Modeling the operation of a distributed high-load monitoring system for a data transmission network in a non-stationary mode

Kirill Shardakov[a], Vladimir Bubnov[a] and Svetlana Kornienko[a]

[a] *Emperor Alexander I St. Petersburg State Transport University, Moskovskiy avenue, 9, Saint-Petersburg, 190031, Russian Federation*

Abstract

The article discusses the numerical-analytical and simulation models of a high-load monitoring system. The examples of modeling are presented and the technical problem of choosing the hardware configuration of the simulated monitoring system is solved, which makes it possible to reduce the time spent by the task in the system by more than 2 times.

Keywords

Non-stationary queuing system, monitoring system, modeling, simulation model, numerical-analytical model

1. Introduction

Automated monitoring systems are an important part of any information system. Monitoring is a continuous process of observing and registering object parameters, processing them, and comparing them with threshold values. This monitoring system must cope with the increasing workload. Zabbix is the most popular and easily scalable for load adaptation free monitoring system for information systems [1].

Most often, the authors consider the stationary mode of operation of such systems in the context of queuing systems, however, it is the non-stationary mode of operation that is of greatest interest. The current state of the issue of non-stationary queuing systems is considered in more detail in [2]. The beginning of the “nonstationary” queueing theory was laid in [3–5] and continued in [6–7].

In [8-9], a model is proposed that allows one to simulate the behavior of such a monitoring system under various loads. The proposed model uses an improved recursive algorithm for generating a list of system states and a matrix of coefficients of an ODE system without constructing a graph of states and transitions of the non-stationary queueing system and deriving the general equation of the ODE system as in [10-13]. On the basis of a parallel-serial model, numerical-analytical [14] and simulation [15] models of a distributed high-load monitoring system for a data transmission network were developed and implemented.

2. Brief description of the simulated monitoring system

As it already known, the monitoring system under consideration allows you to distribute the load and scale it through the use of proxy servers. Each proxy server collects data from a separate set of devices, and then sends the data to the main server that processes it.

Figure 1 shows a general diagram of the interaction of system components. Such a system can also be considered as a queuing system. The queuing system representation is shown in Figure 2.

Models and Methods for Researching Information Systems in Transport, Dec. 11-12, St. Petersburg, Russia

EMAIL: k.shardakov@gmail.com (K.S. Shardakov); bubnov1950@yandex.ru (V.P. Bubnov); sv.diass99@yandex.ru (S.V. Kornienko);
ORCID: 0000-0002-6742-3011 (V.P.Bubnov); ORCID: 0000-0003-2683-0697 (S.V. Kornienko)

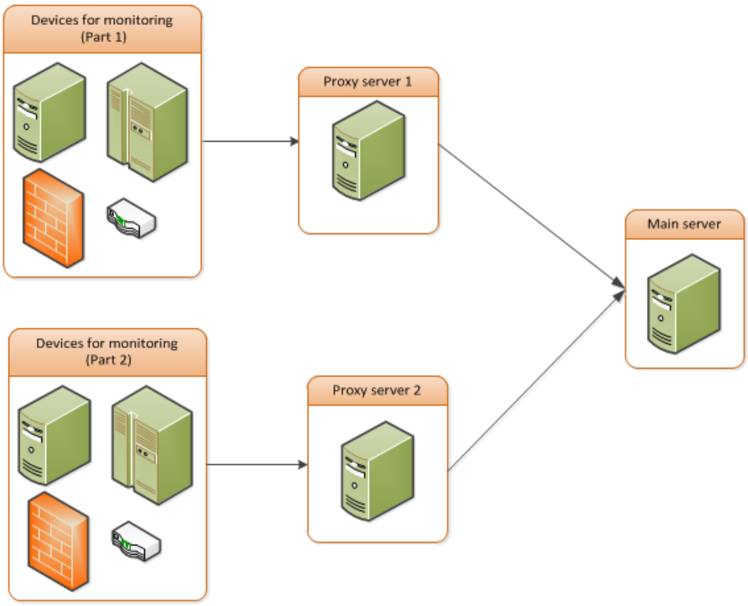

Figure 1: Simplified diagram of the interactions of the monitoring system components

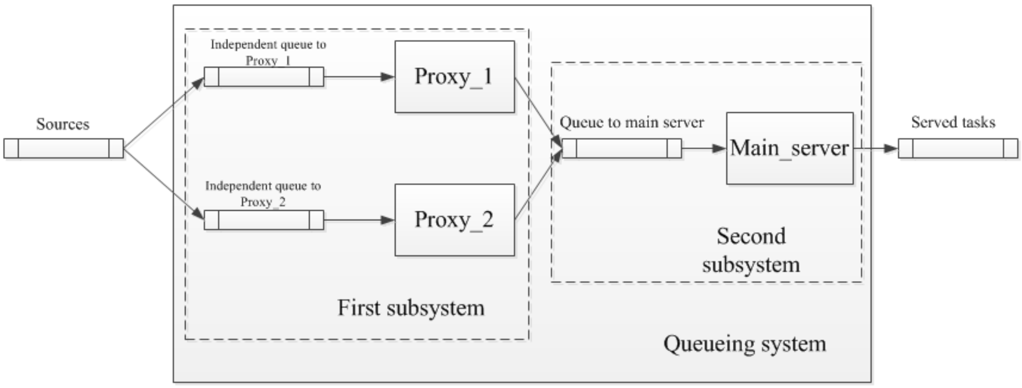

Figure 2: Simplified diagram of the parallel-serial non-stationary queueing system

3. Numerical-analytical modeling of work in a non-stationary mode

Numerical and analytical modeling was carried out using the software package [14]. The result of the operation of such a model is the calculated probabilities of all possible states of the system at given time moments.

Consider a model with the following initial inputs:

1. Amount of proxy servers – 2;
2. Amount of incoming tasks – 10;
3. Amount of calculated time moments, starting from zero – 30;
4. Intensity of tasks arriving – 3;
5. Intensity of processing tasks for proxy – 1;
6. Intensity of processing tasks for main server – 2.

The arrival intensity of tasks higher than the service rates allows one to simulate the non-stationary operation of the system.

As a result of the simulation, it was found that the model with such initial data generates 1001 possible states of the system. This leads to the need to compose and solve the system of Chapman-Kolmogorov differential equations with the same number of equations for calculating the probabilities of the states of the system at given time moments.

As you can see from Figure 3, after time moment 5, the probability of the onset of the absorbing state begins to increase sharply and by time moment 18 is practically equal to 1.

The obtained simulation results allow us to build graphs of the probability distribution of the states of the simulated system at each time moment, which allows us to consider in more detail the probability distribution at the turning points in time, which are clearly expressed in Figure 3. Figure 4 shows the probability distribution at time moment 0, as we can see, this is the initial state of the system and its probability is equal to one, while the probabilities of other states are equal to zero.

As can be seen from Figure 5, the probability distribution of states at time moment 5 allows us to say that at this time moment the final states have the highest probabilities, which indicate that new requests can arrive in the system with the least probability, and those that have already arrived will be processed with a higher probability.

Figure 6 allows us to say that at time moment 18 the absorbing state has the highest probability, while the probability of the system being in other states is negligible.

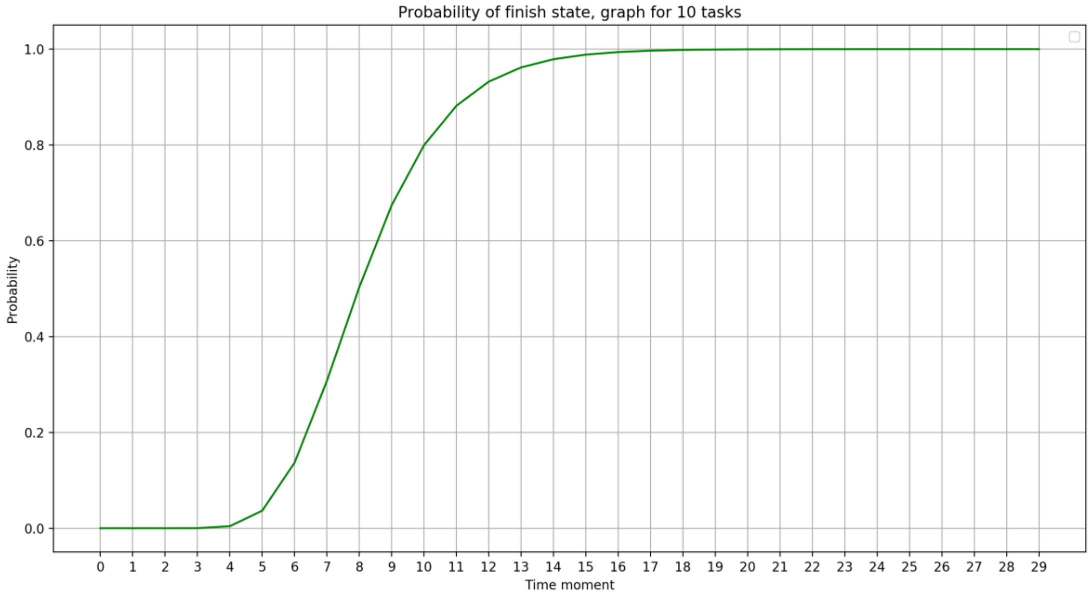

Figure 3: Absorbing state, 2 proxy servers, 10 tasks

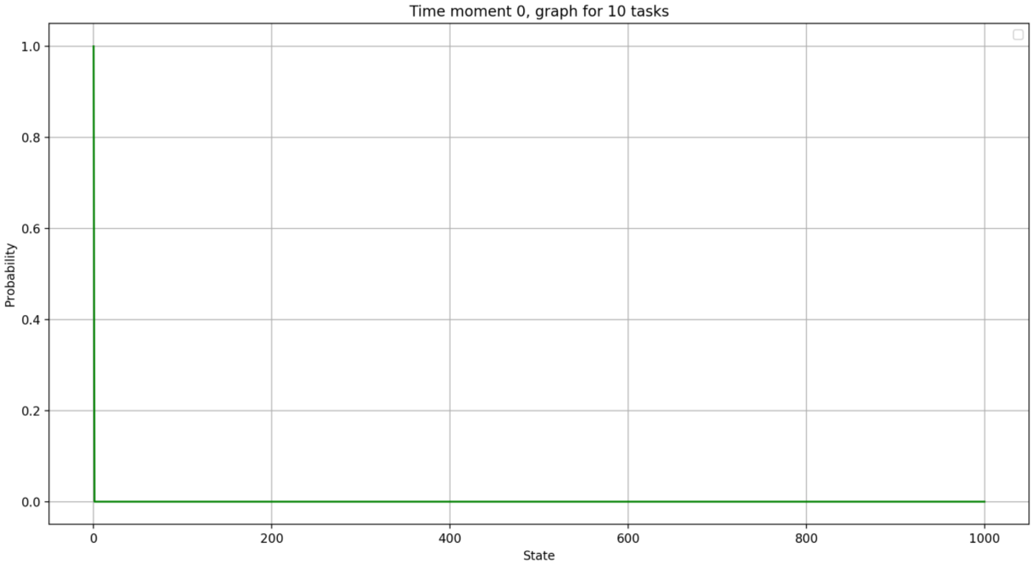

Figure 4: Time moment 0, 2 proxy servers, 10 tasks

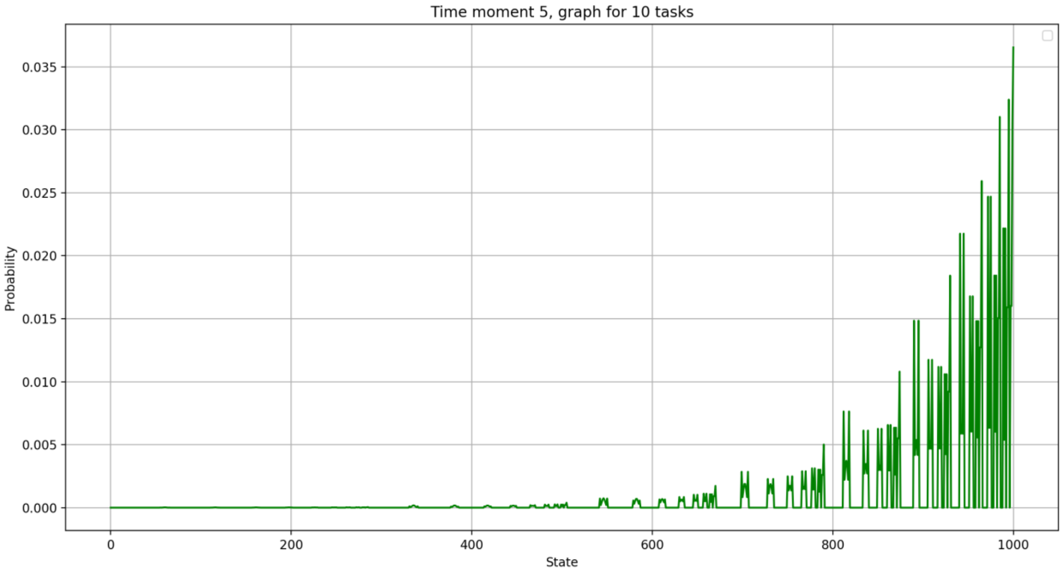

Figure 5: Time moment 5, 2 proxy servers, 10 tasks

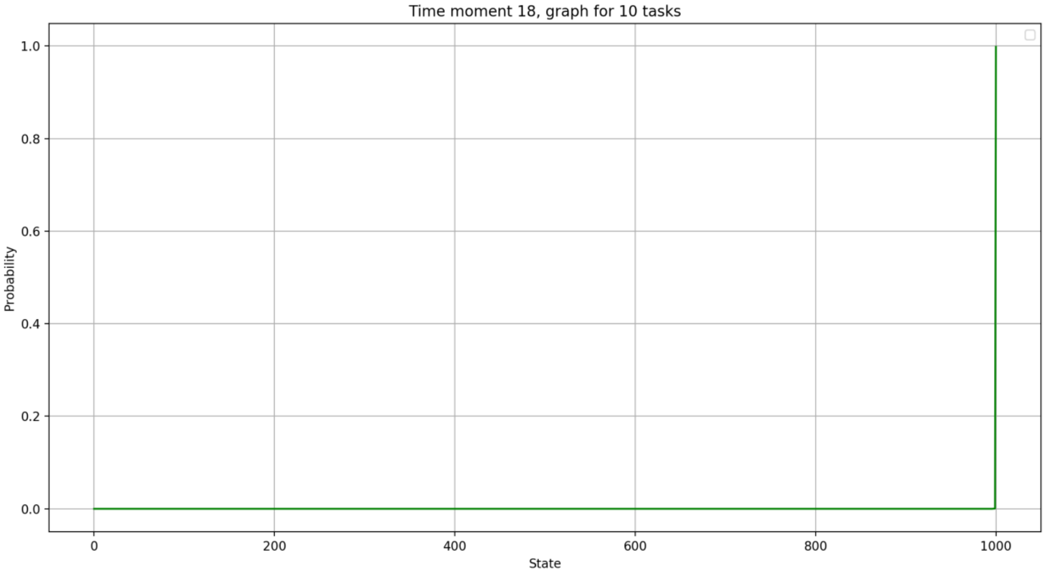

Figure 6: Time moment 18, 2 proxy servers, 10 tasks

With an increase in the number of proxy servers from 2 to 3, in Figure 7 we can see that the time moment with the maximum probability of the onset of the absorbing state shifted from 18 to 16, which, as expected, indicates a slightly higher throughput of such a system compared to the system in which there are 2 proxies.

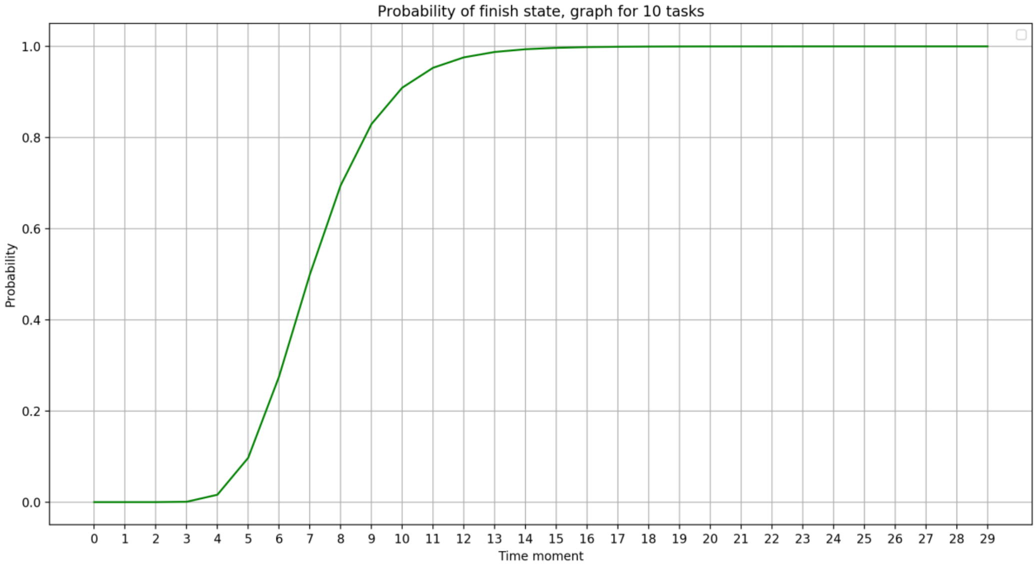

Figure 7: Absorbing state, 3 proxy servers, 10 tasks

So, the numerical-analytical model makes it possible to determine the probabilistic characteristics of the system, as well as to consider their changes dynamically at different time moments.

4. Simulation modeling

Simulation modeling was carried out using the software package [15]. The result of this model is the following statistical data for each application:

1. Time in the queue to the proxy server;
2. Service time on the proxy server;
3. Time in the queue to the main server;
4. Service time on the main server;
5. The total time spent by the task in the system.

Additionally, the model allows tracing the full path through the system state graph.

Consider a model with the following initial inputs:

1. Amount of proxy servers – 2;
2. Amount of incoming tasks – 1000;
3. Intensity of tasks arriving – 3;
4. Intensity of processing tasks for proxy – 1;
5. Intensity of processing tasks for main server – 2.

The arrival intensity of tasks higher than the service rates allows one to simulate the non-stationary operation of the system.

The path of such a model through the state graph was 3001 states. Figure 8 shows the number of tasks in queues to proxy servers and to the main server, as well as the number of tasks they have already served. Values are listed for each condition passed by the system. As we can see, the queue to proxy servers is constantly increasing until all requests has arrive to the system, after which the queue to proxy begins to decreasing rapidly. At the same time, the queue to the main server and the gap in the number of requests served by proxy servers and the main server is minimal.

In Figure 9, we can see that the residence time of each new request in the system increases up to 200 conventional units of time, which is obviously due to the low throughput of proxy servers.

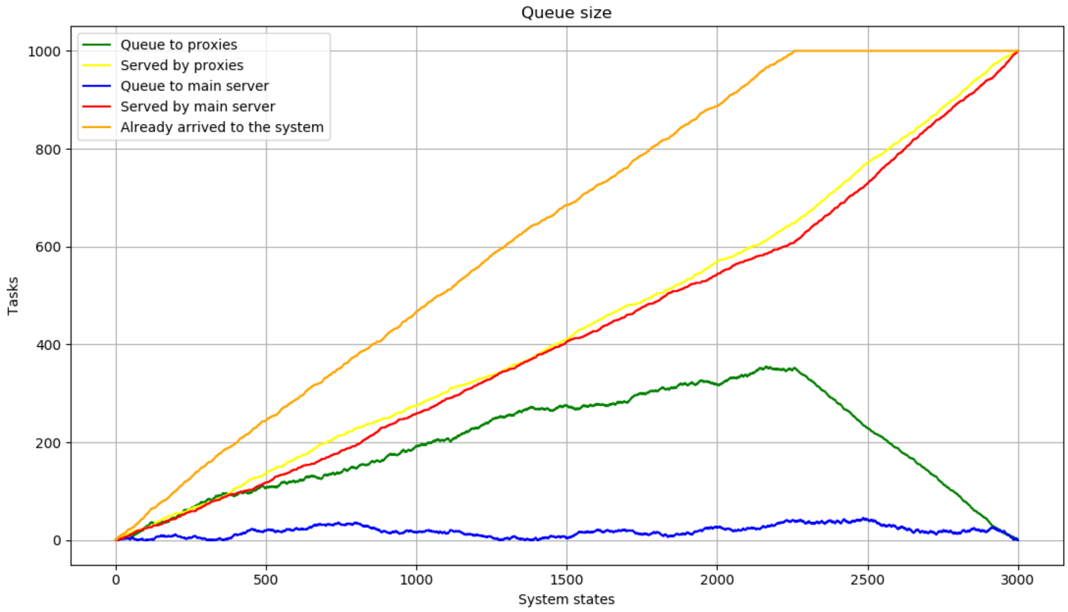

Figure 8: Queue size, 2 proxies, 1000 tasks

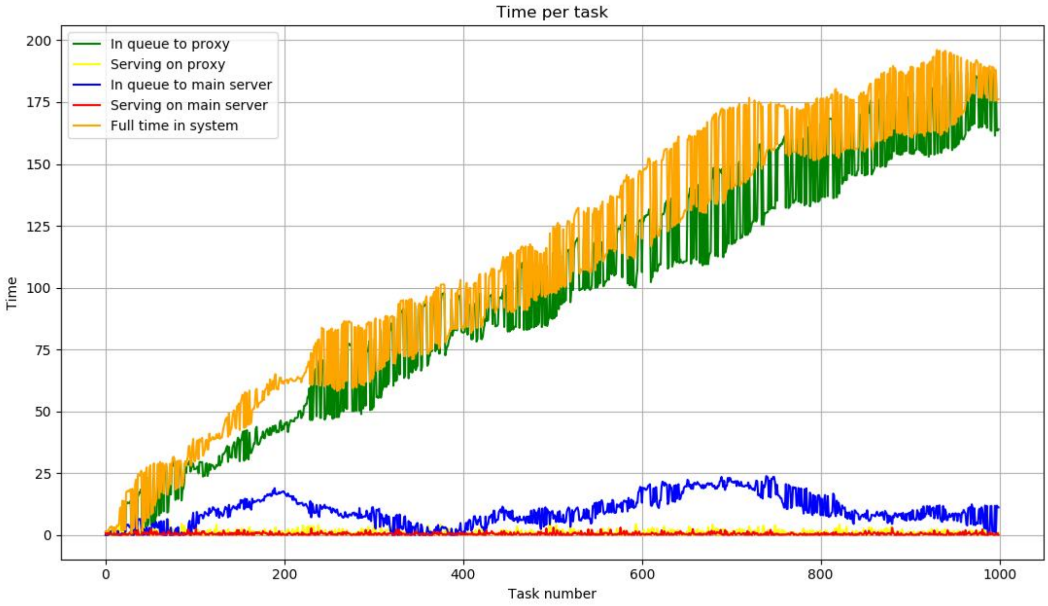

Figure 9: Time, 2 proxies, 1000 tasks

To fix the problem, let's simulate an increase in proxy servers up to 3 with the remaining parameters unchanged. As you can see from Figure 10, now the queue to the proxy is accumulating much less intensively, however, the queue to the main server continues to accumulate even after all requests have arrived to the system.

At the same time, from Figure 11, we can notice that the maximum total time spent by an task in the system is still close to 200. From this we can conclude that the growth of the queue to the proxy in this configuration is insignificant, but it is impossible to reduce their number or performance, since this will reduce the queue to the main server, but increase the queue to the proxy. Thus, the changes are leveled and the total time spent by the order in the system will not change.

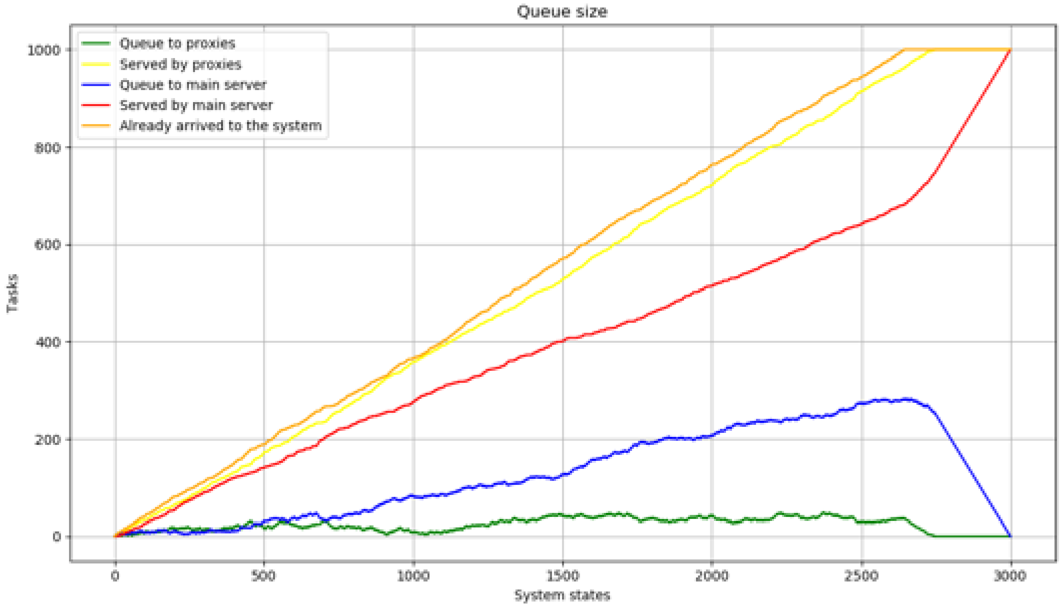

Figure 10: Queue size, 3 proxies, 1000 tasks

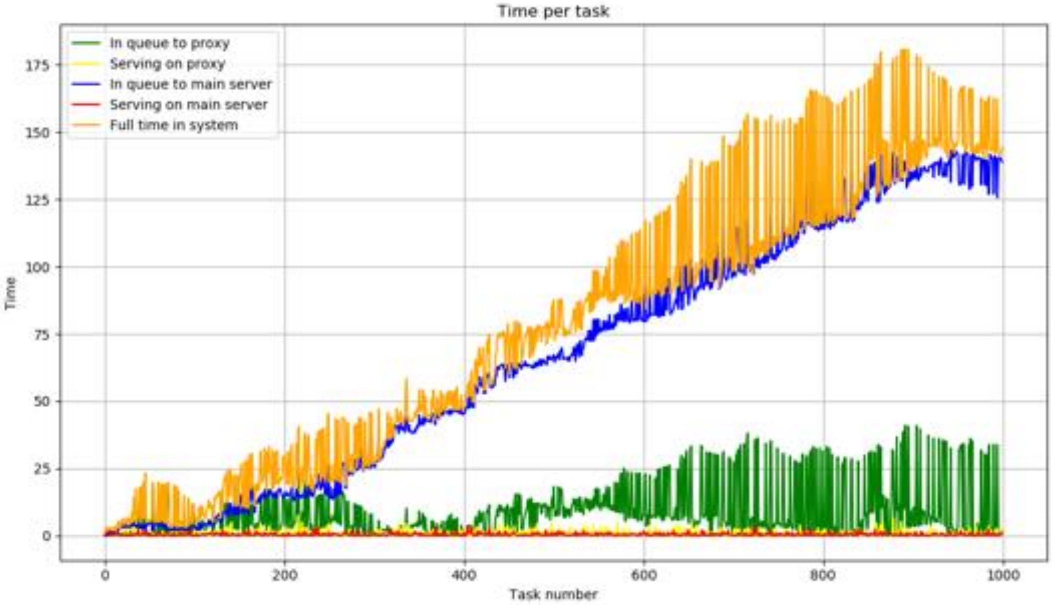

Figure 11: Time, 3 proxies, 1000 tasks

Let's simulate the behavior of the system once again with an increase in the intensity of processing requests on the main server to 2.5.

As it can be seen from Figure 12, queues accumulate almost evenly and are small.

From Figure 13, we can see that the time in the queue to the proxy and to the main server is approximately the same, and the maximum time for a request to be in the system is about 80 conventional units of time, which is more than 2 times less than the values obtained in the simulation with the initial conditions.

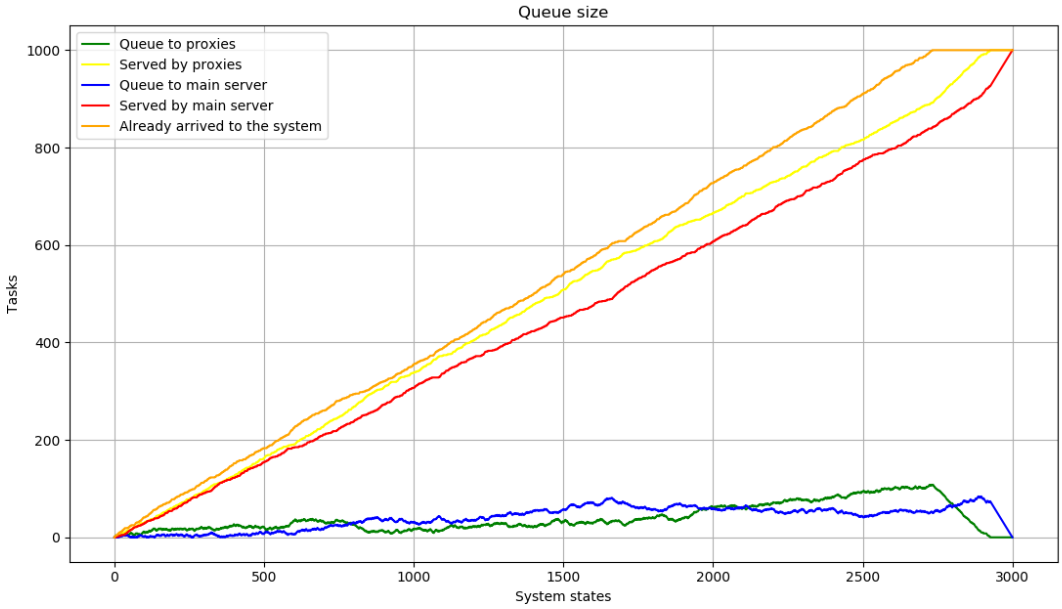

Figure 12: Queue size, 3 proxies, 1000 tasks, 2.5 main server intensity

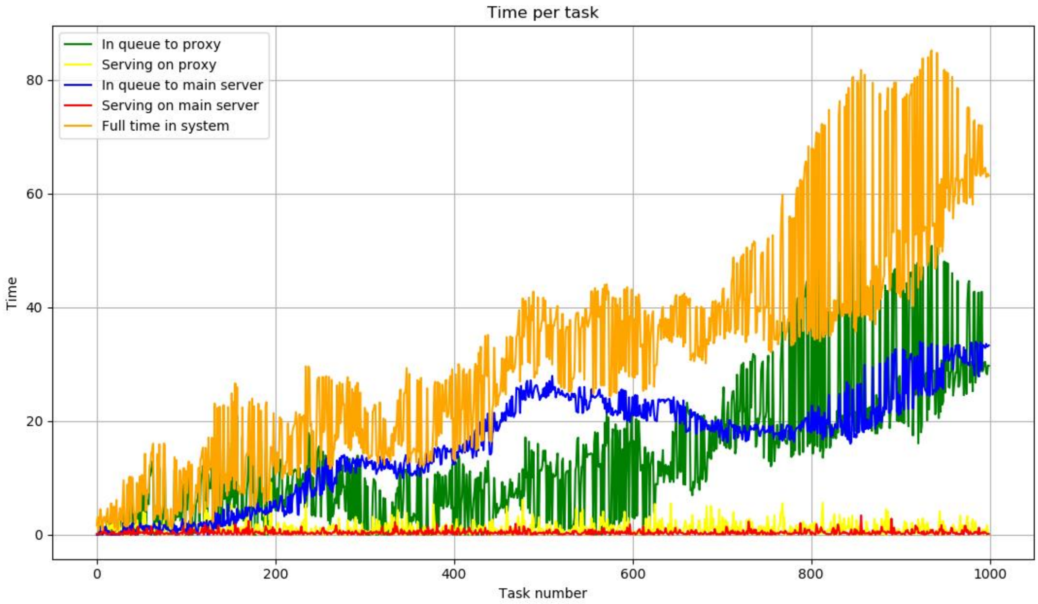

Figure 13: Time, 3 proxies, 1000 tasks, 2.5 main server intensity

So, the simulation model allows solving the technical problem of determining the minimum required hardware configuration of the system to service a finite number of tasks with a certain level of service - the time the task is in the system.

5. Conclusion

The article discusses the numerical-analytical and simulation models of the operation of a high-load distributed monitoring system for a data transmission network in a non-stationary mode. These models make it possible to determine the probabilistic characteristics of the system's behavior, as well as to determine the necessary hardware configuration to maintain the required service level of tasks.

Further development of the topic can be adding priorities to requests and adding the transfer of packages of requests from the proxy to the main server.

Research support. Research carried out on this topic was carried out within the framework of the budget topic No. 0073–2019–0004

References

[1] K. Shardakov, Comparative analysis of the popular monitoring systems for network equipment distributed under the GPL license. Intellectual Technologies on Transport 2018(1), pp. 44–48 (2018). (in Russian)

[2] V. Bubnov, V. Safonov, K. Shardakov, Overview of existing models of non-stationary queuing systems and methods for their calculation. Systems of Control, Communication and Security, 2020, no. 3, pp. 65–121 (in Russian). DOI: 10.24411/2410-9916-2020-10303

[3] I. Kovalenko About the queuing system with the speed of service, depending on the number of requirements in the system, and periodic shutdown of channels. Problems of Information Transmission, 1971, vol. 7, no. 2, pp. 108–114 (in Russian).

[4] I. Ezhov, M. Korneichuk, I. Oliinyk D. Distribution of the number of repair system channels when the flow rate changes in a special way. Kibernetika, 1976, no. 3, pp. 92–97 (in Russian).

[5] L. M. Abol'nikov Non-stationary queuing problem for systems with an infinite number of channels with group receipt of requirements. Problems of Information Transmission, 1968, vol. 4, no. 3, pp. 99-102 (in Russian).

[6] G. Arsenishvili Singleline queuing system with queue-dependent incoming flow rate. Soobshcheniia Akademii nauk Gruzinskoi SSR, 1974, vol. 76, no. 2, pp. 285–288 (in Russian).

[7] B. Conolly Generalized State Dependent Eriangian Queues (speculation about calculating easure of effectiveness). Applied Probability, 1975, no. 2. pp. 358–363.

[8] K. Shardakov, V. Bubnov, Stochastic Model Of A High-Loaded Monitoring System Of Data Transmission Network // Proceedings of Models and Methods of Information Systems Research Workshop. 2019. P. 29–34.

[9] K. Shardakov, V. Bubnov, Non-stationary parallel-serial model of a high-load monitoring system. Information and space, 2020(3), pp. 56-67 (2020). (in Russian)

[10] V. Bubnov, V. Safonov, V. Smagin About the loading of a computing system with a varying intensity of receipt of tasks, Automatic Control and Computer Sciences, 1987, no. 6, pp. 19–22 (in Russian).

[11] V. Bubnov, V. Safonov Development of dynamic models of non-stationary queuing systems. Saint-Petersburg, 1999. 64 p. (in Russian).

[12] V. Bubnov A. Khomonenko, A. Tyrva Software reliability model with coxian distribution of length of intervals between errors detection and fixing moments // International Computer Software and Applications Conference. 2011. pp. 310-314.

[13] V. Bubnov, A. Tyrva, A. Khomonenko Model of reliability of the software with coxian distribution of length of intervals between the moments of detection of errors // International Computer Software and Applications Conference. 34th Annual IEEE International Computer Software and Applications Conference, COMPSAC 2010. 2010. pp. 238-243

[14] K. Shardakov, V. Bubnov The program for the numerical-analytical calculation of the state probabilities of a non-stationary acyclic queueing network with a finite number of tasks. The Certificate on Official Registration of the Computer Program. No. 2020662856, 2020.

[15] K. Shardakov, V. Bubnov The program for simulating the process of

queueing tasks in a non-stationary acyclic queueing network with a finite number of tasks. The Certificate on Official Registration of the Computer Program. No. 2020663070, 2020.