Analysis of the Cluster Operating Time With the Migration of Virtual Machines*

Vladimir Bogatyrev[1,2][0000−0003−0213−0223] and Aleksey Derkach[2][0000−0002−0108−319X]

[1] Saint-Petersburg State University of Aerospace Instrumentation, Saint-Petersburg, Russia
vladimir.bogatyrev@gmail.com
http://new.guap.ru
[2] ITMO University, Saint-Petersburg, Russia
alexitmo1@gmail.com
http://www.ifmo.ru/ru/

Abstract. A Markov model of reliability of a fault-tolerant cluster has been considered, using virtualization technologies that ensure the continuity of the computational process in the event of a failure of the servers' physical resources and the impossibility of recovering from the interruption of the computational process. The probability of maintaining the system's operability under the condition of ensuring the continuity of the computing process for different service organization options was anilized. The mean time time to failure of such sistems was found. The purpose of the work is to increase the functional reliability of computing systems of a cluster architecture while increasing the time to failure, taking into account the requirements for ensuring the continuity of the computing process. A fault tolerance is considered as an object of study. A virtual machine is running on the cluster. The system involves launching a shadow copy of the VM on the backup server, which al-lows after the failure of the primary server to continue its implementation on the backup server.The proposed models can be used to assess the level of system reliability and are important in choosing a system configuration for certain conditions. Assessing the migration of virtual machines in the event of a failure of physical servers will allow you to calculate and evaluate the possible damage when using various models.

Keywords: Virtualization · Cluster · Migration · Virtual machines · Mean time to failure.

1 Introduction

For cluster computing systems, especially real-time, the key is to ensure reliability and fault tolerance while maintaining the continuity of the computing

process. The achievement of high and stable performance indicators, reliability, fault tolerance [1-3] and security [4] of computer systems is facilitated by the use of technologies for consolidation of clustering and virtualization resources [5-6], accompanied by replication and migration of virtual machines between physical servers. Migration and replication of virtual machines speeds up the reconfiguration process after failures of physical resources and contributes to supporting the continuity of the computing process required for managing cyber-physical systems and real-time technological processes.

2 Cluster fault tolerance technology with continuous computing

One of the effective ways to achieve fault tolerance of computing systems and processes is the migration of virtual resources between the physical nodes (servers) of a computing system of a cluster architecture. In a cluster with replication of virtual machines (VMs) on different physical nodes, they can migrate between cluster nodes in the event of failure of physical resources without stopping calculations on servers [7-8, 17-19].

Virtualization allows to optimize the use of computing resources, increases the scalability, fault tolerance and extensibility of the infrastructure, due to the rapid redistribution of the virtual resource [9-10].

Recovery time during VM migration after failures depends on the structure of the data storage.With shared storage for all physical nodes of the cluster, only RAM, virtual processor registers, and VM virtual device states are transferred during migration [5-6]. Information is transferred from hard disks in case of the data storage is localized for each node of the cluster.

Fault tolerance ensures the continuity of the computing process (service) in the cluster after the failure of one of the physical servers with the support of two copies of the VM, which, in RAM, are located on different physical servers, so that in case of failure of one of them, continue to work on the second. During the functioning of the VM on the main servers, the backup copy must support the actual copy of the RAM [12-14] of the active VMs. In this case, the virtual disk images of the VM should be stored on a dedicated or distributed data storage with synchronous data replication. VMware Fault Tolerance, Kemari for Xen and KVM [11, 12] software products support fault tolerance technology.

The purpose of the work is to increase the functional reliability of computing systems of a cluster architecture while increasing the time to failure, taking into account the requirements for ensuring the continuity of the computing process.

By functional reliability, we mean the ability of systems to perform the required functions, taking into account not only the operability of the resources required for their implementation, but also ensuring the necessary conditions for their implementation. Requirements to ensure the continuity of the computational process in the inadmissibility of interruptions of the reservation system at the time of recovery are proposed as conditions of operation. Thus, in the systems under consideration, recovery is possible only if it is combined with the

implementation of the required functions by non-failed nodes. The system enters a non-recoverable state if it is impossible to reconfigure with the activation of the required number of operability resources. For such systems, the reliability indicator is the time to failure, including taking into account violations of the continuity of the computing process, provided that the permissible reconfiguration time, including the costs of migration of virtual machines, is exceeded.

3 Cluster organization and options for its recovery

The cluster architecture computer system contains servers (Fig.1). Each server is connected directly to one local storage device (local server storage device). In the system to ensure automatic reconfiguration, aimed at supporting the continuity of computing processes based on dynamic migration, pairs of physical servers of the primary and secondary are allocated in the cluster. The main server performs the required tasks critical to the continuity of the computing process. The backup server is designed to perform dynamic reconfiguration with ensuring the continuity of the computing process in case of possible failures of the primary server. The backup server, in addition to implementing dynamic system reconfiguration, performs some background tasks that are not critical to the continuity of the computational process and to the time of query execution). If the backup server fails, the background tasks that it performs may be lost or redistributed to the main server if they are performed non-priority in the background.

With the simplest implementation of a fault tolerance cluster, it is equipped with a pair of servers, one of which is designated as the main, and the second as the backup.

Fault tolerance technology involves launching a backup copy of the primary server VM on the backup server and transferring the calculations to the backup server in case of primary server or storage device failure.

Consider system options that provide (option A) and do not provide (option B) the restoration of physical nodes for states in which the continuity of the computing process is ensured during reconfiguration of the cluster, which allows us to select a working server and associated storage device for the implementation of the calculations.

For the options under consideration, in the event of a transition to a failure state with the impossibility of implementing the required functions at least with a minimal workable configuration, it is considered that the computational process is interrupted for a time exceeding the maximum permissible value, which entails a transition to a state of unrecoverable failure.

Let us consider cluster systems while ensuring fault-tolerant functioning with pairwise integration of physical servers into duplicated systems supporting the processes of virtual machine migration and data replication. For each pair of pairs in the cluster interacting to support dynamic reconfiguration of the servers (duplicated system), state and transition diagrams for a variant of organization A and B of a duplicated cluster system with recovery disciplines are shown in

Fig. 2 and 3 . The diagram shows the failure and recovery rates of the server λ_0 and μ_0; disk λ_1, μ_1; commutator λ_2, μ_2. The actual data replica is loaded onto the recovered disk (synchronization of the distributed storage system) with an intensity of 3. The VM startup time on the backup server and the user application loading on it are negligibly small in comparison with the loading of the current data replica, there-fore, in this study, an instant switch between servers is assumed.

The system of differential equations in accordance with the state diagram and transitions in Fig. 2 and have the form:

$$\begin{aligned}
P_0'(t) &= -(2\lambda_0 + \lambda_2 + 2\lambda_1)P_0(t) + \mu_3 P_4(t),\\
P_1'(t) &= -(\lambda_1 + \lambda_0 + \mu_0)P_1(t) + 2\lambda_0 P_0(t) + \lambda_0 P_4(t),\\
P_2'(t) &= -(\lambda_1 + \lambda_0 + \mu_2)P_2(t) + \lambda_2 P_0(t) + \lambda_2 P_4(t),\\
P_3'(t) &= -(\lambda_1 + \lambda_0 + \mu_1)P_3(t) + \lambda_1 P_4(t) + 2\lambda_1 P_0(t),\\
P_4'(t) &= -(2\lambda_0 + \lambda_2 + 2\lambda_1 + \mu_3)P_4(t) + \mu_1 P_3(t) + \mu_0 P_1(t) + \mu_2 P_2(t),\\
P_5'(t) &= -(\lambda_1 + \lambda_0)(P_1(t) + P_2(t) + P_3(t) + P_4(t)).
\end{aligned}$$

For option B:

$$\begin{aligned}
P_0'(t) &= -(2\lambda_0 + \lambda_2 + 2\lambda_1)P_0(t),\\
P_1'(t) &= -(\lambda_1 + \lambda_0)P_1(t) + 2\lambda_0 P_0(t),\\
P_2'(t) &= -(\lambda_1 + \lambda_0)P_2(t) + \lambda_2 P_0(t),\\
P_3'(t) &= -(\lambda_1 + \lambda_0)P_3(t) + 2\lambda_1 P_0(t),\\
P_4'(t) &= -(\lambda_1 + \lambda_0)(P_1(t) + P_2(t) + P_3(t)).
\end{aligned}$$

4 Calculation of the probability of operability of duplicated systems

The presented systems of differential equations make it possible to determine the dependence of the probabilities of all states from time.

The probability of the system working while maintaining the continuity of the computing process for option A and B is defined as:

$$P(t) = \sum_{i=0}^{4} P_i(t),$$

and for option B is defined as:

$$P(t) = \sum_{i=0}^{3} P_i(t).$$

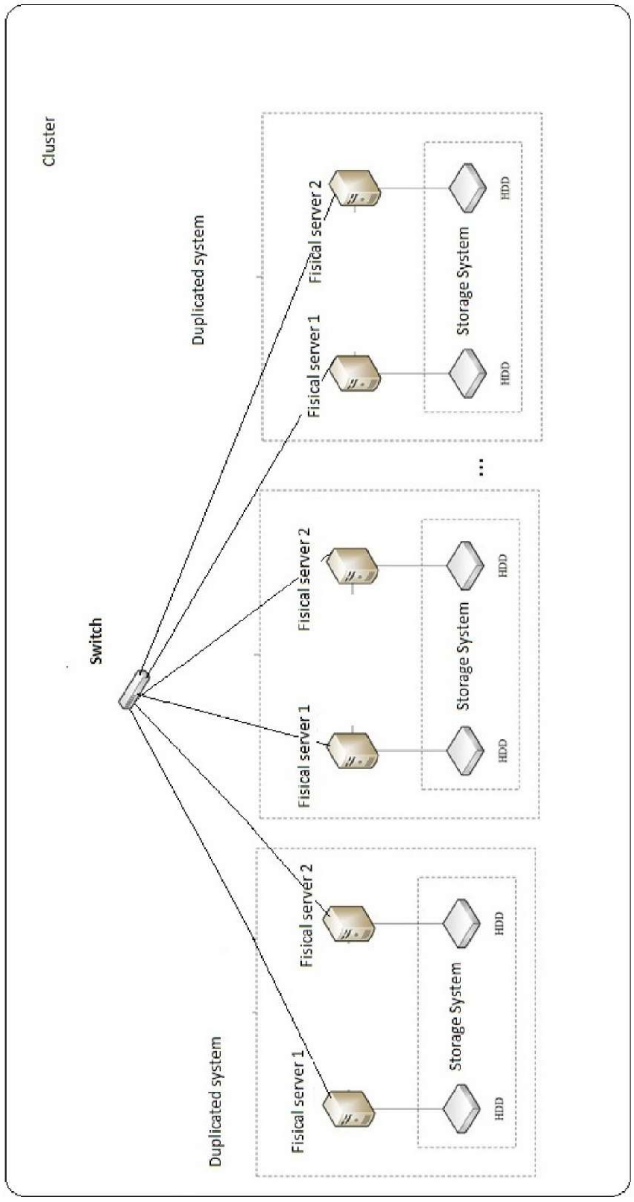

Fig. 1. Cluster model.

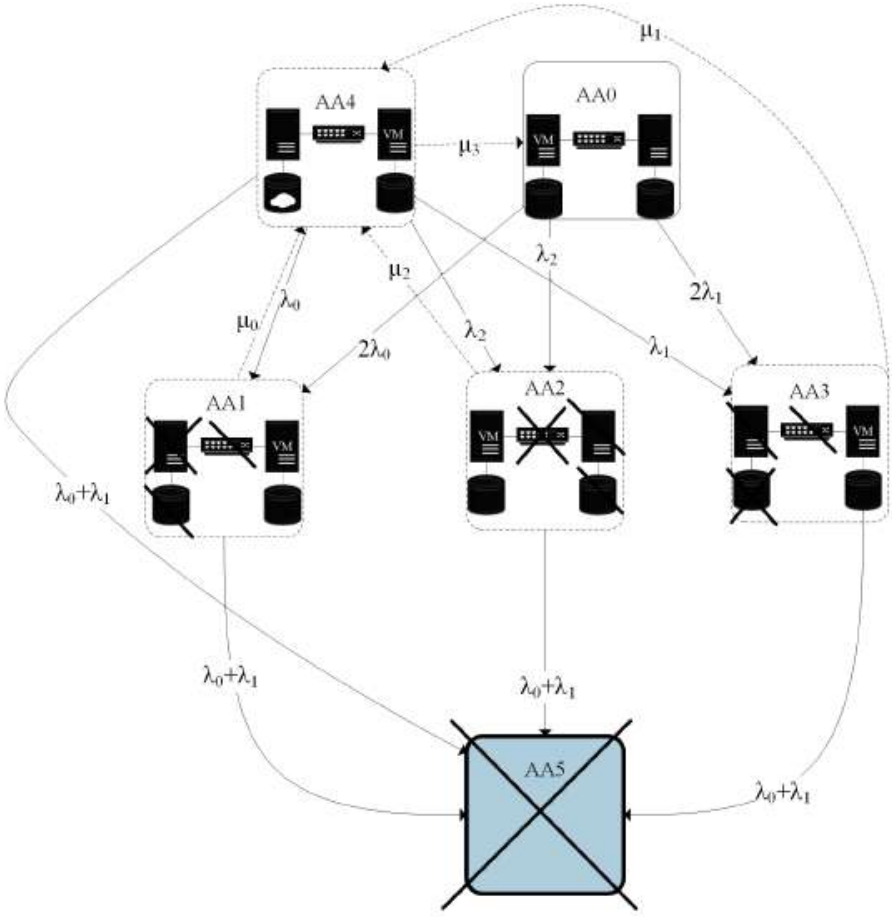

Fig. 2. State and transition graph of a duplicated system with ensuring the continuity of the computing process for organization option A.

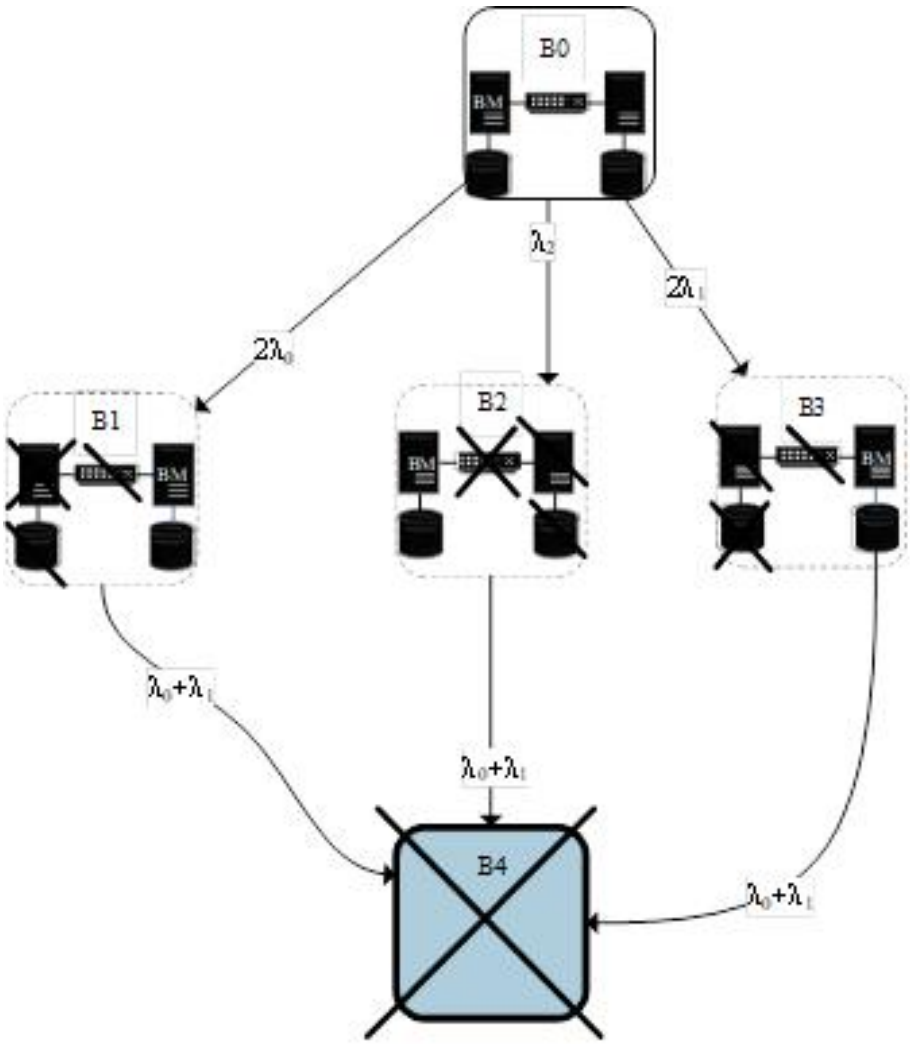

Fig. 3. State and transition graph of a duplicated system with ensuring the continuity of the computing process for organization option B.

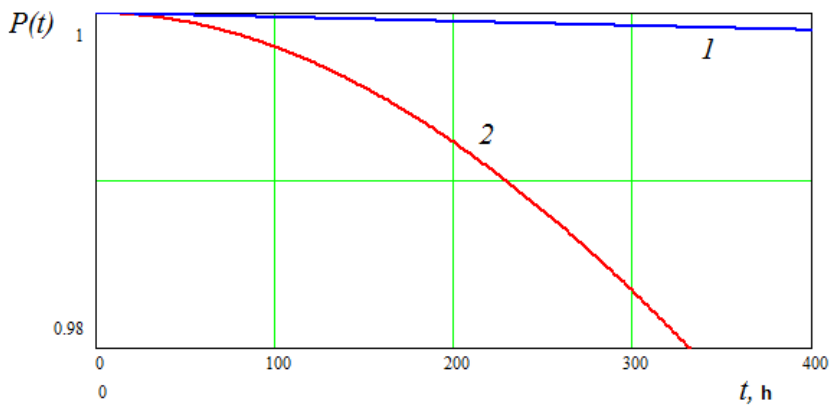

Fig. 4. The probability of maintaining the system's operability under the condition of ensuring the continuity of the computing process for ser-vice organization options A and B.

The results of calculating the probability of duplicated computer systems' operability provided that the computing process is continuous for options A (the curve 1) and B (the curve 2) of the maintaining process' organization are presented in Fig. 3.

The calculations were performed with the following failure rates $\lambda_0 = 1.115 \cdot 10^{-5}\ (1/h)$, $\lambda_1 = 3.425 \cdot 10^{-6}\ (1/h)$, $\lambda_2 = 2.3 \cdot 10^{-6}\ (1/h)$, and recovery rates $\mu_0 = 0.33\ (1/h)$, $\mu_1 = 0.17\ (1/h)$, $\mu_2 = 0.33\ (1/h)$, $\mu_3 = 1\ (1/h)$.

The presented dependences make it possible to evaluate the effect on the probability of maintaining the operability of a duplicated system, restrictions on the inadmissibility of interruption of the computational process, and the impact of restoration work while maintaining the possibility of continuity of the process of performing the required functions.

5 Calculation of the probability of operability of duplicated systems

The mean time between failures and the probability of working without failures are related by the relation [15]:

$$T = \int_0^\infty P(t)dt.$$

The average operating time to failure in accordance with the methodology of [15] is found as follows. The mean time to failure can be obtained by integrating the system of differential equations for a model with an absorbing state, the initial conditions $P_1(0) = 1, \dots P_k(0) = 0, P_n(0) = 0$ for a model with n states.

For the systems under study, integrating the left and right sides of the systems of equations (1), (2) for the models under consideration. Given that in the presence of an absorbing state, $P_i(\infty) = 0$, we have [16]:

$$\begin{gathered}
-(2\lambda_0 + \lambda_2 + 2\lambda_1)T_0 + \mu_3 T_4 = -1, \\
-(\lambda_1 + \lambda_0 + \mu_0)T_1 + 2\lambda_0 T_0 + \lambda_0 T_4 = 0, \\
-(\lambda_1 + \lambda_0 + \mu_2)T_2 + \lambda_2 T_0 + \lambda_2 T_4 = 0, \\
-(\lambda_1 + \lambda_0 + \mu_1)T_3 + \lambda_1 T_4 + 2\lambda_1 T_0 = 0, \\
-(2\lambda_0 + \lambda_2 + 2\lambda_1 + \mu_3)T_4 + \mu_1 T_3 + \mu_0 T_1 + \mu_2 T_2 = 0, \\
-(\lambda_1 + \lambda_0)(T_1 + T_2 + T_3 + T_4) = 0.
\end{gathered}$$

Thus, for the options for organizing system A, we have:
For option B:

$$\begin{gathered}
-(2\lambda_0 + \lambda_2 + 2\lambda_1)T_0 = -1, \\
-(\lambda_1 + \lambda_0)T_1 + 2\lambda_0 T_0 = 0, \\
-(\lambda_1 + \lambda_0)T_2 + \lambda_2 T_0 = 0, \\
-(\lambda_1 + \lambda_0)T_3 + 2\lambda_1 T_0 = 0, \\
-(\lambda_1 + \lambda_0)(T_1 + T_2 + T_3) = 0.
\end{gathered}$$

Where T_i is the average time spent in working condition i when starting work from a operable state. Mean time to failure is determined by summing T_i, for all operational states [16]:

$$T = \sum T_i.$$

For the system under consideration, the time to failure with service discipline $A : T_1 = 3.891 * 10^5$ hours, and B $T_2 = 3,277 * 10^3$ hours.

6 Conclusions

The significance of the impact of ensuring the continuity of the computational process on duplicated systems of the cluster architecture is demonstrated. The result of the study was obtained on the basis of Markov models of reliability of a fault-tolerant cluster with the migration of virtual machines when it is impossible to recover after the interruption of the computational process. The time to the first failure of a duplicated system with recovery and without recovery in failure states of nodes that do not violate the continuity of the computational process to perform the required functional tasks critical to the continuity of the computing process is determined.

References

1. Kopetz H. Real-Time Systems: Design Principles for Distributed Embedded Applications. Springer, 2011.
2. Sorin D. Fault Tolerant Computer Architecture. Morgan Claypool, Madison, 2009.
3. Dudin, A. N., Sun, B.: A multiserver MAP/PH/N system with controlled broadcasting by unreliable servers. Automatic Control and Computer Sciences V. 5, 32-44 (2009)
4. Zhmylev S., Martynchuk I. G., Kireev V. I., Aliev T.: Analytical methods of nonstationary processes modeling. CEUR Workshop Proceedings V. 2344 (2019)
5. Bogatyrev A. V., Bogatyrev V. A., and Bogatyrev S. V.: Multipath Redundant Transmission with Packet Segmentation. 2019 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF) (2019) doi: 10.1109/WECONF.2019.8840643
6. Bogatyrev V. A., Bogatyrev S. V., and Bogatyrev A. V.: Model and Interaction Efficiency of Computer Nodes Based on Transfer Reservation at Mul-tipath Routing. 2019 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF) (2019) doi: 10.1109/WECONF.2019.8840647
7. Jin H., Li D., Wu S., Shi X., Pan X.: Live virtual machine migration with adaptive memory compression. Proc. IEEE International Conf. on Cluster Computing (CLUSTER '09) Art. 5289170 (2009) doi: 10.1109/CLUSTR.2009.5289170
8. Sahni S., Varma V.: A hybrid approach to live migration of virtual machines. Proc. IEEE Int. Conf. on Cloud Computing for Emerging Markets (CCEM 2012), 12-16 (2012) doi: 10.1109/CCEM.2012.6354587
9. Poymanova E. D. Tatarnikova T. M.: Models and Methods for Studying Network Traffic // 2018 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF) (2018) doi: 10.1109/WECONF.2018.8604470
10. Kutuzov O., Tatarnikova T., : On the Acceleration of Simulation Modeling. In 2019 XXII International Conference on Soft Computing and Measurements (SCM) doi: 10.1109/SCM.2019.8903785 (2019)
11. Knowledge sharing portal UNIX/Linux-systems, open source systems, networks, and other related things, http://xgu.ru/wiki/Kemari Last accessed 15 Sep 2019
12. Elizarov E Dell Live Volume: virtualize disk space, http://blog.korphome.ru/2016/06/28/dell-live-volume Last accessed 15 Sep 2019
13. Bogatyrev V. A., Aleksankov S. M., Derkach A. N.: Model of Cluster Reliability with Migration of Virtual Machines and Restoration on Certain Level of System Degradation //2018 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF-2018) 92018)
14. Astakhova T., Shamin A., Verzun N., Kolbanev M. A. Astakhova T., Shamin A., Verzun N., Kolbanev M.: Proceedings of the 10th Majorov International Conference on Software Engineering and Computer Systems. CEUR Workshop Proceedings MICSECS 2018 (2019)
15. Victorova V. S., Stepanjanc A. C.: About reliability indicators of the average operating time type. Reliability 4(51), 27-36 (2014)
16. Victorova V. S., Stepanjanc A. C.: Models and methods for calculating the reliability of technical systems. 2nd edn. URSS LLC Lenand, Moscow (2016)
17. Bogatyrev V. A., Parshutina S. A. : Redundant Distribution of Requests Through the Network by Transferring Them Over Multiple Paths. Communications in Computer and Information Science, 601, 199-207 (2016)

18. Zakoldaev D. A., Shukalov A. V., Zharinov I. O., Zharinov O. O.: Workstations Industry 4.0 for instrument engineering products. IOP Conference Series: Materials Science and Engineering, 1 (665), pp. 012014 (2019)
19. Korobeinikov A. G., Fedosovsky M. E., Zharinov I. O., Polyakov V. I., Shukalov A. V., Gurjanov A. V., Arustamov S. A.: Method for Conceptual Presentation of Subject Tasks in Knowledge Engineering for Computer-Aided Design Systems. Advances in Intelligent Systems and Computing, V. 680, 50-56 (2018)