http://doi.org/10.35784/iapgos.3452

received: 19.01.2023 | revised: 12.03.2023 | accepted: 12.03.2023 | available online: 31.03.2023

DEVELOPMENT OF A MODIFIED METHOD OF NETWORK TRAFFIC FORMING

Valeriy Kozlovskiy[1], Natalia Yakymchuk[2], Yosyp Selepyna[2], Sergiy Moroz[2], Anatolii Tkachuk[2]

[1]National Aviation University, Faculty of Cybernetics, Computer and Software Engineering, Department of Information Security, Kyiv, Ukraine, [2]Lutsk National Technical University, Faculty of Computer and Information Technologies, Department of Electronics and Telecommunications, Lutsk, Ukraine

Abstract. *Based on the analysis of the statistical characteristics of heterogeneous network traffic of "Quadruple Play" mobile subscribers, it is shown that it cannot be represented by Poisson or Erlang distributions. It is shown that for such similar traffic, the rate of growth of the required buffer volume increases as the Hurst parameter increases. A method of adaptive traffic formation using the control of the length of intervals of the intensity of the arrival of data packets has been developed. Congestion management is carried out by changing the frequency of the marker generator (GM) on the basis of the results of forecasting the required bandwidth of the system and the required buffer size. The shaper adapts to changes in the length and instantaneous intensity of packet input in real-time.*

Keywords: communication channels, mass service theory, self-similarity traffic

OPRACOWANIE ZMODYFIKOWANEJ METODY FORMOWANIA RUCHU SIECIOWEGO

Streszczenie. *Na podstawie analizy charakterystyk statystycznych heterogenicznego ruchu sieciowego abonentów telefonii komórkowej "Quadruple Play" wykazano, że nie może on być reprezentowany przez rozkłady Poissona lub Erlanga. Pokazano, że dla takiego podobnego ruchu tempo wzrostu wymaganej objętości bufora wzrasta wraz ze wzrostem parametru Hursta. Opracowano metodę adaptacyjnego kształtowania ruchu wykorzystującą sterowanie długością interwałów natężenia napływu pakietów danych. Zarządzanie ograniczeniami odbywa się poprzez zmianę częstotliwości generatora znaczników (GM) na podstawie wyników prognozowania wymaganej przepustowości systemu oraz wymaganej wielkości bufora. Shaper dostosowuje się do zmian długości i chwilowej intensywności wprowadzanych pakietów w czasie rzeczywistym.*

Słowa kluczowe: kanał komunikacyjny, teoria usług masowych, ruch samopodobieństwa

Introduction

Maintenance of heterogeneous traffic flows. Traffic may include voice, video, data, and traffic of mobile subscribers "Quadruple Play". The statistical characteristics of such heterogeneous network traffic can no longer be represented by Poisson or Erlang distributions. Such traffic is self-similar (fractal) and is described with satisfactory accuracy by models of statistical distributions with heavy tails [13].

A characteristic feature of random processes, which are described by distributions with heavy tails, is the combination of the sufficiently low average intensity of the process with sporadic bursts of instantaneous intensity. In addition, in the presence of a heavy tail of the distribution, the convergence of the sample mean to the mathematical expectation deteriorates.

The self-similarity of real-time traffic is considered in the examples of different components of heterogeneous traffic. One of the most important traffic components of modern telecommunications networks is language traffic. The characteristics of the traffic generated by a single voice source are highly dependent on the language coder (codec) used. Most commonly, voice traffic can be viewed as a superposition of a large number of individual independent ON/OFF sources transmitting at the same intensity but with durations distributed according to a heavy-tailed distribution [2, 8].

Initially, the Poisson model was used to analyze traffic in data networks. In the case of high loading, the Poisson model does not agree with the simulation results, and the combined process is highly correlated and exhibits properties of long-term dependence. The self-similarity assessment conducted using various tests showed that the self-similarity property of the total traffic should be taken into account when modeling in the case of high call volume. Both fractal Gaussian noise and more complex multi-component models can be used to model aggregated VoIP traffic [7].

1. Literature review

When transmitting video traffic, the method of video encoding is important for determining self-similarity. Typically, interframe coding is used. Since adjacent frames are not very different from each other (because the motion is continuous), this results in a significant correlation between adjacent frames. You can protect yourself from transmission errors by periodically transmitting a full frame. It is also known that video traffic is highly correlated and exhibits long-term dependence, which leads to serious estimation drift and difficulties in assessing convergence. Estimating the self-similar properties of video traffic over large intervals using the R/S analysis method has advantages over other methods of estimating self-similarity. The R/S analysis functions themselves exhibit a scaling property and therefore constitute an optimal reference point system on which to track large-scale phenomena [18, 19].

When studying Variable Bit Rate (VBR) traffic, it was established that long-term dependence looks like artificial non-stationarity. And such a process is called a process with a Level Shift (SL). At the same time, the periodicity of the traffic corresponds to taking into account only the contour correlation function of the video. Thus, when describing network traffic using fractal models, care must be taken not to confuse real non-stationarities with stationary fractal behavior with sporadic bursts of traffic intensity [12]. These effects can produce the same results in many statistical tests. Classical methods of mathematical statistics and fractal theory can distinguish non-stationarity and long-term dependence or estimate the Hurst index in the presence of some types of non-stationarity [15].

As for WWW connections, the self-similarity of network traffic depends on user activity. With lower user activity, which results in a lower average speed, behavior is observed that is typically modeled by Poisson processes. Aggregate WWW traffic at low network load (without losses in buffers) is also well modeled by Poisson processes [14].

Self-similar traffic requires more memory resources of buffer devices than the classic M/M/1 model, which is considered the least favorable compared to others (for example, with a constant or Gaussian service time distribution). The rate of growth of the required amount of memory increases with the increase of the Hurst parameter, which is mainly due to the degree of the grouping of homogeneous packets and bursts of network load [10, 11].

It can also be argued that simply increasing buffer memory (hardware or software) is inefficient. In addition, with the expected increase in the share of data traffic in the total volume, the degree of self-similarity will increase, and the dependence $\rho(q_{buff})$ will grow more and more sharply [9].

To eliminate or at least reduce the harmful effect of traffic similarity, methods of regulation or shaping of the incoming flow (policing-shaping) are usually used [1, 17].

2. Researches methodology

As it has been noted, to increase the productivity of information and telecommunication networks, the mass service theory (MST) is used. First, the analytical study of queues concerns the quantification of the phenomenon of waiting for queues using key performance parameters, such as the average length of the queue (average number of customers in the queue; in the analysis of the passage of packets through the switching node – the average number of packets in the buffer memory), the average queue time and average network utilization [5]. Secondly, using simulation modeling methods, the results of the analysis are obtained to solve the following tasks [16]:

a) comparison of waiting time and average number of messages in queues;
b) comparison of parameters of queues M/M/1, M/D/1, and Qd/D/1 and study of the impact of the quality of traffic formation on the probability of blocking and dropping packets for different waiting models.

Therefore, the mathematical model of the switching network node is considered a single-line mass service system (MSS). The statistical characteristics of heterogeneous network traffic at the entrance of such an MSS are determined by the fact that the traffic includes speech, video, data, and traffic of mobile subscribers (Quadruple Play traffic). Such traffic is self-similar by definition (fractal). It is described with satisfactory accuracy by the models G|G|1, G|M|1 or G|G|1|k, G|M|1|k.

When serving requests in a device with a buffer memory of infinite capacity, arbitrary service disciplines can, in principle, give the same results [4, 14]. However, with a limited lifetime of packets T_{live}, a necessary condition for their preservation should be compliance with the inequality $T_{serv} \leq T_{live}$, where T_{serv} is the service time (or the average time in the case of a service model with a random length of time) [6].

Also very important is the question of the acceptability of service disciplines (FIFO, LIFO, etc.) in a discrete-time system in terms of service delay and service failure rates. However, regardless of the specific service discipline, the following statement is valid: the better the degree of smoothing of network traffic dispersion, the slower the queue in the buffer memory of the serving device will grow [3].

3. Results

Let's consider the functionality of packet transmission efficiency with traffic shaping and smoothing. The effectiveness of packet transmission depends not only on the quality of adaptation but also on key network parameters. Fig. 1 shows the scheme of sequential processing of packet flow parameters.

Fig. 1. Sequential flow processing

Classification is carried out according to the hierarchy of priorities; regulation and shaping are intended to transform into a periodic or at least quasi-periodic flow. For systems without feedback, solving the alignment problem data transfer rate can be carried out using "leaky bucket" and "marker bucket" algorithms. The "leaky bucket" algorithm is a traffic control mechanism when a portion of the packet flow that exceeds network bandwidth is simply discarded or marked as redundant or low priority.

When passing through the shapers of the "leaky bucket" or "marker bucket" type, the MSS model and, accordingly, the traffic statistics undergo transformations. In particular, at the output of the "marker bucket" shaper, the traffic acquires a quasi-deterministic (Qd) character with a packet tracking period T_{arr}. It is described by the Qd|G|1|k or Qd|M|1|k models. The packet tracking period takes the form $T_{arr} = T_0 + \xi(n)$, where $\xi(n)$ is a normally distributed random variable with zero mathematical expectation. Root square deviation (RMS) is $\sigma_{\xi} < T_0/6$.

With ideal regulation and shaping of the input flow, the latter becomes deterministic or close to it quasi-deterministic. With a deterministic order of incoming applications and a deterministic processing time, the graph of queue growth is a linear-broken line (Fig. 2). There, for comparison, the dependence of the length of the queue of applications for the M/D/1 model – Poisson flow of applications and deterministic service time is given. Note that in practice both the traffic at the output of the shaper and the packet processing time is quasi-deterministic (Qd).

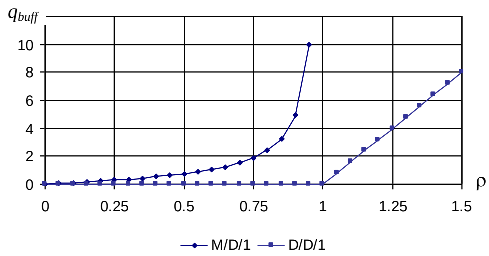

Fig. 2. Dependences of the length of the request queue (required buffer memory) on the utilization factor for incoming traffic models

The formation mechanism follows the rule of placing packets in a queue. Each egress port has a set of queues assigned different priorities or weights. The highest (first) priority is given to packets whose arrival intensity does not exceed C_{IR}. The second priority is assigned to packets whose arrival intensity is higher than C_{IR}, but not higher than E_{IR}. Finally, the lowest (third) priority is assigned to packets whose arrival intensity is higher than E_{IR}. Conditional determination of the priority of packages is carried out by different colors: green (first priority), yellow (second priority), and red (third priority).

In the process of traffic formation, packets are placed in queues with different priorities. Since all queues have a finite length, packets cannot be placed in them indefinitely. If the switching node's buffer is full, new packets are discarded.

Discard rules are different for packages of different colors. Green packets obey the most lenient rules, and it can be expected that most of these packets will be forwarded to the output queue. The rules for yellow packets are stricter, so their number in the output queue is smaller. The strictest rules are for red packets, which are detected when the buffer overflows.

The general concept of building a shaper and assigning priorities is based on key performance parameters. Network elements and sensors used as service resource analyzers are placed in certain nodes of the network infrastructure to collect performance-related data. For example, these can be cumulative counters of protocol events. At regular intervals, in near real-time, this performance information is transmitted to higher-level service and performance management systems.

Now let's develop a modified method of traffic formation using an adaptive multi-speed marker bucket.

Let the flow shaper have variable parameters: the average intensity of the flow – C_{IR} and the limit intensity of the flow – E_{IR} (Fig. 3). They are determined both by the speed and acceleration of the growth of flow intensity and, accordingly, by changes in the speed and acceleration of buffer filling.

Determining the priority of packets and, accordingly, their conditional coloring is as follows:

- if the short-term intensity of the flow does not exceed the C_{IR} value, the packets are assigned a higher priority (colored in green). Certain package delivery guarantees are provided;
- if the short-term flow intensity is higher than the C_{IR}, but not higher than the E_{IR}, the packets are assigned a medium priority (coloring in yellow). Delivery of packages is carried out with the greatest efforts, but without providing guarantees (Best Effort delivery mode);
- if the short-term intensity of the flow exceeds the value of the E_{IR}, the packets are assigned a lower priority (colored in red). Delivery is not guaranteed, some packets are rejected until the value of the short-term intensity drops below E_{IR}. The remaining packets go back into the yellow area.

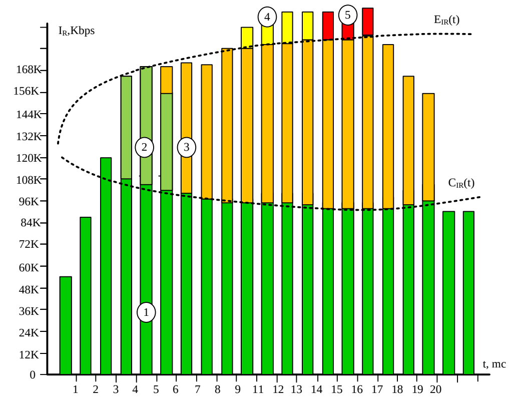

Fig. 3. Load measurement and packet specification management with an adaptively changing "yellow area" window size: 1 – green packets; 2 – packets that remain conditionally green in case of a short-term exceedance of C_{IR} no more than m seconds; 3 – yellow packets whose flow intensity is higher than C_{IR}, but lower than E_{IR}; 4 – packets that remain conditionally yellow in case of a short-term exceedance of E_{IR} no more than n seconds, n > m; 5 – red packets that are discarded first

In contrast to the classic method of traffic formation, the length of the C_{IR} and E_{IR} exceeding intervals is monitored at the same time. Based on the control results, the acceptable values of the number m of conditionally green packets and the number n of conditionally yellow packets are calculated.

Thus, with additional consideration of flow parameters and queue growth rate in the buffer memory, the set of parameters to be optimized expands to two elements, which are described by an additional color: conditionally green and conditionally yellow packets. It is natural to expect a corresponding acceleration of updates of information about loading parameters and a reduction in the reaction time of the marker bucket. These considerations are verified by calculations and computer simulations.

Short-term intensities E_{IR} and C_{IR} will vary within limits that depend on the maximum bandwidth of the switch. Therefore, it is advisable to adapt to the average intensity of packets by changing the intensities of E_{IR} and C_{IR} managing the size of the "yellow range".

The M-th order integrator with variable weighting coefficients $k_1 = k_1(t)$, $k_2 = k_2(t), \ldots k_m = k_m(t)$ can estimate speed, acceleration and higher derivatives of the degree of buffer filling. Controlling signals $y_{db}(t-\tau), y_{tb}(t-\tau), y_{dg}(t-\tau)$, regulate the total size of data and markers in the buffers, as well as the frequency of arrival of markers. The parameters of these signals are determined by traffic parameters, primarily by the degree of its self-similarity. In practice, it does not make sense to calculate derivatives of a higher order than the second (velocity and acceleration), since the accuracy of the statistical estimation deteriorates rapidly.

Figure 4 shows the scheme of the M-th order adaptive shaper. According to the above considerations, you should choose $M \le 2$.

Management of the size of the "yellow area" is carried out by changing the frequency of the marker generator (GM) on the basis of the results of forecasting the required bandwidth of the system and the required buffer size. The scheme of the additional forecasting module developed in is based on the Smith predictor, The modification of the Smith predictor consists in replacing the exponential smoothing function with a power-law smoothing function, which is optimal for smoothing self-similar processes with heavy tails of probability distributions.

The main differences between the multi-speed adaptive shaper and the traditional "marker bucket" shaper are as follows.

1. When the counter counts the number $n(t)$ of packets arriving at the input of the generator, the evaluation period T_{est} changes in accordance with the current frequency $f_{mg}(t_{curr})$, assigned to the token generator:

$$T_{ets}(t_{curr}) = \frac{1}{f_{mg}(t_{curr})}$$

2. The accumulator is a digital adder with saturation. This scheme of the adder was chosen in order to avoid overflow fluctuations that may occur when numbers are represented in the additional code.

3. The problem of fractal traffic is sporadic significant bursts of intensity at a rather insignificant average level of intensity over a relatively long observation interval. Therefore, we offer a new approach - an algorithm for ensuring the quality of service, in the presence of the effect of self-similarity. The idea of this algorithm is to modify the mechanisms of traffic formation to ensure the required QoS classes by introducing an additional module for predicting the required buffer size B_{size} for some time ahead according to changes in the intensity of incoming packets. At the same time, the traffic does not conform to a preset profile, but on the contrary, the system bandwidth adapts to the traffic profile. At the same time, losses are reduced and the use of allocated resources is improved.

4. The formation algorithm to prevent traffic congestion works as follows:

- the receiver measures the loss coefficient and returns it to the sender;
- the sender uses these feedback messages to calculate the Round-Trip Time (RTT)
- sender uses RTT and loss ratio to calculate transmission speed;
- the sender adjusts the sending speed.

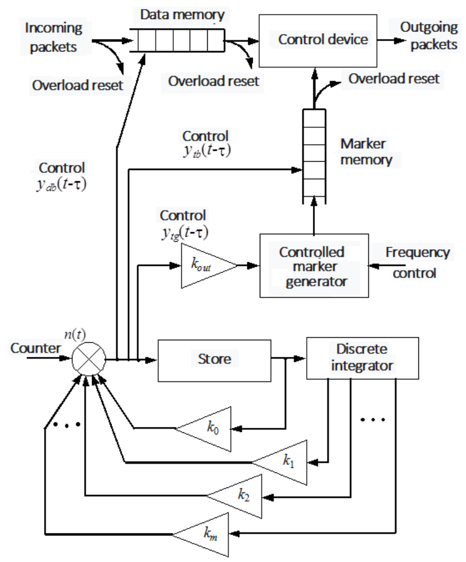

Fig. 4. Scheme of a multi-speed adaptive marker former (the device for controlling the size of the "yellow area" is not shown)

A multi-channel system with a divided buffer memory has been developed, which can be interpreted as a parallel structure with single-channel service systems and the same service disciplines.

Assume that each traffic stream has the form Triple Play or Quadruple Play. Before entering the network, it is formed by a multi-channel former. The guaranteed speed r_i of the i-th flow is determined by the expressions:

$$r_i = \frac{B_{S\max}}{\delta\tau_i + \dfrac{B_{S\max}}{r_p}} \tag{1}$$

where $\delta\tau_i$ is the limit closure of the i-th flow; B_S and $B_{S\max}$ – the current and maximum size of the buffer memory, respectively; r_p is the peak intensity of the i-th flow. Accordingly, the time to fill bucket with E markers will be equal to

$$t_{fE} = \frac{E_{BS}}{E_{IR}} \tag{2}$$

Markers are accumulated over the total time

$$\tau_e = E(\tau_{te} + \tau_{ge}) \tag{3}$$

where τ_{te} is the length of the marker entering bucket E; τ_{ge} – protective interval.

Adaptation to changing the length and instantaneous intensity of packet input can be done as follows:

- by changing the length of the protective interval to a constant length of the marker;
- by changing the length of the marker at a constant length of the security interval;
- by changing the size of the "yellow area";
- by changing the buffer memory of markers.

4. Conclusions

The proposed procedure for forming self-similar network traffic using an adaptive multi-rate marker bucket is quite simple and effective. Simulation results show that it is possible to limit the frequency of the token generator to such a value that all incoming traffic is received and then transmitted without loss or retransmissions. With this method of overload management, elimination of overloads is achieved in a short time, so load fluctuations are also relatively small. Although both congestion and loss of traffic controllability can occur if the duration of bursts exceeds the buffer of dynamic stability of the shaper. For example, if the primary and secondary buffers are full for a period exceeding the allowable timeout, that route or network segment becomes unavailable altogether. Such situations have to be handled with higher-level services.

Asymptotic comparative quality estimates based on the queue length indicator were obtained for classical Poisson and self-similar flows entering a single-channel mass service system (MSS) with class GI/G/1 waiting.

References

[1] Barreiros M., Lundqvist P.: Policing and Shaping. QOS-Enabled Networks: Tools and Foundations. Wiley, 2015 [http://doi.org/10.1002/9781119109136.ch6].

[2] Chakraborty T. et al.: Searching for Heavy-Tailed Probability Distributions for Modeling Real-World Complex Networks. IEEE Access 10, 2022, 115092-115107 [http://doi.org/10.1109/ACCESS.2022.3218631].

[3] Chang C. S. et al.: SDL Constructions of FIFO, LIFO and Absolute Contractors. IEEE INFOCOM 2009, 738–746 [http://doi.org/10.1109/INFCOM.2009.5061982].

[4] Farhadi V. et al.: Service Placement and Request Scheduling for Data-intensive Applications in Edge Clouds. IEEE Conference on Computer Communications – INFOCOM 2019, 1279–1287 [http://doi.org/10.1109/INFOCOM.2019.8737368].

[5] Han C. et al.: Analytical Study of the IEEE 802.11p MAC Sublayer in Vehicular Networks. IEEE Transactions on Intelligent Transportation Systems 13(2), 2012, 873–886 [http://doi.org/10.1109/TITS.2012.2183366].

[6] He J. et al.: Buffer-Aided Relaying for Two-Hop Secure Communication with Limited Packet Lifetime. IEEE 20th International Conference on High Performance Switching and Routing (HPSR), 2019, 1–7 [http://doi.org/10.1109/HPSR.2019.8807995].

[7] Kirichenko L. et al.: Classification of Fractal Time Series Using Recurrence Plots. International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T), Kharkiv, 2018, 719–724 [http://doi.org/10.1109/INFOCOMMST.2018.8632010].

[8] Kotenko I. et al.: A technique for early detection of cyberattacks using the traffic self-similarity property and a statistical approach. 29th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), Valladolid, 2021, 281–284 [http://doi.org/10.1109/PDP52278.2021.00052].

[9] Li Z. et al.: Detecting Saturation Attacks Based on Self-Similarity of OpenFlow Traffic. IEEE Transactions on Network and Service Management 17(1), 2020, 607–621 [http://doi.org/10.1109/TNSM.2019.2959268].

[10] Millán G.: Proposal of an Estimator of the Hurst Parameter for a Self-similar Process Representative of the Degree of Randomness of the Recorded Traffic in IEEE 802.3-2005 Networks. OSF Preprints. 2021 [http://doi.org/10.31219/osf.io/pgbvm].

[11] Moltafet M. et al.: Average Age of Information for a Multi-Source M/M/1 Queueing Model With Packet Management. IEEE International Symposium on Information Theory (ISIT), Los Angeles, 2020, 1765–1769 [http://doi.org/10.1109/ISIT44484.2020.9174099].

[12] Patilkulkarni S. et al.: Programmable Delay and Variable Bit Rate Enabled Video Streaming using C++. Third International Conference on Advances in Electronics, Computers and Communications (ICAECC), Bengaluru, 2020, 1–7 [http://doi.org/10.1109/ICAECC50550.2020.9339499].

[13] Plevyak T., Sahin V.: Management of Triple/Quadruple Play Services from a Telecom Perspective. Next Generation Telecommunications Networks, Services, and Management, 2010, 15–52 [http://doi.org/10.1002/9780470594025.ch2].

[14] Ren D. et al.: Adaptive Request Scheduling and Service Caching for MEC-Assisted IoT Networks: An Online Learning Approach. IEEE Internet of Things Journal 9(18), 2022, 17372–17386 [http://doi.org/10.1109/JIOT.2022.3157677].

[15] Serinaldi F.: Use and misuse of some Hurst parameter estimators applied to stationary and non-stationary financial time series. Physica A: Statistical Mechanics and its Applications 389(14), 2010, 2770–2781 [http://doi.org/10.1016/j.physa.2010.02.044].

[16] Wang W. et al.: Software architecture based on message queue. 6th International Conference on Information Science, Computer Technology and Transportation – ISCTT, Xishuangbanna, 2021, 1–5.

[17] Yakymchuk N. et al.: Monitoring of Link-Level Congestion in Telecommunication Systems Using Information Criteria. Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska – IAPGOS 12(4), 2022, 26–30 [http://doi.org/10.35784/iapgos.3076].

[18] Zablotskyi V. et al.: Method for Evaluation Quality Parameters of Telecommunications Services. Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska – IAPGOS 12(2), 2022, 30–33 [http://doi.org/10.35784/iapgos.2918].

[19] Zhang Y. et al.: Recent Advances on HEVC Inter-Frame Coding: From Optimization to Implementation and Beyond. IEEE Transactions on Circuits and Systems for Video Technology 30(11), 2020, 4321–4339 [http://doi.org/10.1109/TCSVT.2019.2954474].

Ph.D. Valeriy Kozlovskiy
e-mail: kzzi@nau.edu.ua

Research interests: Communications, Programmable logic controllers. Research network security. Features of the communication lines organization and operation.

http://orcid.org/0000-0002-8301-5501

M.Sc. Natalia Yakymchuk
e-mail: n.yakymchuk@lntu.edu.ua

Research interests: diagnostics and control of the telecommunication networks state, end-to-end diagnostics, congestion management.

http://orcid.org/0000-0002-8173-449X

Ph.D. Yosyp Selepyna
e-mail: y.selepyna@lntu.edu.ua

Research interests: Modeling of electronic devices and systems. Digital signal processing and coding in telecommunication systems and networks.

http://orcid.org/0000-0002-2421-1844

Ph.D. Serhii Moroz
e-mail: s.moroz@lntu.edu.ua

Author of more than 100 scientific papers, 7 patents for utility models. Research interests: Methodological bases of construction of sensors of physical quantities and measuring modules, Research of info-communication systems in the context of the development of the concept of the Internet of Things.

http://orcid.org/0000-0003-4677-5170

Ph.D. Anatolii Tkachuk
e-mail: a.tkachuk@lntu.edu.ua

Vice-dean for R&D Faculty of Computer and Information Technologies, Lutsk National Technical University.
Member of European Alliance for Innovation (EAI), International Association for Technological Development & Innovations (IATDI).

http://orcid.org/0000-0001-9085-7777