \cup T_0 \cup T_1 \cup T_2 \cup \\ \cup H_0 \cup H_1 \cup H_2 \cup T_- \cup T_{\{0,2\},1} \cup T_{\{1,2\},0}, g_2 \notin L \cup S_{x+1}.$$

Из определения Sl выполняется $P_3(x) \subset Sl$.

Покажем, что $R_2(P_3(x) \cup f^{P_3(x)}) = P_3$. Если $f^{P_3(x)} \notin Sl$, то, в силу теоремы Слупецкого, все доказано. Пусть $f = f^{P_3(x)} \in Sl$, т.е. $f \notin P_3(x)$, $f \in Sl$. Пусть $|X_f| = n$. Рассмотрим вектор значений $f = (f_1, \dots, f_{3n})$. По определению f множество $\{0, 1, 2\}$ не содержится в $\{f_1, \dots, f_{3n}\}$, следовательно, существует единственное $a \in \{0, 1\}$, такое, что $f_i \in \{a, 2\}$, $i = 1, \dots, 3$. Пусть без ограничения общности $a = 0$. Возможны три случая:

а) Пусть $O_2(f^n) \leq 3^n - 2$, т.е. имеются α_1^n , α_2^n , такие, что $f(\alpha_1^n) = f(\alpha_2^n) = 0$. Перейдем к функции g , такой, что $f \sim g$ и g совпадает с f на всех наборах, кроме α_1^n , на котором $g(\alpha_1^n) = 1$. В силу того, что $1 \notin \{f_1, \dots, f_{3n}\}$, расширение корректно (т.е. $|X_f| = |X_g|$). Заметим, что $g \notin Sl$.

б) Пусть теперь $O_2(f^n) \leq 3^n - 1$, т.е. имеется единственное α_1^n , что $f(\alpha_1^n) = 0$. Заметим, что существует $g \in P_3(x)$, $g = (2, 2, 0)^T$. Рассмотрим функцию $h = g'(f)$. Очевидно, что $|X_f| = |X_h|$. Также заметим, что если $h^n = (h_1, \dots, h_{3n})$, то $h_i \in \{0, 2\}$, $i = 1, \dots, 3n$, и $O_2(h^n) = 1 \leq 3^n - 2$, таким образом пункт б) сводится к пункту а).

с) Если $O_2(f^n) = 0$, то с помощью функции $x + 1 \pmod{3}$ можно построить $f'(\tilde{x}) = f(\tilde{x}) + 1 \pmod{3}$. При этом $f' \in R_2(P_3(x) \cup f^{P_3(x)})$, и для f' выполнено либо условие а), либо условие б), а, значит, найдется $f^{SI} \in R_2(P_3(x) \cup f^{P_3(x)})$, следовательно, $P_3(x)$ – R_2 -предполный класс.

Лемма 3. Если после замены в векторе функции f , $|X_f| \geq 1$ нуля на единицу или единицу на ноль получился вектор функции g , $|X_g| < |X_f|$, то с помощью подстановки констант в функцию f можно получить функцию $h \in P_2(x)$.

Доказательство. Пусть после перехода от f к g переменная x_i стала несущественной, тогда существуют наборы

$$\alpha_1^n = (a_1, \dots, a_{i-1}, a_i^1, a_{i+1}, \dots, a_n), \quad \alpha_2^n = (a_1, \dots, a_{i-1}, a_i^2, a_{i+1}, \dots, a_n), \\ \alpha_3^n = (a_1, \dots, a_{i-1}, a_i^3, a_{i+1}, \dots, a_n),$$

что существует перестановка $\sigma \in S$, $\sigma(\alpha_i^j) = j$, при $j \in E_3$. Заметим, что $f(\alpha_1^n) \sim f(\alpha_2^n) \sim f(\alpha_3^n)$ и $f(\alpha_1^n), f(\alpha_2^n), f(\alpha_3^n) \in E_2$, причем существуют $l, m \in E_3$, что $f(\alpha_1^n) \neq f(\alpha_m^n)$. Получаем, что функция $h(x) = f(a_1, \dots, a_{i-1}, x, a_{i+1}, \dots, a_n)$ будет удовлетворять всем требуемым условиям.

Таким образом, для оператора R_2 доказано, что при $k = 3$ существует 5 предполных классов.

ЛИТЕРАТУРА

1. Яблонский С.В., Гаврилов Г.П., Кудрявцев В.Б. Функции алгебры логики и класса Поста. М.: Наука, 1966.
2. Яблонский С.В. Функциональные построения в k -значной логике // Тр. Матем. инст. им. В.А. Стеклова. 1958. Т. 51.
3. Esin A.A. On function classes in P_3 precomplete with respect to a strengthened closure operator // Math Notes. 2008. V. 83. No. 5. P. 594–603. DOI: 10.1134/S00014f[608050027.
4. Post E.L. Two-valued iterative systems of mathematical logic // Annals of Math. Studies. Princetin Univ. Press. 1941. V. 5.

Математическое и компьютерное моделирование системы согласования материального баланса в нефтепереработке и нефтехимии

Вячеслав Кувыкин, Максим Брюханов

*Нижегородский государственный университет
им. Н.И.Лобачевского, г. Нижний Новгород, Россия
ООО «Ватсон», г. Нижний Новгород, Россия*

Разработана концепция автоматизации согласования материального баланса, построения поточной схемы нефтеперерабатывающего завода с компьютерной визуализацией поточной схемы, автоматического построения модели материальных потоков с применением пакета символьных вычислений. Для согласования баланса использована дополнительная информация по качеству потоков. Приведены результаты практического применения результатов компьютерного моделирования для нефтеперерабатывающего завода.

Введение

Важной задачей цифровизации производств являются вопросы согласования материального баланса, которые возникают в различных отраслях современного производства [1–3]. Одной из проблем является большая размерность модели реального производства, насчитывающего несколько тысяч потоков. Актуальной задачей является определение «узких мест» по оснащению приборами учета и переход к автоматическому сведению баланса в системах большой размерности. Большое внимание уделяется сведению баланса поступления углеводородного сырья и выработки продукции в нефтепереработке [4, 5].

Математическая модель

Пусть \mathbf{x} – вектор, описывающий массу материальных потоков, остатки продукции в резервуарах, а \mathbf{x}_0 – измеренные значения массы. В качестве критерия оптимизации рассматривается минимизация целевой функции $J(\mathbf{x})$, представляющей собой сумму квадратов отклонений, измеренных и получаемых в результате решения [1]

$$\mathbf{x}^* = \arg \min \{ J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_0)^T \mathbf{K}^{-1} (\mathbf{x} - \mathbf{x}_0) \mid \mathbf{x} \in \mathbf{R}^n \} \quad (1)$$

с условием

$$\mathbf{Bx} = 0, \quad (2)$$

где \mathbf{B} – матрица балансовых уравнений размерности $m \times n$, \mathbf{K} – матрица весовых коэффициентов, характеризующих погрешности соответствующих измерений [1].

Если измеряются качественные характеристики потоков, например, при помощи поточных анализаторов, то информацию можно использовать для сведения материального баланса. В данной работе предлагается включить качественные показатели потоков и дополнить систему (2) уравнением

$$F(\mathbf{x}, \mathbf{y}) = 0. \quad (3)$$

Здесь \mathbf{y} – характеристики качества, F – функция, описывающая связь качественных характеристик материальных потоков. Критерий (1) дополняется условием минимизации квадратов отклонений измеренных и согласованных значений по качеству, которые характеризуются вектором \mathbf{y} . В этой расширенной модели число приборов для измерения массы можно уменьшить и соответственно сократить расходы на дорогостоящие средства измерения.

Моделирование и примеры расчета

Для решения задачи размещения приборов и сведения баланса (1), (2) разработана программа с графическим пользовательским интерфейсом GUI (Graphical User Interface), реализованная на языке C# с использованием платформы .Net Framework. Легко интерпретируемый интерфейс GUI дает возможность быстро создавать актуальные схемы потоков, распределить участки по созданию фрагментов поточных схем между группами специалистов. Анализ математической модели осуществляется с применением символьных расчетов. Построение поточной схемы производства, модели баланса и оптимизация размещения приборов объединены в единой информационной системе.

Встроенный редактор позволяет создать графическое изображение объектов модели, вид рабочего окна приведен на рис.1. Визуальная среда позволяет интерактивно манипулировать объектами и связями. Система балансовых уравнений (1) может быть выведена на экран. Расчет числа приборов производится в символьной форме.

Пусть поточная схема такова, что матрица коэффициентов системы уравнений (2) представлена в виде

$$\mathbf{V} = \begin{pmatrix} -1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 & 0 & 0 & 1 \end{pmatrix}. \quad (4)$$

На рис. 1 представлен интерфейс программы для модели смешения. По поточной схеме автоматически сгенерированы уравнения баланса, матрица (4). Расчет показывает, что для построения баланса достаточно 5 приборов. Представлено аналитическое решение задачи о размещении минимального количества измерительных приборов для сведения баланса и установки недостающих приборов.

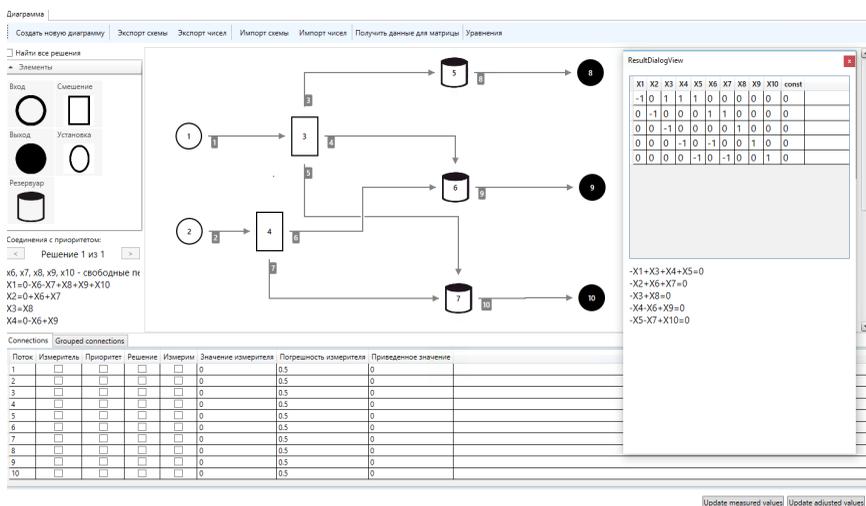


Рис. 1. Интерфейс программы расчета необходимого количества приборов и согласования материального баланса

Предлагаемая методология применялась при модернизации математической модели смешения темных нефтепродуктов на НПЗ ООО «ЛУКОЙЛ-Нижегороднефтеоргсинтез». Для описания качества (3) использована модель Вальтера, которая позволяет произвести расчет вяз-

кости нефтепродуктов при выборе эмпирических коэффициентов на основании экспериментальных данных [6]:

$$x_j \rho_j^{-1} \lg \lg(y_j + c) - \sum_i x_i \rho_i^{-1} \lg \lg(y_i + c) = 0. \quad (5)$$

Здесь индекс j относится к смеси, i – к компонентам, y_i , ρ_i – кинематическая вязкость и плотность i -го потока соответственно, c – эмпирический коэффициент.

Заключение

Предложена концепция согласования баланса и автоматизации расчетов минимального количества приборов учета с использованием информационных систем символьных вычислений. В систему уравнений математического программирования включены переменные, описывающие качество потоков, что дает возможность повысить достоверность баланса и снизить требования по количеству приборов измерения массы.

Предложенная методика и компьютерные программы могут применяться для разработки плана дооснащения измерительными приборами и согласования баланса с учетом качества материальных потоков в нефтеперерабатывающей, нефтехимической и газоперерабатывающей промышленности.

ЛИТЕРАТУРА

1. *Cámara M.M. et al. Numerical Aspects of Data Reconciliation in Industrial Applications // Processes. 2017. No. 5. 56 p.*
2. *Кувыкин В.И. Использование моделей бизнес-процессов НПЗ в системах планирования и учёта // Мир нефтепродуктов. Вестник нефтяных компаний. 2013. № 7. С. 47–48.*
3. *Narasimhan S., Jordache C. Data reconciliation and gross error detection. Houston: Gulf Publishing Company, 2000. 406 p.*
4. *Erokhin V.I., Laptev A.Yu., Lisitsyn N.V. Reconciliation of Material Balance of a Large Petroleum Refinery in Conditions of Incomplete Data // Journal of Computer and Systems Sciences International. 2010. V. 49. No. 2. P. 295–305.*
5. *Кувыкин В.И., Матвеев А.Е., Сычев А.Г., Брыуханов М.В., Кувыкина Е.В. Automation of Calculations of Optimal Disposition of Devices of a Material Balance System // Chemical and Petroleum Engineering. 2018. V. 54(3). P. 266–272.*
6. *Кувыкин В.И., Бреханов М.В., Кувыкина Е.В., Пискунов И.В., Сычев А.Г. Актуализация расчета вязкости смесей темных нефтепродуктов в системе производственного планирования НПЗ // Мир нефтепродуктов. Вестник нефтяных компаний. 2017. № 9. С. 25–31.*

Параметрический анализ математических моделей оптимального планирования нефтепереработки и компьютерное моделирование

Вячеслав Кувыкин¹, Артем Колпаков², Елена Колпакова³

*¹ Нижегородский государственный университет
имени Н.И.Лобачевского, г. Нижний Новгород, Россия*

² ПАО «ЛУКОЙЛ», г. Москва, Россия

³ ООО «ЛУКОЙЛ-Нижегороднефтеоргсинтез», г. Кстово, Россия

Рассматриваются инновационные решения по улучшению математических моделей оптимального планирования нефтепереработки и нефтехимии. Один из подходов к актуальной в период кризиса задаче оптимизации производства связан с использованием специальных ИТ-программ максимизации маржинальной прибыли. Цель работы состоит в разработке методов оценки влияния параметров модели на решение задачи математического программирования. Даны рекомендации по улучшению сопровождения математических моделей реального производства для улучшения бизнес-процессов в нефтегазовой области.

В период резкого падения цен на нефть возникает актуальная задача оптимизации производства. Для оптимального планирования широко используются специальные программы, которые опираются на задачи математического программирования. Модели производства характеризуются большой размерностью и для решения задачи оптимизации разработаны специальные вычислительными методы. Важным элементом практического использования математических моделей является параметрический анализ задачи [1, 2]. Для задач линейного программирования получены условия устойчивости решения при изменении параметров системы [1]. Для параметрического программирования в настоящее время не существует общих методов решения [3].

Задачи нелинейного программирования широко используются в нефтегазовой отрасли при оптимальном планировании производства [4, 5]. В задачах производственного планирования коэффициенты при переменных считаются постоянными. Отметим, что параметры в реальных условиях подвержены изменению под действием случайных факторов и никогда не могут быть точно заданы. В работах [4, 5] было подчеркнуто, что математическая модель может резко менять свое поведение при малом изменении параметров и не отражает реальных процес-

сов. В практических приложениях такая модель требует дополнительного исследования о границах ее применения либо включения в свой состав более сложных динамических процессов. Причем подмножество параметров, при котором происходит разрыв первого рода максимума целевой функции при изменении параметров, наиболее существенно влияет на экономику и представляет несомненный интерес.

Математическая модель

Задачу оптимального планирования производства нефтеперерабатывающего завода можно рассматривать как задачу нелинейного программирования для n переменных (x_1, x_2, \dots, x_n) , которые доставляли бы максимум целевой функции

$$\max L(x_1, x_2, \dots, x_n) \quad (1)$$

и удовлетворяли бы системе из m ограничений

$$g_i(x_1, x_2, \dots, x_n) \leq b_i, i = 1, 2, \dots, m, \quad (2)$$

где переменные $x_j \geq 0, j = 1, 2, \dots, n$; b_i – константы; а либо $L(x_1, x_2, \dots, x_n)$, либо хотя бы одна из функций $g(x_1, x_2, \dots, x_n)$ – нелинейная. Для вектора $\mathbf{X} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ множество допустимых решений задачи обозначим $\mathbf{H} = \{\mathbf{X} \mid g_i(\mathbf{X}) \geq 0, i = 1, \dots, m\}$. Пусть оптимальное решение задачи $\mathbf{X}_p \in \mathbf{H}$

$$\max_{\mathbf{X} \in \mathbf{H}} L(\mathbf{X}) = L_p(\mathbf{X}_p). \quad (3)$$

Предположим, что имеется k параметров λ_k . Введем понятие устойчивости в смысле ключевых параметров. Будем считать, что решение $L_p(\lambda_k)$ системы нелинейных уравнений (1) (2) с параметром λ_k и начальными условиями λ_{0k} устойчиво в смысле ключевых параметров, если для любого $\varepsilon > 0$ найдется число $\delta = \delta(\varepsilon) > 0$, такое, что если $\|\lambda_{0k} - \lambda_k\| < \delta$, то $\|L_p(\lambda_{0k}) - L_p(\lambda_k)\| < \varepsilon$ для всех значений λ_k . В противном случае решение $L_p(\lambda_k)$ назовем неустойчивым.

Дополнительно предположим, что неустойчивость связана со скачками функции $L_p(\lambda_k)$ при изменении параметров и назовем ее неустойчивостью по ключевым параметрам. Значение λ_{0k} , при котором нарушается условие устойчивости, назовем критическим значением ключевого параметра, а сам параметр λ_{0k} – ключевым параметром задачи. Поскольку при малом изменении ключевых параметров изменение максимума целевой функции (3) происходит скачком, то это обстоятельство может привести к серьезным экономическим просчётам и недополученной прибыли. В отличие от известных работ по устойчивости решения

задач линейного программирования [1] в работе изучается подмножество задач, связанное лишь с разрывами первого рода максимума целевой функции при изменении параметров.

Проиллюстрируем сказанное двумя примерами. Проведен расчет оптимального плана производства ООО «ЛУКОЙЛ-Нижегороднефтеоргсинтез» (цены при расчете условны). Нефтеперерабатывающий завод является одним из крупнейших в РФ с мощностью по первичной переработке нефти 17 млн т. Для расчета использовалась система RPMS (Honeywell). Отметим, что в системе RPMS рассчитываются теневые цены. Скачкообразные изменения теневых цен при изменении параметров помогают определить область ключевых параметров.

Первый пример связан с производством автобензинов с октановым числом 100, в выпуске которого нижегородский завод является лидером. Численно рассчитана зависимость максимума целевой функции и теневой цены от октанового числа λ_1 одного из компонентов смешения (катализатора установки риформирования бензинов). Обозначим безразмерную целевую функцию и безразмерную теневую цену L_w и S_w соответственно

$$L_w = L_p / L_0, \quad S_w = S_p / S_0,$$

где L_0 и S_0 – характерные значения маржинальной прибыли и теневой цены.

Результаты компьютерного моделирования представлены на графике (рис. 1). При значении параметра $\lambda_{1p} = 101.5$ происходит скачок максимума целевой функции и теневой цены.

В управлении производственным процессом, поскольку октановое число непостоянно, следует поддерживать технологический режим с запасом по отношению к значению ключевого параметра λ_{1p} . Экономически нецелесообразно, как следует из приведенных расчетов, производить компоненты с большим значением октанового числа, поскольку маржинальная прибыль не возрастает с ростом λ_1 , а при этом износ катализатора увеличивается. С другой стороны, при значениях параметра, меньших ключевого значения, возникают риски существенных потерь операционного дохода.

Как следует из представленного на рис. 1 графика, при численном расчете следует обратиться к анализу теневых цен для идентификации ключевых значений параметров задачи. При обнаружении скачкообразного изменения теневых цен целесообразно рассмотреть поведение решения задачи.

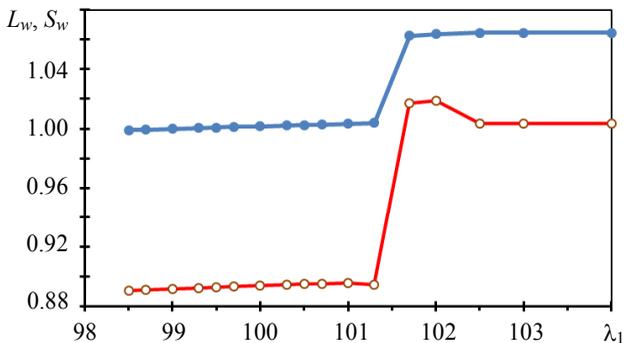


Рис. 1. Зависимость максимума целевой функции L_w (●) и теневой цены S_w (○) от октанового числа одного из компонентов смещения λ_1

Второй пример связан с расчетом кинематической вязкости смеси по известным вязкостям компонентов. Задача теоретически до сих пор не поддается решению и применяются эмпирические формулы. На рис. 2 приведены результаты компьютерного моделирования производства нефтепродуктов при изменении параметра в формуле Вальтера для расчета вязкости смеси [6]. Показаны зависимости максимальной маржинальной прибыли и оптимального объема производства от параметра λ_2 . Максимальное значение целевой функции непрерывно зависит от λ_2 , а оптимальный объем переработки меняется скачкообразно при измене-

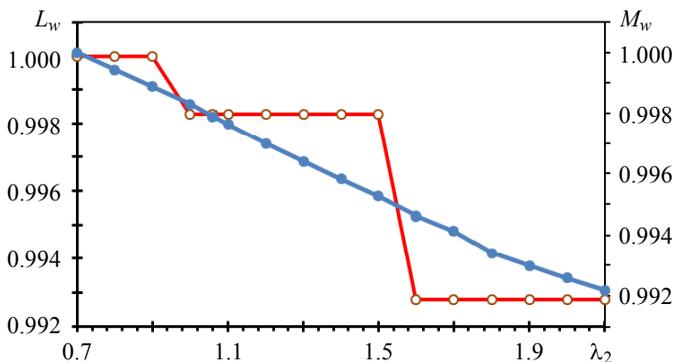


Fig. 2. Зависимость максимума целевой функции L_w (●) и объема переработки M_w (○) от параметра λ_2 при расчете кинематической вязкости

нии параметра. Теневые цены не изменяются. В данном примере неустойчивости по ключевому параметру не обнаружена.

Возникновение неустойчивости при изменении параметров без разрывов максимума целевой функции решения задачи математического программирования можно назвать мягкой параметрической неустойчивостью, при скачкообразном изменении максимума целевой функции – жесткой параметрической неустойчивостью.

Заключение

Исследованы основные закономерности поведения оптимального решения задачи в зависимости от параметров модели. Задачи математического программирования могут обладать особенностями, крайне неприятными для пользователя. Показано, что существуют ключевые значения параметров, в которых максимум целевой функции имеет разрывы первого рода. Именно в точках разрыва требуется провести анализ адекватности математической модели, поскольку существенно изменяется экономический прогноз и возникают проблемы численных расчетов. Результаты работы могут быть использованы при разработке программных комплексов, предназначенных для оптимизации бизнес-процессов.

ЛИТЕРАТУРА

1. *Ашманов С.А.* Условия устойчивости задач линейного программирования // Журн. вычисл. матем. и матем. физ. 1981. Т. 21. № 6. С. 1402–1410.
2. *Zlobec S.* Nondifferentiable Optimization: Parametric Programming // *Floudas C., Pardalos P.* Encyclopedia of Optimization. Boston: Springer, 2008. P. 2607–2615.
3. *Салимоненко Д.А.* Способ решения задачи линейного программирования с переменными коэффициентами в виде параметрических функций // Вестник Башкирского университета. 2015. Т. 20. № 1. С. 25–27.
4. *Кувыкина Е.В., Кувыкин В.И., Петухов М.Ю.* Параметрический анализ математических моделей в задачах линейного программирования // Вестник Нижегородского университета им. Н.И. Лобачевского. 2010. № 3-1. С. 168–172.
5. *Кувыкин В.И., Кувыкина Е.В., Петухов М.Ю.* Анализ оптимальных решений в задачах нелинейного программирования // Вестник Нижегородского университета им. Н.И. Лобачевского. 2011. № 4-5. С. 2285–2286.
6. *Кувыкин В.И., Брюханов М.В., Кувыкина Е.В., Пискунов И.В., Сычев А.Г.* Актуализация расчета вязкости смесей темных нефтепродуктов в системе производственного планирования НПЗ // Мир нефтепродуктов. Вестник нефтяных компаний. 2017. № 9. С. 25–31.

Компьютерное моделирование локализации деформации сыпучей среды в сходящемся канале

Ольга Кузоватова

Сибирский федеральный университет, г. Красноярск, Россия

Рассматривается специальная математическая модель, обобщающая теорию упругости на случай материала, по-разному сопротивляющегося растяжению и сжатию. Основным методом исследования является метод конечных элементов. Разработан алгоритм численного решения задачи о движении сыпучей среды в сходящемся канале.

Математическая модель

Одной из фундаментальных задач геомеханики является задача о движении геоматериалов в сходящихся каналах. К геоматериалам относятся различные горные породы, сыпучие материалы, сухие и водонасыщенные грунты. Анализ поведения геоматериалов имеет значение при добыче полезных ископаемых, строительстве инженерных сооружений [1, 2].

В данной работе рассматривается модель среды с пластическими связями, которая описывает напряженно-деформированное состояние сыпучей среды как разнопрочного материала [3, 4]. Такая среда не деформируется при условии действия сжимающих или растягивающих напряжений, меньших коэффициента сцепления. При достижении предела прочности деформация развивается в соответствии с теорией линейного упрочнения.

На рис. 1 представлена реологическая схема данной модели. Введем следующие обозначения: σ – тензор полных напряжений, σ^c – тензор напряжений в контакте, σ^0 – тензор сцепления, $\sigma^e = E : \varepsilon$ – тензор упругих напряжений, ε – тензор деформаций, E – симметричный положительно определенный тензор модулей упругости.

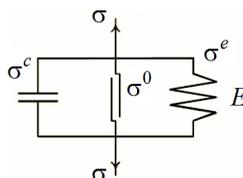


Рис. 1. Реологическая схема

В рамках данной схемы имеем

$$\sigma = \sigma^c + \sigma^0 + \sigma^e.$$

Тензор σ^c удовлетворяет вариационному неравенству

$$\sigma^c : (\tilde{\varepsilon} - \varepsilon) \leq 0, \quad \varepsilon, \tilde{\varepsilon} \in C, \quad (1)$$

где C – конус допустимых деформаций, а именно, $C = \{\varepsilon \mid \kappa\gamma(\varepsilon) \leq \theta(\varepsilon)\}$; κ – параметр дилатансии; $\gamma(\varepsilon) = \sqrt{2\varepsilon' : \varepsilon'}$ – интенсивность сдвига; $\varepsilon' = \varepsilon - \theta\delta/3$ – девиатор деформаций; $\theta(\varepsilon) = \varepsilon : \delta$ – деформация объема; δ – символ Кронекера.

Таким образом, вариационное неравенство (1) имеет вид

$$(E : \varepsilon - \sigma + \sigma^0) : (\tilde{\varepsilon} - \varepsilon) \geq 0, \quad \varepsilon, \tilde{\varepsilon} \in C.$$

Это означает, что

$$\varepsilon = \pi_C [E^{-1}(\sigma - \sigma^0)],$$

где π – оператор проекции на конус C по норме $|\varepsilon| = \sqrt{E : \varepsilon}$.

Пусть Ω – плоская область, $\partial\Omega = \Gamma$ – ее граница. Рассмотрим элемент конструкции из разнопрочного материала, занимающий Ω . Предположим, что $\Gamma = \Gamma_u \cup \Gamma_\sigma$, так что $\Gamma_u \cap \Gamma_\sigma = \emptyset$. На Γ_u отсутствуют перемещения, на Γ_σ задан вектор распределенной нагрузки p . В этом случае уравнения равновесия и граничные условия имеют вид

$$\int_{\Omega} (\nabla \cdot \sigma + f)(\tilde{u} - u) d\Omega = 0; \quad (2)$$

$$u = \tilde{u} = 0 \text{ на } \Gamma_u, \quad \sigma \cdot n = p \text{ на } \Gamma_\sigma. \quad (3)$$

Задача (2), (3) сводится к задаче о минимуме $\min_{\tilde{u} \in U_c} J(\tilde{u}) = J(u)$, где

$$J(u) = \int_{\Omega} \left(\frac{1}{2} \varepsilon : E : \varepsilon + \varepsilon : \sigma^0 - f \cdot u \right) d\Omega - \int_{\Gamma_\sigma} p \cdot u d\Gamma,$$

$$U_c = \left\{ u \in H^1(\Omega) \mid u|_{\Gamma_u} = 0, \quad \varepsilon(u) \in C \right\}.$$

Несложные выкладки показывают, что

$$J(u) = -\max_{\substack{\tilde{u} \in U_c \\ \tilde{u} \neq 0}} \frac{\left[\int_{\Omega} (\varepsilon(\tilde{u}) : \sigma^0 - f \cdot \tilde{u}) d\Omega - \int_{\Gamma_\sigma} p \cdot \tilde{u} d\Gamma \right]_+^2}{2 \int_{\Omega} \varepsilon(\tilde{u}) : E : \varepsilon(\tilde{u}) d\Omega},$$

где

$$z_+ = \begin{cases} z, & z \geq 0, \\ 0, & z < 0. \end{cases}$$

Поле перемещений тождественно равно нулю в том и только в том случае, если

$$\int_{\Omega} (\varepsilon(\tilde{u}) : \sigma^0 - f \cdot \tilde{u}) d\Omega - \int_{\Gamma_\sigma} p \cdot \tilde{u} d\Gamma \leq 0, \quad \forall \tilde{u} \in U_c. \quad (4)$$

Нагрузка (f, p) называется безопасной, если $u \equiv 0$. Пусть $p = 0$, $f = m \cdot f^0$, где m – параметр нагружения. Обозначим через m^* коэффициент запаса. Из (4) следует, что нагрузка является безопасной при изменении параметра m от нуля до его предельного значения:

$$m^* = \min_{\substack{\tilde{u} \in U_c \\ \tilde{u} \neq 0}} \frac{\int_{\Omega} \varepsilon(\tilde{u}) : \sigma^0 d\Omega}{\left[\int_{\Omega} f \cdot \tilde{u} d\Omega \right]_+}. \quad (5)$$

Это утверждение есть формулировка кинематической теоремы о предельном равновесии [5].

Рассмотрим задачу о плоском течении под действием собственного веса сыпучей среды в сходящемся канале, предполагая, что $\alpha > \beta$ (рис. 2).

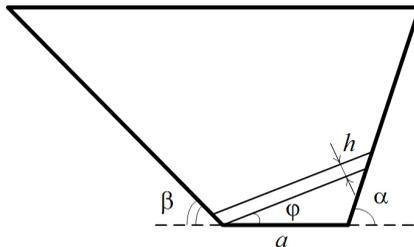


Рис. 2

Локализация деформации простого сдвига с дилатансией происходит в узкой линейной зоне толщины h , наклоненной под углом φ .

Коэффициент запаса (5) в этом случае принимает вид

$$m^* = \frac{2\tau_s}{\kappa g a} \frac{1}{\sin \alpha} \frac{1}{(v \sin \alpha - \cos \alpha)}$$

и зона локализации деформации направлена под углом

$$\varphi = \alpha - \arcsin \frac{1}{\sqrt{v^2 + 1}}, \quad (6)$$

где τ_s – предел текучести и $v = \sqrt{1/\kappa^2 - 4/3}$.

Численное моделирование

Вычислительный алгоритм использует конечно-элементную аппроксимацию модели. Задача определения поля перемещений в разнопрочной среде сводится к решению последовательности статических задач линейной упругости с начальными напряжениями. За начальное приближение итерационного процесса принимается решение упругой задачи. Определяющие уравнения

$$\sigma = E : \varepsilon - \frac{1}{1 + \lambda} \pi(E : \varepsilon - \sigma_0),$$

где $\lambda = \text{const} \geq 0$, заменяются итерационной формулой ($n = 1, 2, 3, \dots$)

$$\sigma^n = E : \varepsilon^n - \frac{1}{1 + \lambda} \pi(E : \varepsilon^{n-1} - \sigma_0).$$

Здесь π – оператор проекции на конус K по норме $|\sigma|_0 = \sqrt{\sigma : E^{-1} : \sigma}$.

Условие завершения итерационного процесса заключается в том, что норма разности двух последовательных решений должна быть меньше 0,0001.

Параметр внутреннего трения разнопрочной среды κ выбирался равным 0.3 для демонстрации результата. Расчеты были проведены и при других значениях $0 < \kappa < \sqrt{3}/2$. Незначительное изменение значения κ влечет незначительное изменение угла φ согласно формуле (6). Параметр регуляризации $\lambda = 0.0001$. Программа написана на языке Borland C++.

Представим графическую интерпретацию результатов вычислительного эксперимента для рассматриваемой задачи (рис. 2).

Обозначим через

$$f_{кр} = m^* \rho g = \frac{2\tau_s}{ka \sin \alpha} \frac{1}{(v \sin \alpha - \cos \alpha)}. \quad (7)$$

При $f < f_{кр}$ локализация деформации не происходит. При $f > f_{кр}$ среда движется. В этом случае должна быть видна зона проскальзывания.

На рис. 3, *a* представлено поперечное сечение несимметричного сходящегося канала: $\alpha = 76.2^\circ$, $\beta = 35^\circ$, $a = 2$ м. На рис. 3, *б* конечно-элементная сетка образца состоит из 1035 узлов и 1922 элементов. Характерный размер стороны треугольника 0.18 м при площади образца 30.92 м^2 . На основе (7) получим $f_{кр} = 0.06\tau_s$, следовательно, по (6) $\varphi = 58.49^\circ$.

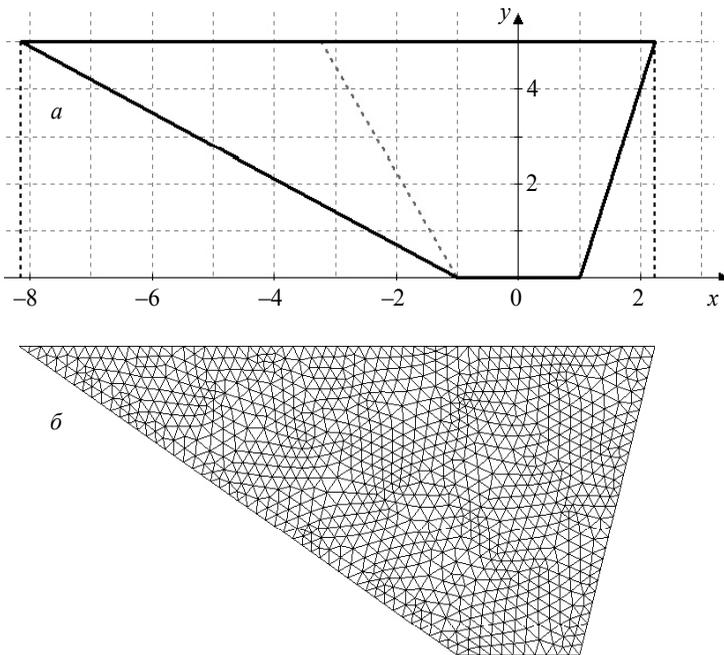


Рис. 3. Несимметричная область

На рис. 4, *a* представлено поле интенсивности сдвига, полученное на основе классической теории упругости, на рис. 4, *б* – на основе модели разнопрочной среды. Прямой черного цвета изображено направление локализации деформаций под углом φ .

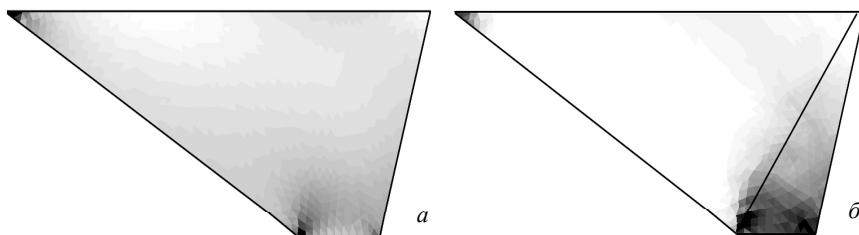


Рис. 4. Интенсивность деформации сдвига

В качестве основного вывода работы можно сказать, что вычислительный эксперимент (для рассматриваемой задачи) подтверждает наличие линейных зон локализации деформаций под углом φ и соответствует рис. 2.

ЛИТЕРАТУРА

1. Лавриков С.В., Ревуженко А.Ф. О расчете локализованных течений сыпучей среды в радиальных каналах // ФТПРПИ. 1990. № 1. С. 3–9.
2. Ревуженко А.Ф., Стажевский С.Б., Шемякин Е.И. Несимметрия пластического течения в сходящихся осесимметричных каналах // ДАН СССР. 1979. Т. 246. № 3. С. 572–574.
3. Мясников В.П., Садовский В.М. Вариационные принципы теории предельного равновесия разнопрочных сред // Прикладная математика и механика. 2004. Т. 68. Вып. 3. С. 488–499.
4. Садовская О.В., Садовский В.М. Математическое моделирование в задачах механики сыпучих сред. М.: Физматлит, 2008. 368 с.
5. Гвоздев А.А. Расчет несущей способности конструкций по методу предельного равновесия. Вып. 1. Сущность метода и его обоснование. М.: Стройиздат, 1949. 280 с.

Реализация механизма сбора и обработки статистических данных потоков заявок в системе имитационного моделирования ODIS

Мария Шкленник, Александр Моисеев

*Национальный исследовательский
Томский государственный университет, г. Томск, Россия*

Программный комплекс ODIS [1, 2], который разрабатывается и используется группой ученых Томского государственного университета, предназначен для имитационного моделирования систем и сетей массового обслуживания. Основными замеряемыми во время имитации показателями в данной системе является число заявок, присутствующих в блоке обслуживания, в очереди, на орбите и т.п. Однако в теории массового обслуживания часто возникают задачи, когда требуется измерять показатели потоков внутри системы [3] или сети, например число заявок, поступивших на вторую фазу обслуживания в многофазной системе за определенный промежуток времени [4] или на блок обслуживания в системах с обратной связью [5, 6]. В дальнейшем собранные данные должны быть подвержены статистической обработке с целью получения эмпирических характеристик, в том числе эмпирического распределения вероятностей (построения полигона частот). Кроме того, несомненный интерес представляют и длины интервалов между последовательными поступлениями заявок в таких потоках, так как их статистический анализ может оказаться ценным для понимания типа потока и оценки его параметров.

Таким образом, была сформулирована задача реализовать в системе имитационного моделирования ODIS компоненты, необходимые для сбора и обработки статистической информации потоков заявок внутри моделируемых систем и сетей.

Накопление и обработка статистической информации

Для решения поставленной задачи был реализован класс StreamStatistics (рис. 1), унаследованный от абстрактного класса StatisticsAccumulator системы ODIS, который является общим для разных конкретных классов сбора статистики в данной системе [7, 8] (на диаграмме для простоты представления опущены операции класса

StatisticsAccumulator, которые не представляют интерес для целей нашей работы, а также их переопределение в классе-потомке).

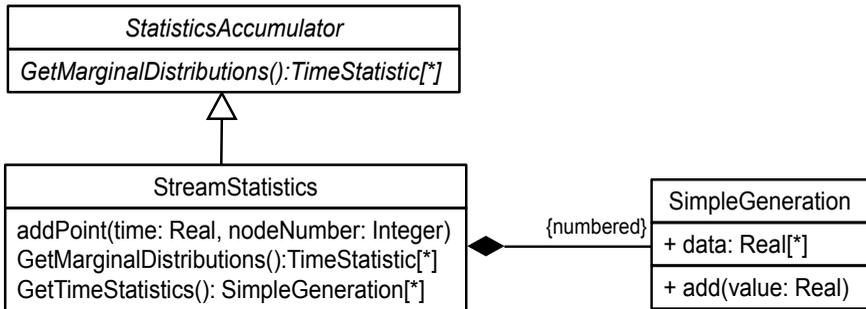


Рис. 1. Разработанный класс StreamStatistics для сбора и обработки статистики потоков заявок

Обязанностью и спецификой класса StreamStatistics является накопление и обработка статистической информации по потокам заявок, поступающих на разные узлы моделируемой сети массового обслуживания (многофазные и однофазные системы здесь также представляются как сеть обслуживания специфической конфигурации). Метод addPoint(...) отвечает за добавление в журнал момента time поступления заявки в узел nodeNumber. Эта информация фиксируется во вложенном объекте класса SimpleGeneration, соответствующем узлу с номером nodeNumber. Класс SimpleGeneration представлен на диаграмме в упрощенном виде, а остальные методы класса StreamStatistics будут описаны ниже.

Для возможности собирать и хранить накопленную статистическую информацию во время процесса моделирования к глобальному объекту SimulationModel (см. [7, 8]) был добавлен атрибут для хранения экземпляра класса StreamStatistics, который инициализируется каждый раз при запуске процесса моделирования. Для сбора статистики потоков были внесены изменения в метод accept(...) класса ServerBlock (рис. 2). Экземпляр класса ServerBlock представляет собой узел сети, а его метод accept(...) вызывается каждый раз при поступлении заявки в данный узел (подробнее см. [8]). В данный метод на входе и выходе (остальная часть метода на диаграмме обозначена как взаимодействие «Process receiving a call») были добавлены вызовы специально созданных методов collectStatisticsBeforeEnter() и collectStatisticsAfterEnter(), которые

позволяют фиксировать состояние узла и собирать статистические данные соответственно в моменты непосредственно до поступления заявки в узел и сразу после. В частности, вызов операции `addPoint(...)` класса `StreamStatistics` для регистрации момента поступления заявки мы добавляем в метод `collectStatisticsBeforeEnter()`, так как нам важно фиксировать сам факт поступления независимо от дальнейшего развития событий (была ли заявка принята узлом или нет). В других случаях полезным окажется метод `collectStatisticsAfterEnter()`, который позволит фиксировать состояние узла в зависимости от дальнейшего развития событий непосредственно после поступления заявки в этот узел.

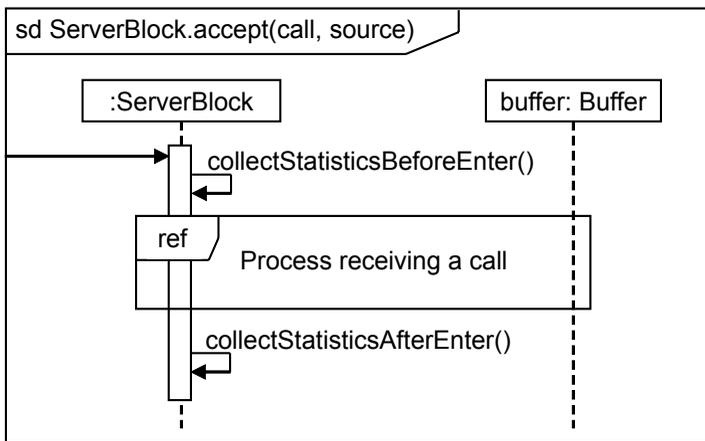


Рис. 2. Модификация реализации метода `accept(...)` класса `ServerBlock`

Аналогичным образом был модифицирован метод `processEvent(...)` класса `ServerBlock`, отвечающий за обработку событий, возникающих внутри узла (например, при окончании времени обслуживания заявки). Для этого в него на входе и выходе добавлены вызовы методов `collectStatisticsBeforeEvent(...)` и `collectStatisticsAfterEvent(...)`, которые соответственно занимаются сбором статистических данных непосредственно до обработки события и сразу после. Для целей нашей работы эти методы не играют роли (так как нам важно фиксировать только поступление заявок в узлы), но их введение позволяет, во-первых, унифицировать подход к реализации механизмов сбора статистической информации в системе ODIS, а во-вторых, обеспечить более гибкую ее настройку.

В итоге по окончании процесса моделирования в экземпляре `StreamStatistics` получаем накопленную информацию по моментам поступления заявок в каждый узел моделируемой сети. Для извлечения, обработки и визуального представления соответствующей статистической информации используем стандартный метод `GetMarginalDistributions()` интерфейса `StatisticsAccumulator`, что позволяет клиентскому коду единообразно работать с целочисленными величинами (в данном случае – с числом заявок, поступивших в каждый узел за заданный интервал времени). А вот для получения информации об интервалах между последовательными поступлениями заявок в каждый узел пришлось реализовать специальный метод `GetTimeStatistics()`, возвращающий массивы из длин интервалов между поступлением заявок для каждого узла. Эти массивы представляют собой экземпляры встроенного в `ODIS` класса `SimpleGeneration` и могут быть обработаны стандартными средствами системы, предназначенными для статистического анализа результатов имитационного моделирования непрерывных случайных величин (построение гистограммы частот, сравнение с распределением и т.д.)

Система `ODIS` реализована на платформе `.NET Framework`, а весь ее исходный код написан на языке программирования `C#`. Поэтому все описанные в работе изменения и нововведения были также реализованы на этом языке программирования и прошли успешные испытания.

Заключение

Таким образом, в результате проделанной работы был спроектирован и реализован механизм сбора и обработки статистических данных потоков заявок в системе имитационного моделирования `ODIS`, реализованный в виде специального класса `StreamStatistics` в иерархии классов, предназначенных для накопления и обработки статистической информации в данной системе. Кроме того, были внесены изменения в существующий класс `ServerBlock`, представляющий узел моделируемой сети массового обслуживания, и класс `SimulationModel`, представляющий моделируемую систему в целом. Эти изменения позволили не только решить поставленную задачу, но и сделать всю систему более гибкой и устойчивой к изменениям, касающимся способов сбора статистической информации.

ЛИТЕРАТУРА

1. *Моисеев А.Н., Сияяков М.В.* Разработка объектно-ориентированной модели системы имитационного моделирования процессов массового обслуживания // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2010. № 1. С. 89–93.
2. *Moiseev A., Demin A., Dorofeev V., Sorokin V.* Discrete-event approach to simulation of queueing networks // Key Engineering Materials. 2016. V. 685. P. 939–942.
3. *Nazarov A., Dammer D.* Methods of limiting decomposition and Markovian summation in queueing system with infinite number of servers // CCIS. 2018. V. 912. P. 71–82.
4. *Shklennik M., Moiseeva S., Moiseev A.* Optimization of two-level discount values using queueing tandem model with feedback // Communications in Computer and Information Science. 2018. V. 912. P. 321–332.
5. *Melikov A., Zadiranova L., Moiseev A.* Two asymptotic conditions in queue with MMPP arrivals and feedback // Communications in Computer and Information Science. 2016. V. 678. P. 231–240.
6. *Жидкова Л.А., Моисеева С.П.* Исследование системы $GI|M|\infty$ с повторными обращениями. // Труды Томского государственного университета. Серия физико-математическая: Математическое и программное обеспечение информационных, технических и экономических систем: материалы II Всероссийской молодежной научной конференции. Т. 295. Томск: Издательский Дом Томского государственного университета, 2014. С. 94–100.
7. *Мещеряков Р.В., Моисеев А.Н., Демин А.Ю., Дорофеев В.А., Матвеев С.А.* Применение параллельных вычислений в имитационном моделировании сетей массового обслуживания // Известия Томского политехнического университета. 2014. Т. 325. № 5. С. 99–109.
8. *Моисеев А.Н.* Исследование математических моделей систем и сетей массового обслуживания с высокоинтенсивными непуассоновскими входящими потоками: дис. ... д-ра физ.-мат. наук: Томск, 2016. 333 с.

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ И ВИЗУАЛИЗАЦИЯ

INTELLIGENT DATA ANALYSIS AND VISUALIZATION

Methods of spelling correction in information retrieval systems

Alyona Borisovskaya

Russian University of Transport (MIIT), Moscow, Russia

This paper deals with basic spelling errors correction algorithms in full-text information retrieval systems.

The main purpose of information retrieval system is to process big array of strings to find set of words or character sequence to satisfy user's information need more. This includes word or character sequences to be expressed as search query. The main relevance criteria are precision (proportion of relevant results among those returned by the system) and recall (proportion of returned results among all relevant) [1].

The methods of spelling correction are applied widely in typographical error robust systems [2] and systems with noisy data channels[3].

Problem statement

Incorrect query is a query containing typographical errors like letter transposition, missed or added letters. The main purpose of retrieval in that case is to find the most relevant term q' in vocabulary V based on misspelled user's query.

There are two approaches to solve the problem of spelling correction: isolated-term correction and context-sensitive correction. In this paper we will focus on isolated-term correction, which is to correct spelling errors in one term at each moment, even if query contains several terms.

The problem is being solved in two stages: the first one is based on editing distance, the second one – on k -gram overlap. The first step is to select the «nearest» alternative term spelling; the concept of distance

between terms will be introduced below. Then, if two correctly spelled terms are tied, the most common is selected. As a criterion of the term frequency, the number of occurrences of the term in the dictionary or the frequency of its occurrence in the queries of other users can be used.

Edit distance (Damerau–Levenstein edit distance)

The edit distance between two strings S_1 and S_2 is a minimal amount of editing operations necessary to transform S_1 to S_2 .

Editing operations include:

- 1) Single letter insertion.
- 2) Single letter deletion.
- 3) Replacing one character with another.
- 4) Adjacent characters transposition.

Damerau–Levenstein editing distance [4] is an editing distance in the case of the above editing operations. It is calculated by the following recurrent formula:

$$d(S_1, S_2) = D(M, N), \quad (4)$$

$$\text{where } D(i, j) = \begin{cases} 0, & i = 0, j = 0, \\ i, & j = 0, i > 0, \\ j, & i = 0, j > 0, \\ \min\{ & \\ & D(i, j-1) + 1, \\ & D(i-1, j) + 1, \quad j > 0, i > 0, \\ & D(i-1, j-1) + m(S_1[i], S_2[i]) \\ & \}, \end{cases} \quad (5)$$

where M, N – the length of each string; $S_1[i], S_2[i]$ – character with an ordinal number i in each string; $m(S_1[i], S_2[i]) = 0$, if $S_1[i] = S_2[i]$, otherwise $m(S_1[i], S_2[i]) = 1$.

The calculation of edit distance is performed by Wagner–Fisher dynamic programming method. The complexity of this algorithm is $O(M * N)$.

Heuristic «nearest» string searching algorithms

Since analysis of all terms in the dictionary in order to find «nearest» strings is expensive, in practice, different heuristic algorithms are successfully applied [5].

The simplest method is to search terms in restricted part of the dictionary consists of terms beginning with the same letter as the query term. Most often spelling errors do not occur in the first character of the query. Another, more complex variant is to use permuterm index in which the end of line symbol «\$» is excluded. For each rotation r from the set of all rotations in query string on each step in B-tree of permuterm index the algorithm should retrieve terms that have a rotation beginning with r . As a modification on this algorithm a suffix of l characters can be omitted.

Jaro–Winkler distance

Jaro–Winkler distance [6] between strings can be calculated as the minimal amount of edit operations, but besides it uses scaling coefficient p due to which strings with the same prefix of length l have higher score.

The result is normalized, therefore, the minimal Jaro–Winkler distance d_w is 0, the maximal is 1.

Jaro–Winkler distance can be calculated by the following formulas:

Jaro distance:

$$d_j(S_1, S_2) = \begin{cases} 0, & m = 0, \\ \frac{1}{3} \left(\frac{m}{M} + \frac{m}{N} + \frac{m-t}{m} \right), & \text{otherwise,} \end{cases} \quad (6)$$

where M, N – the length of each string, m – the amount of matching symbols, t – the half of the total number of transpositions.

Jaro–Winkler distance:

$$d_w(S_1, S_2) = d_j(S_1, S_2) + (lp(1 - d_j(S_1, S_2))), \quad (7)$$

where $d_j(S_1, S_2)$ – Jaro distance for strings S_1 and S_2 ; l – the length of common prefix; p – scaling coefficient, necessary to increase the rating of the terms with common prefixes (<0.25).

K-gram overlap. Jaccard coefficient

K-gram is the sequence of k symbols. In k -gram index dictionary contains all k -grams that appear in dictionary terms. Each k -gram is assigned with linked list of terms from dictionary in which it is contained.

In fact, the k -gram index is used to retrieve dictionary terms that have much in common with the query k -grams [7]. Linear set overlap can be modified with the Jaccard coefficient, which measures the intersection of two sets of the query and the dictionary term k -grams. It can be calculated as

$|A \cap B| / |A \cup B|$. While proceeding from one dictionary term to the next, Jaccard coefficient is dynamically calculated as ratio between q and t . If the coefficient riches the set value, t is added to search result, otherwise the next word is analyzed.

Elasticsearch implementation

Elasticsearch is distributed full-text search engine with web interface and JSON documents. Recently, the Elasticsearch system has been very popular due to its compatibility with the main tools for monitoring and visualizing the state of a distributed system (Elasticsearch, Logstash, Kibana stack), modern flexible API, and serialization using schema-free JSON documents. Starting from version 0.90.0 Beta1, Elasticsearch allows us to use the suggest API to correct user spelling mistakes. The `term` suggester suggests terms based on edit distance. The provided suggest text is analyzed before terms are suggested. The suggested terms are provided per analyzed suggest text token. The `term` suggester doesn't take the query into account that is part of request. HTTP JSON request allows to define parameters such as:

size	The maximum corrections to be returned per suggest text token
max_edits	The maximum edit distance candidate suggestions can have in order to be considered as a suggestion. Can only be a value between 1 and 2. Any other value results in a bad request error being thrown. Defaults to 2.
prefix_length	The number of minimal prefix characters that must match in order to be a candidate for suggestions. Defaults to 1. Increasing this number improves spellcheck performance. Usually misspellings don't occur in the beginning of terms. (Old name "prefix_len" is deprecated)
string_distance	Which string distance implementation to use for comparing how similar suggested terms are. Five possible values can be specified: <ul style="list-style-type: none"> - internal: The default based on <code>damerau_levenshtein</code> but highly optimized for comparing string distance for terms inside the index. - damerau_levenshtein: String distance algorithm based on Damerau–Levenshtein algorithm. - levenshtein: String distance algorithm based on Levenshtein edit distance algorithm. - jaro_winkler: String distance algorithm based on Jaro–Winkler algorithm. - ngram: String distance algorithm based on character n-grams.

sort	Defines how suggestions should be sorted per suggest text term. Two possible values: - <code>score</code> : Sort by score first, then document frequency and then the term itself. - <code>frequency</code> : Sort by document frequency first, then similarity score and then the term itself.
-------------	--

Conclusions

In this paper we considered the problem of spelling correction in user's queries in full-text information retrieval systems as well as main approaches to its solution. Each algorithm measures the distance between query and dictionary term strings can be implemented in the well-known modern search engine Elasticsearch by specifying a special parameter in REST HTTP query.

REFERENCES

1. *Lipani A.* Fairness of information retrieval // Association for Computing Machinery. 2016. P. 1–3.
2. *Gregory V. Bard* Spelling-Error Tolerant, Order-Independent Pass-Phrases via the Damerau–Levenshtein String-Edit Distance Metric // Dept of Applied Mathematics and Scientific Computation, University of Maryland at College Park. 2007. P. 1–4.
3. *Flachs S., Lacroix O., Søgaard A.* Noisy Channel for Low Resource Grammatical Error Correction // Educational Applications Association for Computational Linguistics. 2019. P.191–192.
4. *Frederic P. Miller, Agnes F. Vandome* String metric, Damerau–Levenshtein distance, Spell checker, Hamming distance // Alpha Press. 2009. P. 1–3.
5. *Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze* An introduction to information retrieval // Cambridge UP. 2009. P. 57–61.
6. *Feigenbaum J.* JAROWINKLER: Stata module to calculate the Jaro–Winkler distance between strings // Boston College Department of Economics. 2014. P. 1–5.
7. *Jumanov I.I., Karshiev Kh.B.* Mechanisms for optimization of detection and correction of text errors based on combining multilevel morphological analysis with n-gram models // Journal of Physics: Conference Series. 2020. V. 1546. P. 1–3.

Recurrent neural networks to analyze the quality of natural gas

Ivan Brokarev¹, Mais Farkhadov², Sergey Vaskovskii²

¹ *National University of Oil and Gas «Gubkin University», Moscow, Russia*

² *V. A. Trapeznikov Institute of Control Sciences
of Russian Academy of Sciences, Moscow, Russia*

There are a large number of methods and tools to analyze the quality indicators of natural gas [1]. The most promising direction in the field is the correlation methods that use statistical models, in particular, neural networks. This is because it is possible to train these models. As a result, it is possible to get a clearer definition of the relationship between the model input and output parameters and to determine subsequently the required target parameters of natural gas.

Currently lacks a general algorithm to determine the architecture and parameters of neural network models. Therefore, a comparative analysis of various models was previously carried out to solve the problem. Based on the results of such analysis, it was concluded that recurrent neural networks are main statistical models in this problem. But in contrast to the study mentioned above, this paper considers a recurrent neural network with a more complex architecture, namely, a network with a controlled recurrent block.

Proposed recurrent neural network

Based on the results of correlation analysis and neural network models investigated in previous works [3, 4], we choose the speed of sound, thermal conductivity coefficient and concentration of carbon dioxide as the model input parameters. As the output parameters, we choose the concentrations of the components included in the investigated gas mixture. The Levenberg–Marquardt algorithm was chosen to train the model.

To solve this problem, we use a recurrent neural network with a gated recurrent unit [5] to model our system. Gated recurrent units are a specialized gate mechanism for recurrent neural networks.

The structure of a gated recurrent unit is shown on Fig. 1. The structure of a neural network with this block is similar to the previously presented simple recurrent network [4].

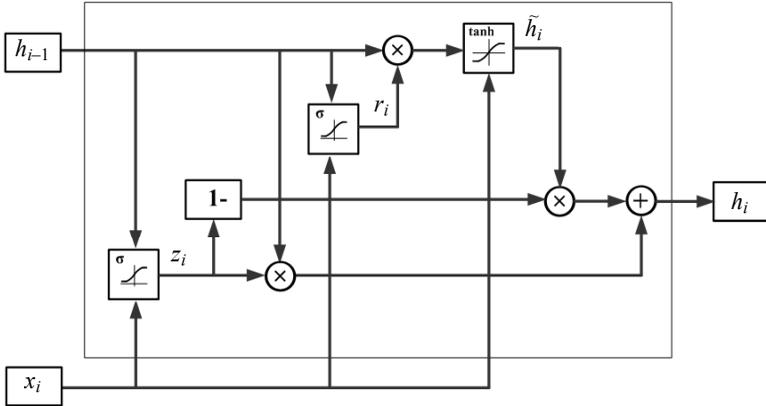


Fig. 1. The architecture of gated recurrent unit

This neural network architecture includes an update gate, a reset gate and a current memory container. Both gates are functions of the input vector and the state of the hidden layer on the previous step. The update gate determines the amount of information from the previous state of the hidden layer that should affect its current state. The reset gate determines the part of information from the previous state that will not be taken into account when we calculate the current state. The current memory container is used to store the relevant information from the previous state and depends on the reset gate.

Let us investigate how the gated recurrent unit operates. When the vector x_i is fed to the block, the vector is multiplied by the vector of weights W_z of the update gate. The vector of the previous state of the hidden layer h_{i-1} , that stores information about the previous $i-1$ steps, is multiplied by the vector of weights U_z of the update gate. Then the two products are added and fed to the sigmoidal activation function σ . As a result, the output value of the update gate z_i is calculated as

$$z_i = \sigma(W_z x_i + U_z h_{i-1}). \quad (1)$$

The formula to calculate the output value of the reset gate r_i is similar to the formula for the update gate, except for the weight vectors W_r and U_r :

$$r_i = \sigma(W_r x_i + U_r h_{i-1}). \quad (2)$$

We use the value of the reset gate to calculate the current value of the memory container. First, the corresponding vectors of the weights W and U .

multiply the input vector and the previous state vector of the hidden layer. Then the Hadamard product (*) of the state reset filter value and the Uh_{i-1} value is calculated. This is necessary to determine the amount of information from the previous steps we do not consider. Then, we sum up the two previously obtained terms and apply the nonlinear hyperbolic tangent activation function (tanh) to obtain the value of the current memory container \tilde{h}_i :

$$\tilde{h}_i = \tanh(Wx_i + r_i * Uh_{i-1}). \quad (3)$$

If the value of the reset gate is close to one, then the memory container saves information from the previous hidden state. If the value of the reset gate is close to zero, then the memory container does not take into account the previous hidden state.

At the final stage of the block, the vector h_i is calculated; the vector contains the information about the current state of the hidden layer and is transmitted to the next layer of the neural network. For this procedure, we use an update gate; the filter determines how much information to take from the current memory container and from the previous state of the hidden layer. To do this, the Hadamard product of the update gate value and the vector of the previous state of the hidden layer, and the Hadamard product of the value of the current memory container and the value $(1-z_i)$ are calculated. Finally, we sum up the results of the two products to get the block output value:

$$h_i = z_i * h_{i-1} + (1 - z_i) * \tilde{h}_i. \quad (4)$$

If the update gate value is close to one, then the current memory container is not taken into account, and the block output equals the previous value of the hidden layer. If the update gate value is close to zero, then the previous state of the hidden layer is not taken into account, and the block output depends only on the current memory container.

The main advantage of a managed recurrent block is the ability to store and filter information using update and reset gates. Such system has no vanishing gradients as the model stores the relevant information and passes the information on to the next time steps of the network.

Testing of neural network model

Based on the results of this study, we obtain a neural network model with a gated recurrent unit to analyze the quality of natural gas by measured physical parameters. To determine if the proposed neural network model is

adequate, we studied a test sample of data. The results and accuracy of our model are shown in Table. The ranges of gas mixtures for the test sample were as follows: 85–100% for methane, 0–5% for propane, nitrogen and carbon dioxide. We took the concentration of carbon dioxide to be zero, since this parameter is an input for the model and is considered a known value.

Accuracy of our RNN model, tested of the gas component composition

Component	Max absolute deviation, %	Average absolute deviation, %	Mean square deviation	Determination coefficient
Methane	0.73	0.33	0.32	0.9999
Propane	0.21	0.12	0.13	0.9999
Nitrogen	0.19	0.08	0.10	0.9999

Conclusions

Proposed neural network model with a gated recurrent unit determines the quality of natural gas by the values of the gas measured parameters. Our model is adequate to determine accurately the composition of the gas for the specified problem. In further research, we plan to test the developed model on the real measurement data of gas mixture parameters and to study how reliable the indicators of the developed system are to analyze the quality of natural gas.

REFERENCES

1. *Brokarev I.A., Vaskovskii S.V., Farkhadov M.P.* Modern methods and tools to analyze the quality of natural gas // *Gas industry*. 2020. No. 9. – P. 38–43.
2. *Brokarev I.A., Vaskovskii S.V.* Distributed data gathering system to analyze natural gas composition // *Advances in Systems Science and Applications*. 2019. V. 19. No. 4. P. 14–24.
3. *Koturbash T.T., Brokarev I.A.* Method for determining the properties and composition of natural gas from the gas physical properties // *Sensors and Systems*. 2018. № 6. P. 43–50.
4. *Vaskovskii S.V., Brokarev I.A.* Neural networks to analyze natural gas composition. The proceedings of 18th Terpugov international conference. Tomsk: NTL Publ., 2019. Part 1. P. 21–26.
5. *Chung, J., Gulcehre, C., Cho, K., Bengio, Y.* Empirical evaluation of gated recurrent neural networks on sequence modeling // *arXiv preprint arXiv:1412.3555*. 2014.

Analysis of reliability of gas analysis system based on vector Wiener process

Ivan Brokarev¹, Sergey Vaskovskii²

¹ National University of Oil and Gas «Gubkin University», Moscow, Russia

² V. A. Trapeznikov Institute of Control Sciences
of Russian Academy of Sciences, Moscow, Russia

The object of research in this work is one of the most important parts of an integrated system for analyzing the quality of natural gas – a system of sensors of individual gas components. Failure of any sensor is a difficultly detected event and its failure actually does not lead to the termination of the system operation, but to the incorrect determination of the components in the gas composition (which, in terms of consequences, is much more critical than an explicit failure). Therefore, the reliability of the sensor system in this work will be considered from the point of view of ensuring the specified measurement accuracy of the components included in the composition of the gas mixture.

To measure each individual component, two independent sensors are used, the readings from which are fed into the system at discrete times. We will assume that if the sensors are operating normally, then the readings from the sensors will give approximately the same result. We will observe a discrete sequence of some random values: $\{\xi_i\}$, $i = 0, 1, 2, \dots$, which we will interpret as the modulus of the difference between the readings of two independent sensors: $\{|\xi_i^{(1)} - \xi_i^{(2)}|\}$, $i = 0, 1, 2, \dots$. It is natural to assume that if the readings of the sensors approximately coincide, then the values of this sequence will fluctuate around zero. If the values deviate from 0, then obviously there is some anomaly in the system. At the same time, one-time deviations, after which the system returns to normal functioning, can be considered insignificant, but a systematically repeated error will already indicate regular systemic violations. Therefore, it is more convenient to consider the accumulated sum of errors $X(k) = \xi_1 + \dots + \xi_k$, and the refusal to record when the sum is reached by the sum of a certain critical level A (Fig. 1). Some substantiation of the transition from the discrete to the continuous model will be given below.

This approach is new and differs significantly from the generally accepted schemes for calculating structural reliability. The work will give some new definitions of reliability for gas analysis systems and propose new

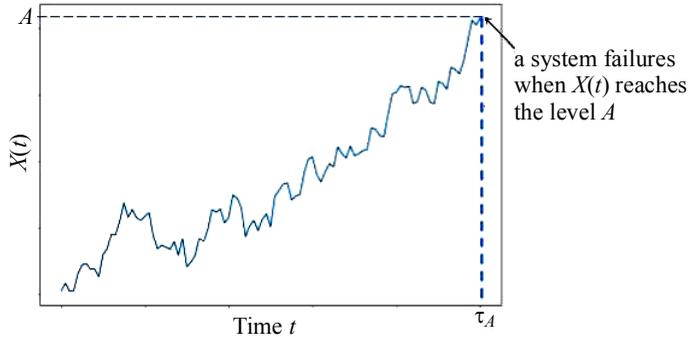


Fig. 1. Subsystem runtime for one component

models for its assessment. A similar approach was used in constructing a model and solving an optimization problem for a system of sensors in [1–3]. We propose a similar reliability model for a sensor system based on a vector wine process [3].

Determining the reliability of the sensor system

Methods for calculating structural reliability based on failure rate data are well known. Examples of building and optimizing similar models for distributed computing systems and network management systems can be found, for example, in [4–7]. However, as applied to the problem of gas analysis, this approach has drawbacks: the concept of structural reliability as «the ability of a system to perform specified functions for a given time interval» in a gas analysis problem is too general and requires the specification of which functions are important in this case.

The first priority is the function of detecting N components that make up natural gas, while characterization of this function of the system through the discrete finite space of the state of the system (for example, «works» / «does not work») does not allow building a model of the system applicable for practical use. since the concept of measurement accuracy is a characteristic that changes continuously, and can vary depending on the task. If the failure of the system for collecting, processing and analyzing information can be detected explicitly (this is, in fact, a failure of the computer), then with sensors this cannot be detected immediately. If the sensors fail, then there may be no obvious external signs of this phenomenon. In this way. this phenomenon is unobservable, and we can only judge about it by the output data [3]. Let's introduce the following definitions:

Definition 1. By the reliability of a measuring system we mean its ability to perform specified functions during a given time interval. In this case, the «given function of the system» will be considered its ability to measure gas components with a given accuracy.

Definition 2. A measurement for the i -th component will be considered accurate if the difference between the values obtained from the first and second sensors lies in a certain preselected range $(-A_i, A_i)$. In the case when it is necessary to determine N components, then we will deal with a vector of positive values:

$$A = (A_1, \dots, A_N), A_i > 0 \forall i = 1, \dots, N,$$

those, for each i -th sensor, a certain range of permissible values $(-A_i, A_i)$ is set.

Definition 3. By system failure, we mean the accuracy indicator going beyond the permissible limits. At the same time, in practice, the concept of failure must be clarified each time when solving a specific problem: for example, in the case of a N -component mixture, failure will mean going beyond the specified measurement accuracy for k sensors from N , where k can take values from 1 to N (depending on operating conditions and others, for example, economic factors, operating conditions of the system, customer requirements, etc.). In the case $k = 1$, the most stringent requirements are imposed on the sensor subsystem, i.e. the impossibility to measure exactly at least one component is already considered a failure, in the case $k = N$ for the operability of the sensor subsystem it is sufficient to be able to measure at least one component. In this report, we present the results for these two extreme scenarios.

Mathematical model and main results

Consider a random process in continuous time $X(t)$, which can be described by a simple stochastic differential equation [9]:

$$dX(t) = \mu dt + \sigma dW(t). \quad (1)$$

This approach can also be interpreted in the following way: the error (difference) can be modeled by a one-dimensional Brownian motion $X(t) = W(t)$, or, formally, $dX(t) = dW(t)$. Generally,

$$dX(t) = \sigma(t, X(t)) dW(t),$$

where the coefficient σ in the general case may not be a constant, but depend on the state of the system. Those. we assume that there is some randomness in our measurements, so the error can both increase and decrease. In addition, we assume that the sensor is worn out over time, so the error grows on average at a certain rate μ . Those. at the intuitive level, the model for

changing the error X is described by a random walk, which, when the time step tends to zero, will tend to the process determined by equation (1). Further, without loss of generality, we will assume that σ and μ are some positive constants. Now, suppose that a certain permissible range of values $(-A, A)$, $A > 0$ is set for the sensor, the scope of which determines the normal functioning of the system [2].

Definition. The system is operational if $X(t) \in (-A, A)$.

Definition. A system failure is understood as an event – the exit of the process $X(t)$ to the boundary of the sets: $X(t) \leq -A$ or $X(t) \geq A$.

Definition. MTBF is the first moment of reaching level A :

$$\tau_A = \inf\{t \geq 0 : |X(t)| \geq A\} =: \tau.$$

Definition. The system reliability function (in the general case) is defined as

$$R(t) = P\{\tau > t\}.$$

Definition. The distribution function of the operating time of the system (in the general case):

$$F(t) = 1 - R(t) = P\{\tau \leq t\}.$$

Theorem 1. The distribution function of the operating time of the subsystem $F_1(t)$ for the scenario when the system fails when all subsystems fail (Fig. 2) is found by the following formula:

$$F_1(t) = \prod_{i=1}^5 \left(1 - \Phi\left(\frac{A_i - \mu_i t}{\sigma\sqrt{t}}\right) + \exp\left\{\frac{2\mu_i - A_i}{\sigma^2}\right\} \Phi\left(\frac{-A_i - \mu_i t}{\sigma\sqrt{t}}\right) \right). \quad (2)$$

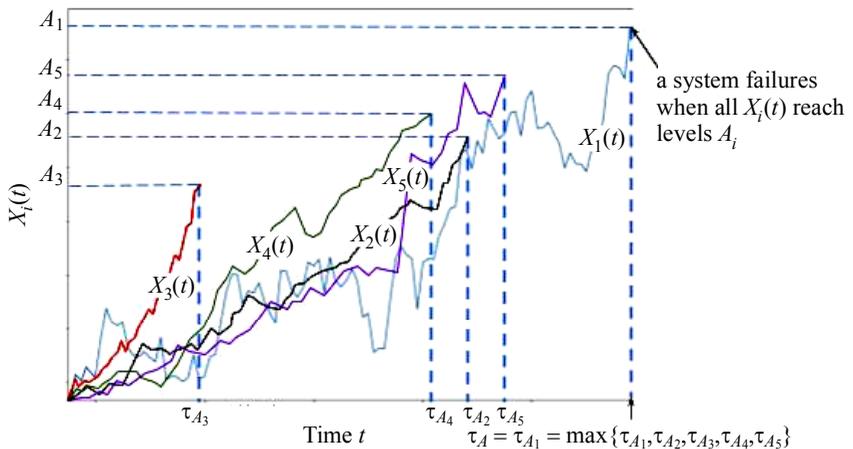


Fig. 2. The function of distributing the operating time of a subsystem for a scenario, when a system failure occurs when all subsystems fail

Theorem 2. The distribution function of the operating time of the subsystem $F_2(t)$, the reliability function and the mean time to failure for the scenario when the system fails when at least one of the subsystems fails (Fig. 3), is found by the following formulas:

$$F_2(t) = \prod_{i=1}^n \left(1 - \Phi \left(\frac{A_i - \mu_i t}{\sigma \sqrt{t}} \right) + \exp \left\{ \frac{2\mu_i - A_i}{\sigma^2} \right\} \Phi \left(\frac{-A_i - \mu_i t}{\sigma \sqrt{t}} \right) \right),$$

$$R_2(t) = 1 - F_2(t),$$

$$MTTF_2 = \int_0^{\infty} R_2(t) dt.$$
(3)

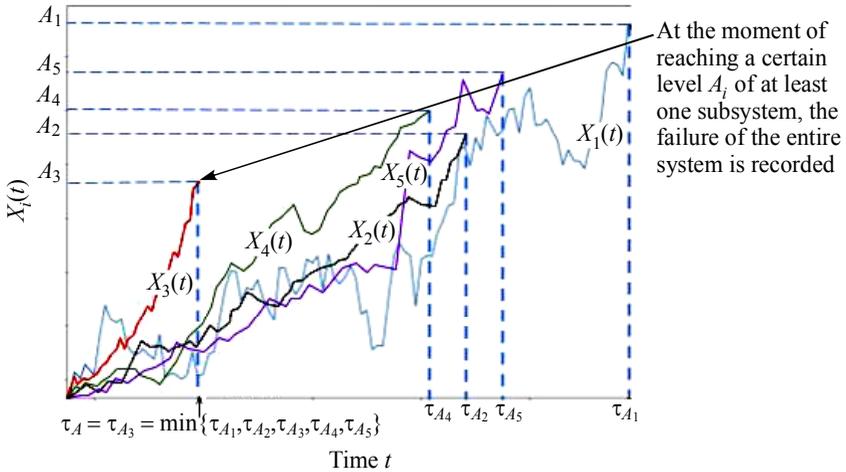


Fig. 3. The function of distributing the operating time of a subsystem for a scenario when a system failure occurs when at least one of the subsystems fails

Conclusions

The paper gives a definition of reliability for the investigated gas analysis system in the problem under consideration, and proposes a model based on the vector Wiener process to assess the reliability of the sensor subsystem. The approach developed in the report has not been previously applied for this class of target systems. Its application in the design of systems for determining the component composition of natural gas will make it possible to evaluate the reliability characteristics without carrying out full-scale and semi-natural tests, based solely on the results of modeling. Analytical

assessments of the average operating time of the sensor until the moment of its replacement allow us to determine the required number of sensors (and their replacements) during a given period of system operation, determine the required amount of spare sensors for a certain period, draw up an operational plan for servicing and replacing the system, which in turn will optimize the performance of the system.

REFERENCES

1. *Wolfgang Granig, Lisa-Marie Faller, Hubert Zangl*. Sensor system optimization to meet reliability targets // *Microelectronics Reliability*. 2018. V. 87. DOI: 10.1016/j.microrel.2018.06.005.
2. *Shubin Si, Lirong Cui, Shubin Si*. Reliability and availability analysis of stochastic degradation systems based on bivariate Wiener processes // *Applied Mathematical Modelling*. October, 2019. DOI: 10.1016/j.apm.2019.10.044.
3. *Junliang L., Yueliang Chen, Yong Zhang*. System availability modelling and optimization considering multigeneral quality characteristics // *Hindawi Mathematical Problems in Engineering*. V. 2020. P. 1–9. <https://doi.org/10.1155/2020/6409838>.
4. *Kalimulina E.Y.* Math modeling of the reliability control and monitoring system of complex network platforms // *Intelligent Systems Design and Applications. ISDA-2018. Advances in Intelligent Systems and Computing*. V. 941. Springer, 2018.
5. *Kalimulina E.Y.* Analysis of unreliable open queueing network with dynamic routing // *Distributed Computer and Communication Networks. DCCN 2017. Communications in Computer and Information Science*. V. 700. Springer. 2017.
6. *Kalimulina E.Y.* Analysis of system reliability with control, dependent failures, and arbitrary repair times // *International Journal of System Assurance Engineering*. 2017. V. 8. doi.org/10.1007/s13198-016-0520-5.
7. *Kalimulina E.Y.* A new approach for dependability planning of network systems // *International Journal of System Assurance Engineering*. 2013. V.4. P. 215. <https://doi.org/10.1007/s13198-013-0185-2>.
8. *Osipov L.V., Petrov V.V.* Ob ocenke ostatochnogo chlena v central'noj predel'noj teoreme // *Theory Probab. Appl.* 1967. V. 12. No. 2. P. 281–286.
9. *Oksendal Bernt*. *Stochastic Differential Equations. An Introduction with Applications*. Fifth Edition, Corrected Printing. Berlin; Heidelberg: Springer-Verlag, 2003.
10. *Koralov L., Sinai Y.G.* *Theory of Probability and Random Processes*. Springer Science & Business Media, Aug 10, 2007 – Mathematics. 358 p.
11. *Ibragimov I.A., Hasminsky R.Z.*, *Statistical Estimation. Asymptotic Theory, Applications of Mathematics*. New York; Heidelberg; Berlin: Springer Verlag, 1981.

Analysis of business processes of construction and operation of highways on a toll basis using BIM tools

Victoria Shamraeva

Financial University under the Government of the Russian Federation, Moscow, Russia

The article discusses the business processes of construction and operation of linearly extended sections of highways and analyzes the geographic information system (GIS) of Avtodor state company (developer «Indor-Soft», Tomsk). The main tasks assigned to the GIS of the Avtodor are listed on the site <https://avtodor-eng.ru/services/arenda-spetstekhniki/geoinformatsionnye-sistemy>. The primary task is to use the GIS of Avtodor group for the purpose of operation and maintenance of transport infrastructure facilities on linearly extended sections of highways. The object of the study is the construction of the Far Western bypass of Krasnodar project (FWBK) and its implementation by means of BIM modeling.

The main components of the digital transformation of the road industry are: mobility; safety; efficiency; use of information modeling when building models of linear-extended road sections [1]. Issues of standardization and creation of a database of information models in urban planning and road construction are currently priority areas in the Russian Federation. The international non-profit organization buildingSMART based on the NAICS Association, has been dealing with these tasks since 2017 in Russia. National BIM standards have not yet been fully developed and there is a lack of software compatibility in the road sector, as not only domestic developers say [2]. The development of a regulatory framework is essential and should contribute to the successful implementation of BIM information modeling in Russia. The use of digital building models at various stages of the life cycle of the capital construction object is regulated by the decree N 1431 of September 15, 2020 «On approval of the rules for forming and maintaining the information model of the capital construction object, the composition of information, documents and materials included in the information model of the capital construction object and presented in the form of electronic documents, and the requirements for the form of these electronic documents, as well as on making changes to item 6 regulations on performing engineering surveys for the preparation of project documentation, construction, and reconstruction of capital construction projects», signed by Mikhail Mishustin, Chairman of the government of the Russian Federation

(see <https://ipbd.ru/doc/0001202009220002/>). A number of road organizations are also working on the use of BIM tools: road information models, databases, and automated systems for processing road data banks, including on the Web.

Let's list the differences between the information models of capital construction objects and linear-extended transport infrastructure objects. In the life cycle of a road object, there are the following stages:

1. Investment and territorial planning.
2. The layout of the territories.
3. Design.
4. Construction.
5. Operation.
6. Reconstruction.

In [3], a more detailed description of the options for the road life cycle and the role of GIS, CAD, and database technologies at different stages is presented. For transport objects, design and construction is carried out in separate sections that allow for different implementation periods, i.e. they are at different stages of the life cycle at the same time (for example, construction, operation and repair can be performed on the same road simultaneously), and the project itself is usually performed by several contractors. Therefore, it is optimal to use GIS at the operational stage of a complete BIM model of a highway [4].

At the same time, it is necessary to take into account the requirements of current regulatory documents regarding import substitution and the gradual transition to the preferential use of domestic software. Avtodor state company business processes and software products have significant potential in optimizing and automating production activities at the stage of construction and operation of highways (action plan of December 18, 2017 in the direction of "Formation of research competencies and technological reserves" of the Digital economy of the Russian Federation program), Event 03.00.000.001.18.11-Creation of a software package for instrumental monitoring of the diagnostic state of roads, intended for collecting, accumulating, storing, consolidating, analyzing and interactive visual representation of data on the state of roads received from various measuring and control equipment, management bodies and the population (https://www.sbras.ru/files/news/docs/tex_zadely_d_e_plan.pdf). The approach to the formation of GIS of Avtodor state company serves as an example of interaction between the state and business and will allow developing common standards and regulations for information exchange.

BIM modeling in research and design

Transport infrastructure is an object designed for safe and comfortable movement of its users. One of the main conditions for the formation of transport infrastructure is its digitalization [5]. Obviously, with a common data environment (CDE), BIM designers will develop project documentation much faster and at a lower cost. To manage transport infrastructure objects, GIS such as ESRI's ArcGIS, MapInfo of MapInfo Corp., Indorroad of IndorSoft, freely distributed QGIS, DorGIS of Intelnova and others are used [4]. Avtodor state company develops its own approach to the use of BIM modeling and GIS technologies in the provision of construction control services.

Geo-information systems (GIS) are tools for managing the state of transport infrastructure objects at the operational stage: Building Information Modeling (BIM-modeling) and GIS are interrelated, since they are based on an information model and CDE. The need for the exchange of 2D and 3D information between computer-aided design (CAD) systems has existed for a long time, as well as in a standard that records changes in an infrastructure object throughout all stages of its life cycle [6].

The use of BIM tools in the implementation of transport infrastructure development projects will significantly save time and money when planning and implementing activities at the stage of facility operation. The main functionality and examples of using of the Avtodor state company IndorRoad GIS for data entry, accounting and analysis are disclosed in [7], to solve practical problems of road maintenance.

GIS of Avtodor state company keeps records and certification of the road network, as well as management of operation and maintenance of the life cycle of highways. Currently, this company is implementing a project for the construction of the FWBK highway, which is part of the «North-South» transport corridor. The analysis of business processes of Avtodor state company was made taking into account the use of GIS for the purposes of operation and maintenance of objects of the transport infrastructure of FWBK.

Namely, the following are considered:

- ability to use GIS functionality on the site;
- asynchronous editing, and selecting data without access to the Internet;
- an appropriate set of object data;
- updating the database.

To justify the investment in the project, a consolidated model of FWBK was created (Project model_2108018_dop_attributes.nwd) in the Autodesk

Navisworks environment, using the NWD file format, which is a data exchange format for summary models. This format is static, meaning that the data contained in the file cannot be changed and used to create 2D drawings.

The individual models that make up the BIM summary model correspond to different sections of the project and are integrated using different formats (including dwg, dwf and fbx). Not all of these formats allow you to load feature attribute data into a BIM model.

The geometric parameters of the information model elements are completely correct (see Fig.1). The level of detail of objects is very high.

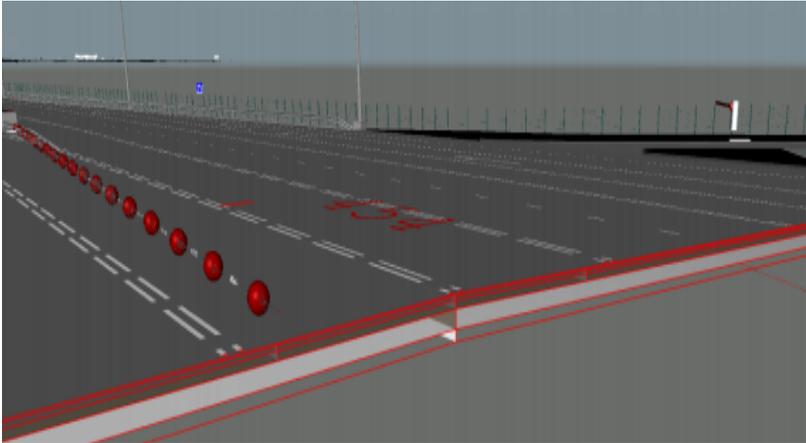


Fig. 1. Section of the BIM model on a road section

Common data environment for project implementation

CDE and the BIM model at the operational stage are described in [8]. An information model in GIS is generated for roads that are being operated using CDE during surveys, routine diagnostics, and engineering surveys. A BIM model is formed for roads under construction and under design. These processes will continue to develop in parallel for quite a long time due to the significant length of the road network. Thus, only the use of CDE will allow avoiding contradictions and ensuring the reliability of the information model data in both cases. The description of aspects of the organization of CDE is given in the UK BIM standard [9]. In [4], we present evidence – based arguments that the GIS of highways is at the same time CDE and BIM model of roads at the stage of operation.

Unfortunately, at the moment, there are certain problems associated with the lack of powerful domestic CAD systems that allow you to automate the execution of design and survey work. These systems should be the source of the main data set for building full-fledged BIM models of highways. If in the field of industrial and civil construction, the element base is significantly developed, then in the field of road construction, such developments are quite small. The designer often has to redraw and calculate the same standard structures without using the capabilities of computer-aided design.

Another problem with using BIM modeling is working out the requirements for detailing the information model at the stage of developing working documentation. On the one hand, the detail of the information model should not be excessive, i.e. it should not have elements that will not be used at the construction and operation stages. On the other hand, the details of the information model should be sufficient to perform construction and installation works, carry out construction control and acceptance of such works, as well as subsequent operation of the object.

When planning a project using BIM tools and 3D coordination with the subsequent transfer of the necessary information, the basic levels of detail of information model elements are determined, the so-called Level of Development (LOD). Each of the LOD levels contains information about the execution of project tasks that correspond to a particular level of detail. In [10] the attributes of five LOD levels of detail are presented as the project develops: LOD100, LOD200, LOD300, LOD400, and LOD500. An analysis of the business processes of construction and operation of linear sections of highways shows that LOD500 is the most optimal for making reasonable and clearly formulated customer requirements for information models. It is also optimal in the sense of minimum requirements for information saturation of model elements by the project team before starting the project. This level of model development is intended for further transmission to the maintenance service, and therefore the LOD 500 model is the most complete with all the actual dimensions, clear links, and bindings, and has the most complete data for all its elements (see Fig. 2).

As a rule, BIM models are developed for the design stage, and therefore they should not be used for the purpose of operation and maintenance of the highway. Since the model intended for use by the operating organization should reflect the actual constructed object. In the current state, BIM models can be used to get an idea of construction and installation works and structures under construction. However, they contain a large amount of 3D data and very little attribute data, which makes it impossible to use

Model levels in detail						
Investment and territorial planning		The layout of the territories		Design		Construction Operation
LOD 100	⇒	LOD 200	⇒	LOD 300	⇒	LOD 400 LOD 500
Attributes: <ul style="list-style-type: none"> • Length • Width • Height • Volume • Area <p>.....</p> Result: territory diagrams with models of existing and planned highways (2-D model)		Attributes: <ul style="list-style-type: none"> • Length • Width • Height • Volume • Area • Shape • Orientation <p>.....</p> Result: an area plan with the planned highway in the form of a highway model (number of lanes, configuration of junctions and intersections, etc.)		Attributes: <ul style="list-style-type: none"> • Length • Width • Height • Volume • Area • Profile • Material • Weight • Sectional • Manufacturer <p>.....</p> Result: engineering model of the highway route, structural lines, road surface and roadbed, road markings, artificial structures, construction site development plan, etc.		Attributes: <ul style="list-style-type: none"> • Length • Width • Height • Volume • Area • Profile • Reference values for profile axes <p>.....</p> Result: <ol style="list-style-type: none"> 1) a Production model with a detailed work schedule, logistics for the supply of products and materials, etc. (including 3D models for automated control systems for road construction machines) 2) Executive model (models of laser scanning of the highway, etc.) 3) Operational model (formed on the basis of the Executive model)

Fig. 2. LOD levels in detail as the project is developed

the project BIM model in its current state for operation and maintenance tasks. In the end, the BIM model should be filled with information at the entire stage of the project life cycle and the need for its re-development in certain sections of the object due to insufficient detail should be minimized.

Based on the initially designed BIM model, a CIM model for optimizing construction and installation works should be created through the integration

of the financing and execution schedule, and, subsequently, an operational AIM model supplemented with data on the design characteristics of the performance and wear of structural elements and installed equipment. These works can also be done using GIS technologies.

Applying these BIM tools in the road industry will not only expand them, but also form a common approach to the structure and industry library of elements of the digital road model. This means that the road GIS reaches the highest level of maturity (in the British classification) – iBIM [4] and demonstrates the capabilities of BIM modeling in practical application in road infrastructure.

The use of the web-based general data environment (geoportals) GIS allows online use of a digital terrain model with sub-metric accuracy for unmanned vehicles. The development of unmanned vehicles and "convoys" technology for transport trains requires interaction between the road infrastructure, on-Board computers and car computer vision: Avtodor state company prepares GIS and CDE interfaces for interaction with such vehicles based on V2I (Vehicle to Infrastructure-vehicle to infrastructure) [7].

Domestic experts develop competitive software solutions that bring road activities to the new environment of economic development. In 2020, specialists of JSC "Institute Stroyproekt" implemented a BIM model of FWBK in the C-INFO environment (<https://sinfo.tech/ru/portfolio/dalnij-zapadnyj-obhod-krasnodara>). Experts determine whether this model is suitable for breaking down construction and installation work into stages and getting a general idea of the implementation of construction and installation work. They also specify to what extent it can be used for the operation and maintenance of infrastructure facilities. In the current state, the BIM model can be used to get an idea of construction and installation works and structures under construction, update data on the volume of earthworks and control geodetic works.

A digital model of the site with an orthophoto would be an additional advantage. However, using this model without developing C-INFO to divide construction and installation work into stages is more difficult, even impossible. A breakdown of the work into stages usually involves the separation of the original objects. It is not possible to perform a repeated breakdown of the selected pavement structure (see Fig. 3).

If you plan to build a roadbed in two stages, you will not be able to display only part of the selected element. Using the model to break down construction and installation work into stages requires access to the source models of each element. Temporary structures were not modeled, which

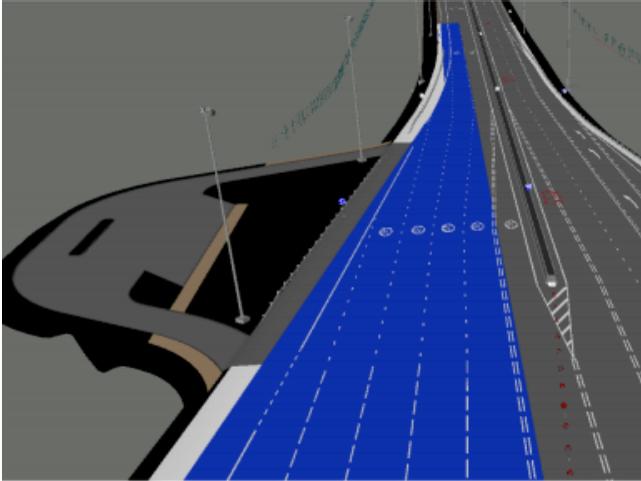


Fig. 3. View of the model. The roadway area is highlighted in blue

made it impossible to control their influence on the boundaries of other structures. A large amount of data must be loaded into the data model associated with objects, mainly design and construction data, and earth structures (depressions and embankments) must be modeled.

Let's take a closer look at the functional part of the GIS of Avtodor state company and its use on the object under consideration. The saturation of the GIS of Avtodor state company can be judged by the number of objects entered in the database and the quality of data entry (indivisible polygons, at first glance). It contains a large number of photos of infrastructure objects with a 360° view (see Fig. 4).

The navigation interface in photos with a 360° view is very fast, which allows you to significantly improve your work efficiency. For more effective management of road facilities, CDE is needed. This resource is designed to provide reliable and timely information [6], namely:

- Consolidation of data on the current situation on the roads on a single resource;
- Storage of databases on the state of roads and mobile objects for the preparation of analytical reports and forecasts;
- Prompt provision of up-to-date information about road conditions to a wide range of consumers;
- Receive feedback from road users about an emergency situation, the state of the roadway, and other important information.

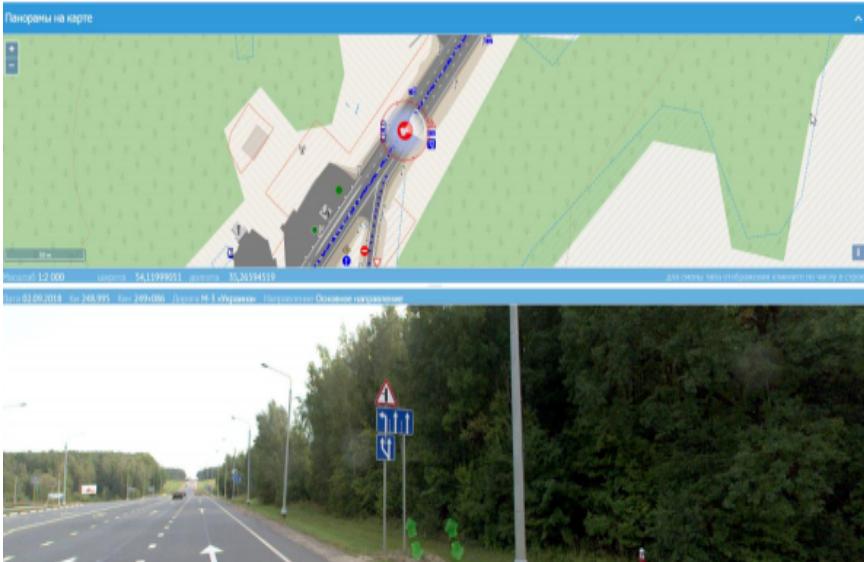


Fig. 4. Screenshot of the screen with the navigation interface in photographs with a 360° overview

The object information access portal has an adaptive web interface that is optimized for display on your computer and mobile devices. Elements contained in the GIS database are requested by selecting a multi-level menu or selecting objects on the map and going to the section of the desired object. There are also a number of disadvantages:

1. Since it is difficult to select elements on the map using the cursor when navigating on mobile devices, it is necessary to provide an interface that is more adapted to mobile devices and configured to meet the needs of the specialist working on the site.

2. It is not possible to enter information online (perhaps this is due to access rights granted by governmental to audit). Obviously, if data is to be uploaded on the site, it is important to ensure that the person who creates the data can enter it.

3. 3G/4G network Coverage is not available in all areas where specialists work on the site. Having only the web version restricts the use of the system. Therefore, you should consider a solution for extracting data for a specific work area. In addition, if data must be entered on an object in a zone without Internet access, asynchronous database operation will be required.

4. Attribute Tables contain a fairly large set of data (more than 250 unique properties of database objects, in addition to geometric characteristics and symbols). However, there is a lack of data for using GIS to obtain information about current and future construction projects or as a basis for the operation and maintenance of infrastructure facilities. First, there is no information about the validity of the data: the data source, the corresponding accuracy, and so on. For example, it is not possible to know whether the position of a pipe crossing a road was determined during design or by surveying. Secondly, the number of documents on artificial structures is quite limited and often only photos are available.

5. Finally, no data on technical surveys of structures is often available. If there is data, it is contained in a PDF file corresponding to a specific structure, and not presented as object metadata.

Thus, it is not possible to use the database to determine the number of objects that need to be maintained and maintained in X years. It is recommended to change the data management for accounting for maintenance and maintenance of structures.

Updating the database must be performed using a special program. IndorRoad currently implements the iFC data exchange format for integrating data from BIM models. This function is necessary in the future for both new construction projects and infrastructure reconstruction projects. To update plans and network schedules, the ETL (Extract Transform Load) program can be used when connecting directly to the CDE without using IndorRoad. This issue should be addressed in order to limit human intervention in data entry.

Thus, BIM model of the FWBK is a project that does not implement all BIM processes. the creation of a data center follows parallel paths:

- a GIS («operational BIM») is formed for the transport infrastructure in operation, which reflects the actual state of the object and does not require significant costs for its updating and filling;
- for all road construction projects starting from 2021, a BIM model is formed that covers the entire life cycle of the road.

However, currently there are domestic software solutions only for the design stage, since the construction stage requires a compulsory decision of the state customer to apply BIM and CDE by all participants in the process (the designer, contractor and construction control) with the choice of specific software solutions.

Upon completion of the construction phase of the FWBK, it is planned to conduct acceptance diagnostics using unmanned aerial vehicles (UAVs) and

mobile laboratories of the 5th generation equipped with lidars and mobile laser scanning systems (LOD400). The integrated laboratory allows you to create 3D models of roads in a cross-section. Data is obtained using the scanner at speeds up to 100 km/h from a seven-meter-wide road surface, as well as from a depth of seven meters below the surface level. The collected data is automatically processed using special software and uploaded to the CDE. From there, interested organizations and individuals will be able to draw information for the design, construction, maintenance and repair of the Russian transport network. Not only engineers and specially trained operators can join the process of recording information, but also motorists. Any violations on the roads – holes, bumps, incorrect signs-can be photographed and sent to the General database via a special app on your smartphone. The project involves the accumulation of huge amounts of information-hundreds of terabytes.

The considered BIM model of the FWBK was developed for the design and development stages of working documentation. For the purposes of operation and maintenance of the highway, it will require significant changes, since the information model intended for use by the operating organization describes the actual built object. Updating of original models is required. It is necessary to provide for new modeling in areas where significant changes were made to the project during the construction and installation stage. Such developments are among the priority areas for the development of the information society, as evidenced by presidential decree N 203 of may 9, 2017 «On the information society development strategy in the Russian Federation for 2017–2030» and support national sovereignty over information flows in the transport sector of the Russian Federation.

Conclusions

At this stage of road digitalization in our country, the practical application of BIM modeling for road infrastructure objects plays an important role. The tools and software products that are used and developed for road diagnostics make up the information base that can be integrated with intelligent transport systems. The use of BIM tools in the implementation of transport infrastructure development projects will significantly save time and money when planning and implementing activities at the stage of facility operation.

GIS of Avtodor state company is an example of the experience of interaction between the state and business, which will allow you to establish not only legally significant data on transport infrastructure and vehicles, but

also help to develop common standards and regulations for information exchange. This is another advantage of using BIM modeling in the design of transport infrastructure. When implementing projects for the development of transport infrastructure, the main tasks set for the road area are solved – optimization of time and money when planning and implementing measures at the stage of operation of infrastructure facilities, as well as a high level of maintenance of highways and the safety of the existing road network.

REFERENCES

1. *Shamraeva V.V.* Predicting the effectiveness of implementation of financial models for toll collection on the road // *Transport business of Russia*. 2020. No. 1. P. 37–41.
2. *Sandeep L.* Challenges with BIM implementation: A review of literature // *53rd ASC Annual International Conference Proceedings of the Associated Schools of Construction*. 2017. URL: https://www.researchgate.net/publication/317842173_Challenges_with_BIM_Implementation_A_Review_of_Literature.
3. *Skvorzov A.V., Sarychev D.S.* Road projects life cycle within the context of information modeling // *CAD & GIS for Roads*. 2015. No. 1(4). DOI: 10.17273/CADGIS.2015.1.1.
4. *Shamraeva V.V., Kuzovlev E.G., Barannik S.V.* Realizatsiya geoinformatsionnykh sistem v dorozhnoy oblasti kak odnogo iz napravleniy informatsionnogo modelirovaniya [Implementation of geographic information systems in the road sector as one of the directions of information modeling] // *Vestnik komp'yuternykh i informatsionnykh tekhnologii*. 2018. No. 6(168). P. 20–26. DOI: 10.14489/vkit.2018.06. P. 020–026 (in Russian).
5. *Evstigneev I.A.* Road infrastructure and highly automated vehicles // *CAD & GIS for Roads*. 2019. No. 2(13). P. 44–50. DOI: 10.17273/CADGIS.2019.2.7.
6. *Znobishchev S., Shamraeva V.* Prakticheskoye primeneniye BIM modelirovaniya dlya ob'yektov dorozhnoy infrastruktury [Practical use of BIM modeling for road infrastructure facilities] // *Architecture and Engineering*, 2019. V. 4. No. 3. P. 49–54.
7. *Dmitrienko V.E., Kuzovlev E.G., Shamraev L.G.* GIS application of the avtodor state company for solving practical tasks of road maintenance // *CAD & GIS for Roads*. 2019. No. 2(13). P. 52–59.
8. *Draft standard for information management in the phase of capital construction using information modeling*. Standard No. PAS 1192-2:2013. URL: <http://bimstandart.tilda.ws/pas1192-2>.
9. *British standard. Joint production of architectural, engineering and design information – norms and rules*. Standard No. BS 1192:2007. URL: http://browniesoft.com/uploads/files/BS_1192_2007_new_ru_demo.pdf.
10. *E203–2013 Building Information Modeling and Digital Data Exhibit*. The American Institute of Architects, 2013. 7 p

Применение метода двудольных множеств событий в задачах регрессионного анализа многомерных разнотипных данных

Ирина Баранова

Сибирский федеральный университет, г. Красноярск, Россия

Исследуется зависимость между многомерными случайными величинами, представляющими собой векторы признаков, которые могут принимать числовые, качественные или множественные значения. Задача выявления зависимости между подобными данными является очень важной и актуальной. Во многих областях современной научной и практической деятельности при проведении различных исследований довольно часто возникает необходимость анализа многомерных разнотипных данных, часть которых принимает числовые значения, а часть – нечисловые (качественные или множественные значения). К числу таких областей деятельности можно отнести здравоохранение, экономику, финансы, социологию, биологию, лингвистику, педагогику и другие.

Для выявления характера связи между рассматриваемыми величинами будет использоваться регрессия. Как известно, регрессия является одним из основных понятий теории вероятностей и математической статистики и выражает зависимость среднего значения случайной величины от значения другой случайной величины или от нескольких случайных величин.

В работах [1, 2] автором был предложен метод, позволяющий успешно решать задачи анализа разнотипных данных. Данный метод базируется на понятии двудольного множества событий. Под двудольным множеством событий понимается объединение двух множеств событий, первая доля которого определяется случайными величинами, а вторая – случайными множествами. Для исследования зависимости между разнотипными данными, классическая множественная регрессия не подходит, поскольку речь идет о выявлении связи не только между случайными величинами, но и между множествами, которые могут состоять из любого количества случайных величин и случайных множеств. Поэтому

му в данной работе предлагается новое понятие статистической регрессии между двудольными множествами событий и демонстрируется вид функции регрессии.

Двудольное множество случайных событий

Пусть задано вероятностное пространство (Ω, F, P) , где F – алгебра событий, события $x, y \in F$ – элементы алгебры событий, на которой определена вероятность P . Множество избранных событий $X \in F$.

Случайное множество событий под X определяется как измеримое отображение

$$K : (\Omega, F, P) \rightarrow (2^X, 2^{2^X}), \quad (1)$$

где 2^X – множество всех подмножеств множества X [1].

Множество случайных элементов $\{\xi, K\}$, представимое в виде объединения двух множеств: множества случайных величин ξ и множества K случайных событий, будем называть *двудольным множеством случайных элементов*:

$$\{\xi, K\} = \xi \cup K = \{\xi_a, a \in A, K_\beta, \beta \in B\}. \quad (2)$$

Здесь A – множество индексов случайных величин, B – множество индексов случайных множеств событий, $A+B = m$.

Двудольное множество случайных событий представляет собой объединение двух множеств – множества событий, которое определяется случайными величинами, и множества событий, которое определяется случайными множествами событий:

$$Z = \{Y, X\} = \{Y_a, X_\beta, a \in A, \beta \in B\}. \quad (3)$$

Пусть задано двудольное множество случайных событий s , являющееся подмножеством двудольного множества событий Z :

$$s = \{Y_{k_A}, X_{k_B}, k_A \subseteq A, k_B \subseteq B\} \subseteq Z, \quad (4)$$

где каждое множество событий является подмножеством соответствующего множества избранных событий $Y_{k_A} \subseteq Y_A$ или $X_{k_B} \subseteq X_B$.

В работе [1] был получен вид распределения вероятностей событий двудольного множества s .

**Условное двудольное множество событий
и условные вероятности двудольного множества событий**

Пусть задано двудольное множество событий $s \subseteq Z$. Событие-терраска двудольного множества событий s представляет собой набор непересекающихся событий, причем каждое его событие является подмножеством соответствующего множества событий Y_a or X_β :

$$\begin{aligned} \text{ter}(s) &= \{Y_{s_A}, X_{s_B}\} = \bigcap_{a \in s_A} \text{ter}(Y_a) \bigcap_{\beta \in s_B} \text{ter}(X_\beta) = \\ &= \bigcap_{a \in s_A} Y_a(r_a) \bigcap_{\beta \in s_B} \left(\bigcap_{x_\beta \in X_\beta} x_\beta \bigcap_{x_\beta \in X_\beta^c} x_\beta^c \right), \end{aligned} \quad (5)$$

где $s_A \subseteq A$, $s_B \subseteq B$, $r_a \in \mathfrak{R}_a$, $X \in X_\beta$. Здесь множество \mathfrak{R}_a – конечное множество возможных значений случайной величины ξ_a :

$$\mathfrak{R}_a = \{r_{a_1}, \dots, r_{a_{N_a}}\} \subset \mathbb{R}.$$

Условное двудольное множество событий $s | x$ представляет собой событие, означающее, что наступление событий из двудольного множества s происходит при условии, что выполнилось событие x (событие $x \in Y_a$ или $x \in X_\beta$).

Условная вероятность наступления двудольного множества событий (вероятность наступления условного двудольного множества событий $s | x$) может быть задана следующей формулой:

$$P(\text{ter}(s) | x) = p_{|x}(\text{ter}(s)) = \frac{P(\text{ter}(s) \cap x)}{P(x)}, \quad (6)$$

где $\text{ter}(s)$ находится по вышеприведенной формуле (5).

Пусть s, t – двудольные множества событий, являющиеся подмножествами двудольного множества событий Z (т.е. $s, t \subseteq Z$):

$$s = \{Y_{s_A}, X_{s_B}, s_A \subseteq A, s_B \subseteq B\}, \quad t = \{Y_{t_A}, X_{t_B}, t_A \subseteq A, t_B \subseteq B\}.$$

Условное двудольное множество событий $t | s$ представляет собой событие, означающее наступление событий из двудольного множества t при условии наступления всех событий из фиксированного двудольного множества s .

Условную вероятность наступления двудольного множества событий (вероятность наступления условного двудольного множества событий $t | s$) можно определить следующим образом:

$$P(ter(t) | ter(s)) = p_{ter(s)}(ter(t)) = \frac{P(ter(t) \cap ter(s))}{p(ter(s))}, \quad (7)$$

где

$$p(ter(s)) = \sum_{ter(s) \subseteq Z \subseteq Z} P(Z),$$

а $ter(t)$ и $ter(s)$ находятся по формуле (5).

Регрессия между двудольными множествами событий

В работе [3] О.Ю. Воробьев предложил формулу эвентологической регрессии одного множества событий на другое множество. Данная регрессия между множеством событий X и множеством событий Y , связывающая события-терраски первого множества с событиями-террасками второго, выглядит следующим образом:

$$ter(Y) = \varphi(ter(X)). \quad (8)$$

Обобщим данное определение на более сложный случай регрессии двудольных множеств событий.

Регрессией двудольного множества событий $s \subseteq Z$ на другое двудольное множество событий $t \subseteq Z$, связывающей события-терраски $ter(s)$, $s \subseteq Z$ с событиями-террасками $ter(t)$, $t \subseteq Z$, называется статистическая функция следующего вида:

$$ter(t) = \varphi(ter(s)), s \subseteq Z. \quad (9)$$

Поставим задачу статистической оценки функции теоретической регрессии на уровне α , значение которого определяется как условный α -квантиль двудольного множества t .

Под условным α -квантилем двудольного множества случайных событий t при наступлении событий из двудольного множества s будет пониматься такое множество событий из двудольного множества t (то есть события-терраски $ter(t)$), условные вероятности которых больше заданного значения α :

$$Q_\alpha(t | ter(s)) = \{ter(t) \subseteq Z : P(ter(t) | ter(s)) \geq \alpha\}. \quad (10)$$

Таким образом, функция теоретической регрессии между двудольным множеством событий $s \subseteq Z$ и двудольным множеством событий

$t \subseteq Z$ представляет собой отображение событий множества s на множество t . Чтобы определить, какие элементы $ter(t)$ принадлежат области значений функции регрессии, необходимо найти события, входящие в условный α -квантиль двудольного множества случайных событий по вышеприведенной формуле (10). Условные вероятности $P(ter(t)|ter(s))$ из определения условного α -квантиля двудольного множества случайных событий будут вычисляться по формуле (7).

Для оценивания значений функции теоретической регрессии двудольных множеств событий необходимо иметь статистику для оценки вероятностей пересечений событий (из числителя формулы (7)) и вероятностей распределения $p(ter(s))$ (из знаменателя формулы (7)). После сравнения найденных оценок условных вероятностей с порядком условного α -квантиля можно получить оценки множеств-значений функции эвентологической регрессии для каждого события $ter(t)$, $t \subseteq Z$.

На рис. 1 показан график функции теоретической регрессии двудольных множеств событий для элементарного случая, когда двудольный набор событий состоит только из одной множественной части. На графике кружками обозначены события, попавшие в график функции теоретической регрессии.

Предложенный инструмент может быть эффективно применен для решения широкого круга практически важных задач анализа данных, в том числе в кластерном анализе, классификации, ковариационном и факторном анализе.

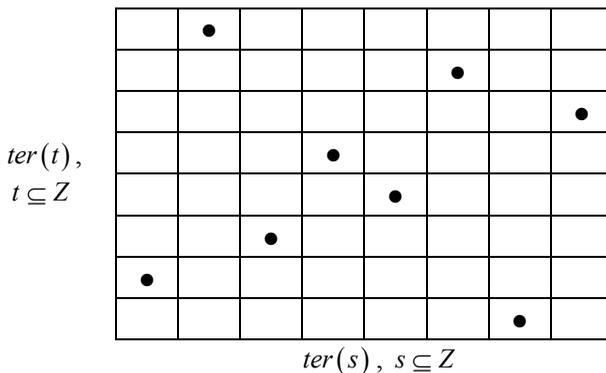


Рис. 1. График функции теоретической регрессии

Заключение

Предлагается новое понятие статистической регрессии между двудольными множествами событий и демонстрируется вид функции регрессии. Также рассмотрены понятия условной вероятности двудольного множества событий и события-терраски двудольного множества событий.

ЛИТЕРАТУРА

1. *Баранова И.В.* Метод двудольных множеств событий в эвентологическом анализе социально-экономических систем // Вестник Красноярского государственного университета: физико-математические науки. 2006. Вып. 1. С. 142–152.
2. *Baranova I.* Eventological approach to factor analysis of complex systems // Proceedings of the IASTED International Conference on Automation, Control, and Information Technology (ACIT-CDA 2010). Novosibirsk, Russia, 2010. P. 341–344.
3. *Воробьев О.Ю.* Эвентология. Красноярск: СФУ, 2007. 512 с.
4. *Baranova I.* Sets of bipartite sets of events and their application // Communications in Computer and Information Science. Switzerland, Springer International Publishing, 2014. V. 487. P. 26–33.
5. *Баранова И.В.* Применение метода двудольных множеств событий для кластеризации разнотипных данных // Информационные технологии и математическое моделирование (ИТММ-2018): материалы XVII Международной конференции имени А.Ф. Терпугова. Томск: Изд-во НТЛ, 2018. С. 281–286.

Алгоритм репрезентации кастомизированных диалектологических корпусов для Саратовского диалектологического корпуса русского языка

Инна Батраева, Александра Крючкова

Саратовский национальный исследовательский университет имени Н.Г. Чернышевского, г. Саратов, Россия

В настоящее время весьма ограничено представлены мультимедийные лингвокультурологические корпуса русской диалектной речи, полно репрезентирующие частные диалектные системы и основные типы русской диалектной речи. Вследствие интенсивного развития технологий и методологий корпусных лингвистических исследований создание электронных исследовательских корпусов диалектной коммуникации, предоставляющих репрезентативный материал для многостороннего изучения специфики диалектных подсистем, является актуальной задачей [1, 2].

Саратовский диалектологический корпус (СарДК) разрабатывается как электронно-исследовательский диалектологический корпус русского языка сотрудниками кафедры теории, истории языка и прикладной лингвистики и кафедры технологий программирования СГУ имени Н.Г. Чернышевского. Корпус-менеджер предоставляет пользователям инструменты для ведения функциональных корпусных исследований диалектной речи, многокритериальной обработки и разностороннего исследования говоров, а также обеспечивает моделирование мультимедийных электронных диалектологических подкорпусов, сопровождаемых аудиовизуальными материалами.

Корпус-менеджер СарДК разработан в виде веб-приложения с использованием возможностей фреймворка ASP.NET MVC 5. Данная платформа позволяет создавать клиент-серверное приложение на основе шаблона «Модель – Представление – Контроллер» (MVC). Главным преимуществом использования шаблона MVC является обеспечение уровня абстракции, при котором слои бизнес-логики и доступ к данным не зависят от представления, что позволяет переиспользовать их в ходе разработки. Для реализации представлений был выбран движок Razor 4.0.0, так как данный инструмент позволяет генерировать HTML-элементы посредством HTML-хелперов, а также использовать управляющие конструкции на языке C# для вывода данных и элементов на страницу, что существенно ускоряет разработку приложения. В качест-

ве фреймворка для работы с данными в приложении использовался Entity Framework. Преимущество данного фреймворка заключается в возможности использования его как объектно-реляционного модуля сопоставления, позволяющего работать с базой данных с помощью объектов .NET и устранившего необходимость в написании кода, требуемого для доступа к данным. В качестве СУБД было решено использовать MS SQL Server 2014. Данная система обеспечивает интеграцию БД в проект ASP.NET MVC 5.0 с использованием ADO.NET. ADO.NET позволяет программировать на основе концептуальных моделей EDM.

Уникальной функциональностью корпуса, отличающего его от существующих корпусов диалектной речи, является возможность создания подкорпусов для авторизованных пользователей. Данный модуль является важнейшей базой для проведения лингвистических исследований, поскольку исследователи могут формировать целевые подкорпуса по следующим критериям:

- 1) поговору;
- 2) по характеристикам интервьюируемых (пол, уровень образования, год рождения, год записи);
- 3) по теме;
- 4) по жанру.

Также пользователь может задать определенный идиолект, то есть подколлекцию, представляющую совокупность семантических и стилистических особенностей речи определенного информанта.

Репрезентация пользовательского подкорпуса обеспечивается архитектурой модели базы данных, учитывающей как лексико-грамматические характеристики текстов, так и метаданные об информантах [3]. Благодаря хранению зависимостей интервьюируемого и лингвокультурными материалами, исследователи имеют возможность формировать целевые подкорпуса.

Алгоритм организации подколлекции состоит из нескольких этапов. На первом этапе из всех материалов корпуса выбираются документы, соответствующие выбранному говору. Если данный критерий не указан пользователем, то во внимание берутся все материалы корпуса. После из получившегося множества выбираются только те тексты, которые принадлежат искомым информантам, то есть соответствуют набору метапризнаков: пол, год рождения, год записи, образование. Указание метаданных об информантах не является обязательным, поэтому в случае пропуска данных критериев рассматриваемая коллекция остается неизменной. Далее сформированная коллекция фильтруется в соответствии с выбранными темой и/или жанром. Стоит отметить, что жанровая раз-

новидность и тематическая принадлежность не требуют обязательного заполнения. Если пользователь пропустил данные поля, то сформированная текстовая коллекция на предыдущих шагах остается неизменной. Необходимо подчеркнуть, что только результирующее множество материалов участвует в лексико-грамматическом поиске, поиске по расстоянию, маске, точным формам, работе со словарем, документами и аудиоматериалами.

Выбор идиолекта, по совокупности семантических и стилистических особенностей речи и текстов отдельного носителя данного языка, определенного информанта осуществляется следующим образом. Сначала необходимо выбрать говор – данный шаг является обязательным для выполнения. Затем пользователю будет доступен выбор информанта из списка. Если этот шаг будет пропущен, система выведет предупреждающее сообщение. В процессе составления идиолекта пользователь может просмотреть сведения о диалектоносителе. Полученные сведения об информанте будут доступны только для чтения. После этого можно задать тему и/или жанр текстов.

Данный алгоритм был разработан с использованием возможностей LINQ и делегатов Predicate<T>. Последние позволяют обеспечить бесперебойную обработку как заполненных полей формы, так и пропущенных.

В корпусе имеются ограничения при работе с подколлекциями. Исследователь может задать только один подкорпус и удалить его. После создания пользовательского корпуса все поисковые механизмы, словарь и документы с аудиоматериалами работают только в его рамках. Сформированный подкорпус закрепляется за пользователем только в течение одной сессии. Если пользователь вышел из системы, то созданный им подкорпус автоматически удалится.

Предложенные метод построения кастомизированных подкорпусов и его реализация может быть легко модифицированы в случае изменения множества критериев составления.

ЛИТЕРАТУРА

1. *McEnery T., Hardie A.* Corpus linguistics: method, theory and practice. Cambridge: Cambridge University Press, 2012.
2. *Батраева И.А., Крючкова А.А.* Алгоритмы тематического и лексико-грамматического поиска для Саратовского диалектологического корпуса русского языка // Информационные технологии и математическое моделирование (ИТММ-2019): материалы Междунар. науч. конф., Саратов. Томск: Изд-во НТЛ, 2019. С. 9–14.
3. *Крючкова О.Ю., Гольдин В.Е., Батраева И.А.* Принципы организации электронной базы диалектных текстов // Русский язык: исторические судьбы и современность: труды III Междунар. конгресса. М.: МАКС Пресс, 2007. С. 620–621.

Прогнозирование частных показателей индекса активного долголетия*

Светлана Гагарина¹, Юрий Гагарин²

¹Калужский государственный университет им. К.Э. Циолковского, г. Калуга, Россия,

²Калужский филиал МГТУ им. Н.Э. Баумана, г. Калуга, Россия

Проблема старения населения является глобальным демографическим вызовом современного этапа развития и характеризуется повышением медианного возраста и удельного веса лиц старших возрастных групп в общей численности населения, не только развитых, но и развивающихся стран [2]. В настоящее время в 34 странах мира ожидаемая продолжительность жизни составляет 80 лет и более [5].

Для России характерны мировые демографические тенденции. По данным официальной статистики на 1 января 2020 года каждый пятый россиянин находится в возрасте 60 лет и более (32806 тыс. чел. или 22.4% жителей страны). Численность данной возрастной группы за пятилетний период увеличилась на 3012 тыс. чел. или на 2.1%. Ожидаемая продолжительность жизни при рождении в 2019 г. составила 73.34 года [4]. В этой связи важным приоритетом для органов государственной власти является разработка и реализация политики активного долголетия в Российской Федерации.

Для анализа и мониторинга социально-экономических явлений и процессов широко используется индексный метод, который обеспечивает комплексный подход [2]. В рамках Евросоюза с 2012 года рассчитывается индекс активного долголетия. С целью проведения межстрановых сопоставлений, с учетом положений руководства по измерению Индекса активного долголетия, изданного под эгидой Европейской экономической комиссии ООН, в 2019 году Федеральной службой государственной статистики разработана методика расчета данного индекса в России на основе данных официальной статистической информации [1].

Анализ текущей ситуации по реализации политики активного долголетия является основой для определения перспектив развития, поскольку, как известно, управление на всех уровнях требует научного предвидения на основе современных технологий, в частности прогнозирования.

* Исследования проведены при финансовой поддержке Российского фонда фундаментальных исследований (проект № 18-010-00672А).

В прогнозировании экономических показателей широко используются методы математической статистики, например регрессионный анализ. Использование метода наименьших квадратов (МНК), при оценивании параметров функциональных зависимостей, предполагает, что аргументы являются детерминированными значениями. При нахождении оценок параметров функций имеется ряд подходов, в которых учитывается случайный характер аргументов функциональных зависимостей. При стохастических исходных данных оценивание параметров экономических моделей предлагается осуществлять методами конфлюэнтного анализа [3].

Математическая модель

Рассмотрим задачу нахождения прогнозных значений модели следующего вида: $y = a + bx$. Прогнозные значения будем определять двумя способами: в первом оценивание параметров a и b осуществляется традиционным методом – МНК, во втором – воспользуемся методами конфлюэнтного анализа, т.е. будем учитывать погрешности x, y . При таком подходе модель оценивания параметров линейной функции имеет вид

$$\begin{cases} y_i = a + b\xi_i + \varepsilon_i \\ x_i = \xi_i + \delta_i \end{cases}, \quad i = \overline{1, n},$$

где ε и δ_i – ошибки значений y_i и x_i ; ξ_i – точные (неизвестные) значения x_i .

При оценивании параметров методами конфлюэнтного анализа введем предположение, что ошибки измерений ε и δ_i являются нормально распределенными случайными величинами с нулевыми средними значениями, с дисперсиями $\sigma^2(y_i)$ и $\sigma^2(x_i)$ и коэффициентом корреляции $\rho_i = 0$.

Минимизируемый функционал для нахождения оценок параметров линейной модели имеет вид

$$F = \frac{1}{2} \sum_{i=1}^n \frac{(y_i - a - bx_i)^2}{\sigma^2(y_i) + b^2 \sigma^2(x_i)}.$$

Неизвестные значения ξ_i определяются формулой

$$\xi_i = \frac{\sigma^2(y_i)x_i + b\sigma^2(x_i)(y_i - a)}{\sigma^2(y_i) + b^2\sigma^2(x_i)}.$$

В качестве исходных данных используем официальную статистическую информацию о динамике уровня занятости населения в возрасте от 60 до 64 лет за период 2013–2018 гг. [6]. В табл. 1 приведены исходные и расчетные данные. Столбец 1 соответствует порядковому номеру года исследуемого периода (переменная x), столбец 2 – уровню занятости населения (переменная y), столбец 3 – значения y , полученные МНК, столбец 4 – значения y , полученные методами конфлюэнтного анализа.

Таблица 1

Исходные и расчетные данные

Порядковый номер года (x)	Уровень занятости населения, % (y)	y – МНК	y – конф. анализ
1	2	3	4
1	3.5	3.56	3.52
2	3.7	3.69	3.67
3	3.9	3.82	3.82
4	4.0	3.95	3.97
5	4.0	4.08	4.12
6	4.2	4.2	4.27

Оценки параметров, полученные МНК, дают возможность получить следующее уравнение:

$$y = 3,43 + 0.13x .$$

Значения y , вычисленные методами конфлюэнтного анализа, приведены в табл. 1 (столбец 4). Уравнение прямой имеет вид

$$y = 3,37 + 0.15x .$$

В табл. 2 приведены прогнозные значения уровня занятости населения на последующие два года.

Таблица 2

Прогнозные значения уровня занятости населения

Порядковый номер года (x)	y – МНК	y – конф. анализ
7	4.33	4.43
8	4.46	4.58

Заключение

В данном исследовании с целью выявления преимуществ учета случайного характера исходной информации при прогнозировании одного из частных показателей индекса активного долголетия на примере простейшей линейной модели предложен способ оценивания параметров, который базируется на методах конфлюэнтного анализа и позволяет учитывать неопределенность значений функций и аргументов при получении несмещенных оценок параметров функциональных зависимостей.

Не вызывает трудностей использование предложенного метода при решении задач, связанных с прогнозированием социально-экономических показателей, а именно частных показателей индекса активного долголетия, а учет неопределенности значений функций и аргументов способствует повышению качества прогноза.

В дальнейшем предполагается построение многофакторной модели по всем индикаторам, входящим в интегральный индекс активного долголетия.

ЛИТЕРАТУРА

1. Приказ Росстата от 31.10.2019 N 634 «Об утверждении методики расчета Индекса активного долголетия в Российской Федерации». URL: http://www.consultant.ru/document/cons_doc_LAW_337013/
2. Бурцева Т.А., Гагарина С.Н., Чаусов Н.Ю. Оценка качества жизни населения старших возрастов при обосновании стратегий активного долголетия в условиях структурных демографических изменений // Вестник университета. 2019. № 2. С. 5–12.
3. Гагарин Ю.Е., Гагарина С.Н. Прогнозирование показателей деятельности предприятий с учетом неопределенности исходных данных // Вестник университета. 2019. № 1. С.94–99.
4. Официальный сайт Федеральной службы государственной статистики. URL: <https://rosstat.gov.ru/folder/12781>
5. Рейтинг стран мира по уровню продолжительности жизни. URL: <https://gtmarket.ru/ratings/life-expectancy-index>
6. Социальное положение и уровень жизни населения России 2019 г. URL: https://rosstat.gov.ru/bgd/regl/b19_44/Main.htm

Решение задачи распознавания образов при помощи алгоритма гибридной СММ-нейросети

Степан Гилян

Сибирский федеральный университет, г. Красноярск, Россия.

Решается частный случай задачи распознавания образов – задача распознавания слов в аудиозаписях. Разработка систем распознавания речи является одним из самых актуальных и востребованных направлений в области искусственного интеллекта, машинного обучения и анализа данных. На данный момент существующие технологии распознавания речи позволили упростить поиск информации в базах данных (или поисковых системах) с помощью голосовых запросов и управление различными техническими приспособлениями с помощью голосовых наборов команд. Также они способствовали появлению систем интерактивного речевого самообслуживания, приложений для диктовки текстов, искусственных помощников на мобильных устройствах, онлайн-консультантов на web-сайтах и по многим другим направлениям IT-технологий.

Предлагается собственный алгоритм гибридной СММ-нейросети, представляющий собой синтез алгоритмов скрытой марковской модели и нейронной сети. Данный алгоритм применяется для решения задачи распознавания слов в аудиозаписях. Также создан программный модуль, реализующий работу предложенного алгоритма гибридной СММ-нейросети.

Математическая постановка задачи

Формулировка задачи распознавания образов выглядит следующим образом: имеется множество объектов $X = \{x_1, \dots, x_n\}$, где $x: x \in X$ – объект распознавания. Тогда $g(x): X \rightarrow M, M = \{1, \dots, m\}$ – индикаторная функция, разбивающая пространство образов X на m непересекающихся классов. Индикаторная функция неизвестна.

Пусть F – пространство признаков, функция $f(x): F \rightarrow X$ ставит в соответствие каждому объекту $x_i \in X$ точку $f(x)$. Вектор $f(x)$ называется образом объекта $x \in X$. В пространстве образов признаков опреде-

лены непересекающиеся множества точек $K_i \in F, i = \overline{1, m}$, соответствующие образам одного класса.

Функция $\hat{g}(f): F \rightarrow M$ называется решающим правилом, оно представляет собой оценку для $g(x)$ на основании $f(x)$, то есть $\hat{g}(f) = \hat{g}(f(x))$.

Пусть $f_j = f(x_j), j = \overline{1, N}$, – доступная информация о функциях $f(x)$ и $g(x)$, но сами эти функции неизвестны. Тогда (g_j, f_j) – есть множество прецедентов.

Требуется построить такое решающее правило $\hat{g}(f)$, чтобы распознавание и классификация проводилось с минимальным числом ошибок.

Задача распознавания речи является частным случаем таких задач, как распознавание образов, классификация и машинное обучение. В текущей работе рассматривается практический пример задачи распознавания речи – задача распознавания слов в аудиозаписях.

Сложность этой задачи обусловлена многообразием возможных орфоэпических, лексических и небольших интонационных отличий речи для каждого человека.

В настоящее время распознавание речи сводится к решению трех типов задач:

1. Распознавание отдельно произносимых слов;
2. Распознавание слитной речи;
3. Идентификация по образцу речи.

В работе решается задача распознавания слитной речи с разделением файла на отдельные слова. В поставленной задаче распознавания речи в качестве множества объектов выступают слова, записанные на диктофон в аудиофайлах.

Таким образом, имеется множество слов $X = \{x_1, \dots, x_n\}$, где $x_j \in X$ – слово, которое необходимо распознать.

Как уже было сказано ранее, каждому объекту $x_j \in X$ ставится в соответствие вектор $f(x_j) = (f_j^1, f_j^2, \dots, f_j^n)$ – элемент векторного пространства.

Функция $f(x)$ выражает качество речевого сигнала, выраженное в количестве искажений и помех, а также интенсивность, амплитуду, частоту и т.д. Данная функция представляет собой формальную запись

фреймов, из которых состоит аудиофайл, т.е. $x_j = \{f_j^i\}^{i=1,N} \sim \{f_j^1, \dots, f_j^N\}$, где x_j – j -й фрейм файла.

Фрейм – структура, содержащая описание объекта в виде атрибутов и их значений.

В качестве значений фрейма используются элементы громкости и частоты отрезка звукового файла.

Методы распознавания речи

Для нашей задачи распознавания слов в аудиозаписях в качестве множества объектов берутся записанные на диктофон аудиофайлы.

Основу создания систем распознавания речи составляют методы, которые разделяют и описывают классы образов, где класс описывается списком входящих в него членов, а система распознавания может базироваться на принципе принадлежности к списку классов. Класс описывается некими общими свойствами, свойственными всем его членам. Система базируется на принципе общности свойств. Если при анализе класса обнаруживается тенденция к формированию конечной группы объектов, которые образуют компактную область в пространстве образов, то система основывается на принципе кластеризации.

На данный момент существует широкий спектр методов распознавания речи. Множество авторов разделяют методы на параметрические, непараметрические и эвристические.

Эвристические методы. Основой эвристического подхода является интуиция и полученный опыт; в нем используются основы перечисления членов класса и общности свойств. Основным методом данного класса является перцептрон.

Математические методы. В основу математического подхода положены правила классификации, которые формулируются с помощью принципов общности свойств и кластеризации. Примерами этой классификации могут послужить следующие методы:

1. Метод линейного анализа.
2. Метод подобия.
3. Метод дискриминантных функций.

Структурные методы. Данные методы основываются на использовании специальных грамматик производящих языков, с помощью которых можно описывать совокупность свойств объектов, которые распознаются. К структурным методам относятся следующие методы:

1. Скрытые марковские модели.

2. Нейросети.

3. Метод плавающего окна.

В данной работе были рассмотрены наиболее популярные и эффективные методы, а именно скрытые марковские модели и нейросети.

Этапы решения задачи распознавания образов. Задача распознавания речи является частным случаем таких задач, как распознавание образов, классификация и машинное обучение.

Перечислим основные этапы решения задачи распознавания речи – выявление объектов некоторых заданных классов на аудиофайлах:

1. Реализация считывания речи для получения входных данных.
2. Оценка качества речевого сигнала.
3. Акустическая адаптация.
4. Выделение участка, содержащего речь, и оценка параметров речи.
5. Нахождение соответствия между промежутками звукового сигнала и окнами эталонных слов.
6. Поступление промежутков звукового сигнала в декодер.
7. Вывод результата решения задачи.

Гибридная СММ-нейросеть

Рассмотрим практический пример задачи распознавания речи – задачу распознавания слов в аудиозаписях. В процессе решения данной задачи был предложен алгоритм гибридной СММ-нейросети, который построен на основе нейронных сетей и Скрытых Марковских Моделей.

Искусственная нейронная сеть (ИНС) – это математическая модель, а также устройства параллельных вычислений, представляющие собой систему соединённых и взаимодействующих между собой простых процессоров (искусственных нейронов).

ИНС представляет собой систему соединённых и взаимодействующих между собой простых процессоров (искусственных нейронов). Такие процессоры обычно довольно просты (особенно в сравнении с процессорами, используемыми в персональных компьютерах). Каждый процессор подобной сети имеет дело только с сигналами, которые он периодически получает, и сигналами, которые он периодически посылает другим процессорам. И тем не менее будучи соединёнными в достаточно большую сеть с управляемым взаимодействием такие по отдельности простые процессоры вместе способны выполнять довольно сложные задачи.

Скрытая марковская модель (СММ) – статистическая модель, имитирующая работу процесса, похожего на марковский процесс с неиз-

вестными параметрами, и задачей ставится разгадывание неизвестных параметров на основе наблюдаемых.

Марковский процесс – случайный процесс, эволюция которого после любого заданного значения временного параметра t не зависит от эволюции, предшествовавшей t , при условии, что значение процесса в этот момент фиксировано.

Обозначим полный набор параметров СММ как $\Theta = (\pi, A, B)$, где A – матрица переходных вероятностей, B – матрица вероятностей для наблюдения выходных значений, π – вектор вероятностей начальных состояний.

В нашем случае мы анализируем речь, преобразовывая её в подобие схем, или набора символов алфавита с помощью СММ для дальнейшего распознавания нейросетью.

При оценке распознанных слов звукового файла в работе используется метрика WER (World Error Rate).

Вычисляется WER по следующей формуле:

$$WER = \frac{S + D + I}{N}, \quad (1)$$

где S – количество замен фреймов; D – количество удалений фреймов; I – количество вставок фреймов; N – количество слов.

Алгоритм гибридной СММ-нейросети для решения задачи распознавания слов в аудиозаписях:

1. Через устройство записи речи приложение получает объект распознавания.
2. Производится вычисление уровня помех и искажений в аудиофайле.
3. Выполняется выделение участка, содержащего речь, разбивка на отдельные слова при помощи промежутков тишины.
4. Начинается оценивание параметров полученных промежутков звукозаписи.
5. Выполняется преобразование звуковых промежутков в символьные схемы с помощью СММ.
6. Производится декодинг полученных схем с помощью обученного алгоритма нейросети.
7. Для проверки правильности распознавания используется метрика WER для оценки правильности распознавания слов.
8. Если пункт 7 не дает желаемого результата, то промежутки возвращаются к пункту 5.

В процессе решения практической задачи создано программное обеспечение на языке C#, в котором реализована работа метода гибридной СММ-нейросети. Программное обеспечение предназначено для решения практической задачи распознавания речи, а именно задачи распознавания слов в аудиозаписях. На вход программы подается аудиофайл, а на выходе выдается текстовый файл или текст выводится на экран.

При оценке объемности задачи потребовалось собрать соответствующую обучающую выборку. Объем данных был взят из популярной видеоплатформы YouTube. При отборе спикеров использовались новостные и образовательные видео, которые мы преобразовывали в формат аудио.

Количество классов составило 300 единиц, т.е. 300 различных слов, где использовалось по 8 вариантов звучания каждого слова.

Итого объем выборки составил 2400 слов в 120 аудиозаписях. Для простоты и быстроты обучения в одной аудиозаписи подавались словосочетания из 2–5 слов. Для повышения качества распознавания каждое слово обучающей выборки записывалось в «неторопливой» скорости произношения.

В качестве исходных данных используются аудиозаписи одного наиболее распространенного формата: WAV. Данный формат был выбран, так как в нем используется, как правило, кодировка для хранения несжатого звука в импульсно-кодовой модуляции, что позволяет получить более чистый звук для дальнейшей обработки.

Для удобства работы алгоритма используется ограничение на запись аудиофайла, которое составляет 3 минуты.

Запись, как и в обычных диктофонах или других портативных устройствах, ведется в режиме моно. В данном случае стерео плохо подходит, так как из-за разбивки на каналы звука – ухудшается восприятие записи машиной.

Для более высокого качества записи использовался битрейт 128 кбит/с и выше.

Битрейт – это число бит или количество данных, которые обрабатываются в течение определенного периода времени.

Единственным минусом высокого битрейта является большой размер аудиофайлов. Поэтому после считывания и записи данных при обучении алгоритма – файл удалялся с устройства.

В работе был выполнен сравнительный анализ разработанного алгоритма распознавания и двух популярных приложений распознавания речи по следующим характеристикам: скорость распознавания, точ-

ность распознавания, количество ложных положительных и ложных отрицательных срабатываний (т.н. ошибок первого и второго рода).

Для вычислительного эксперимента были выбраны самые популярные разработанные приложения с инструментами распознавания речи. Yandex SpeechKit является самым точным приложением для распознавания русской речи. В настоящее время Google считается самым популярным интернет-поисковиком в мире. Также у них есть своя версия платформы голосового ввода, но в русской речи он показывает себя с проигрывающей стороны по сравнению с Yandex SpeechKit. Эти факторы повлияли на решение взять для сравнения обе эти платформы.

Вычислительный эксперимент производился при одинаковых условиях для каждого программного модуля. В каждый модуль подавалось по 50 одинаковых аудиозаписей, каждое слово уже находилось среди обучающей выборки разработанного модуля. Производились замеры скорости работы приложений, количества ошибок первого и второго рода, количества верно распознанных слов из набора в 200 слов.

Yandex SpeechKit позволяет распознать или озвучить любой текст на нескольких языках. Структура сервиса спроектирована с учётом высоких нагрузок, чтобы обеспечивать доступность и безотказную работу системы даже при большом количестве одновременных обращений. Сервис работает с аудио и текстами на трёх языках: русском, английском и турецком.

Gboard является экранной клавиатурой, созданной Google для устройств под управлением Android и iOS. GBoard сочетает в себе непрерывный ввод, поиск Google, голосовую диктовку и многие другие функции. Голосовая диктовка занимает относительно много времени, так как приложение многозадачно и не сосредоточено только на распознавании речи.

Исходя из результатов сравнения, можно сделать вывод, что Gboard уступает в скорости распознавания оставшимся двум алгоритмам, а также вероятность появления ошибок распознавания выше. Это обуславливается многозадачностью приложения и тем, что приложение не настроено на конкретно русскоязычную аудиторию. Абсолютным лидером является Yandex SpeechKit, так как при каждом использовании модуля он непрерывно обучается. Стоит отметить, что для увеличения показателей гибридной СММ-нейросети в дальнейшей поддержке модуля требуется увеличить обучающую выборку, обучив алгоритм большему количеству классов. В дальнейшем также планируется реализовать алгоритм непрерывного обучения при использовании программного модуля.

Заключение

Рассмотрена проблема распознавания слов в аудиозаписи. Это частный случай распознавания образов. Были описаны методы решения этой проблемы. Программное приложение разработано на языке С#, в котором реализована работа двух наиболее эффективных алгоритмов распознавания речи: нейронных сетей и скрытых марковских моделей. Решена практическая задача распознавания слов в аудиозаписях и проведен анализ результатов.

ЛИТЕРАТУРА

1. *Воронцов К.В.* Лекции по алгоритмам кластеризации и многомерного шкалирования. М.: МГУ, 2007. 18 с.
2. *Дуда Р., Харт П.* Распознавание образов и анализ сцен: пер. с англ. Г.Г. Вайнштейна, А.М. Васьковского / под ред. В.Л. Стефанюка. М.: Мир, 1976. 502 с.
3. *Еремин Д.М., Гарцев И.Б.* Искусственные нейронные сети в интеллектуальных системах управления: учеб. пособие. М.: МГИРЭА, 2004. 34 с.
4. *Загоруйко Н.Г.* Прикладные методы анализа данных и знаний. Новосибирск: ИМ СО РАН, 1999. 270 с.
5. *Зенин А.В.* Анализ методов распознавания образов // Молодой ученый. 2017. № 16. С. 125–130.
6. *Лепский А.Е., Броневиц А.Г.* Математические методы распознавания образов. Таганрог: ЮФУ, 2009. 155 с.
7. *Местецкий Л.М.* Математические методы распознавания образов. М.: МГУ, 2004. 85 с.

Нечёткий анализ формальных понятий: метод α -сечения

Валерий Голышев, Дарья Семенова

Сибирский федеральный университет, г. Красноярск, Россия

Интеллектуальный анализ данных нацелен на выявление, анализ и описание понятий, взаимосвязей между ними, а также ограничений, которым они должны удовлетворять. В таких задачах изучаемая предметная область, как правило, описывается объектно-признаковой таблицей, в которой каждый столбец соответствует некоторому признаку, а каждая строка определяет признаковое описание отдельного объекта. В случае представления объектно-признаковой таблицы в виде 0,1-матрицы мощным инструментом для извлечения знаний является анализ формальных понятий (АФП, англ. Formal Concept Analysis), который был заложен в 1982 году Рудольфом Вилле, а позже более полно описан в 1999 году совместно с Бернандом Гантером [8]. Анализ формальных понятий нашел свое место в многочисленных приложениях, таких, как машинное обучение, анализ текстов, менеджмент знаний, семантическая паутина, онтология, химия, биология и медицина, генетика, социология и др.

Сегодня активно развивается нечёткий анализ формальных понятий, использующий аппарат нечёткой логики. Данное обобщение позволяет уйти от бинарного представления формальных контекстов и рассматривать в качестве нечёткого контекста матрицу нечёткого бинарного отношения между множествами объектов и признаков [4]. В работе Офицерова и Смирнова [5] обозначены два подхода к извлечению формальных понятий из нечёткого контекста: использование оператора замыкания нечёткого множества [1, 2, 6, 7] и метод α -сечения, позволяющий строить чёткие [10] и нечёткие понятия [10, 11]. В данной работе рассмотрим подход α -сечений для построения нечётких формальных понятий [11].

Нечёткий анализ формальных понятий: метод α -сечения

Нечёткий формальный контекст будем определять как тройку $\tilde{K} = (G, M, \tilde{I})$, где G – множество объектов, M – множество признаков, а $\tilde{I} = \{((g, m), \mu(g, m)) \mid g \in G, m \in M, \mu : G \times M \rightarrow [0, 1]\}$ – нечёткое отноше-

ние между G и M , ставящее в соответствие каждой паре (g, m) степень принадлежности $\mu(g, m)$ [11]. В дальнейшем для краткости будем писать «нечёткий контекст» вместо «нечёткий формальный контекст».

Для заданного нечёткого контекста $\tilde{K} = (G, M, \tilde{I})$ его α -сечением ($\alpha \in [0, 1]$) будем называть формальный контекст $K_\alpha = (G, M, I_\alpha)$, в котором отношение инцидентности $I_\alpha = \{(g, m) \in G \times M \mid \tilde{I}(g, m) \geq \alpha\}$ представлено 0,1-матрицей.

Для $\alpha \in [0, 1]$ и произвольных $g \in G$, $m \in M$ из \tilde{K} определим соответствия Галуа [11] φ_α и ψ_α следующим образом:

$$\varphi_\alpha(g) = \{m \in M \mid \mu(g, m) \geq \alpha\}; \quad (1)$$

$$\psi_\alpha(m) = \{g \in G \mid \mu(g, m) \geq \alpha\}. \quad (2)$$

Отображения φ_α и ψ_α обобщаются на множества $A \subseteq G$ и $B \subseteq M$:

$$\varphi_\alpha(A) = \bigcap_{g \in A} \varphi_\alpha(g) = \{m \in M \mid \forall g \in A: \mu(g, m) \geq \alpha\}; \quad (3)$$

$$\psi_\alpha(B) = \bigcap_{m \in B} \psi_\alpha(m) = \{g \in G \mid \forall m \in B: \mu(g, m) \geq \alpha\}. \quad (4)$$

Для отображений φ_α и ψ_α будем использовать единое обозначение $(\cdot)'_\alpha$, аналогичное обозначению в классическом анализе формальных понятий [3, 8].

Нечётким формальным понятием уровня α (или нечётким α -уровневым формальным понятием) нечёткого контекста \tilde{K} называют пару $(\tilde{A}, B)_\alpha$, где $\tilde{A} = \{(g, \mu_g) \mid g \in A, \mu_g = \min_{m \in B} \mu(g, m)\}$ – нечёткое множество объектов и справедливы равенства $A'_\alpha = B$ и $B'_\alpha = A$, $A \subseteq G$ и $B \subseteq M$. В дальнейшем для краткости слово «формальное» перед словом «понятие» будем опускать.

Множество всех нечётких понятий уровня α будем обозначать \widetilde{FC}_α . Множество \widetilde{FC}_α с заданным отношением частичного порядка образует α -решетку формальных понятий \tilde{L}_α [11].

Проиллюстрируем введенные определения на примере нечёткого контекста $\tilde{K} = (G, M, \tilde{I})$, представленного в табл. 1. В табл. 2 приведен пример α -сечения $K_{0.5}$ нечёткого контекста \tilde{K} при $\alpha = 0.5$.

Таблица 1

Нечёткий формальный контекст

	m_1	m_2	m_3
g_1	1	0.5	0.7
g_2	0.8	0.1	0.3
g_3	0	0.4	0.7

Таблица 2

α -сечение исходного контекста при $\alpha = 0.5$

	m_1	m_2	m_3
g_1	1	1	1
g_2	1	0	0
g_3	0	0	1

Рассмотрим ситуацию при $\alpha = 0.3$. Формулы (3) и (4) для $A = \{g_1, g_2\}$ и $B = \{m_1, m_3\}$ примут вид

$$\Phi_{0.3}(A) = \bigcap_{g \in A} \Phi_{0.3}(g) = \Phi_{0.3}(g_1) \cap \Phi_{0.3}(g_2) = \{m_1, m_2, m_3\} \cap \{m_1, m_3\} = \{m_1, m_3\},$$

$$\Psi_{0.3}(B) = \bigcap_{m \in B} \Psi_{0.3}(m) = \Psi_{0.3}(m_1) \cap \Psi_{0.3}(m_3) = \{g_1, g_2\} \cap \{g_1, g_2, g_3\} = \{g_1, g_2\}.$$

Отсюда $A'_{0.3} = \{m_1, m_3\} = B$ и $B'_{0.3} = \{g_1, g_2\} = A$. Найдем \tilde{A} . Для этого вычислим μ_{g_1} и μ_{g_2} :

$$\mu_{g_1} = \min_{m \in B} \mu(g_1, m) = \min \{\mu(g_1, m_1), \mu(g_1, m_3)\} = \min \{1, 0.7\} = 0.7,$$

$$\mu_{g_2} = \min_{m \in B} \mu(g_2, m) = \min \{\mu(g_2, m_1), \mu(g_2, m_3)\} = \min \{0.8, 0.3\} = 0.3.$$

Итак, $(\{(g_1, 0.7), (g_2, 0.3)\}, \{m_1, m_3\})_{0.3}$ – нечёткое понятие уровня 0.3.

Заметим, что при $\alpha = 0.5$ мы получим $A'_{0.5} = \{m_1\} \neq B$, $B'_{0.5} = \{g_1\} \neq A$, т.е. нечёткое понятие при таком значении α построить нельзя.

Сформулируем и докажем следующее утверждение.

Предложение 1. Для нечёткого контекста $\tilde{K} = (G, M, \tilde{I})$ пара множеств $(\tilde{A}, B)_\alpha$ является нечётким формальным понятием уровня $\alpha \in [0, 1]$ тогда и только тогда, когда (A, B) является классическим формальным понятием в бинарном контексте $K_\alpha = (G, M, I_\alpha)$.

Доказательство. Заметим, что для бинарного контекста $K_\alpha = (G, M, I_\alpha)$ отношение I_α имеет вид

$$I_\alpha(g, m) = \begin{cases} 0, & \mu(g, m) < \alpha, \\ 1, & \mu(g, m) \geq \alpha. \end{cases}$$

Из формул (3) и (4) следует, что отображения $\varphi_\alpha(A)$ и $\psi_\alpha(B)$ в K_α есть классические соответствия Галуа [3, 8]:

$$\varphi(A) = \bigcap_{g \in A} \varphi(g) = \{m \in M : \forall g \in A (g, m) \in I_\alpha\},$$

$$\psi(B) = \bigcap_{m \in B} \psi(m) = \{g \in G : \forall m \in B (g, m) \in I_\alpha\}.$$

Тогда, если выполняется $\varphi(A) = B$ и $\psi(B) = A$, то также верно и $\varphi_\alpha(A) = B$ и $\psi_\alpha(B) = A$. С другой стороны, из $\varphi_\alpha(A) = B$ и $\psi_\alpha(B) = A$ следует, что $\varphi(A) = B$ и $\psi(B) = A$. Значит, $(\tilde{A}, B)_\alpha$ – нечёткое формальное понятие уровня α в $\tilde{K} \Leftrightarrow (A, B)$ – формальное понятие в K_α .

Алгоритм Fuzzy CbO нахождения нечётких α -уровневых понятий

Авторами [11] разработан алгоритм извлечения нечётких формальных понятий уровня α , сложность которого составляет $O(|\widetilde{FC}_\alpha| \cdot |G|^3 \cdot |M|)$.

Для определенности в данной работе будем называть его алгоритмом Янга. На основе доказанного предложения нами разработан алгоритм Fuzzy CbO для нахождения нечётких α -уровневых понятий. На первом этапе алгоритма Fuzzy CbO мы находим все формальные понятия контекста $K_\alpha = (G, M, I_\alpha)$ с помощью алгоритма Close-by-One [9], на втором этапе вычисляем функции принадлежности для каждого объекта объема понятия методом, предложенном в [11]. Сложность алгоритма FuzzyCbO составляет $O(|\widetilde{FC}_\alpha| \cdot |G|^2 \cdot |M|)$.

Вычислительные эксперименты

В данной работе был реализован Fuzzy CbO и алгоритм Янга [11]. Сравнение алгоритмов проведено на случайных нечётких контекстах размерности 100×50 и с разным значением α . Данные контексты заполнены значениями $\{0, 0.1, 0.2, \dots, 0.9, 1\}$. Результаты экспериментов приведены в табл. 3 и на рис. 1.

Результаты экспериментов

	α	$ \widehat{FC}_\alpha $	t, c
Yang's algorithm	0	1	8
Fuzzy CbO			0.001
Yang's algorithm	0.3	42250	54215
Fuzzy CbO			419
Yang's algorithm	0.4	19833	11910
Fuzzy CbO			216
Yang's algorithm	0.5	9043	2434
Fuzzy CbO			102
Yang's algorithm	0.6	4507	614
Fuzzy CbO			53
Yang's algorithm	0.8	858	35
Fuzzy CbO			11
Yang's algorithm	1	129	3
Fuzzy CbO			2

На рис. 1 показано время работы двух алгоритмов для разного количества понятий. По результатам экспериментов видно, что количество понятий зависит от значения параметра α . Также заметна разница в производительности двух реализованных алгоритмов, которая только возрастает с увеличением размерности задачи.

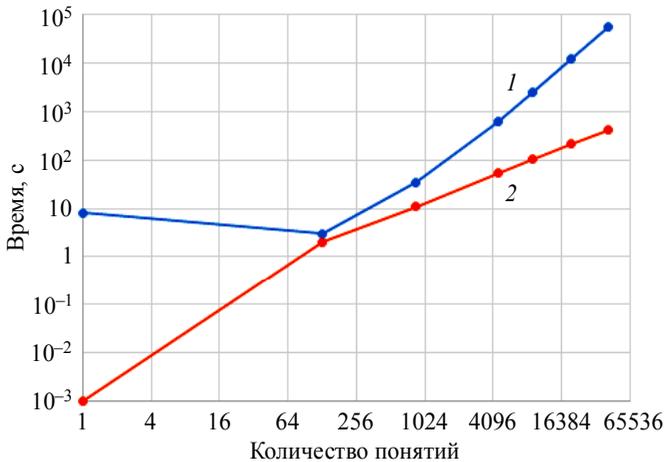


Рис. 1. Сравнение времени работы алгоритмов Янга (кр. 1) и Fuzzy CbO (кр. 2)

Заключение

Метод α -сечений позволяет извлекать нечёткие формальные понятия уровня α из нечёткого контекста. Мы можем получать различные понятия в зависимости от значения параметра α , что было продемонстрировано в примерах. В ходе работы было доказано предложение, позволяющее упростить процесс поиска нечётких понятий, разработан и реализован алгоритм Fuzzy CbO. Также был запрограммирован алгоритм Янга и проведены вычислительные эксперименты.

ЛИТЕРАТУРА

1. *Гольшев В.К.* Извлечение формальных понятий из нечёткого контекста // Информационные технологии и математическое моделирование (ИТММ-2019): материалы XVIII Международной конференции имени А.Ф. Терпугова. Томск: Изд-во НТЛ, 2019. Часть 1. С. 120–125.
2. *Гольшев В.К.* Методы нечёткой логики в анализе формальных понятий // Материалы VII Международной молодежной научной конференции «Математическое и программное обеспечение информационных, технических и экономических систем». Томск, 2019. С. 269–276.
3. *Игнатов Д.И., Яворский Р.Э.* Анализ формальных понятий: от теории к практике // Анализ изображений сетей и текстов. Екатеринбург, 2012. С. 1–12.
4. *Кофман А.* Введение в теорию нечетких множеств: пер. с франц. М.: Радио и связь, 1982. 432 с.
5. *Офицеров В.П., Смирнов С.В.* Нечёткий анализ формальных понятий при разработке онтологий // Онтология проектирования. 2017. № 4(26).
6. *Belohlavek R.* Algorithms for fuzzy concept lattices. 2002.
7. *Belohlavek R., Vychodil V., Funiokova T.* Galois connections with hedges.
8. *Ganter B., Wille R.* Formal Concept Analysis: Mathematical Foundations. Springer, 1999. P. 284.
9. *Kuznetsov S.O., Obiedkov S.A.* Comparing performance of algorithms for generating concept lattices // Stability of a Formal Concept. 2002. No. 15. P. 189–216.
10. *Tho Q.T., Hui S.C., Fong A.C.M., Cao T.H.* Automatic fuzzy ontology generation for semantic Web // IEEE Transactions on Knowledge and Data Engineering. 2006. V. 18. No. 6. P. 842–856.
11. *Yang K.M., Kim E.H., Hwang S.H., Choi S.H.* Fuzzy concept mining based on formal concept analysis // Int. J. Computers. 2008. V. 2. Iss. 3. P. 279–290.

Распознавание k -кластеризуемости знаковых графов

Эллада Ибрагимова, Дарья Семенова

Сибирский федеральный университет, г. Красноярск, Россия

Знаковый граф – это граф, каждое ребро которого помечено знаком «+» или «-». Данные графы возникли в социальной психологии в теории структурного баланса Хайдера [1], где изучается система Р-О-Х, где Р и О – субъекты, а Х – некоторый объект. Отношения между данными субъектами могут быть положительными или отрицательными. Если Р имеет отношение приязни к О, и О ответственен за Х, то у Р будет наблюдаться тенденция одобрять Х. Но если сам О не одобряет Х, то вся система Р-О-Х оказывается в состоянии дисбаланса и будет стремиться исправить это. Состояние баланса достигается изменением отношения приязни между Р и О, ответственности между О и Х или оценки между Р и Х.

Наиболее полный обзор возможных приложений знаковых графов приведен в работе [2]. Отметим, что в настоящее время знаковые графы активно применяются в задачах кластеризации [3] и тематического моделирования текстов [4]. Кластерный анализ – это задача о разбиении множества на подгруппы схожих элементов. Существуют различные формулировки задачи о кластеризации. Количество кластеров k может быть задано изначально – тогда решается задача о разбиении ровно на k кластеров или не более чем на k кластеров, или может быть не задано [5]. Данная работа посвящена задаче распознавания k -кластеризуемости знакового графа.

Основные определения

Формально знаковым графом называется пара $\Sigma = (G, \sigma)$, где $G = (V, E)$ является неориентированным графом, на ребрах которого задана функция знака $\sigma: E \rightarrow \{+, -\}$, $n = |V| \geq 2$, $m = |E| \geq 1$.

Знаковый граф $\Sigma = (G, \sigma)$ называется 2-сбалансированным, если любой его простой цикл положительный. По определению знак цикла равен произведению знаков входящих в него ребер. Задачи, связанные с 2-сбалансированностью знаковых графов, рассматривались в работах [6–8].

Картрайтом и Харари [1] было установлено, что знаковый граф $\Sigma = (G, \sigma)$ 2-сбалансирован тогда и только тогда, когда множество его

вершин V можно разбить на две доли A и B , одна из которых может быть пустой таким образом, что $A \cup B = V$, $A \cap B = \emptyset$ и любое ребро, соединяющее вершины из одной доли, имеет знак «+», а ребра, соединяющие вершины из разных долей, имеют знак «-».

Обобщением 2-сбалансированности знакового графа является понятие k -кластеризуемости. Знаковый граф называется k -кластеризуемым (k -сбалансированным), если множество его вершин можно разбить на k попарно непересекающихся непустых кластеров так, что все положительные рёбра находятся внутри кластеров, а отрицательные – между кластерами.

В работе Дэвиса [9] показано, что знаковый граф кластеризуем тогда и только тогда, когда он не содержит простых циклов с ровно одним отрицательным ребром.

Распознавание k -кластеризуемости знакового графа

Задача распознавания k -кластеризуемости знакового графа формулируется следующим образом.

Условие. Задан знаковый граф $\Sigma = (G, \sigma)$, где $G = (V, E)$ – простой неориентированный граф, $n = |V| \geq 2$, $m = |E| \geq 1$.

Вопрос. При каких k граф $\Sigma = (G, \sigma)$ является k -кластеризуемым?

Рассматриваемую задачу можно сформулировать как задачу дискретной оптимизации. Рассмотрим некоторое разбиение (кластеризацию) множества вершин знакового графа. Будем говорить, что система множеств $C_k = \{V_1, V_2, \dots, V_k\}$ образует разбиение множества вершин V на k кластеров, если множества непусты ($V_i \neq \emptyset$, $i = 1 \dots k$), попарно не пересекаются $V_i \cap V_j = \emptyset$, $i \neq j$ и их объединение равно $V = V_1 \cup V_2 \cup \dots \cup V_k$. Множество всех возможных разбиений V будем обозначать $\Phi = \bigcup_{k=1}^n \Phi_k$, где Φ_k – множество возможных разбиений V на k кластеров. Заметим, что мощность пространства решений Φ равна количеству всех возможных разбиений множества вершин графа V , то есть числу Белла B_n , где $n = |V|$.

Пусть C_k – какое-либо разбиение графа на k кластеров. Под отрицательной ошибкой будем понимать число отрицательных ребер внутри кластера $N(C_k)$, а положительной $P(C_k)$ – число положительных ребер между кластерами. В качестве критерия качества кластеризации будем рассматривать функцию суммарной ошибки $E_\alpha(C_k)$ – выпуклую ком-

бинацию положительной и отрицательной ошибок

$$E_{\alpha}(C_k) = \alpha N(C_k) + (1 - \alpha)P(C_k), \alpha \in [0, 1].$$

Коэффициент α показывает, какую ошибку важнее учитывать при решении задачи: межкластерную $P(C_k)$ или внутрикластерную $N(C_k)$.

Тогда задача распознавания k -кластеризуемости знакового графа может быть сформулирована следующим образом. Для заданного $\alpha \in [0, 1]$ требуется найти такое разбиение C^* множества вершин V знакового графа Σ , для которого суммарная ошибка минимальна:

$$C^* = \arg \min_{C \in \Phi} E_{\alpha}(C).$$

Ответом на вопрос задачи будет число $k = |C^*|$. В работе [3] доказано, что задача минимизации суммарной ошибки является NP-полной.

Рассмотрим задачу при $\alpha = 0$, то есть найдём такое $k = |C^*|$, что

$$C^* = \arg \min_{C \in \Phi} P(C).$$

Очевидно, что положительная ошибка достигает минимума на разбиении, для которого $P(C^*) = 0$.

Для решения данной задачи был разработан алгоритм kClusterability, который основан на поиске разбиений с нулевой положительной ошибкой.

Алгоритм kClusterability

Вход: количество вершин и ребер графа Σ , список ребер с указанием знака для каждого ребра.

- 1: построить граф Σ' удалением из графа Σ отрицательных ребер;
 - 2: найти все компоненты связности в графе Σ' ;
 - 3: проверить наличие отрицательных ребер внутри кластеров;
 - 4: объединить компоненты связности, между которыми нет отрицательных ребер.
-

Выход: k , при которых граф k -кластеризуем, кластеры.

На первом шаге для получения графа без отрицательных ребер Σ' и втором шаге для выделения всех компонент связности графа Σ' используется алгоритм поиска в ширину, вычислительная сложность которого составляет $O(m + n)$. На третьем шаге в компонентах связности проверяется наличие отрицательных ребер путём попарного сравнения вер-

шин, входящих в эти компоненты. На последнем шаге формируются кластеры путём объединения компонент связности, между вершинами которых в исходном графе Σ нет отрицательных ребер. Операция объединения требует попарного сравнения полученных компонент связности между собой. Нетрудно показать, что вычислительная сложность предложенного алгоритма составляет $O(2m + n^4)$.

Результаты вычислительных экспериментов

Были проведены несколько серий вычислительных экспериментов на графах, для которых заранее была известна k -сбалансированность. Эксперименты проводились на ПК с процессором Intel® Core™ i7-10510U CPU @ 1.80ГГц и оперативной памятью объемом 8 ГБ. Результаты экспериментов представлены в таблице.

Результаты вычислительных экспериментов

Размерность графа		Сбалансированность	Минимальное количество кластеров	Максимальное количество кластеров	Время работы алгоритма, мс
n	m				
9	9	1- и 3-сбалансирован	1	3	менее 1
9	72	3-сбалансирован	3	3	менее 1
9	72	несбалансирован	-	-	1
25	50	1- или 5-сбалансирован	1	5	1
25	575	5-сбалансирован	5	5	менее 1
25	575	несбалансирован	-	-	1
36	90	1- или 6-сбалансирован	1	6	1
36	1206	6-сбалансирован	6	6	1
36	1206	несбалансирован	-	-	2
100	450	1- или 10-сбалансирован	1	10	3
100	9550	10-сбалансирован	10	10	4
100	9550	несбалансирован	-	-	4

Заключение

Для решения задачи распознавания k -кластеризуемости знакового графа для случая минимизации положительной ошибки ($\alpha = 0$) нами разработан, реализован и протестирован алгоритм kClusterability. В дальнейшем планируется решение задачи кластеризации знакового графа при различных значениях параметра α суммарной ошибки кластеризации.

ЛИТЕРАТУРА

1. *Kartwright D., Harary F.* Structural balance: a generalization of Heider's theory // *The Psychological Review*. 1956. V. 63. No. 5. P. 277–293.
2. *Zaslavsky T.* Balance and Clustering in Signed Graphs: Unpublished manuscript. New York: Binghamton University, 2010.
3. *Doreian P., Mrvar A.* Structural balance and partitioning signed graphs // A. Ferligoj & A. Kramberger (Eds.). *Developments in data analysis (Metodoloski zvezki 12, p. 195–208)*. Ljubljana, Slovenia: FDV.
4. *Nikhil Bansal, Avrim Blum, Shuchi Chawla.* Correlation clustering // *Machine Learning*. 2004. V. 56. P. 89–113.
5. *Ильев В.П., Ильева С.Д., Моршин А.В.* Алгоритм приближенного решения одной задачи кластеризации графа // *Прикладная дискретная математика*. 2019. V. 45. P. 64–77.
6. *Ибрагимова Э.И.* Сбалансированные графы и их применение // *Информационные технологии и математическое моделирование (ИТММ-2019): материалы XVIII Международной конференции имени А.Ф. Терпугова. Томск: Изд-во НТЛ, 2019. Часть 2. С. 15–20.*
7. *Ибрагимова Э.И.* Исследование 2-сбалансированных знаковых графов // *Материалы VII Международной молодежной научной конференции «Математическое и программное обеспечение информационных, технических и экономических систем»*. Труды Томского государственного университета. 2019. Т. 303. С. 276–281.
8. *Figueiredo R., Frota Y.* The maximum balanced subgraph of a signed graph: Applications and solution approaches // *European Journal of Operational Research*. 2014. V. 236. P. 473–487.
9. James A. Davis, Clustering and structural balance in graphs // *Human Relations*. 1967. P. 181–187.

Первичный концептуальный анализ сестринского дела для экспертной советующей системы*

Анна Ивлева, Сергей Смирнов

*Самарский федеральный исследовательский центр РАН,
Институт проблем управления сложными системами РАН, г. Самара, Россия*

Сегодня особенно важным и актуальным становится применение технологий для быстрого и эффективного удаленного доступа медицинских работников и населения к большому объему знаний, терминов о диагностике, лечении и уходе за больными. Фундаментальная проблема, препятствующая выводу программного инструментария медицинских сестер (МС) с уровня лишь административных средств автоматизации документооборота в лечебном учреждении на конструктивный уровень помощника в сестринском деле, заключается в отсутствии полной единой отраслевой базы знаний, структурирующей и систематизирующей накопленные знания и опыт сестринского ухода за пациентами. Тенденции и предпосылки к созданию такой базы знаний в последнее время проявляются все больше, в том числе и в связи со становлением доказательной сестринской практики [1]. Одним из свидетельств этому является востребованность и актуальность электронных баз научных статей по сестринскому делу [2].

В настоящей работе описаны первые шаги совместного с отраслевыми экспертами исследования, направленного на решение задач системного анализа процесса сестринского ухода и формирования онтологии, которая позволит создать информационного помощника, реализующего объясняющий искусственный интеллект при диагностике, планировании, проведении и контроле сестринского ухода.

Особенности применения информационных технологий в сестринском деле

В настоящее время онтологии представления и интеграции предметных знаний в медицине активно развиваются и получают широкое распространение и практическое применение [3]. В последние годы на-

* Исследование выполнено при частичной финансовой поддержке Министерства науки и высшего образования Российской Федерации, регистрационный номер АААА-А19-119030190053-2.

блюдается рост интереса к медицинским компьютерным онтологиям, которые являются удобным инструментом для описания предметной области, в том числе для создания систем и приложений с веб-интерфейсом. В сочетании с технологиями машинного обучения, их применение позволяет создавать комплексные медицинские информационные системы [3, 4], диагностические экспертные системы [5], автоматизированные системы поиска и хранения информации, системы медицинского администрирования, телемедицинские системы и т.д. Сегодня рынок перечисленных продуктов в России наиболее широко представлен комплексными программными продуктами, позволяющими автоматизировать документооборот в медицинском учреждении. Они включают сбор финансово-административной информации, данных медицинских исследований и мониторинг состояния пациента с медицинских приборов в цифровой форме, являются средствами общения между сотрудниками, системами поддержки принятия врачебных решений. Среди наиболее распространенных продуктов MEDODS, Medesk, MedElement, Smart Медицина, Медицинский помощник, Clinic365, Infoclinica, IDENT, MedWork, ClinicIQ, Renovatio, МедОфис, Инфоклиника, Augmedix. Что же касается специализированных приложений и модулей, посвященных обязанностям МС, то они представляют собой преимущественно административные средства автоматизации документооборота в лечебном учреждении. Однако не представлены в достаточной мере информационные подсистемы, посвященные сестринскому процессу, который отличается от врачебного и дополняет его. Немного лучше дело в этой сфере обстоит с зарубежными помощниками медицинских сестер. Существующие англоязычные примеры интеллектуальных систем (AI-based nursing assistants, virtual nurses и др.) в этой сфере имеют несколько более практичную направленность, связанную с непосредственным уходом за пациентами. Часть из них (GreatCall, Angel) ориентирована в основном на пользователей, – это чат-боты или голосовые помощники, выполняющие функции справочника по симптомам и менеджера ежедневных активностей/действий, связанных со здоровьем. При этом, если предусмотрена возможность измерения объективных данных пациента, приложения также могут собирать данные о пациенте и передавать их медицинской сестре или родственникам. Другой, намного более дорогостоящий, вид систем связан с непосредственной заменой части функций медицинских сестер интеллектуальными роботами.

В целом, существующие модули медицинских информационных систем, относящиеся к сестринскому делу, играют значительную роль в повышении скорости и эффективности работы МС, но пока еще не содержат баз знаний, включающих конструктивные логические взаимосвязи концептов, например сестринских диагнозов с целями и планами ухода, нормативной документацией и данными научных исследований.

Концептуальный анализ сестринского дела

Для разработки прототипа информационного помощника при диагностике, планировании, проведении и контроле сестринского ухода первостепенной задачей исследования является концептуальный анализ и структурирование объектно-признаковых данных в отрасли.

В ходе извлечения знаний из специальных текстов, включающих научные статьи и учебники [6–10], сделан вывод о том, что ключевым понятием в практике медицинского дела является сестринский процесс (СП). Поэтому при выявлении основных концептов, а также эксплицитных и имплицитных связей между ними выделена подсистема понятий, непосредственно связанных с пятью основными этапами СП. Остальные понятия в исследуемой отрасли являются не менее важными и фундаментальными, но скорее относятся к созданию контекста для осуществления СП и обобщаются единым понятием философии СД. Часто понятия, связанные с организацией инфекционного контроля и безопасности СП выделяют в отдельную группу. Это, скорее всего, связано с ее исключительной важностью для жизни пациентов и медицинских работников. Однако, следуя принципам конструктивности и целостности, мы будем рассматривать данную группу понятий непосредственно как часть СП на всех его этапах.

Подход и перспективы разработки отраслевой базы знаний

Единая отраслевая база знаний СД может стать платформой для реализации различных модулей (рис. 1).

Для разработки базы знаний СД необходимым является выявление связей между концептами исследуемой отрасли. Разработка тезауруса знаний сестринского дела как ресурса концептуального типа, то есть семиотической сети понятий с мультимедийными компонентами будет осуществляться с помощью специализированной литературы и экспертных знаний. Для описания структурных связей между объектами в исследуемой предметной области в контекст задачи анализа формаль-

ных понятий потребуются ссылочные признаки, что, в конечном счете, позволит получить формальные онтологии в сестринском деле с произвольными отношениями между концептами.

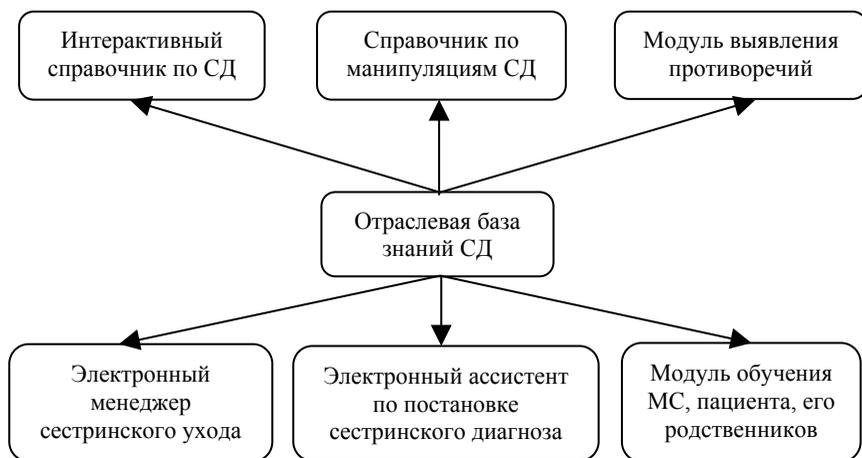


Рис. 1. Модули на основе отраслевой базе знаний СД

Создание формальной онтологии будет осуществлено введением формального языка – дескриптивной логики на выявленных и проанализированных концептах и отношениях. Введение стандартного языка для компьютерных систем позволит создать на базе полученной онтологии информационного помощника МС, который, используя заложенную и обновляемую информацию, сможет преобразовывать первичные данные в более информативную форму, обеспечивающую помощь при диагностике, планировании, проведении и контроле сестринского ухода.

Выводы

Проведенное исследование позволило увидеть контуры проекта адекватной и обновляемой базы знаний по процессу сестринского ухода за пациентами, имеющая научную и прикладную значимость для медицинских онтологий, теории, философии сестринского дела, обучения медицинских работников. Предполагается, что база знаний будет включать нормативные документы; медицинскую документацию лечебного учреждения; научные исследования; анамнез пациента; назначения пациента с указанием дозировок, частоты, времени приема; стандарты до-

зировок препаратов, соответствующие диагнозу и анамнезу; алгоритмы манипуляций, их особенности и варианты; показания и противопоказания к проведению того или иного сестринского вмешательства; жизненные потребности, сестринские диагнозы, проблемы пациента, цели и оценки ухода, а также связи между перечисленными элементами.

Междисциплинарные связи, состоящие во взаимодействии инженеров знаний, программистов и медицинских работников обеспечат реализовать всесторонний практический подход к повышению скорости и эффективности персонализированного ухода за пациентами.

ЛИТЕРАТУРА

1. *Бальдассарре Ф.* Сестринская практика, основанная на принципах доказательной медицины: для чего и для кого? // Медицинское образование и профессиональное развитие. 2012. № 1(7). С. 51–58.
2. *Торчинский Н.В., Бражников А.Ю.* Доказательная сестринская практика. Оценка достоверности диагностических методов // Медицинская сестра. 2010. № 6. С. 27–33.
3. *Нефедов Ю.В., Цыленкова В.А.* Основные тенденции и особенности развития медицинских онтологий // Терминология и стандартизация. 2018. № 4. С. 6–19.
4. *Подкольный Н.Л., Подкольная О.А.* Онтологии в биоинформатике и системной биологии // Вавиловский журнал генетики и селекции. 2015. № 9(6). С. 652–660.
5. *Грибова В.В., Петряева М.В., Окунь Д.Б., Шалфеева Е.А.* Онтология медицинской диагностики для интеллектуальных систем поддержки принятия решений // Онтология проектирования. 2018. № 1(27). Т. 8. С. 58–73.
6. *Борисов Ю.Ю.* Теория сестринского дела. Учебник для студентов вузов, обучающихся по направлению подготовки 34.03.01 Сестринское дело. Краснодар, 2014. 121 с.
7. *Двойников С.И., Карасева Л.А., Пономарева Л.А.* Теория сестринского дела. Учебное пособие для студентов факультета высшего сестринского образования. Самара: ГП «Перспектива», 2002. 160 с.
8. *Камынина Н.Н., Островская И.В., Пьяных А.В.* Теория сестринского дела. М.: ИНФРА-М, 2016. 224с.
9. *Обуховец Т.П.* Основы сестринского дела. Учебное пособие. М.: Феникс, 2016. 768 с.
10. *Островская И.В., Широкова Н.В.* Основы сестринского дела. М.: ГЭОТАР-Медиа, 2016. 320 с.

Алгоритм HGFC нахождения формальных понятий

Александр Солдатенко, Дарья Семенова

Сибирский федеральный университет, г. Красноярск, Россия

В настоящей работе рассматривается задача выделения формальных понятий из формального контекста, представленного гиперграфом. Такое представление позволяет применять методы теории графов для работы с формальными контекстами, а также использовать полученные результаты для анализа гиперграфов.

Анализ формальных понятий используется при построении и проектировании онтологий и концептуальных моделей предметных областей [1]. При этом исходный контекст представляется $(0, 1)$ -матрицей, что позволяет применять различные методы комбинаторной оптимизации. Так задача выделения всех формальных понятий сводится к задаче построения конфигураций – максимальных единичных подматриц [2]. К аналогичной задаче сводится задача поиска всех клик графа. Заметим, что задача поиска всех формальных понятий является $\#P$ -полной [1].

Основные определения и алгоритм нахождения формальных понятий HGFC

Введем необходимые понятия и определения теории гиперграфов [3] и анализа формальных понятий [1].

Пусть задан (n, m) -гиперграф $H = (X, U)$, где X – конечное множество вершин и U – конечное семейство гиперребер гиперграфа, при этом $n = |X| \geq 1$, $m = |U| \geq 1$ и всякое гиперребро гиперграфа – некоторое подмножество множества X . Пусть $X(u)$ – множество всех вершин, инцидентных гиперребру $u \in U$, а $U(x)$ – множество гиперребер, инцидентных вершине $x \in X$. Двойственным к гиперграфу $H = (X, U)$, называют гиперграф $H^* = (U, X)$.

Основной задачей анализа формальных понятий (АФП) является выделение формальных понятий из формального контекста $K = (G, M, I)$, где G – множество объектов, M – множество признаков, I – непустое отношение инцидентности $I \subset G \times M$. Поскольку контекст определяется $(0, 1)$ -матрицей, то ему можно взаимно-однозначно сопоставить гиперграф $H = (X, U)$, при этом множество объектов G совпадает

с множеством вершин X , а множество признаков M – с множеством гиперребер U . Исходя из этого, отношение инцидентности I контекста задает одноименную матрицу инцидентности для гиперграфа $H = (X, U)$. Также отметим, что для двойственного гиперграфа H^* данная матрица будет транспонирована $I^* = I^T$.

Анализ формальных понятий основан на соответствиях Галуа, которые определяются отображениями $\varphi(x)$ и $\psi(u)$. Данные отображения определяют набор признаков для объекта и объекты, обладающие признаками соответственно [4]. Отображения можно описать с помощью теории гиперграфов следующим образом:

$$\varphi(x) = U(x), \quad \psi(u) = X(u),$$

а для множеств $A \subset X$ и $B \subset U$:

$$\varphi(A) = \bigcap_{x \in A} U(x), \quad \psi(B) = \bigcap_{u \in B} X(u).$$

Формальным понятием контекста $K = (G, M, I)$ является пара (A, B) , такая, что $\varphi(A) = B$, $\psi(B) = A$, где $A \subset G$, $B \subset M$. Пусть FC множество всех формальных понятий. Пусть $(A_1, B_1), (A_2, B_2) \in FC$. Множество FC частично упорядочено отношением $(A_1, B_1) \preceq (A_2, B_2)$ тогда и только тогда, когда $A_1 \subset A_2$, что эквивалентно $B_2 \subset B_1$. При этом говорят, что формальное понятие (A_2, B_2) является более общим, чем формальное понятие (A_1, B_1) . Отметим, что упорядоченное множество (FC, \preceq) образует решетку формальных понятий [1, 4, 7].

Опишем формальное понятие на языке $(0, 1)$ -матриц. Для этого введем определения полной подматрицы и максимально полной подматрицы [2].

Определение. Полной подматрицей $(0, 1)$ -матрицы называется подматрица, все без исключения элементы которой равны 1.

Определение. Максимально полной подматрицей называется полная подматрица, не содержащая никакой другой максимально полной матрицы.

Последнее можно трактовать следующим образом: максимально полной подматрицей называется полная подматрица, не входящая ни в какую другую полную подматрицу. Исходя из определений отображений $\varphi(x)$, $\psi(u)$ и формального понятия, легко заметить, что в терминах

(0, 1)-матриц максимально полная подматрица является формальным понятием.

Основываясь на введенных выше определениях, опишем алгоритм нахождения формальных понятий HyperGraphFormalConcept (HGFC).

Алгоритм принимает на вход гиперграф $H = (X, U)$ без кратных гиперребер и вершин. Для простоты изложения будем считать, что $\max_{u \in U} |X(u)| \leq \max_{x \in X} |U(x)|$. Это соответствует тому, что максимальная степень гиперребра меньше максимальной степени вершины. Идея алгоритма заключается в генерации всех сочетаний на каждом уровне решетки формальных понятий.

Уровнем l решетки формальных понятий будем называть формальные понятия (A, B) , такие, что $|A| = l$. При этом количество уровней составляет $l = 1, \dots, \max_{u \in U} |X(u)|$, поскольку в матрице инциденции нельзя найти полной матрицы, которая превосходит матрицу с числом строк $\max_{u \in U} |X(u)|$.

Для каждого уровня $l = 1, \dots, \max_{u \in U} |X(u)|$, процесс генерации состоит из двух этапов.

Этап 1. На первом этапе для заданного l строится множество всех сочетаний $C_X^l = \{S : S \subseteq X, |S| = l\}$. Согласно множеству C_X^l , формируется множество пар $P_l = \{(S, U(S)) : S \in C_X^l\}$.

Этап 2. Для заданного l производится преобразование множества P_l . Из множества P_l удаляются пары с пустой правой частью $|U(S)| = \emptyset$, а также все пары, у которых совпадают правые части $U(S)$. Очевидно, что такие пары не являются формальными понятиями, поскольку могут быть включены в другую максимально полную подматрицу.

Следующим шагом алгоритма является объединение всех полученных множеств P_l в единое множество $P = \bigcup_{l=1, \dots, \max_{u \in U} |X(u)|} P_l$, при этом мно-

жество P упорядочивается согласно лексикографическому порядку и размеру $|S|$ по убыванию. Таким образом, множество P состоит из пар $(S, U(S))$, которые имеют уникальные непустые правые части $U(S)$ для каждого уровня $|S| = l$.

Далее сформируем множество формальных понятий FC следующим образом. Последовательно просматривая множество P , будем добавлять в FC только такие пары $(S, U(S))$, для которых $U(S)$ не совпадает с правой частью ранее добавленных пар. Очевидно, что полученное множество FC содержит все значащие формальные понятия.

Отметим, что в случае, когда $\max_{u \in U} |X(u)| \geq \max_{x \in X} |U(x)|$ – максимальная степень вершины меньше максимальной степени ребра, алгоритм HGFC можно применять к двойственному гиперграфу H^* с целью уменьшения количества генераций.

Результаты вычислительных экспериментов

Предложенный алгоритм HyperGraphFormalConcept (HGFC) сравнивался с известным алгоритмом нахождения множества всех формальных понятий Close-By-One (CbO). Исходя из описания HGFC очевидно, что узким местом алгоритма является генерация всех сочетаний, поэтому вычислительные эксперименты проводятся на сильно разреженных гиперграфах с фиксированной степенью вершин. Такие гиперграфы возникают при анализе текстов и обработке телекоммуникационных сетей [5, 6].

Были проведены несколько серий экспериментов, в каждой из которых генерировались случайные гиперграфы с фиксированной степенью вершин равной 4. Такой подход к генерации использовался в работе [7]. Размеры множества вершин гиперграфов варьировались от 100 до 50 000 вершин. Размеры множества гиперребер принимали значения 100, 200 и 500 соответственно. В таблице, на пересечении строк и столбцов параметров, приведены средние значения процессорного времени, требуемого для получения множества всех формальных понятий FC .

Результаты сравнения по процессорному времени алгоритмов HGFC и CbO

Значения параметров		$ X = 100$	$ X = 1000$	$ X = 10000$	$ X = 50000$
$ U = 100$	<i>HGFC</i>	1.1	47.3	2655	48049.75
	<i>CbO</i>	4.3	421.2	50486.3	2376248
$ U = 200$	<i>HGFC</i>	2	62.3	3993.4	92344.5
	<i>CbO</i>	4.2	278.2	43881.5	946274
$ U = 500$	<i>HGFC</i>	1.3	70.2	6073.7	198608.5
	<i>CbO</i>	3.2	252.5	26219.2	1007647

Эксперименты проводились на ПК с процессором Intel® Core™ i7-7700K 4.20 ГГц и оперативной памятью объемом 16 ГБ. Один процессорный такт занимает 0.001 с. Из таблицы видно, что при увеличении числа вершин гиперграфа алгоритм HGFC дает значительное преимущество перед алгоритмом CbO.

Однако, как показали эксперименты, не приведенные в данной работе, при случайном генерировании гиперграфа без фиксированной степени вершины, предложенный алгоритм HGFC проигрывает известному алгоритму CbO. Это связано с этапом генерации всех сочетаний.

Заключение

Предложенный в работе алгоритм нахождения всех формальных понятий дает значительный выигрыш по времени для частного случая формальных контекстов и соответствующих им гиперграфов. Дальнейшее исследование алгоритма позволит установить точные границы его применимости. Данные исследования перспективны, поскольку данный частный случай имеет тесную связь с анализом текстов и анализом классических графов.

СПИСОК ЛИТЕРАТУРЫ

1. *Ganter B., Wille R.* Formal Concept Analysis. Springer, 1999. 284 p.
2. *Тарканов В.Е.* Комбинаторные задачи и (0, 1)-матрицы. М.: Наука, 1985. 192 с.
3. *Емеличев В.А., Мельников О.И., Сарванов В.И., Тышкевич Р.И.* Лекции по теории графов. М.: Книжный дом «Либроком», 2012. 390 с.
4. *Быкова В.В., Монгуш Ч.М.* Декомпозиционный подход к исследованию формальных контекстов // ПДМ. 2019. № 44. С. 113–126.
5. *Dengyong Z., Jiayuan H., Schölkopf B.* Learning with Hypergraphs: Clustering, Classification, and Embedding // Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference. 2007. V. 19. P. 1601–1608.
6. *Hongseon Y., Youngjoong K., Jungyun. S.* Unsupervised-learning-based keyphrase extraction from a single document by the effective combination of the graph-based model and the modified C-value method // Computer Speech & Language. 2019. V. 58. P. 304–318.
7. *Kuznetsov S., Obiedkov S.* Comparing performance of algorithms for generating concept lattices // J. Exp. Theor. Artif. Intell. 2002. V. 14. P. 189–216.

СВЕДЕНИЯ ОБ АВТОРАХ

Achyutha Krishnamoorthy, Professor, Department of Mathematics, CMS College, Kottayam, Cochin University of Science and Technology, Kochi, India, achyuthacusat@gmail.com

Adou Kpangny Yves Berenger, PhD student, Peoples' Friendship University of Russia, Moscow, Russia, 1042205051@rudn.university

Ageev Kirill, PhD student, Peoples' Friendship University of Russia, Moscow, Russia, kaageyev@gmail.com

Aliyeva Sevinj, Baku State University, Baku, Azerbaijan, s@aliyeva.info

Ambily P. Mathew, Assistant Professor, Department of Mathematics, CMS College, Kottayam, India, ambilypm@cmscollege.ac.in

Anilkumar M P., Assistant Professor, Dept. of Mathematics, T. M. Govt. College, Tirur-676502, Kerala, India, anilkumarm77@gmail.com

Beena P., Assistant Professor, P Govt. Sanskrit College, Pattambi, Palakkad-679303, Kerala, India, beenapathari@gmail.com

Begre Mary Michel, PhD student, Peoples' Friendship University of Russia, Moscow, Russia, 1042195017@rudn.ru

Daraseilya Anastasia PhD student, Peoples' Friendship University of Russia, Moscow, Russia, avdaraseilya@sci.pfu.edu.ru

Dhanya Babu, Department of Mathematics, CMS College, Kottayam, India, dhanyababu@cmscollege.ac.in

Divya V., Assistant Professor, N.S.S. College, Cherthala-688556, Kerala, India, divyavelayudhannair@gmail.com

Gaidamaka Yuliya, Professor, Department of Mathematics and Informatics, Peoples' Friendship University of Russia, Moscow, Russia, gaidamaka_yuv@rudn.ru

Gopakumar B. Government Engineering College, Thrissur, India, bhaskarangopakumar@gmail.com

Jose K. P. Associate Professor, Dept.of Mathematics, St. Peter's College, Kolenchery-682311, Kerala, India, kpjpsc@gmail.com

Khamis Abdullah Khamis AL Maqbali, Department of Mathematics, CMS College Kottayam, Kerala, India, khamis@cmscollege.ac.in

Kolosova Irina, PhD student, Peoples' Friendship University of Russia, Moscow, Russia, 1042160102@rudn.ru

Korshikov Maxim PhD student, Peoples' Friendship University of Russia, Moscow, Russia, 1042200075@rudn.university

Korshok Evgeniya, PhD student, Peoples' Friendship University of Russia, Moscow, Russia, 1042150139@rudn.ru

Lebedev Eugene, Head of Department of Applied Statistics, Full Professor, Faculty of Computer Science and Cybernetics, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, leb@unicyb.kiev.ua

Lisovskaya Ekaterina, junior researcher, Peoples' Friendship University of Russia, Moscow, Russia, lisovskaya-eyu@rudn.ru

Markova Ekaterina, Associate Professor, Peoples' Friendship University of Russia, Moscow, Russia, markova_ev@pfur.ru

Melikov Agassi, Professor, Head of Department of Teletraffic Theory, Institute of Control Systems, Baku, Azerbaijan, agassi.melikov@gmail.com

Moskaleva Faina, PhD student, Peoples' Friendship University of Russia, Moscow, Russia, moskaleva-fa@rudn.ru

Nemouchi Hamza, PhD student, Department of Computer Science, University of Debrecen, Debrecen, Hungary, nemouchih@gmail.com

Nisha Mathew, Department of Mathematics, B.K College Amalagiri, Kottayam, India, nishatmathew@gmail.com

Pichugin Sergey, Rocket and Space Corporation Energia named by S.P. Korolev, Korolyov, Russia, sergey.pichugin@rsce.ru

Ponomarov Vadim, Junior Scientific Researcher, PhD Taras Shevchenko National University of Kyiv, Kyiv, Ukraine, vponomarov@gmail.com

Pryshchepa Oksana, National University of Water and Environmental Engineering, Rivne, Ukraine, o.v.pryshchepa@nuwm.edu.ua

Ranjith K R, Government Engineering College, Thrissur, India, kr.ranjith3@gmail.com

Sajeev Nair, Government Engineering College, Thrissur, India, sajeev@gectcr.ac.in

Sandhya E, Research Scholar, Government College, Chittur, University of Calicut, India, esandhya1729@gmail.com

Smija Skaria, Research Scholar, Government Engineering College, Thrissur, APJ Abdul Kalam Technological University, Kerala, India, smijashano@gmail.com

Sopin Eduard, Assistant Professor, Peoples' Friendship University of Russia, Moscow, Russia, sopin_es@pfur.ru

Sreenivasan C, Government Victoria College, Palakad, India, sreenikoonathara@gmail.com

Sztrik János, Full Professor, Department of Informatics Systems and Networks, University of Debrecen, Debrecen, Hungary, sztrik.janos@inf.unideb.hu

Takhirov Jozil, Professor, Institute of Mathematics, Tashkent, Uzbekistan, prof.takhirov@yahoo.com

Tóth Ádám, PhD student, University of Debrecen, Faculty of Informatics, Debrecen, Hungary, toth.adamg@inf.unideb.hu

Tuan Phung-Duc, Associate Professor, College of Policy and Planning Sciences, Tsukuba, Japan, tuan@sk.tsukuba.ac.jp
Varghese C. Joshua, Associate Professor, Department of Mathematics, CMS College, Kottayam, India, vcjcms@gmail.com
Vasilyeva Darya, PhD student, Peoples' Friendship University of Russia, Moscow, Russia, 1032202191@rudn.ru
Warren Kassy Dougg Feussi, PhD student, Peoples' Friendship University of Russia, Moscow, Russia, 1042195041@rudn.ru
Zaghouani Mohamed Hedi, University of Debrecen, Debrecen, Hungary, zaghoueni.hedi@gmail.com
Абдразакова Анжела Романовна, студентка, Новосибирский государственный университет, Новосибирск, Россия, abdrzakova96@inbox.ru
Акоста Даниэль Перес, аспирант, Российский университет дружбы народов, Москва, Россия, 1042185137@rudn.ru
Бабанов Алексей Михайлович, доцент, Томский государственный университет, Томск, Россия, babanov2000@mail.ru
Барабанова Елизавета Александровна, ведущий научный сотрудник, Институт проблем управления им. В.А. Трапезникова РАН, Москва, Россия, ElizavetaAlexB@yandex.ru
Баранова Ирина Владимировна, доцент, Сибирский федеральный университет, Красноярск, Россия, irinabar@yandex.ru
Батраева Инна Александровна, заведующая кафедрой, Саратовский национальный исследовательский государственный университет им. Н.Г. Чернышевского, Саратов, Россия, tlanaufp@yandex.ru
Беляков Никита, студент, Сибирский федеральный университет, Красноярск, Россия
Бикмурзин Рустам Васильевич, Сибирский федеральный университет, Красноярск, Россия, sfu_brvt@mail.ru
Благинин Алексей Леонидович, студент, Томский государственный университет, Томск, Россия, alex-b.l@yandex.ru
Бобкова Ольга Сергеевна, аспирантка, Томский государственный университет, Томск, Россия, osia153@yandex.ru
Борисовская Алена Алексеевна, Российский университет транспорта, Москва, Россия, borisovsk0101@gmail.com
Бояркина Анна Александровна, аспирантка, Томский государственный университет, Томск, Россия, boyarkina.93@inbox.ru
Брокарев Иван Андреевич, ассистент, РГУ нефти и газа (НИУ) имени И.М. Губкина, Москва, Россия, brokarev.i@gubkin.ru
Буатта Мохамед Адель, аспирант, Российский университет дружбы народов, Москва, Россия, 1042185006@rudn.ru

Булгакова Татьяна Евгеньевна, Новосибирский государственный университет, Новосибирск, Россия, tatyana.bulgakova@gmail.com

Бушкова Татьяна Владимировна, аспирантка, Томский государственный университет, bushkova70@mail.ru

Вавилов Вячеслав Анатольевич, доцент, Томский государственный университет, Томск, Россия, vavilovv@yandex.ru

Ванькович Иван, студент, Белорусский государственный университет, Минск, Беларусь, ivan.vankovich98@gmail.com

Васильев Сергей Анатольевич, доцент, Российский университет дружбы народов, Москва, Россия, vasilyev-sa@rudn.ru

Васьковский Сергей Владимирович, старший научный сотрудник, Институт проблем управления им. В.А. Трапезникова РАН, Москва, Россия, v63v@yandex.ru

Вишневский Владимир Миронович, Институт проблем управления им. В.А. Трапезникова РАН, Москва, Россия, vishn@inbox.ru

Войтиков Константин Юрьевич, доцент, Московский физико-технический университет, Москва, Россия, voytikovk@gmail.com

Войтишек Антон Вацлавович, ведущий научный сотрудник, Институт вычислительной математики и математической геофизики СО РАН, Новосибирск, Россия, anton.voytishek@yandex.ru

Воронина Наталья Михайловна, аспирантка, Томский политехнический университет, Томск, Россия, vnm@tpu.ru

Вытовтов Константин Анатольевич, ведущий научный сотрудник, Институт проблем управления им. В.А. Трапезникова РАН, Москва, Россия, vytovtov_konstan@mail.ru

Вячеслав Федорченко, аспирант, Российский университет дружбы народов, Москва, Россия, 1042165109@rudn.ru

Гагарин Юрий Евгеньевич, заведующий кафедрой, Калужский филиал МГТУ им. Н.Э. Баумана, Калуга, Россия, g_ug@mail.ru

Гагарина Светлана Николаевна, доцент, Калужский государственный университет им. К.Э. Циолковского, Калуга, Россия, g_ug@mail.ru

Гайнутдинов Марат Фаридович, студент, Томский государственный университет, Томск, Россия, marat.gainutdinov.main@gmail.com

Галилейская Анастасия Александровна, аспирантка, Томский государственный университет, Томск, Россия, n.galileyskaya@bk.ru

Гендрина Ирина Юрьевна, доцент, Томский государственный университет, Томск, Россия, igendrina@bk.ru

Гилин Степан Валентинович, аспирант, Сибирский федеральный университет, Красноярск, Россия, gilin.stepan@mail.ru

Гольшев Валерий Константинович, студент, Сибирский федеральный университет, Красноярск, Россия, valeriygolyshev@mail.ru

Данилюк Елена Юрьевна, доцент, Томский государственный университет, Томск, Россия, daniluc.elena.yu@gmail.com

Демиденко Людмила Руслановна, студентка, Томский государственный университет, Томск, Россия, mikamikadem@gmail.com

Дудин Александр Николаевич, профессор, Белорусский государственный университет, Минск, Беларусь dudin@bsu.by

Есин Антон Анатольевич, Институт проблем управления им. В.А. Трапезникова Российской академии наук, Москва, Россия, ae@incarnet.ru

Жарков Максим Леонидович, Институт динамики систем и теории управления имени В.М. Матросова, Иркутск, Россия, zharkm@mail.ru

Жуков Игорь Андреевич, студент, Томский государственный университет, Томск, Россия, igzhukov963@yandex.ru

Задорожный Владимир Николаевич, профессор, Омский государственный технический университет, Омск, Россия, zwn2015@yandex.ru

Захаренкова Татьяна Романовна, Омский государственный технический университет, Омск, Россия, zakharenkovatatiana@gmail.com

Змеев Денис Олегович, ассистент, Томский государственный университет, Томск, Россия, denis.zmееv@accounts.tsu.ru

Змеев Олег Алексеевич, проректор по цифровой трансформации, Томский государственный университет, Томск, Россия, ozmееv@mail.tsu.ru

Зорин Андрей Владимирович, профессор, Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия, zoav1602@gmail.com

Ибрагимова Эллада Ибрагимовна, студентка, Сибирский федеральный университет, Красноярск, Россия, ibragimowa_lada@mail.ru

Иванова Лидия Сергеевна, ассистент, Томский государственный университет, Томск, Россия, lidiya.ivanova@persona.tsu.ru

Ивлева Анна Игоревна, старший научный сотрудник, Институт проблем управления сложными системами Российской академии наук, Самара, Россия, annushka199@bk.ru

Измайлова Яна Евгеньевна, доцент, Томский государственный университет, Томск, Россия, evgenevna.92@mail.ru

Калимулина Эльмира Юрьевна, старший научный сотрудник ИПУ РАН, Москва, Россия, elmira.yu.k@gmail.com

Канзитдинов Шахмурад, аспирант, Российский университет дружбы народов, Москва, Россия, 1042180082@rudn.ru

Качалов Дмитрий Леонидович, Институт проблем управления им. В.А. Трапезникова РАН, Москва, Россия, comdcompdim@mail.ru

Квач Елена Сергеевна, аспирантка, Томский государственный университет, Томск, Россия, kvachelena93@gmail.com

Кетова Татьяна Семеновна, преподаватель, Томский государственный университет, Томск, Россия, kts1966@mail.ru

Клименок Валентина Ивановна, профессор, Белорусский государственный университет, Минск, Беларусь, vklimenok@yandex.ru

Ключникова Полина Николаевна, студентка, Томский государственный университет, Томск, Россия, polya.kluch@yandex.ru

Копать Дмитрий Ярославович, аспирант, Гродненский государственный университет, Гродно, Беларусь, dk80395@mail.ru

Костюк Юрий Леонидович, профессор, Томский государственный университет, Томск, Россия, kostyuk_y_1@sibmail.com

Крючкова Александра Андреевна, студентка, Саратовский национальный исследовательский государственный университет им. Н.Г. Чернышевского, Саратов, Россия, kruchkovaaleksandra@yandex.ru

Кувыкин Вячеслав Иванович, доцент, Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия, mmer@iee.unn.ru

Кузоватова Ольга Игоревна, доцент, Сибирский федеральный университет, Красноярск, Россия, OKuzovatova@sfu-kras.ru

Куликова Яна Владимировна, аспирантка, Волгоградский государственный технический университет, Волгоград, Россия, cul.jana@yandex.ru

Лапатын Иван Леонидович, доцент, Томский государственный университет, Томск, Россия, ilapatin@mail.ru

Лебедева Яна Евгеньевна, студентка, Томский государственный университет, Томск, Россия, jana.lebedeva@go.ru

Левичев Игорь, аспирант, Российский университет дружбы народов, Москва, Россия, 1042190197@rudn.ru

Лизюра Ольга Дмитриевна, аспирантка, Томский государственный университет, Томск, Россия, oliztsu@mail.ru

Маталыцкий Михаил Алексеевич, профессор, Гродненский государственный университет, Гродно, Беларусь, m.matalytski@gmail.com

Моисеев Александр Николаевич, заведующий кафедрой, Томский государственный университет, Томск, Россия, moiseev.tsu@gmail.ru

Моисеева Светлана Петровна, профессор, Томский государственный университет, Томск, Россия, smoiseeva@mail.ru

Морозов Евсей Викторович, профессор, Институт прикладных математических исследований КарНЦ РАН, Петрозаводск, Россия, emorozov@karelia.ru

Назаров Анатолий Андреевич, заведующий кафедрой теории вероятностей и математической статистики, Томский государственный университет, Томск, Россия, nazarov.tsu@gmail.com

Павидис Михаил Максимович, аспирант, Иркутский государственный университет путей сообщения, Иркутск, Россия, pavidismiha1994@mail.ru

Павлова Екатерина Алексеевна, аспирантка, Томский государственный университет, Томск, Россия, pavlovakaty_2010@mail.ru

Пагано Микеле, профессор, Пизанский Университет, Пиза, Италия, m.pagano@iet.unipi.it

Пауль Светлана Владимировна, доцент, Томский государственный университет, Томск, Россия, paulsv82@mail.ru

Пичугин Сергей Борисович, ведущий научный сотрудник, ПАО «Ракетно-космическая корпорация “Энергия” им. С.П. Королева», Королев, Россия, sergey.pichugin@rsce.ru

Пичугина Анастасия Александровна, преподаватель, Томский государственный университет, Томск, Россия, aapichugina@gmail.com

Полин Евгений Павлович, ассистент, Томский государственный университет, Томск, Россия, polin_evgeny@mail.ru

Полховская Анна Васильевна, студентка, Томский государственный университет, Томск, Россия, anua.polhovskaya00@mail.ru

Постовалов Ярослав Сергеевич, учащийся, Лицей № 130 города Новосибирска, Новосибирск, Россия, postovalovya@gmail.com

Приступа Павел Викторович, аспирант, Томский государственный университет, Томск, Россия, pristupa@gmail.com

Пройдакова Екатерина Вадимовна, доцент, Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия, rev_1@mail.ru

Протасевич Юлия Алексеевна, студентка, Томский государственный университет, Томск, Россия, protasevich.yuliya@gmail.com

Рафикова Руфина Рамильевна, студентка, Томский государственный университет, Томск, Россия, rafikova0rr@gmail.com

Рогозин Степан Сергеевич, преподаватель, Петрозаводский государственный университет, Петрозаводск, Россия, рреха@mail.ru

Рожкова Светлана Владимировна, профессор, Томский политехнический университет, Томск, Россия, rozhkova@tpu.ru

Самородова Мария Владимировна, аспирантка, Томский государственный университет, Томск, Россия, samorodova21@gmail.com

Санникова Виктория Олеговна, Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, v.sannikova@yandex.ru

Семашко Александра Викторовна, студент, Томский государственный университет, Томск, Россия, AlexandraSV18@mail.ru

Семенова Дарья Владиславовна, доцент, Сибирский федеральный университет, Красноярск, Россия, dariasdv@gmail.com

Сизова Ксения Олеговна, Нижегородский государственный университет им. Н.И. Лобачевского, Нижний Новгород, Россия, sizovaksuniy@yandex.ru

Смирнов Иван Валентинович, главный научный сотрудник, Институт проблем управления сложными системами РАН, Самара, Россия, smirnov@iccs.ru

Соколов Данила Александрович, ассистент, студент, Томский государственный университет, Томск, Россия, danila.sokolov@accounts.tsu.ru

Соколова Евгения Валерьевна, студентка, Томский государственный университет, Томск, Россия, evgenia.596@mail.ru

Солдатенко Александр Александрович, аспирант, Сибирский федеральный университет, Красноярск, Россия, asoldatenko@sfu-kras.ru

Станкевич Елена Петровна, доцент, Саратовский национальный исследовательский государственный университет им. Н.Г. Чернышевского, Саратов, Россия, stankevichelena@mail.ru

Сущенко Сергей Петрович, профессор, заведующий кафедрой, Томский государственный университет, Томск, Россия, ssp.inf.tsu@gmail.com

Тананко Игорь Евстафьевич, заведующий кафедрой, Саратовский национальный исследовательский государственный университет им. Н.Г. Чернышевского, Саратов, Россия, tanankoie@info.sgu.ru

Титаренко Екатерина Юрьевна, аспирантка, Томский политехнический университет, Томск, Россия, teu@tpu.ru

Туренова Ирина Алексеевна, доцент, Томский государственный университет, Томск, Россия, irenaturena@mail.ru

Фархадов Манс Паша оглы, заведующий лабораторией, Институт проблем управления им. В.А. Трапезникова РАН, Москва, Россия, mais@ipu.ru

Федченко Дмитрий Петрович, доцент, Сибирский федеральный университет, Красноярск, Россия, DFedchenko@sfu-kras.ru

Цициашвили Гурами Шалвович, главный научный сотрудник, Институт прикладной математики ДВО РАН, Владивосток, Россия, guram@iam.dvo.ru

Черкашин Данил Андреевич, учащийся, Лицей № 130 города Новосибирска, Новосибирск, Россия, cherkashin.daniel@gmail.com

Шамраева Виктория Викторовна, доцент, Финансовый университет при Правительстве России, Москва, Россия, vvshamraeva@fa.ru

Шкленник Мария Александровна, ассистент, Томский государственный университет, Томск, Россия, shklennikm@yandex.ru

Шкуркин Алексей Сергеевич, доцент, Томский государственный университет, Томск, Россия, shkurkin@mail.ru

Шульгина Ксения Сергеевна, студентка, Томский государственный университет, Томск, Россия, shulgina19991999@mail.ru

СОДЕРЖАНИЕ

ТЕОРИЯ МАССОВОГО ОБСЛУЖИВАНИЯ И ЕЕ ПРИЛОЖЕНИЯ

QUEUING THEORY AND APPLICATION

<i>Yves Adou, Ekaterina Markova.</i> To queueing system model performance measures analysis under network slicing.....	5
<i>Kirill Ageev, Eduard Sopin.</i> Analysis of the simplified network slicing model	11
<i>Anilkumar M.P., K.P. Jose.</i> An eigen value approach to a discrete-time queueing model with N -policy on two modes of service.....	17
<i>P. Beena, K.P. Jose.</i> A MAP/PH(1), PH(2)/2 inventory system with production, multiple servers and vacations.....	24
<i>Anastasia Daraseliya, Eduard Sopin.</i> Optimization of task offloading thresholds in the fog computing system	31
<i>Dhanya Babu, Varghese. C. Joshua, Achyutha Krishnamoorthy.</i> A queueing system with probabilistic joining strategy for priority customers	37
<i>Elmira Kalimulina.</i> On convergence of queueing network with changing structure to stationary distribution.....	43
<i>Maksim Korshikov, Eduard Sopin.</i> Analysis of the processor sharing systems with random serving rate coefficients	46
<i>Achyutha Krishnamoorthy, Varghese C. Joshua, Ambily P. Mathew.</i> A reliability problem with Interdependent Lifetimes	52
<i>Eugene Lebedev, Vadim Ponomarov, Oksana Pryshchepa.</i> The exact formulas for state-dependent Markov retrial queues	58
<i>Eugene Lebedev, Hanna Livinska.</i> Gaussian approximation and reducing of dimension for a general-type multichannel network	64
<i>Khamis Abdullah Khamis AL Maqbali, Varghese C. Joshua, Achyutha Krishnamoorthy.</i> On A single server queueing inventory system with common life time for inventoried items	70
<i>Agassi Melikov¹, V. Divya, Sevinc Aliyeva.</i> Analyses of feedback queue with positive server setup time and impatient calls.....	77
<i>Faina Moskaleva, Ekaterina Lisovskaya, Yuliya Gaidamaka.</i> A two-class service system for performance analysis of network slicing with QoS Isolation	82

<i>Anatoly Nazarov, Tuan Phung-Duc, Yana Izmailova. Asymptotic-diffusion analysis of multiserver retrial queueing system with priority customers.....</i>	88
<i>Anatoly Nazarov, Tuan Phung-Duc, Svetlana Paul, Olga Lizyura, Ksenia Shulgina. Asymptotic analysis of Markovian retrial queue with unreliable server and two-way communication under low rate of retrials condition</i>	99
<i>Anatoly Nazarov, Maria Samorodova. Asymptotic waiting time analysis of a M/M/1 retrial queueing system</i>	105
<i>Hamza Nemouchi, Mohamed Hedi Zaghouni, János Sztrik. Simulation analysis in cognitive radio networks with unreliability and abandonment.....</i>	110
<i>Nisha Mathew, Varghese Joshua, Achyutha Krishnamoorthy. On a MMAP/(PH,PH)/1/(∞,N) queueing-inventory system.....</i>	115
<i>K.R. Ranjith, Achyutha Krishnamoorthy, B. Gopakumar. Analysis of a PH/PH/1 queue with interdependence.....</i>	122
<i>Stepan Rogozin, Evsey Morozov. Stability condition of a modified Erlang loss system with different service rates</i>	126
<i>Sandhya E., C. Sreenivasan, Sajeev S. Nair. An explicit solution for an inventory model with positive lead time and backlogs.....</i>	131
<i>Smija Skaria, Sajeev S. Nair. Transient analysis of an inventory model with instantaneous replenishment and catastrophes</i>	138
<i>János Sztrik, Ádám Tóth, Elena Danilyuk, Svetlana Moiseeva. Simulation of retrial queueing system M/G/1 with impatient customers, collisions and unreliable server</i>	145
<i>János Sztrik, Ádám Tóth. Some special features of finite-source retrial queues with collisions, an unreliable server and impatient customers in the orbit</i>	152
<i>Алексей Благинин, Иван Лапатын, Анатолий Назаров. Исследование двумерного выходящего потока марковской модели узла обработки запросов с повторными обращениями и вызываемыми заявками</i>	159
<i>Анна Бояркина, Светлана Моисеева, Ирина Туренова, Алексей Шкуркин. СМО вида $GI^{(k)}/GI/\infty$ с групповым обслуживанием.....</i>	166
<i>Татьяна Бушкова, Анастасия Галилейская Екатерина Лисовская, Светлана Моисеева. Асимптотический анализ ресурсной гетерогенной СМО $(MMPP+2M)^{(v)}/M/\infty$</i>	172
<i>Константин Вытовтов, Елизавета Барабанова, Владимир Вишневецкий. Аналитический метод анализа случайных процессов с</i>	

непрерывным временем и дискретными состояниями при времязависимых вероятностях переходов.....	178
<i>Максим Жарков, Михаил Пavidис.</i> Об использовании четырех- фазных систем массового обслуживания для описания работы грузовых и сортировочных железнодорожных станций.....	184
<i>Владимир Задорожный, Татьяна Захаренкова.</i> Метод бесконеч- ных разметок в системах с неизвестным временем обслужива- ния поступающих заявок	188
<i>Владимир Задорожный, Микеле Пагано, Татьяна Захаренкова.</i> Применение метода бесконечных разметок к сетям с коммута- цией пакетов.....	194
<i>Андрей Зорин, Ксения Сизова.</i> Метод решения стационарных уравнений для процесса приоритетного обслуживания с раз- делением времени в случайной среде.....	200
<i>Валентина Клименок, Александр Дудин, Иван Ванькович.</i> Стацио- нарные характеристики системы массового обслуживания с повторными вызовами и поиском на орбите	205
<i>Дмитрий Копать, Михаил Матальцкий.</i> Анализ ожидаемого до- хода в открытой сети с ограниченным числом заявок и обхо- дами ими систем обслуживания.....	211
<i>Анатолий Назаров, Екатерина Павлова.</i> Исследование СМО вида $MMPP M N$ с обратной связью методом асимптотически диф- фузионного анализа.....	217
<i>Анатолий Назаров, Светлана Рожкова, Екатерина Титаренко.</i> Исследование системы с обратной связью, рекуррентным об- служиванием и неординарным пуассоновским входящим по- током.....	223
<i>Анна Полховская, Ольга Бобкова, Светлана Моисеева.</i> Ресурсная RQ-система с коллизиями.....	228
<i>Павел Приступа, Павел Михеев, Сергей Суценко.</i> Прямая кор- рекция ошибок на внутрисегментном уровне транспортного протокола.....	232
<i>Екатерина Пройдакова, Виктория Санникова.</i> Математическое моделирование и исследование приоритетной управляющей системы с непостоянной интенсивностью обслуживания тре- бований	238
<i>Светлана Рожкова, Наталья Воронина, Александра Семашко.</i> Исследование RQ-системы $M/M/1$ с ненадежным прибором асимптотическим и матричным методами	244

<i>Елена Станкевич, Игорь Тананко.</i> Метод анализа замкнутых сетей массового обслуживания с системами типа $M_a/M^{[x,y]}/1$	251
<i>Елена Станкевич, Игорь Тананко.</i> Приближенный метод анализа замкнутых сетей массового обслуживания с ненадежными системами и групповым обслуживанием.....	255
<i>Гурами Цициашвили, Анатолий Назаров, Александр Мусеев.</i> Асимптотическая оценка интенсивности сборки пуассоновских потоков	258

ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ПРОГРАММНАЯ ИНЖЕНЕРИЯ

INFORMATION TECHNOLOGIES AND SOFTWARE ENGINEERING

<i>Marat Gainutdinov, Aleksey Shkurkin, Anastasia Pichugina.</i> Development of back-end of the service for internationalization of web-applications.....	265
<i>Алексей Бабанов, Елена Квач.</i> «IS-A»-отношение, как способ представления взаимосвязи обобщенных и специализированных понятий.....	271
<i>Людмила Демиденко.</i> Проектирование базовой архитектуры модуля «Расписание» системы Alterum Med.....	278
<i>Игорь Жуков, Юрий Костюк.</i> Программная реализация заданий по программированию с многовариантными решениями.....	285
<i>Денис Змеев, Лидия Иванова, Руфина Рафикова.</i> О представлении прогресса проекта по разработке программного обеспечения в форме динамической байесовской сети	291
<i>Олег Змеев, Юлия Протасевич, Данила Соколов.</i> Поддержка настраиваемых типов проектов в системе автоматизации управления Git-репозиториями для использования в процессе обучения.....	298
<i>Татьяна Кетова, Евгения Соколова.</i> Формальная модель образовательной программы в области компьютерных наук с точки зрения международного стандарта АСМ и IEEE	303
<i>Яна Куликова, Дмитрий Качалов, Маис Паша Оглы Фархадов.</i> Сценарии управления беспилотными транспортными средствами в среде «Умного города».....	308
<i>Яна Лебедева, Вячеслав Вавилов.</i> Разработка системы автоматизации процессов обращения кассовой техники в банковской организации	314

<i>Евгений Полин, Александр Моисеев, Константин Войтиков.</i> Имитационное моделирование СМО с входящими потоками, параметры которых зависят от состояния системы.....	320
<i>Вадим Тренькаев.</i> Обзор исследований по проблеме достижения высокой производительности протокола OPC UA.....	324

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ ТЕХНОЛОГИЧЕСКИХ ПРОЦЕССОВ

MATHEMATICAL AND COMPUTER MODELLING OF TECHNOLOGICAL PROCESSES

<i>Mary Michel Begre, Warren Kassy Dougg Feussi, Shakhmurad Kanzitdinov, Sergey Vasilyev.</i> Stability analysis and motion of the Kapitsa pendulum.....	330
<i>Mohamed Adel Bouatta, Irina Kolosova, Evgeniya Korshok, Darya Vasilyeva.</i> Kadshevsky equation numerical analysis with periodic boundary conditions on adaptive grids.....	336
<i>Jozil Takhirov.</i> A reaction-diffusion-advection competition model with a free boundary.....	339
<i>Sergey Pichugin.</i> Problem definition for LEO system switching technique development.....	345
<i>Анжела Абдразакова, Татьяна Булгакова, Антон Войтишек.</i> Об особенностях выбора ортонормированных систем функций в рандомизированных численных проекционных функциональ- ных алгоритмах.....	350
<i>Даниэль Перес Аюста, Сергей Васильев, Шахмурад Канзитди- нов, Игорь Левичев.</i> Построение решений задач оптимального управления динамическими системами в бесконечномерных пространствах с малым параметром.....	356
<i>Антон Войтишек, Ярослав Поставалов, Данил Черкашин.</i> Систе- ма численного моделирования одномерных случайных вели- чин NMPUD: формирование банка плотностей, автоматизация математических выкладок и приложения.....	363
<i>Мохамед Адель Буатта, Сергей Васильев, Вячеслав Федорченко.</i> Численный анализ на адаптивных сетках многомерного урав- нения Фоккера – Планка с малым параметром.....	369
<i>Никита Беляков, Рустам Бикмурзин, Дмитрий Федченко.</i> Об ис- пользовании конечных автоматов при моделировании наност- руктур.....	373

<i>Ирина Гендрина.</i> Использование метода фиктивных переменных для исследования пространственной характеристики систем видения через атмосферу.....	377
<i>Антон Есин.</i> Исследование принципов применения моделей многозначной логики в современных приложениях	383
<i>Антон Есин.</i> Теоретические аспекты построения современных систем управления на базе многозначной логики.....	389
<i>Вячеслав Кувыкин, Максим Брюханов.</i> Математическое и компьютерное моделирование системы согласования материального баланса в нефтепереработке и нефтехимии.....	394
<i>Вячеслав Кувыкин, Артем Колпаков, Елена Колпакова.</i> Параметрический анализ математических моделей оптимального планирования нефтепереработки и компьютерное моделирование.....	398
<i>Ольга Кузоватова.</i> Компьютерное моделирование локализации деформации сыпучей среды в сходящемся канале	403
<i>Мария Шкленник, Александр Мусеев.</i> Реализация механизма сбора и обработки статистических данных потоков заявок в системе имитационного моделирования ODIS.....	409

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ И ВИЗУАЛИЗАЦИЯ

INTELLIGENT DATA ANALYSIS AND VISUALIZATION

<i>Alyona Borisovskaya.</i> Methods of spelling correction in information retrieval systems	414
<i>Ivan Brokarev, Mais Farkhadov, Sergey Vaskovskii.</i> Recurrent neural networks to analyze the quality of natural gas	419
<i>Ivan Brokarev, Sergey Vaskovskii.</i> Analysis of reliability of gas analysis system based on vector Wiener process	423
<i>Victoria Shamraeva.</i> Analysis of business processes of construction and operation of highways on a toll basis using BIM tools.....	429
<i>Ирина Баранова.</i> Применение метода двудольных множеств событий в задачах регрессионного анализа многомерных разнотипных данных	441
<i>Инна Батраева, Александра Крючкова.</i> Алгоритм репрезентации кастомизированных диалектологических корпусов для Саратовского диалектологического корпуса русского языка	447
<i>Светлана Гагарина, Юрий Гагарин.</i> Прогнозирование частных показателей индекса активного долголетия	450

<i>Степан Гилин. Решение задачи распознавания образов при помощи алгоритма гибридной СММ-нейросети</i>	<i>454</i>
<i>Валерий Гольшев, Дарья Семенова. Нечёткий анализ формальных понятий: метод α-сечения.....</i>	<i>462</i>
<i>Эллада Ибрагимова, Дарья Семенова. Распознавание k-кластеризуемости знаковых графов.....</i>	<i>468</i>
<i>Анна Ивлева, Сергей Смирнов. Первичный концептуальный анализ сестринского дела для экспертной советующей системы.....</i>	<i>473</i>
<i>Александр Солдатенко, Дарья Семенова. Алгоритм HGFC нахождения формальных понятий.....</i>	<i>478</i>
СВЕДЕНИЯ ОБ АВТОРАХ	483

Научное издание

**ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ
И МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ
(ИТММ-2020)**

**МАТЕРИАЛЫ
XIX Международной конференции
имени А. Ф. Терпугова
2–5 декабря 2020 г.**

Редактор *Т.С. Портнова*
Дизайн, верстка *Д.В. Фортеса*

ООО «Издательство научно-технической литературы»
634034, г. Томск, ул. Студенческая, 4, тел. (3822) 53-10-35

Изд. лиц. ИД № 04000 от 12.02.2001. Подписано к печати 24.02.2021.
Формат 60 × 84 ¹/₁₆. Бумага офсетная. Печать офсетная. Гарнитура «Таймс».
Усл. п. л. 28.95. Уч.-изд. л. 32.42. Тираж 100 экз. Заказ № 4.
