

SIMULATION OF IT INFRASTRUCTURE WITH CONSIDERATION OF CRITICAL ASPECTS FOR QUALITY OF SERVICE MANAGEMENT

Annotation. Testing hypotheses about service quality management in IT infrastructure requires large and complex data centers with sufficient resources to explore various possible scenarios of infrastructure operation during the provisioning of IT services. For testing purposes, dozens of solutions already exist, but all of them don't consider critical aspect of IT infrastructure. In order to solve this issue general mathematical model for quality of service management in critical infrastructures was introduced. Based on the proposed model simplest set of tools was developed for creating heavy simulations which can cover criticality during functioning.

Keywords: simulation; modelling; IT infrastructure; quality of service management; QoS; critical IT infrastructure.

Introduction. Information technologies have evolved significantly today which made it possible to create ubiquitous and widespread services available literally at the fingertips. Such breakthrough could be possible due to emergence of big clusters of computing units, interconnected with each other into grids, clouds as part of big data centers. Those units can consist not only stationary processing machines but also storage units, network units, and other hardware and software abstractions which are located under the hood, or even can be named as a back-end, where front-end is an IT service itself, available for ordinary user or business unit.

Actually, initial concept of IT service has emerged with introduction of the novel way for developing new software via paradigm of service-oriented computing (SOC) which provides an approach for the development of rapid, low-cost, interoperable, evolvable, and massively distributed applications [1].

With that in mind the field of IT services also introduces a bunch of problems related to quality management. Quite commonly these problems as a phenomenon was addressed in [2]. Furthermore, this work should be considered as a foundation of this research and continuation. It contains basic management model of IT services quality within IT infrastructure. Briefly, it introduces the decomposition-compensation approach to quality management and correspondingly performs a de-

composition of tasks for service level management and compensate a negative impact of different factors by allocation of additional resources to critical applications [2].

In order to perform a research in such broad area it is important to have access to a pool of resources which can provide means for developing some testbed environment for experiments. However, this can be expensive and not effective because of old resources used at the institutions. Moreover, cloud providers propose to use their resources but only with limited capacity which can not afford to conduct any desirable test suit but only small ones. Another way to handle experiment is to create a simulation environment which is not something new but can save a lot of efforts during experiment implementation.

A comprehensive overview regarding available simulators and possible issues which those are solving as for today is given in [3]. This research gives a broader view on the field of cloud simulators which are actually a subclass of IT infrastructures and concludes that most tools focus on energy modelling and performance, but lacks on security aspects, which is crucial for critical IT infrastructure. Also, [3] has comparison results for 33 simulation toolkits each of which covers some peculiarity of system functioning, therefore for complex problems a couple of them should be used together which introduces additional time for learning curve.

This research proposes to consider novel framework, which can cover critical IT infrastructure simulation issues and provide space for improvement of covered features of critical aspects.

Problem Statement. As for today, a lot of empirical experiments are being conducted in the area of data center modelling, network modelling and cloud simulations. However, the term IT infrastructure is quite rare in this field due to more widespread and specific technical terms mentioned previously. Also, big part of studies uses some hypothesis testing without providing any general implementation which could be used for further improvements and therefore could reduce the number of projects and researches started from scratch. The main goals for this research are to introduce the term IT infrastructure as a generalization of already known models which are basically a subclasses of IT infrastructure models, investigate known solutions and define a novel way to simulate IT infrastructure. Most of problems related to simulating a data center functioning are related to optimization of resources consolidation, reducing energy footprint of a data center and comply with service level agreements during simulation. However, those studies concern only some specific aspects of the IT infrastructure performance. Another objective of this

research is to take into account real data during the simulation and add the possibility to simulate different scenarios for quality of service maintenance and management alongside critical plane of IT infrastructure. Basic model of IT infrastructure is going to be tested as well in order to show feasibility of application of provided simulation framework.

Overview of Critical Infrastructure Simulation Definitions. The majority of complex actual system requires quite expensive analytical models to be created and therefore such models then widely investigated via simulation. Worth mentioning that complexity of many real-world systems involves unaffordable analytical models, and consequently, such systems are commonly studied by means of simulation. However, a simulation term should be clarified before continuing.

In [4, 5] simulation is described as “an experiment to determine characteristics of a system empirically. It is a modeling method that mimics or emulates the behavior of a system over time”. So, for creating a simulation environment it is necessary to design a model of real or hypothetical system, running this model on a computer and consequently analyzing the output with the help of statistical methods. In such circumstances the state of the modelled system is being modified by a simulation program and therefore reproduces the way how actual system is evolving over time.

According to [6] Shannon has formulated a term of simulation as “the process of designing a model of a real system and conducting experiments with this model for the purpose either of understanding the behavior of the system or of evaluating various strategies (within the limits imposed by a criterion or a set of criteria) for the operation of the system”.

Despite the fact that modern IT infrastructures and interconnection inside of them have derived from telecommunication industry it still relevant to use similar approaches for analysis for the first ones. According to [4] three phases of telecom networks design can be distinguished: math analysis with simple models is performed and as a result some numeric data is obtained, simulation phase which in contrast to simple models in math phase results in data closer to the true worlds rather than data returned on analysis and real setup, which gives empirical data.

For assessing the availability, the primal approaches stand on measurement and modelling methods [7]. Approaches which depend on models are rapid and reasonable in contrast to the methods based on measurements. Among the methods for system simulation there are analytical models, discrete-event simulation or combination of two approaches.

For clarity M. Rybnicek et al. [8] has made an overview on approaches for modelling a critical infrastructure, among which are: agent-based, unified modelling language (UML), graphical modelling, etc. The most performable among those is agent-based modelling and simulation approach.

In order to model critical IT infrastructure is to follow similar approach for critical infrastructure, which stand on using agents in model. Shortly, agent is object or subject which can perceive it's surrounding and act upon the circumstances happened in the environment. Worth mentioning that each agent must have sensors and effectors in order to be able to do something depending on the conditions. In [8] authors have used defined attributes for each agent which has made it identifiable, situated, goal-directed, self-directed and autonomous.

The primary goal during critical infrastructure simulation is to investigate dynamic effects introduced from attacks or disruptions. The use of techniques to perform a simulation allows to understand deeply the domino and cascading effects. Because of interconnection of unique services in critical infrastructure and unavailability or malfunctioning of any can lead to significant consequences. Moreover, simulation assists in providing new knowledge about complex systems which allows to advance redundancies planning and development of incident response strategies, which is stated in [8].

Authors of [9] have introduced an approach for enhancing the protection of critical infrastructures which involves a bunch of steps to facilitate the simulation of such infrastructure. They have covered such steps like model development of services, context description, dependency and interdependency identification, probabilistic and deterministic models and final Monte Carlo simulation stage for exploratory analysis. Some of these steps are also important during simulation of critical IT infrastructure, because despite the fact that boundaries of IT infrastructure are bit different but threats are similar.

As a continuation of [9] researches in [10] have introduced the description for the consequences of cyber-attacks on critical infrastructure. In case of attack on power system of infrastructure the impact can be in range from 0 to growing losses of not providing energy to some customers, which can influence on country in general. To address critical aspect authors in [9–11] are referring to interdependencies issues. Among the reasons for dependencies between services there are: functional dependencies like inputs for one service has come from the outputs of other service; analogous components which contain common source of failures; and common envi-

ronment used for different services, which in case of failure will result in failure of all connected services.

Another approach for calculating interdependencies is proposed in [12] which is based on quality of service (QoS) indices, performance indicators and dependence indices. A performance indicator implies a measured value of system requirement which represents some characteristic of a system. Objective function which involves all performance indicators should be also defined because sometime increase in one indicator can lead to decline in other.

On other hand, quality of service is nothing more than objective function for the service of interest from the mathematical point of view and which is also implies function based on a user perspective of a service.

The last one, which [12] contains regarding interdependencies metrics is a dependence index, which is “a numerical measurement of the degree to which an activity (or service) depends on another activity (service), system or physical or human component.”

IT infrastructure management tasks. According to [13] there are three generalized areas of management, two of which cover technical side of all management functions in IT infrastructure: operative or automatic quality level maintenance for IT services and reasonable utilization of resources.

Those tasks should be considered in tight connection. For critical IT infrastructures reasonability in resource utilization are not relevant, because reliability indicators are becoming more prevailing. Moreover, allocation of significant amount of resources for reservation aims is not only justified but necessary.

General management task can be defined as follows. Let the state in IT infrastructure be described with a variable S from the set of possible state \mathbf{S} . The state of IT infrastructure at some moment in time can depend on control impacts like in (1):

$$U \in U; S = f(U) \quad (1)$$

Supposing that there a functional $F(U, S)$ which is defined on product set of U and S , that defines efficiency of IT infrastructure functioning. By knowing that, efficiency management indicator can be determined by (2):

$$K(U) = F(U, f(U)) \quad (2)$$

Then the objective of IT infrastructure management is reduced to finding such acceptable control impact that maximizes value of management efficiency indicator. But it is true only when reaction of IT infrastructure (1) to control impacts is known:

$$K(U) \rightarrow \max_{U \in U} \quad (3)$$

The most important task of IT infrastructure is delivery of qualified services to the end users. With this in mind the efficiency of management in this case can be estimated by the quality Q of provided services. With operative management the primary objective is being achieved by maintenance of service quality on specified level with taking into account availability and reliability indicators. Then maximal efficiency of control can be achieved by choosing such control impact that allows to approach agreed level of quality Q_{agr} and maximal availability.

$$Availability(Q(U) > Q_{agr}) > \max_{U \in \Omega} \quad (4)$$

In turn, quality of all services Q is determined by the quality of all services:

$$Q_j, j = \overline{1, N}; Q = f(Q_1, \dots, Q_N) \quad (5)$$

Therefore, control impacts must support given level of quality for each service with ensuring maximum level of availability:

$$Availability(Q_j(U) > Q_j^{agr}) > \max_{U \in \Omega}, j = \overline{1, N} \quad (6)$$

From the user's perspective management efficiency criteria for maintenance of the quality of j -th service can be selection of such control impact which brings minimal time for handling i -th request to the application A_j :

$$\min_{U \in \Omega} (T_{R_{i,j}} = (t_{R_{i,j}} - t_{A_{i,j}})) \quad (7)$$

Where $t_{A_{i,j}}$ is time when a request has been acquired by the system and $t_{R_{i,j}}$ when the system has handled it and send back to user. As a result, in order to provide reasonable models for IT infrastructure with considering critical indicators it was necessary to define all the values for each part of management process from (1) to (7), which allows to proceed with implementation of simulation models.

IT Infrastructure Simulation Framework Implementation. For developing a simulation framework R language was used. Extension package written for R, called *simmer* was used. According to [14], this package introduces discrete-event simulations to R. It is developed as universal process-oriented framework, which utilizes core written with C++. Moreover, it allows to monitor all processing during simulation in automatic manner. It adopts the concept of trajectory which is “a common path in the simulation model for entities of the same type” [14].

For further research it is suggested to use models from [15] to simulate network traffic. Authors have introduced a mathematical procedure to define equations of bimodal traffic distributions. According to them, Uniform distribution is usually considered if there is no a priori information about the packet's attributes. On other hand, when some additional info exists then the Normal and Beta distributions are

considered. As a result, application of new information obtained from this research can cover denial-of-service attacks issues, improve the performance metrics of routers, as well as provide means to application level traffic classification.

A few common entities were developed for simulation framework: IT infrastructure, scheduler, server and request. As for now, request was marked as http but it is very abstract so far, therefore can be considered as any request. Future work in this direction will elaborate peculiarities of http request and extend types of requests supported by IT infrastructure.

In order to produce some simulated events inside the simulation, it is necessary to get some empirical data about say waiting and run times of jobs or requests from real data sets. For this purpose, the AuverGrid [16] dataset from Grid Workload Archives [17] was used. Also, authors in [17] have described the generic structure of all datasets and common properties collected in different systems. The dataset contains info about wait time and run time of each request, a lot of meta information, like user id or group id, queue id, etc. The most interesting for dynamic behavior is the wait and run times. The last one which is also very important is interarrival time, which can be calculated additionally by grouping all requests into hourly intervals, get a sum of all requests in each hour which will refer to as arrival rate according to (8). It shows that arrival rate for the whole dataset is going to be a vector of values.

$$Arrival\ Rate = \sum_1 R_1 \quad \dots \quad \sum_N R_N \quad (8)$$

And then to get interarrival time it is necessary to divide 1 by arrival rate and multiple with 3600 which is seconds equivalent of an hour in order to get time between arrivals in seconds, as shown on (9).

$$Interarrival\ Time = \frac{3600}{Arrival\ Rate} \quad (9)$$

Now, once having all the required parameters to obtain from real data, it is suggested to use empirical cumulative distribution function, which is a relation of cumulative sum of all values from the vector and division by sum of all values which will give a relative scale of cumulative changes of some statistical property. With this in mind it is easy to get wait time probability density and empirical cumulative distribution, which is shown on Fig. 1.

After obtaining data of probability density and cumulative distribution of wait time the next step will be to create a generator function, which can produce values with the same probability as original wait time has. That can be done by calculating probability density, which is a standard function in R and then by means of linear

approximation and cumulative distribution a new random generator can be created. That possible by using reversed cumulative distribution function, which allows to map values in range from zero to one (0; 1) to actual values of some property, in this case wait time. It is obvious that in order to get all values of wait time vector with the same probability the Uniform distribution can be used on the range of (0; 1). But before moving to the application of created generator of random values some validation should be performed to make sure that produced values from created generator fits into original distributions.

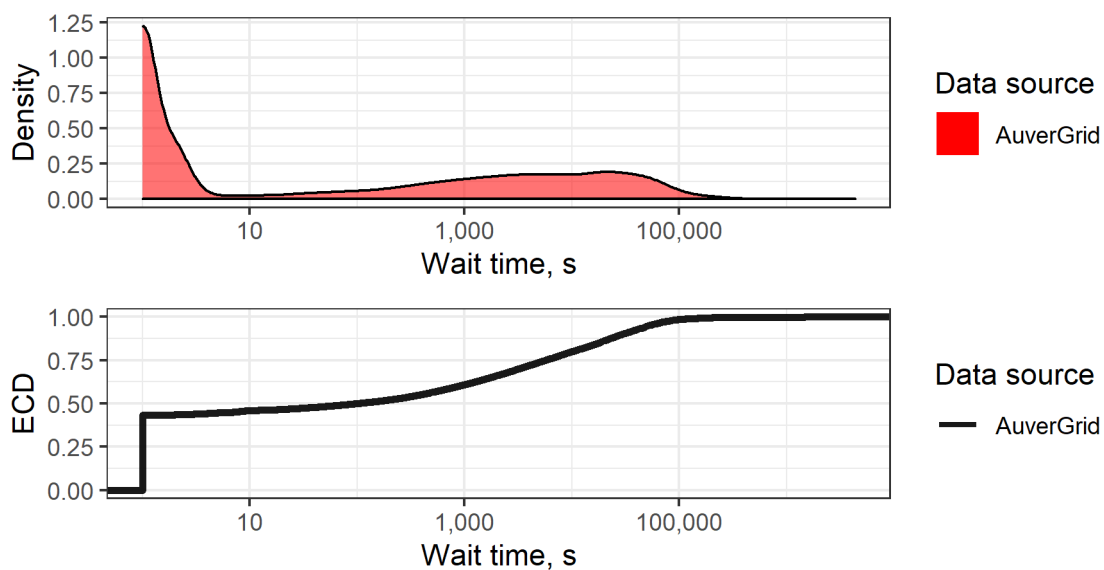


Figure 1 – Density and empirical cumulative distribution for wait time among all requests in the system

Let compare values from both sources, original dataset and random generator based on density functions of original data (Fig. 2). It is important to use almost the same number of values as original dataset contains, as far as convergence of the produced density should follow a law of big numbers.

From the picture can be seen small gap between real data and random data. Due to approximations of course it is possible to lose some values from the sample, but in total cumulative results are showing better picture, which can reflect adequacy of this model of behavior. Next step is to develop appropriate objects for simulation. Simmer contains some basic tools which then can be extended for solving any specific problems. For simulating an IT infrastructure provided entities were extended as follows: initial resource in simulation was wrapped into few entities, like Scheduler, Server, ResultNode and IT infrastructure itself. Those are also described as: IT infrastructure object, which is actually a simmer environment with ad-

ditional class; Scheduler, Server, ResultNode, which are resources with some capacity and queue; and HttpRequest – extended trajectory which is following from Scheduler, through Server and to the ResultNode.

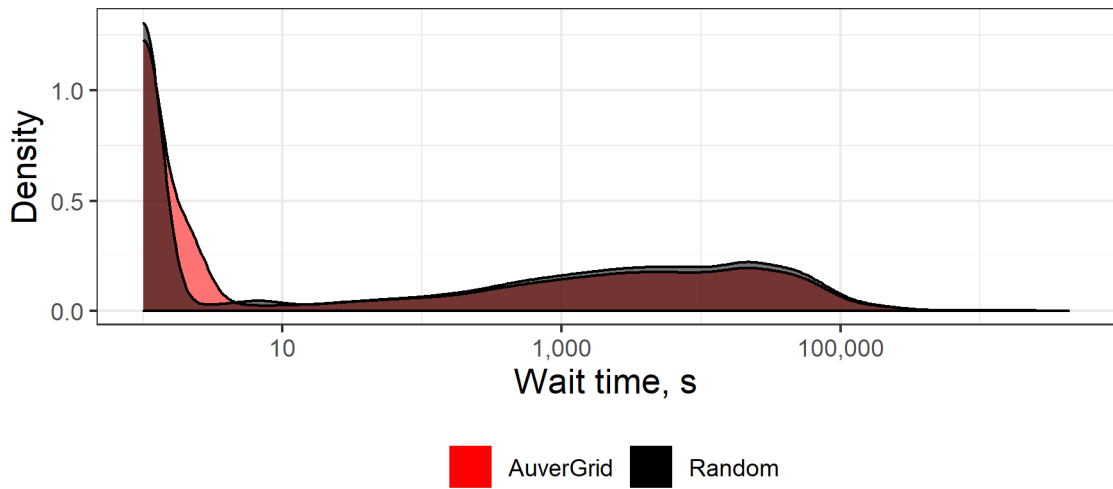


Figure 2 – Fitness of randomly generated values and original data

The final step in after all those investigations is the simple simulation. The research implies usage of three main characteristics obtained from the AuverGrid dataset, which are wait time, run time and interarrival time to run some scenario. After setting up the environment next results were obtained: usage of servers in IT infrastructure (Fig. 3) and utilization of schedulers and servers (Fig. 4).

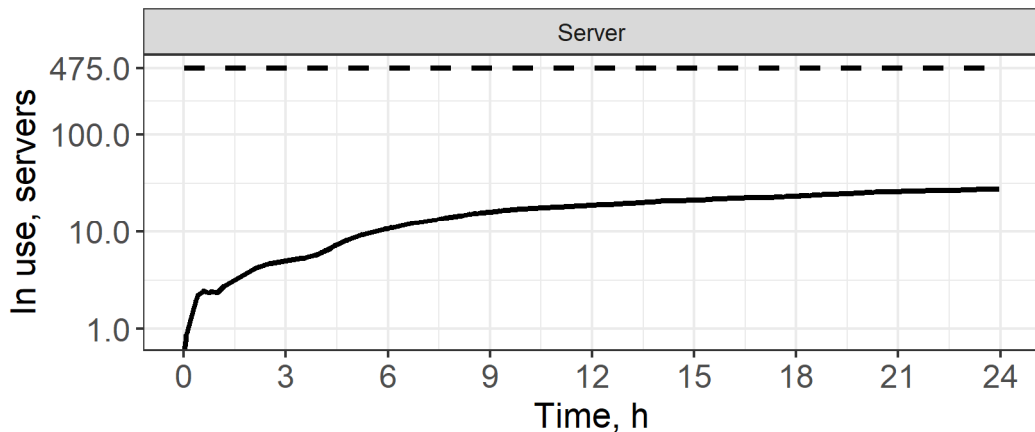


Figure 3 – The usage of servers in simulated environment

From the obtained results with the simulation parameters, related to the dataset very distinct behavior of the system can be observed which allows to make a conclusion that the system is overwhelmed with scheduling resource but not the processing resource.

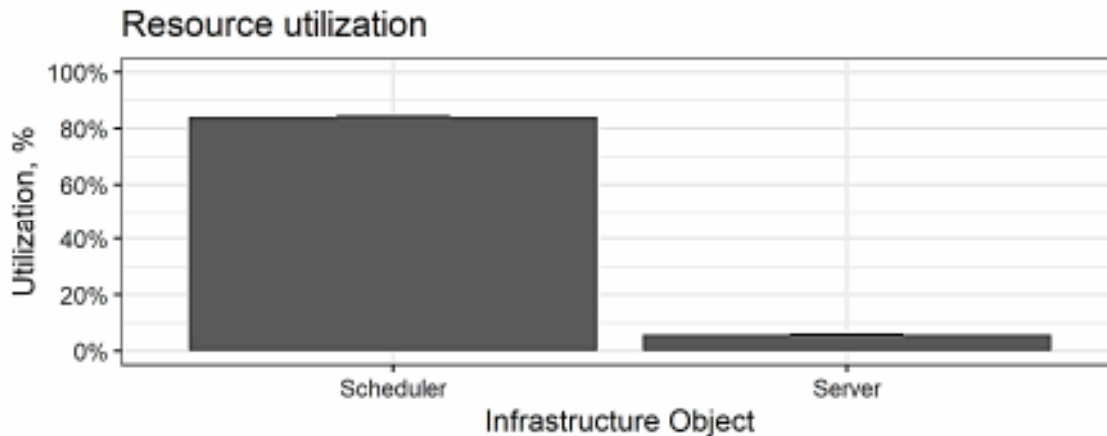


Figure 4 – The utilization of all entities in simulated environment

Original report on AuverGrid also shows that the system's overall utilization during a year period differs between average and maximum values, which are correspondent to 58 and 100 per cents. By knowing this we can make an assumption that the model is adequate for simulating similar to AuverGrid environment and perform some testing of management methods.

Conclusion. As a result of this research some models were discovered and connected with each other as an ontological description of some processes in IT infrastructure. Also, overview of common tools for simulating IT infrastructure has been performed.

The main difference from already existing simulation solutions is in putting new restriction on IT infrastructure as a research object which is a criticality aspect. It allows to construct basic scenarios of IT infrastructure functioning with making some blocking and attacking possible in order to replicate similar circumstances to real world.

Further research will focus on details of network functioning, cyber attacks peculiarities, resource management issues and user's perception issues during calculating quality of service values and checking SLA compliances.

REFERENCES

1. Papazoglou M. P. Service-Oriented Computing: State of the Art and Research Challenges / M. P. Papazoglou, P. Traverso, S. Dustdar, F. Leymann // Computer, vol. 40, no. 11. – Nov. 2007. – pp. 38-45. doi: 10.1109/MC.2007.400.
2. Rolik O. Decomposition-Compensation Method for IT Service Management / O. Rolik, V. Kolesnik, D. Halushko // Advances in Intelligent Systems and Computing, Advances in Intelligent Systems and Computing, 2017. – pp. 89–107. doi: 10.1007/978-3-319-44260-0_6

3. Mansouri N. Cloud computing simulators: A comprehensive review, *Simulation Modelling Practice and Theory* / Mansouri, R. Ghafari, B. Mohammad Hasani Zade // *Simulation Modelling Practice and Theory*, 2020. – Vo. 104. – 102144. ISSN 1569-190X, doi:10.1016/j.simpat.2020.102144.
4. Zola E. Discrete Event Simulation of Wireless Cellular Networks / E. Zola, I. Martin-Escalona, F. Barcelo-Arroyo // *Discrete Event Simulations*. – 2010. doi: 10.5772/9903.
5. Szczerbicka H. Discrete Event Simulation with Application to Computer Communication Systems Performance / H. Szczerbicka, K. Trivedi, P. Choudhary // *Information Technology*, 2004. – pp. 271–304. doi: 10.1007/1-4020-8159-6_10.
6. Shannon R. Systems Simulation: The Art and Science // R. Shannon, J. Johannes // *IEEE Transactions on Systems, Man, and Cybernetics*, 1976. – vol. -6. – no. 10. – pp. 723–724. Available: 10.1109/tsmc.1976.4309432.
7. Nguyen T. Stochastic Reward Net-based Modeling Approach for Availability Quantification of Data Center Systems / T. Nguyen, D. Min, E. Choi // *Dependability Engineering*, 2018. doi: 10.5772/intechopen.74306.
8. Rybnicek M. A Generic Approach to Critical Infrastructure Modeling and Simulation / M. Rybnicek, R. Poisel, M. Ruzicka, S. Tjoa // *2012 International Conference on Cyber Security*, 2012. doi: 10.1109/cybersecurity.2012.25.
9. Popov P. Preliminary Interdependency Analysis (PIA): Method and Tool Support / P. Popov // *Lecture Notes in Computer Science*, 2011. – pp. 1–8, doi: 10.1007/978-3-642-24124-6_1.
10. Netkachov O. Model-Based Evaluation of the Resilience of Critical Infrastructures Under Cyber Attacks / O. Netkachov, P. Popov and K. Salako // *Critical Information Infrastructures Security*, 2016. – pp. 231–243, doi: 10.1007/978-3-319-31664-2_24.
11. Bloomfield R. Preliminary interdependency analysis: An approach to support critical-infrastructure risk-assessment / R. Bloomfield, P. Popov, K. Salako, V. Stankovic, D. Wright // *Reliability Engineering & System Safety*, 2017. – vol. 167. – pp. 198–217, doi: 10.1016/j.ress.2017.05.030.
12. De Nicola A. A methodology for modeling and measuring interdependencies of information and communications systems used for public administration and eGovernment services // A. De Nicola, M. Villani, M. Brugnoli and G. D'Agostino // *International Journal of Critical Infrastructure Protection*, 2016. – vol. 14. – pp. 18–27. Available: 10.1016/j.ijcip.2016.06.001.
13. Rolik A. I. Upravlenie korporativnoy IT-infrastrukturoy / A.I. Rolik, S.F. Telenyk, M.V. Yasochka. – Kiev: Naukova dumka, 2018. – 576 p.
14. Ucar I. simmer: Discrete-Event Simulation for R / I. Ucar, B. Smeets, A. Azcorra // *Journal of Statistical Software*, 2019. – vol. 90. – no. 2. doi: 10.18637/jss.v090.i02.

15. Castro E.R.S. Probability Density Functions of the Packet Length for Computer Networks with Bimodal Traffic / E.R.S. Castro, M.S. Alencar, I.E. Fonseca, // International journal of Computer Networks & Communications, 2013. – vol. 5. – no. 3. – pp. 17–31. Available: 10.5121/ijcnc.2013.5302.
16. GWA-T-4 AuverGrid, Gwa.ewi.tudelft.nl, 2021. [Online]. Available: <http://gwa.ewi.tudelft.nl/datasets/gwa-t-4-auvergrid>. [Accessed: 01- Feb- 2021].
17. Iosup A. The Grid Workloads Archive /A. Iosup // Future Generation Computer Systems, 2008.– vol.24. – no.7. –pp. 672-686, Available: 10.1016/j.future.2008.02.003.

Received 12.01.2021.

Accepted 14.01.2021.

Симуляція IT-інфраструктури із розглядом аспектів критичності для управління якістю послуг

Перевірка гіпотез щодо управління якістю послуг в IT-інфраструктурі вимагає наявність великих і складних центрів обробки даних з достатньою кількістю ресурсів для дослідження різних можливих сценаріїв роботи інфраструктури під час надання IT-послуг. Для спрощення на сьогодні вже існує кілька десятків рішень для симуляції центрів обробки даних, проте більша частина з них зосереджена на продуктивності або на енергетичній складовій, проте не існує рішень, які можуть дозволити накладати додаткові обмеження та дослідити суб'єктивний показник якості послуг.

Метою дослідження було виявлення математичних моделей для накладання обмежень на моделі симуляції у вигляді критичності і запропонувати новий засіб для проведення симуляції IT-інфраструктури.

У якості методики дослідження виступало визначення існуючих способів моделювання центрів обробки даних та синтез модифікованого способу засобами моделювання та симуляції мовою програмування R.

Розроблений інструмент симуляції IT-інфраструктури дозволяє відтворити сценарії, подібні до реального життя, оскільки базуються на використанні значень, згенерованих випадковим чином згідно з щільністю випадкового розподілу значень з даних моніторингу реальної системи. Спрощений сценарій симуляції показав, що протягом симуляції тривалістю 24 умовних годин використання ресурсів планувальника зайняло майже 80 відсотків часу, при цьому обчислювальний ресурс простоював, що може розглядатися як відповідність до звіту використання справжніх даних.

*У висновку, для моделювання поведінки реальної системи окрім використання аналітичних моделей можна використовувати емпіричні моделі. При цьому, з невеликим похибками можливо отримати чіткий імовірнісний розподіл значень того чи іншого параметру реальної системи. Використання розширення *simmer* в мові програмування R дозволяє розширити ідею використання емпіричних даних для моделювання поведінки.*

Simulation of IT infrastructure with consideration of critical aspects for quality of service management

Testing hypotheses about service quality management in IT infrastructure requires large and complex data centers with sufficient resources to explore various possible scenarios of infrastructure operation during the provisioning of IT services. To simplify, there are already dozens

of solutions for simulating data centers, but most of them focus on performance or the energy component, but there are no solutions that can impose additional constraints and explore the subjective quality of services.

The objective of the research is to identify models for constraints on simulation models in the form of criticality and propose a new tool for conducting IT infrastructure simulations.

The primary method of the research is to determine of existing methods of modeling data centers and synthesis of a modified method by means of modeling and simulation with R programming language.

The developed IT infrastructure simulation tool allows to reproduce real-life scenarios, as they are based on the use of values generated randomly according to the density of random distribution of values from the monitoring data of the real system. The simplified simulation scenario has shown that during the simulation of 24-hour period, the scheduler's resource usage took up almost 80 percent of the time, with the computing resource idle, which can be considered consistent with the report on real data.

As a conclusion in addition to the use of analytical models, empirical models can be used to model the behavior of a real system. In this case, with small errors it is possible to obtain a clear probability distribution of the values of a parameter of the real system. The use of the simmer extension in the R programming language allows us to extend the idea of using empirical data to model behavior.

Ролік Олександр Іванович – завідувач кафедри автоматички та управління в технічних системах, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», д.т.н., професор.

Колеснік Валерій Миколайович – аспірант кафедри автоматички та управління в технічних системах, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»

Ролик Александр Иванович – заведующий кафедрой автоматички и управления в технических системах, Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского», д.т.н., профессор.

Колесник Валерий Николаевич – аспирант кафедры автоматички и управления в технических системах, Национальный технический университет Украины «Киевский политехнический институт имени Игоря Сикорского».

Rolik Oleksandr Ivanovych – the head of the department of Automatics and Control in Technical systems, National technical university of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”, Doctor of Science in engineering sciences, full professor.

Kolesnik Valerii Mykolaiovych – PhD student at department of Automatics and Control in Technical systems, National technical university of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”.