Conditionally infinite telecommunication resource for subscribers

Larysa Globa [1[0000-0003-3231-3012]], Mariia Skulysh [1[0000-0002-5141-1382]] and Siemens Eduard [2[0000-0002-2986-3614]]

[1] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" Kyiv, Ukraine
lgloba@its.kpi.ua, mskulysh@gmail.com
[2] Anhalt University of Applied Sciences Anhalt, Germany
eduard.siemens@hs-anhalt.de

Abstract. Modern communication systems should provide the end user with a conditionally infinite telecommunication resource. To ensure guaranteed quality, services are combined into groups. A group of similar services with the same service process requirements is called a slice. Using software-defined networks (SDN) allows to deploy a management system slices and telecommunications resources that are allocated for their maintenance. This is possible only using the latest telecommunication technologies, and seamless user connection management. The article proposes an original methodology for managing the servicing streams process in network service providers tunnels. The methodology uses the allowable service node load calculation, overload prediction and live migration algorithms to seamlessly change the service node for the flow.

Keywords: SDN, micro operator network, 5G, network slicing, communication networks, smart migration system.

1 Introduction

Nowadays, the number of Internet users are rapidly increasing along with the number of devices we use to get access to the Internet. The growing number of network services, more strict requirements to the quality of service and to the speed of the fixed and mobile Internet-access have caused the modernization of the network architecture, based on new requirements, standards and recommendations.

Due to the active development of the Internet of Things concept, devices connected to the Internet generate huge volumes of network traffic, which makes the further usage of the networks built only on routers and switches impossible.

It becomes one of the most important reasons of active development of the Software Defined Networks (SDN) and Network Functions Virtualization Services (NFVS), which are integral elements of the fifth generation network concept [1].

With these technologies, 5G networks can provide the necessary programmability, flexibility, scalability, etc., that are needed to build different logical (virtual) networks for a specific type of task, without changing the infrastructure platform. These logical

networks with shared resources are network slices. The purpose of this work is to analyze the current models and methods of network slicing in modern telecommunication networks, and their implementations.

A significant part of the new operators prefers to cooperate with small micro-telecommunication networks that cover the interior after the onset of the 5G era. Since these operators can provide various networks such as 3G, 4G, 5G and even Wi-Fi [1,2].

Up-to-day technology use more and more Network Function [7] that is a functional unit within a network infrastructure that has clearly defined external interfaces and well-defined functional behaviour. In practice, a network function is today a network node or physical device.

Innovative industry groups such as the ETSI ISG group (Industry Specification Group) for the NFV and ONF (the Open Networking Foundation) organization for the SDN created reference architectures, substantiated usage scenarios and changed the requirements for the components that are an integral part of NFV and SDN. The SDN network architecture supports the principles of network slicing because SDN allows you to manage a common infrastructure network and effectively support multiple client instances of the network [8].

Fig. 1. illustrates an example of integrating SDN and NFV technologies for implementing slicing in 5G communication networks [9].

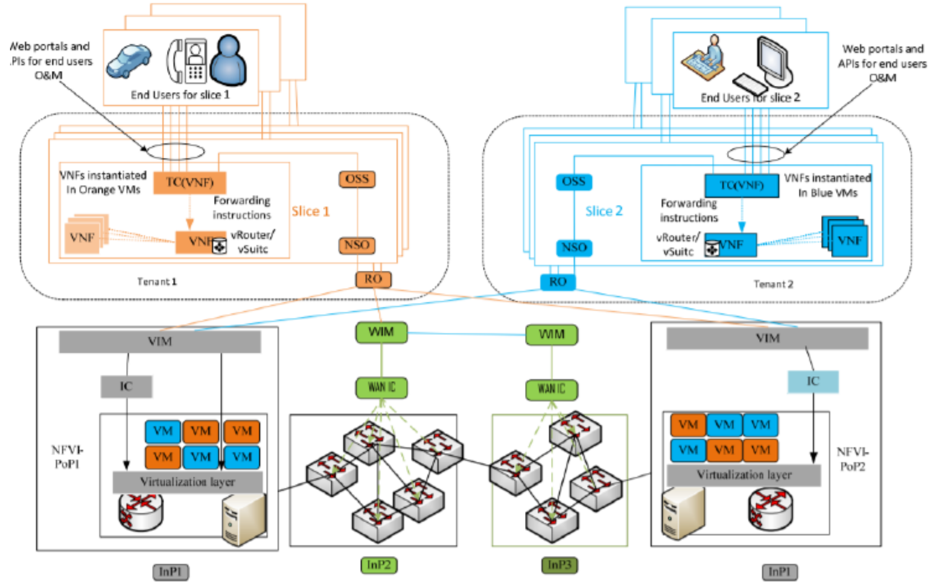

Fig. 1. Network slice architecture for SDN and NFV

2 STATE OF THE ART AND BACKGROUND

The task is to organize the maintenance of independent slices within the existing computation and telecommunication infrastructure. To ensure the maintenance of independent slices, the following features must be considered:

- it is required to provide each with sufficient amount of resources, both telecommunication and computing ones, for servicing virtualized network functions in order to provide service at a given quality level.
- it is necessary to take into account the nature of the load change in each slice for optimal allocation of resources between slices during the day.
- determine the conditions for the migration of slices in the telecommunications infrastructure, which will ensure the smooth operation of the system.

With regard to infrastructure design, there SDN/NFV network technology and tunnels to slice services are used. Infrastructure allows users to connect multiple interfaces using tunneling technology and running fast network connection to effectively strengthen the relationship networks.

In response to the demand telecom operator network resource distribution that allows users to gain access of nearby network resources, the paper [8] proposes network selection mechanism for a Micro Operator and uses decision tree theory to serve as the reference in determining the SDN traffic flows path. The proposed method disadvantage is that traffic will be distributed without taking into account the all network resources load.

The method shown in Fig. 2 application will allow to predict the moments network tunels overload and to migrate the slice's sessions between network tunels in time, which will allow to provide conditionally infinite bandwidth for each network slice.

The “The flows and node resources use monitoring” block involves the accumulation the communication channels congestion information, and information about the resource use dynamics by each service.

For the ediction of the possibility of exceeding the permissible value λ (the admissible input stream intensity), the method proposed in [15] is used. The basic method idea is to formulate requirements for the average input load on the basis of ergodic distribution for the possible states of the system, which will allow to make the most efficient use of the available physical resources of servicing the incoming application flow.

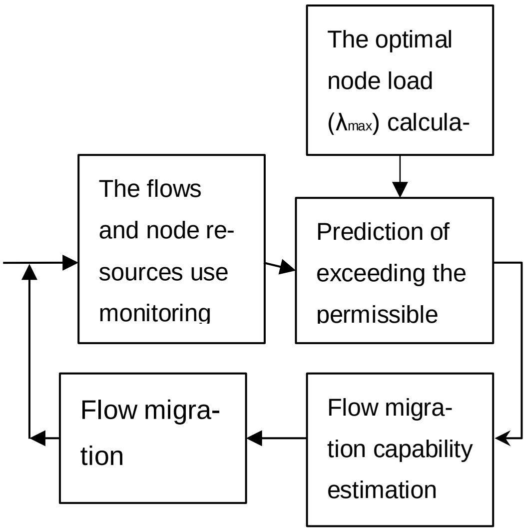

Fig. 2. Dynamic Flow Control Algorithm

For prediction of exceeding the permissible value λ we propose to use the method [7] consists of two stages: the calculation of the prediction interval based on the operation servicing node statistics and directly periodic forecasting of the load and the control of the sufficiency resources.

If periodic forecasting of the load showed that it overload is expected and the available slice resources are not enough to provide services at a given level, then the migration mechanism starts.

3 THE METHOD OF FORMING INPUT LOAD FLOW FOR EFFICIENT USE OF SERVICE RESOURCES

The main method objective is to formulate requirements for the average input load on the basis of ergodic distribution for the possible system states, which will allow to make the most efficient available physical resources of servicing the incoming flow.
The servicing process is modeled as an n-channel servicing device, the service time in the channel is a random variable distributed by Poisson law.

A. Input data

n is number of channels for simultaneous requests service.

μ is the service intensity,

G – is the number of resources involved in servicing slices,

v^g – the amount of g-resource required to serve in a single request, $g = \overline{1, G}$)

V^g – д is the available volume of the g-th resource shared by requests.

s is the allowable number of service queue requests.

R is the percentage of applications served on the system not exceeding the allowable delay time.

l is the number of queued requests that block requests from being sent to the system, according to early overload prevention algorithms.

B. Problem

It's necessury to find the recommended value for the input stream intensity (λ)

Application of the proposed method consists of two steps.

Step 1. According to K. Zhernovyi's models, find ergodic distribution of the number of applications in the system for multi-channel service system, by the formulas:

$$p_0 = \frac{1-\beta^{s-l}}{A_n(\alpha,\beta)}, \qquad \beta \neq 1, \qquad \alpha = \lambda/\mu, \qquad \beta = \lambda/n$$

$$A_n(\alpha,\beta) = (1-\beta^{s-l})\sum_{k=0}^{n}\frac{\alpha^k}{k!} + \frac{\alpha^n}{n!}\left(\frac{\beta-\beta^{s-l+1}}{1-\beta} - (s-l)\beta^{s+1}\right)$$

$$p_k = \frac{\alpha^k}{k!}p_0 \qquad (k=\overline{1,n})$$

$$p_{n+k} = \frac{\alpha^n}{n!}\beta^k p_0 \quad (k=\overline{1,l})$$

$$p_{n+k} = \frac{\alpha^n}{n!}\frac{\beta^k-\beta^s}{1-\beta^{s-l}}p_0 \; (k=\overline{l+1,s-1})$$

$$p_{n+s} = \frac{\alpha^n}{n!}\frac{(1-\beta)\beta^s}{A_n(\alpha,\beta)} \;\; (k=\overline{n+l+1,n+s-1})$$

where $\beta = 1, \quad \alpha = \lambda/\mu, \;\; \beta = \lambda/n$, and

$p_k = \frac{n^k}{k!}p_0 \;\; (k=\overline{1,n}); \;\; p_{n+k} = \frac{n^n}{k!}p_0 \;\; (k=\overline{1,l});$

$p_{n+s} = \frac{n^n}{n!}p_0 \;\; (s-l-1)p_{n+s} \Longrightarrow$

$$\Longrightarrow p_{n+s} = \frac{n^n}{n!\,(s-l)}p_0 \;\; \left(k=\overline{n+l+1,n+s-1}\right);$$

$$p_{n+k} = \frac{n^k}{n!}p_0 - \;(k-l)p_{n+s} = \frac{n^n}{n!}\,\frac{s-k}{s-l}p_0 \qquad (k=\overline{l+1,s-1})$$

Step 2. Solving the optimization problem of finding the maximum load, which will ensure the fulfilment of conditions for an acceptable amount of service resources.

$$\lambda \to max$$

$$\begin{cases} 4 * \left(\sum_{i=1}^{n} i v_k^g \, \mathrm{p}_i + n v_k^g \sum_{i=n+1}^{s} \mathrm{p}_i\right) \leq V^g, g = \overline{1,G} \\ \sum_{i=1}^{s} \mathrm{p}_i \leq R \end{cases}$$

Incoming stream balancing in a billing system is performed using a modified Round Robin scheme. To calculate the recommended input intensity value of the entire system, it is sufficient to analyze the metrics from one DOCS server and scale the obtained values to the entire subsystem.

Calculation of the optimum value of the input stream intensity for an existing system

To calculate the optimal value of λ, we analyzed the metrics available in the billing subsystem for the busiest day in 2017 (11/24/2017 Black Friday). The maximum input stream intensity for a known number of sessions was calculated at peak hours, at the time when the service failure metrics (REJECTS) were signaled. The analysis the service intensity of the application (μ) and the concurrent value of the input stream intensity (λ) for each record was calculated. Based on the data obtained, we can conclude that the maximum allowable value of the input stream intensity for a single OCS process in which no service degradation occurs (no REJECTS) corresponds to **λ = 860** at **μ = 430**.

A graphical representation of the calculated application service intensity (μ) and the simultaneous input stream intensity value (λ) for a single OCS process is shown in Fig.3

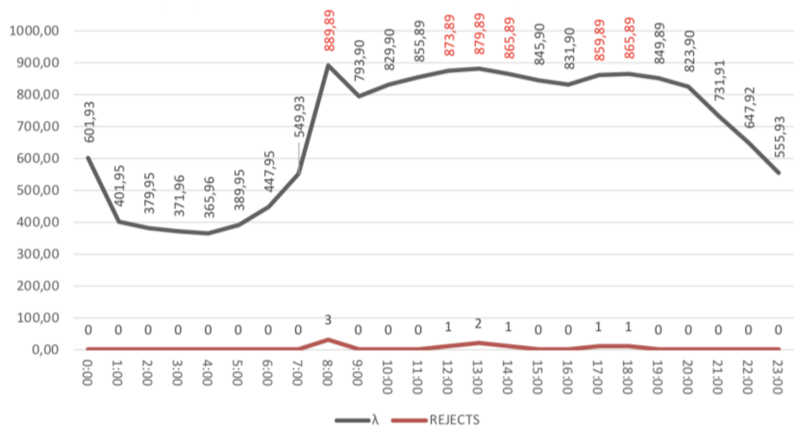

Fig. 3. Graph of the intensity value of λ for 11/24/2017

Short-term load forecasting method

Short-term scheduling is an advanced ARIMA prediction method - an autoregressive method with a rolling expectation. However, unlike the known method, it is suggested to solve the problem of finding a minimum rolling interval, the use of which will satisfy the requirements. This will minimize the number of floating point operations to perform the prediction, which will provide the optimum speed of the prediction execution.

The proposed method consists of two steps - the calculation of the forecasting interval based on the statistics of operation of this service node and directly periodic load forecasting and control of resource adequacy.

Incoming data:

- $T_{п}$ - the time interval for which the forecast is required.
- λ_i - number of applications per 1ms, ($i\epsilon 0,..N$), $N=T_{inf}/1ms$, $\lambda_i \epsilon \Lambda$, where Λ - the set of statistics for the number of applications received over time T_{inf} (it is

first specified, then corrected in step 2 of the method) before the forecast, $|\Lambda|= N$.

- T_{for} - forecasting period, the time after which, runs forecasting algorithm.
- M – the maximum number of requests that can be served with a given configuration of the serving device.
- P – the probability of a prediction error.

Short-term statistics is collected locally with the operating device and stored no longer than $T_{inf} + T_{for}$, a sampling interval of 1 ms.

Output data:

- $z\epsilon\{0,1\}$– z=0 do not change the configuration; z=1 change the configuration.

Method algorithm:

Preparatory stage. Learning the system based on statistics. Finding the minimum T_{inf} (Information Collection Time Interval)

$T_{inf}\rightarrow min$,

for which the limit is performed:

$\bar{\lambda}_{Tfor} + 3\sigma > M$,

where $\bar{\lambda}_{Tfor}$ and σ – are calculated according to the main step.

The restriction is satisfied for P statistical samples obtained at different time intervals.

Solution method: validate values for a sequence formed by a principle $T_{nf}^{k+1}= T_{inf}^{k}+\Delta$; $T_{inf}^{0}= T_{for}$.

The main stage of dynamic control.

1. Analysis of statistics $\lambda_{i,}$ for the time interval $T_{інф}$, preceding the moment of calculation. Construction of coefficients â and $\hat{b}$ by the least square method:

$\lambda = ai + b$

2. Calculate $\bar{\lambda}_{Tfor}$ =âT_{for}+$\hat{b}$.
3. If $\bar{\lambda}_{Tfor} + 3\sigma \leq M$, than z=0, else z=1.

4 METHOD OF AUTOMATIC LIVE MIGRATION

The main task is to provide automatic balancing of the load of physical resources in telecommunication nodes. In this case, it is necessary to avoid too much load on one node and inefficient resources use of the second one. This balancing mechanism is called “smart migration”.

We formulate the main tasks of the system:

1. Make the migration process invisible to users.
2. Use the migration mechanism in order to optimize the state of the telecommunication network.

3. Ensure a minimum migration time, since too long process will negatively affect the state of the system.
4. The operative and backup channel of one slice should be located in different physical nodes. For this purpose, it is necessary to ensure high availability of the telecommunication node.
5. Provide anti-loop protection to prevent endless migration of the same slice.
6. Provide protection against failures.
7. Work in cluster mode in the case of multiple migrations.

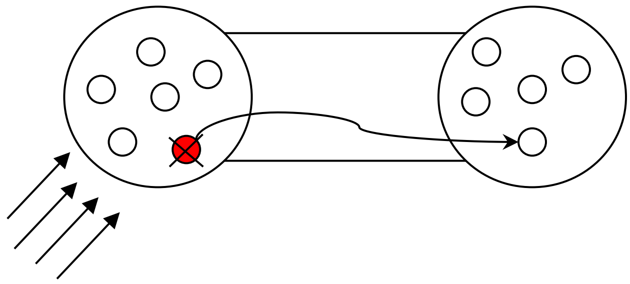

Fig. 4. Migration of flows between the neighboring micro-operator networks

Migration System Architecture

Each slice forms data from the load statistics on the telecommunication node. They are periodically read and converted to metrics, which are placed in a specific storage. Having access to such a repository, one can analyze the dynamics of resource consumption in each telecommunication node. It is also possible to get information about the amount of physical servers' resources.

The control unit makes a decision about the need for migration and distributes the load between nodes (Fig. 4). The sequence of the process is as follows:

1. First, candidates for migration are selected and placed in a distributed queue.
2. The queue is processed by a special process that analyzes the number of elements in it and selects the one to be transferred.

Physical migration is a synchronous process without breaks and returns. A special mechanism allows to avoid delays and errors. The main task of the process is to choose the right candidate for migration and transfer it to the desired system node, ensuring its optimal functioning. For this purpose, it is necessary to solve the problem of multidimensional optimization. There are several algorithms for solving the problem.

Simple algorithm:

1. Analyze the load on each node.
2. Select the node that uses resources most optimally.
3. Move the slice to this node.

This algorithm will work effectively with a small number of telecommunication nodes. However, for a large network, it is not optimal. In this case, you will need a more complex algorithm with hard and soft restrictions. Its essence is to determine two kinds

of rules: those that cannot be violated in any case, and those that can be neglected under certain circumstances. You also need to identify three types of problem solutions:

1. Possible solutions – bad decisions that violate hard rules.

2. Feasible solutions - violate some of the soft rules, but do not violate hard rules.

3. Optimal solutions - satisfy both hard and soft restrictions. These are the best decisions received as soon as possible.

First, we define hard and soft restrictions for a telecommunication system that needs to be optimized:

Hard restrictions:

1. The target node must have enough resources to switch the slice. You also need to provide redundancy resources.

2. It is not allowed to move a slice to its own physical node.

3. A physical node cannot contain flows of the same slice. Thus, in the event of a system failure, there will be minimal data loss.

Soft restrictions:

1. It is necessary to ensure the migration of the most loaded slice streams.

2. The target physical node must have minimal load.

This algorithm has one significant disadvantage: it does not have tools that will take into account trends in the rate of consumption of resources by different slices. We conducted a study on the effectiveness of processing statistical data during the selection of the migration flow and the target node. For the correct choice of the node to which the flows will be redirected, it is necessary to evaluate the dynamics of the resources usage of a particular server. This takes into account the load created by the flow of migrating slices.

To determine the moment of migration, it is necessary to evaluate the current statistics of resource use with a given time interval and build a statistical trend, taking into account the number of served requests. Based on this trend, the probability that the content of the containers of the evaluated node will exceed the available amount of resources is estimated. Then the migration process begins. A method for assessing the adequacy of resources is presented in [10].

Based on the statistics received, the sum of the flows of the selected slices will be generated. After evaluating the upper limit of the node capacity, it will be possible to make a decision on the need for migration.

Conclusion

This document uses SDN and NFV technologies as the basis and combines network streaming and tunneling technologies to create a network infrastructure model for MSO using a smart migration mechanism. This model allows users of different MOs to connect using tunneling technology, and then implement a fast network connection to effectively improve network interaction, while balancing the load between all nodes of a given network. To meet the needs of the regional micro-operator service, this article proposes a DTBFR mechanism that uses decision tree theory as the basis for making SDN-based traffic decisions. As a distribution and control of the load on the nodes, we use the method of slices working together in the existing telecommunication foreign

infrastructure, which ensures the automatic distribution of telecommunication and computing resources of the system depending on the load and allows solving the problem of peak loads and idle resources. This method of automatic load balancing of telecommunication nodes (“smart migration”) does not allow overloading one node and downtime of another node. The functions used by the regional micro-operator service can effectively reduce the load on the datacenter on the Internet and accelerate the development of the regional computer service in the future 5G network. And the “smart migration” method will allow rational use of network resources.

References

1. A. Raschellà, F. Bouhafs, G. C. Deepak, and M. Mackay, “QoS aware radio access technology selection framework in heterogeneous networks using SDN,” Journal of Communications and Networks, vol. 19, no. 6, pp. 577–586, 2017.
2. M. Matinmikko, M. Latva-aho, P. Ahokangas, S. Yrjölä, and T. Koivumäki, “Micro operators to boost local service delivery in 5G,” Wireless Personal Communications, vol. 95, no. 1, pp. 69–82, 2017.
3. J. S. Walia, H. Hammainen, and M. Matinmikko, “5G Micro-operators for the future campus: A techno-economic study,” in Proceedings of the 2017 Internet of Things - Business Models, Users, and Networks, pp. 1–8, Copenhagen, Denmark, November 2017.
4. M. Matinmikko-Blue and M. Latva-aho, “Micro operators accelerating 5G deployment,” in Proceedings of the 2017 IEEE International Conference on Industrial and Information Systems (ICIIS), pp. 1–5, Peradeniya, Sri Lanka, December 2017.
5. P. Ahokangas, S. Moqaddamerad, and M. Matinmikko, “Future micro operators business models in 5G,” The Business and Management Review, vol. 7, no. 5, pp. 143–149, 2016.
6. P. Mach and Z. Becvar, “Mobile Edge Computing: A Survey on Architecture and Computation Offloading,” IEEE Communications Surveys & Tutorials, vol. 19, no. 3, pp. 1628–1656, 2017
7. Globa, L., Skulysh, M., Romanov, O., & Nesterenko, M. (2018, November). Quality Control for Mobile Communication Management Services in Hybrid Environment. In The International Conference on Information and Telecommunication Technologies and Radio Electronics (pp. 76-100). Springer, Cham.
8. Skulysh, M. A., Romanov, O. I., Globa, L. S., & Husyeva, I. I. (2018, September). Managing the Process of Servicing Hybrid Telecommunications Services. Quality Control and Interaction Procedure of Service Subsystems. In International Multi-Conference on Advanced Computer Systems (pp. 244-256). Springer, Cham.
9. Chia-Wei Tseng, Yu-Kai Huang, Fan-Hsun Tseng, Yao-Tsung Yang, Chien-Chang Liu, and Li-Der Chou, “Micro Operator Design Pattern in 5G SDN/NFV Network,” Wireless Communications and Mobile Computing, vol. 2018, Article ID 3471610, 14 pages, 2018.
10. Semenova, O., Semenov, A., & Voitsekhovska, O. (2019, July). Neuro-Fuzzy Controller for Handover Operation in 5G Heterogeneous Networks. In 2019 3rd International Conference on Advanced Information and Communications Technologies (AICT) (pp. 382-386). IEEE.
11. Skulysh, M. (2017, September). The method of resources involvement scheduling based on the long-term statistics ensuring quality and performance parameters. In 2017 International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo) (pp. 1-4). IEEE.

12. M. Skulysh; O. Romanov. The structure of a mobile provider network with network functions virtualization. In: 2018 14th International Conference on Advanced Trends in Radioelecrtronics, Telecommunications and Computer Engineering (TCSET). IEEE, 2018. p. 1032-1034.
13. Globa, L., Kurdecha, V., Ishchenko, I., Zakharchuk, A., & Kunieva, N. (2018, June). The intellectual IoT-system for monitoring the base station quality of service. In 2018 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom) (pp. 1-5). IEEE.
14. Romanov, O. I., Hordashnyk, Y. S., & Dong, T. T. (2017, September). Method for calculating the energy loss of a light signal in a telecommunication Li-Fi system. In 2017 international conference on information and telecommunication technologies and radio electronics (UkrMiCo) (pp. 1-7). IEEE.
15. M. Skulysh, F. Shilov and A. Safaryan. " Investigation of the method of computing resources optimal choice for billing systems effectiveness." CONTROL, NAVIGATION AND COMMUNICATION SYSTEMS. ACADEMIC JOURNAL 3.49 (2018): 147-152. DOI: https://doi.org/10.26906/SUNZ.2018.3.147
16. Globa, L. S., Ishchenko, I., Kurdecha, V., Zakharchuk, A., & Zvonarov, O. (2016, June). An approach to the Internet of Things system with nomadic units developing. In 2016 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom) (pp. 1-5). IEEE.