

*Розглянута схема контролю перевантажень з використанням зворотного зв'язку по знаку функції чутливості продуктивності телекомунікаційної мережі. Для визначення даної функції запропоновано використання простої нейронної мережної моделі динамічної системи. Контроль наявності або загрози перевантаження здійснюється на основі аналізу довжини черги на стороні приймача інформації. Для аналізу системи визначена функція вартості як цільова функція наявності перевантаження. Запропонований алгоритм оптимального управління забезпечує формування керуючого сигналу таким чином, щоб вихід системи відповідав як можна ближче заздалегідь встановленим характеристикам – ключовим показникам ефективності мережі. Розроблена схема контролю перевантаження зі зворотним зв'язком по знаку чутливості функції продуктивності системи. Знак чутливості продуктивності надає оптимальний напрям для налаштування швидкості джерела даних.*

*Запропонована нейронна модель для багатокрокового передбачення стану черги на стороні приймача телекомунікаційної мережі. Якщо нейронна мережа налаштована на відслідковування динаміки системи і показує, що квадратична помилка є незначною, вважається, що виконаний крок відповідає наперед передбаченому виходу системи.*

*Запропонований алгоритм адитивного збільшення/множинного зменшення, який визначає зміну швидкості джерела даних в залежності від знаку функції чутливості показника продуктивності. Даний алгоритм є альтернативою системи прогнозування перевантаження і керування потоком, заснованої на пороговому заповненні черги.*

*Проведено порівняльний аналіз ефективності керуючих схем виявлення перевантаження на основі черг і на основі функції чутливості продуктивності телекомунікаційної мережі. Показано, що величина черги і коливання швидкості джерела менші для схеми на основі чутливості, ніж для схеми на основі черги.*

*Результати моделювання продуктивності запропонованої схеми показують, що схема на основі функції чутливості має кращі ключові показники ефективності в порівнянні зі звичайною схемою вибору порога черги*

*Ключові слова: телекомунікаційна мережа, функція чутливості, нейронна мережа, динамічна система, управління чергою*

# DEVELOPMENT OF THE METHOD TO CONTROL TELECOMMUNICATION NETWORK CONGESTION BASED ON A NEURAL MODEL

**N. Vinogradov**

Doctor of Technical Sciences, Professor\*

**M. Stepanov**

Doctor of Technical Sciences, Senior Researcher  
Department of Radio Reception and Signal Processing  
National Technical University of Ukraine  
«Igor Sikorsky Kyiv Polytechnic Institute»  
Peremohy ave., 37, Kyiv, Ukraine, 03056

**Ya. Toroshanko**

PhD, Senior Researcher\*\*  
E-mail: toroshanko.ya@gmail.com

**V. Cherevyk**

PhD, Associate Professor\*\*

**A. Savchenko**

PhD, Associate Professor\*

**V. Hladkykh**

PhD\*\*\*

**O. Toroshanko**

Lecturer\*\*\*

**T. Uvarova**

PhD

Center of Military and Strategic Studies  
Ivan Chernyakhovsky National Defense University of Ukraine  
Povitroflotskyi ave., 28, Kyiv, Ukraine, 03049

\*Department of Computer Information Technologies  
National Aviation University  
Kosmonavta Komarova ave., 1, Kyiv, Ukraine, 03680

\*\*Department of Computer Sciences  
State University of Telecommunications  
Solomianska str., 7, Kyiv, Ukraine, 03110

\*\*\*Department of Telecommunication  
O. S. Popov Odessa National Academy of Telecommunications  
Kovalska str., 1, Odessa, Ukraine, 65029

## 1. Introduction

The quality of services in the telecommunication network is largely determined by algorithms of routing, data flow control and functioning under congestion conditions. Different routing methods, static or dynamic,

local or centralized, deterministic or stochastic, try to direct messages from the source to the destination, so that:

- data delay in the network should be minimal;
- data traffic control should ensure avoidance or minimizing the appearance of congestion.

A telecommunication network is a totality of resources that are used by competing users [1–3]. Resources of such networks can be divided into two main classes:

- main: buffer memory, throughput, processing time;
- side: name space, table entries, logical channels, etc.

The totality of resources has limited (finite) capabilities that cause appearance of conflicts between the users of a system. These conflicts can cause a decrease in system performance to such a point when a system becomes «noisy». The reason for this «noisiness» is repeated duplication of data and a sharp increase in technological controlling information. As a result of it, the throughput decreases significantly, perhaps, to the zero mark. This is a typical behavior of «competing» system [3–6]. Such a situation can lead to the collapse of a network.

Networks cannot service the entire proposed traffic without some control. There should be the rules that govern external traffic and flow coordination within a network.

There are several definitions of congestion, which, however, do not contradict one another. We will use the following definition [7–9]: «congestion is a data loss by a user, caused by an increase in load in the network».

Therefore, congestion control can be determined as a set of the mechanisms that prevent or reduce such decline. If a network does not prevent a loss of user's data, you need to try to maximally limit the losses, and, in the future, try to be fair to all users who suffered.

Congestion has a significant impact on key performance indicators of a telecommunication network and users' service quality. The above determines the relevance and the need for research in this direction.

---

## 2. Literature review and problem statement

---

Papers [10, 11] considered the RED method of dealing with congestion, which is used in the TCP (Transmission Control Protocol) of the Internet. The expanded classification of the methods and algorithms for dealing with congestion that are used today was proposed. The characteristics of algorithms were described, advantages and disadvantages of their use under certain conditions of network operation were determined. However, the mentioned papers [10, 11] are, to a certain extent, reviewing in nature. They do not present quantitative comparative estimates of efficiency/complexity of the RED method and other techniques, including Tail Drop, WRED (weighted random early detection), etc. In addition, the analytical materials concerning the traffic distribution by the types of packets are not given.

Articles [8, 12] studied the TCP Reno algorithm. The algorithm is quite common for fighting congestion and is able to work effectively in both wired and wireless networks. It is known that the algorithm of TCP Reno was created for wireless networks with a high share of lost packets. It tries to separate the losses, not related to the congestion so as not to include the algorithms of combating congestion where it is not required. Meanwhile, the method of recognition of the nature of losses by the algorithm TCP Reno, in fact, is quite trivial (a linear classifier). It is not mentioned in articles [8, 12], obviously, because of the impossibility of detailed analysis due to the lack of sufficient statistics of large volume.

Paper [4] deals with the method to prevent congestion by increasing the amount of memory of input buffers. However, the problem of Bufferbloat was not analyzed in this paper.

At an increase in the amount of buffer memory, the number of unprocessed packets increases, and, what is more, so does waiting time for their processing. This can lead to exceeding the permissible norm of timeout duration, leading to a further decrease in useful throughput of a network. Usually, this can cause the avalanche process: bufferbloat causes the loss of packets that will have to be transmitted again or even several times. Thus, the computation node of the router-sender receives the excessive parasitic load, which can lead to an increase in negative consequences associated with congestion.

An important role in data flow control and congestion prediction, as well as in evaluating and optimizing key performance indicators of telecommunication networks is played by the indicators of sensitivity of complex systems. In the fundamental research [13], the category of sensitivity of a complex system as a mathematical indicator was determined and logarithmic sensitivity functions were proposed. However, at a high degree of abstraction and strictness of the mathematical apparatus, paper [13] does not provide the approaches to use of sensitivity function as a tool for applied analysis of technical systems.

In articles [14, 15], the advisability of using the mathematical apparatus for analysis of sensitivity of complex systems in the problems of congestion control in computer telecommunication networks was substantiated. The results, presented in publications [14, 15], act as recommendations of a general nature. Following them, of course, it is possible to develop specific methods of dealing with congestion, but it requires new unconventional approaches to the solution of the problem as a whole. One of such approaches is the use of artificial intelligence methods. First of all, these are neural networks of improved architecture with explicit and hidden layers and current optimization of parameters by results of learning through back error propagation.

Article [16] examined the ways and methods for determining sensitivity of output characteristics of telecommunication networks as systems of mass service. These methods are based on the models of queue control for the adaptation of the controlled access of external traffic to the system in order to obtain the expected limits of performance.

Papers [14, 15] do not specify the definitions of the functions of sensitivity of a telecommunication network and the methods of their identification in network control systems with explicit or indirect feedback.

In article [16], models and methods for queue control are based on the functions of sensitivity of output characteristics of telecommunication networks as systems of mass servicing. However, the asymptotic characteristics of sensitivity functions are not determined. In addition, the expression in a closed form for the functional or statistical connection of parameters of sensitivity functions and corresponding parameters of the queue control system were not obtained.

Study [17], focusing on the control of the telecommunication network congestion, proposed the algorithm of active distribution of attempts of simultaneous access to a slot with a uniform time distribution. This approach is fair for the Least Favorable Distribution (LFD) of a request flow. In modern telecommunication networks, distributions of request flows are far from uniform, so obtained asymptotic estimates will lead to unrealistically optimistic conclusions. Realistic estimates can be obtained when applying neural network models, which must adapt to load jumps and variations of probabilistic distributions of request flows. In this work, an attempt to solve these problems was made.

### 3. The aim and objectives of the study

The aim of this study is to improve and to develop the method for control and prediction of congestion in the telecommunication network with the use of explicit feedback by the sign of the function of sensitivity of network performance.

To achieve the set aim, the following tasks were solved:

- to analyze the methods for congestion detection based on the control of quantitative value of the length of a queue on the side of the receiver, to identify their advantages and disadvantages;
- to develop and substantiate the circuit multi-step prediction of the queue state and congestion threat based on the neural model;
- to develop the algorithm of neural network learning and of the formation of the feature of network congestion.

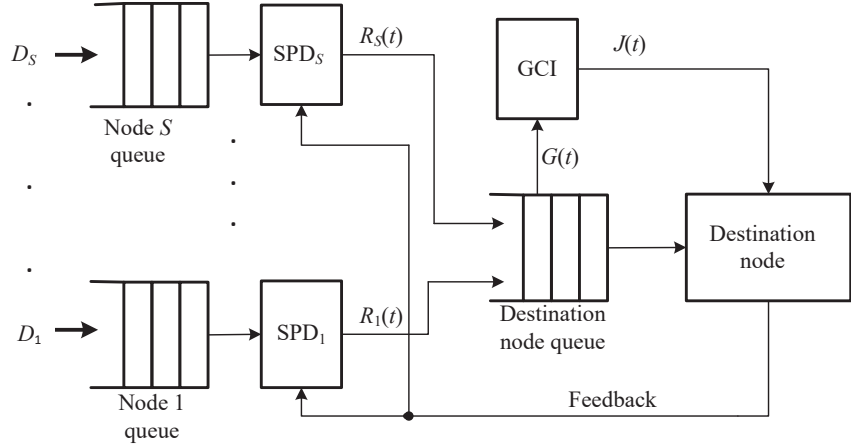


Fig. 1. Establishment of congestion bit based on queue length

### 4. Control of congestion and multi-step prediction of the state of telecommunication network

#### 4.1. Congestion control based on a queue length indicator

The most effective methods and algorithms for congestion control are implemented based on the explicit feedback principle [1, 2, 16].

One of the most common ways of data flow control by the feedback principle is binary bit control [18, 19]. Feedback is implemented between the nodes, through which each separate session of information exchange is performed. We will note that the feedback circuit can be implemented between two adjacent nodes or between two end nodes, between which there are some transit nodes along the information exchange route (end-to-end control).

The network provides binary indication about whether there were congestions or the threats of their occurrence during connection. The value of «0» of the congestion indicator is set in the header of the packet by a data source, that is, on the side of the node-sender. Any node along the path can set (or prove) the value of congestion bit in a packet as «1» to indicate the existence or a congestion threat in this node. The end system (node-receiver) controls the bits of congestion of received packets and returns a feedback message to the source with the information about the existence of congestion. If a feedback message reports on the absence of congestion, the information source increases the input traffic (rate of packet receiving, data stream) – additive feedback. If the feedback message reports on the existence or the threat of congestion, the input traffic from the information sources decreases – multiple (multiplicative) feedback.

The detection and indication of congestion based on a queue length is a simple and commonly used feedback-based circuit of information flow control (Fig. 1).

At the simultaneous implementation of  $S$  of information processes (datagram way of transmission), the packets  $D_1...D_S$  from source  $S$  arrive at appropriate service queues, from which enter the queue of destination node

through the circuits of packet delivery  $SPD_1...SPD_S$ . The generator of congestion indicator (GCI), based on the information about the state of queue  $N(t)$ , provides the data on the existence or a threat of congestion  $J(t)$  to the end user or the transit node.

The congestion state is traditionally determined by the queue length  $G(t)$  in a transit communication node or at the destination of the final data recipient at moment  $t$ . When a queue length reaches a preliminarily determined boundary value  $Q$ , the packets that pass through the queue will have an indication bit that is set in the state «congestion confirmed».

The main advantage of the queue-based circuit is its low complexity, because the absolute queue length can be controlled using one counter. However, this method is not effective when using the method of control of network segments with transit nodes of information transmission. Using one congestion bit for the whole segment does not make it possible to localize the congestion place with the precision to a node. This method can create long queues in network nodes and cause a great delay of feedback information. Detection of the created congestion should be delayed for the time, required to create the queue. Similarly, the identification of the congestion decision is also delayed for some time needed to process the queue. To control «additive increase/multiple decrease» in the source rate, we have:

$$R(t+1) = \begin{cases} R(t) + F^+, & G(t-t_d) < Q \text{ or } E(t-t_d) > 0, \\ F^- R(t), & G(t-t_d) \geq Q \text{ or } E(t-t_d) \leq 0, \end{cases} \quad (1)$$

where  $R$  is the rate of filling the queue at the entrance of destination node;  $F^+ > 0$  is the factor of additive increase;  $0 < F^- < 1$  is the factor of multiple decrease;  $E(t-\tau) = Q - G(t-\tau)$  is the error function;  $t_d$  is the time of delay of transmission of the state of congestion queue to the source.

#### 4.2. Neural network model for congestion detection

Let us consider the circuit of congestion control with feedback by the sign of sensitivity of the function of system performance. A sign of sensitivity of performance provides the optimal direction to adjust the data source rate [15]. The proposed circuit for determining the sensitivity function uses a simple neural network model of dynamic system.

Neural networks are mathematical structures, capable to self-learning on the basis of external information [20].

Let the dynamics of a service system be expressed with the following input-output equation:

$$G(t+1) = f \left[ \begin{matrix} G(t), G(t-\tau), \dots, G(t-l\tau), R(t), \\ R(t-\tau), \dots, R(t-m\tau) \end{matrix} \right], \quad (2)$$

where  $G(t)$  is the scalar output: the queue length or service delay at the moment of time  $t$ ;  $R(t)$  is the scalar input: instantaneous rate of the queue at the input at moment  $t$ ;  $f[\cdot]$  is the unknown function, evaluated with the help of neural network;  $l, m$  are, respectively, orders  $\{N(t), R(t)\}$ ;  $\tau$  is the period of time counts, period of clock frequency of the system.

The aim of the algorithm of optimal control is the selection of control signal  $R(t)$  so that the output of system  $G(t)$  should meet as much as possible the characteristics set beforehand  $Q(t)$  (as a rule,  $Q(t) = Q = \text{const}$ ). Sequential-parallel neural network model [16] of the unknown system (2) can be represented as:

$$\begin{aligned} \hat{G}^i(t+i\tau) &= \\ &= f^i[G(t), \dots, G(t-l\tau), R(t), \dots, R(t-m\tau)], \quad i = 1, 2, \dots, L, \end{aligned} \quad (3)$$

where  $\hat{G}^i(t+i\tau)$  is the output of the neural network;  $f^i$  is the evaluation of function  $f$ ;  $L$  is the prediction horizon.

Fig. 2 shows the circuit of multi-step prediction of the queue state based on neural network.

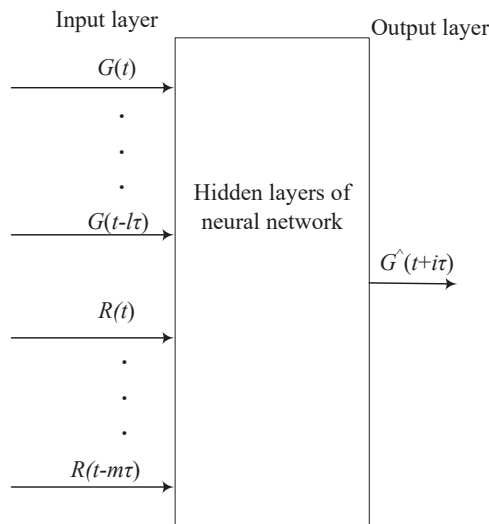


Fig. 2. Neural model of queue state prediction

If the neural network is set up to track the dynamics of the system and shows that the quadratic error:

$$[G(t+i\tau) - \hat{G}^i(t+i\tau)]^2 = \varepsilon \quad (4)$$

is insignificant, it is considered that the  $i$ -th step corresponds to the system output predicted beforehand (2). As a result, a controlling signal can be selected so that  $\hat{G}^i(t+i\tau)$  should be as close to  $Q$  as possible.

We will determine the cost function as objective function of existence of congestion as follows:

$$J = \frac{1}{2} e^2(t+i\tau) = \frac{1}{2} [Q - \hat{G}^i(t+i\tau)]^2, \quad i = 1, 2, \dots, L. \quad (5)$$

Control signal  $R(t)$  (that is, the rate of data source) must be selected so as to minimize  $J$ . In the discrete case, the controlling variable is updated according to the following rule of gradient descent:

$$R(t+\tau) = R(t) + \Delta R(t) = R(t) - \eta \frac{\partial J}{\partial R(t)}, \quad (6)$$

where  $\eta$  is the size of control pitch.

Obviously, the weakest assumption for characteristics of the loss function, which enables using these methods, is the unimodality of the loss function. For actual telecommunication systems and networks this assumption is fairly realistic [8, 22].

You can see that to determine the appropriate control signal, the process of minimizing (6) is based on the approximation made by a neural network. Therefore, it is necessary that  $\hat{G}^i(t+i\tau)$  should approximate to the output of the actual system  $G(t+i\tau)$  asymptotically. This can be achieved by supporting the neural network learning online.

Differentiating (5) for function  $R(t)$ , we obtain:

$$\Delta J(t) = \frac{\partial J}{\partial R(t)} = -e(t+i\tau) \frac{\partial \hat{G}^i(t+i\tau)}{\partial R(t)}, \quad i = 1, 2, \dots, L, \quad (7)$$

where expression  $\partial \hat{G}^i(t+i\tau) / \partial R(t)$  is known as sensitivity or gradient of the system [7, 15].

Substituting (7) into (6), we obtain:

$$R(t+\tau) = R(t) + \eta e(t+i\tau) \frac{\partial \hat{G}^i(t+i\tau)}{\partial R(t)}, \quad i = 1, 2, \dots, L. \quad (8)$$

Gradient of the system can be analytically evaluated using the known structure of the neural network [20].

Expression (6) can be represented as:

$$R(t+\tau) = R(t) - \eta \text{sgn}[\Delta J(t)], \quad (9)$$

where  $\text{sign}[\Delta J(t)]$  denotes  $\text{sign} \Delta J(t)$  (that can be positive or negative).

Thus, we can conclude that (7) implements the rule of control of additive increase/ multiple decrease in the source rate. Therefore an attractive alternative to the circuit of congestion indicator generation, based on the threshold queue filling is the algorithm of additive increase/multiple decrease. The algorithm determines the change in the rate of data source  $R(t)$ , depending on the sign of sensitivity of performance indicator  $\Delta J(t-t_d)$ . That is, congestion indicator  $B(t)$  is formed depending on the gradient of system  $\Delta J(t)$  at time moment  $t$ :

$$\begin{cases} B(t) = 0, & \text{if } \Delta J(n) < 0, \\ B(t) = 1, & \text{if } \Delta J(n) \geq 0. \end{cases}$$

The value of  $\Delta J$ , computed using formula (7), gives the optimal direction to adjust the source rate. In short, only the sign, but not the value of  $\Delta J$  matters in this case. The circuit of the regulation of the input data stream (packets reception rate) using the neural network for analyzing sensitivity is shown in Fig. 3.

Current values of the rates of regulated input flows  $R_1(t) \dots R_5(t)$  enter the input layer of the neural network HM. Based on analysis of values  $R_1(t) \dots R_5(t)$  and the current value of the length of network queue  $G(t)$ , the output layer of

neural network tracks down the magnitude of deviation  $\Delta J$  of objective controlled function  $J$  (existence or congestion threat). The sign of deviation magnitude  $\text{sgn}\Delta J$  is taken into account by the destination node during formation of the feedback signal.

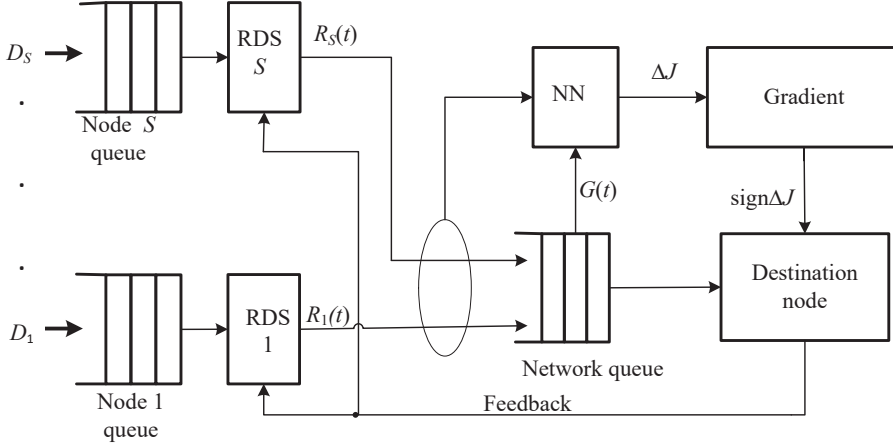


Fig. 3. Optimization circuit of congestion control

For the algorithm of control of additive increase/multiple decrease in the source rate, we have:

$$R(t+\tau) = \begin{cases} R(t) + F^+, & \Delta J(n-\tau) < 0, \\ R(t) - F^-, & \Delta J(n-\tau) \geq 0. \end{cases} \quad (10)$$

#### 4. 3. Algorithm of neural network learning and formation of congestion feature

Model of neural network learning can change over time, online learning is necessary in the controlling system of load prediction based on sensitivity function. Because of using an explicit feedback signal by sensitivity function, the chosen method for neural network learning based on inverse error distribution is applied in this work as the most accurate and statistically steady [22].

Neural network learning algorithm and the formation of congestion feature is as follows.

*Initial state:* Time moment  $t$ .

*Step 1:* Obtained dimensions of the system at the inlet and outlet:  $R(t)$  and  $G(t)$ .

Available history of the system:

$$R(t-\tau), \dots, R(t-m\tau); \quad G(t-\tau), \dots, G(t-l\tau).$$

History of neural network:

$$G^\wedge(t), \delta_i^{(p)}(t-\tau), x_i^{(p)}(t-\tau),$$

for all layers  $p$  and all nodes  $i$ .

*Step 2:* Computation of the predicted value by the neural network:

$$G^\wedge(t-i\tau).$$

Computation of the feedback value by the neural network:

$$\frac{\partial G^\wedge(t-i\tau)}{\partial R(t)}.$$

*Step 3:* Computation of gradient function:

$$\Delta J(t) = \frac{\partial J}{\partial R(t)}.$$

Formation of congestion indicator:

$$B(t) = \begin{cases} 1, & \text{if } \Delta J \geq 0, \\ 0 & \text{other.} \end{cases}$$

*Step 4:* Using the expression:

$$[G^\wedge(t) - G(t)]$$

to update the weights of the neural network, we perform computation:

$$\begin{aligned} \Delta w_{ij}^{(p)} = & -[G^\wedge(t) - G(t)] \times \\ & \times \delta_i^{p-1}(t-\tau) x_j^{p-1}(t-\tau) + \\ & + \gamma \Delta w_{ij}^{(p)}(t-\tau), \end{aligned}$$

$$w_{ij}^{(p)} \leftarrow w_{ij}^{(p)} + \Delta w_{ij}^{(p)}(t)$$

for

$$p = p_s, p_{s-1}, \dots, 1; \quad i = 1, 2, \dots, m_{(p)}; \quad j = 1, 2, \dots, m_{p-1}.$$

$\beta$  is the rate of neural network learning;  $\gamma$  is the constant pulse of neural network.

*Step 5:* The following interval:  $t \leftarrow t + \tau$ . Proceed to step 1.

*Note:* Steps 3 and 4 can be performed in parallel.

We will determine the following performance indicators for modeling, where  $T$  is the time to perform modeling:

–  $G_{\max} = \max\{G(t): 0 \leq t \leq T\}$ . Maximum value  $G(t)$  displays the buffer size, which is required in a narrow place to avoid the packet loss;

– average time ( $f$ ) to reach the queue size and the source rate:

$$f^- = \frac{1}{T} \int_0^T f(t) dt;$$

– variance  $\sigma^2(f)$  of the queue size and source rate:

$$\sigma^2(f) = \frac{1}{T} \int_0^T [f(t) - \tilde{f}]^2 dt.$$

Let us make a comparative analysis of efficiency of controlling circuits for congestion detection based on the function of sensitivity of telecommunication network productivity. Consider one connection with the following set of parameters: peak rate of source  $R_{\max} = 100$  packets/time unit, time unit  $\tau = 0,25$  ms, minimum rate  $R_{\min} = 100$  packets/time unit, coefficient of additive increase  $F^+ = R_{\max}/16$ , coefficient of multiple decrease  $F^- = 15/16$ . Congestion threshold is set at  $Q = 500$  packets.

Consider the rounding delay from 6 time units –  $6\tau$ . Two bottlenecks can be taken into account in the study:

– sinusoidal integer  $[35(1 + \sin(2\pi\tau)) = 10]$  packets/time unit;

– random (maximum value of 80 packets/time unit).

The architecture of the neural network is as follows: 3-layer neural network (8 inputs, 8 input neurons, 8 hidden

neurons, 1 output neuron); the order of queue length  $l=3$ ; the order of the rate of queue filling  $m=3$ . To determine the state of the bottleneck of the queue, two different prediction horizons are considered in this study: prediction step 1 and 3.

A short list of modeling results is shown in Table 1, where  $G_{av}$  and  $R_{av}$  denote time average queue size and source rate.

List of modeling results

Approach	Servicing type	$G_{max}$	$G_{av}$	$\sigma^2(G)$	Index $G_{max}$	$R_{av}$	$\sigma^2(R)$
Queue-based	Random	1560	724.5	187569.3	757	45.1	1370.1
	Sinusoidal	1556	713.8	179305.5	281	45.7	1211.4
Gradient: 1-Step	Random	1293	666.2	108986.4	480, 481, 482	44.4	878.3
	Sinusoidal	1428	653.3	86115.5	48	44.4	1066.9
Gradient: 3-Step	Random	1307	622.9	88526.7	800	44.2	929.9
	Sinusoidal	1443	628.7	89936.5	49	44.5	1139.7

Modeling results, presented in Table 1, show that the magnitude of the queue and fluctuations in the source rate are smaller for the circuit, which is controlled based on sensitivity function than for the circuit that is controlled based on the queue analysis. Sensitivity-based circuit with the three-step prediction of state provides better productivity than the corresponding circuit with one-step prediction. This is because for a three-step prediction, the delay of controlling feedback signals (that is, congestion indicators), received in the data source from a network queue, is insignificant. That is, the congestion indicator more closely reflects the network conditions implying the process. In the feedback-based load control circuits with considerable delays in the distribution, control signals received in the sources can be out of date. For feedback control, they come into action within the network after some delay control. It means that it will be possible to predict more accurately the queue state with a significant delay.

Thus, the disadvantage of the queue state prediction is that with the more delay we make predictions, the harder it is to get predictions of admissible errors.

The queue-based circuit is more sensitive to changes in the queue service rate than for the sensitivity-based circuit. It should be noted that in the queue-based circuit at a significant decrease in the queue service rate within a certain period of time, the queue size grows extensively beyond previously observed values. An increase in the queue size is steadier under the same conditions for sensitivity-based circuits. The value of difference gives an idea of the magnitudes observed.

To minimize the errors in the network state prediction, it is necessary to follow the obvious rule: the delay in delivery of official and customer information must be of the same order with the time of reaction of a switching node. The problem of accurate determining of delays and the time of reaction, development of the methods for delay regulation, depending on the state of the network load and reaction time of switching nodes is by no means trivial. Its complete solution, specifically, with obtaining quantitative estimates, requires thorough research of theoretical and practical nature. Certain aspects of this problem were considered in some previous research of the authors of [4, 15].

### 5. Discussion of results of studying the method of control and prediction of network congestion

In this work, the neural network was applied to control the congestion of the telecommunication network as a complex system with random delays of service (management) and user information. Unlike the traditional systems of congestion control by the changes of state and parameters of queues, the proposed system operates by optimal algorithms of configurations of weight parameters. This increases the accuracy of determining control signals, decreases the influence of their delays and, as a result, minimizes average consumption of resource.

Examining the neural network as a system of control of telecommunication network, the assumption about the used neural network belonging to the class of dynamic neural networks is made by default. In fact, this is a static neural network, which introduces a feedback through an element of one-bit delay. This assumption for the packet telecommunication networks is rather logical.

In addition, gradient of functional is minimized by the vector of weight coefficients of a neural network. This gradient is considered as a set of gradients by matrices of weight coefficients of individual layers. In this case, it turns out that the results of calculation by the matrix of weights of the above lying layer can be essentially used to calculate the gradient by the matrix of weights of below lying layer. Thanks to the proposed architecture of the dynamic neural network, firstly, the current process of selection of optimal weight coefficients is accelerated considerably, and, secondly, the procedure of network self-learning using the method of back error propagation is simplified.

To construct the optimal algorithms of configuration of neural network parameters, it is necessary to calculate separately both the gradient of vector output of neural network, and the gradient of functional by the vector of discrepancy. Calculation by the vector of discrepancy causes no especial difficulties. Calculation of the gradient of the vector output of the neural network can be made by using the method, which is close to the method of inverse error distribution. However, in this case, it is necessary to carefully control the values of the roots of characteristic polynomial of the dynamic neural network, which requires additional computation resource.

It is possible to show that the expression for the gradient of network output for the vector of weight coefficients of the  $i$ -th layer is represented in the form of a rectangular matrix. At the same time, it is known that gradient takes the simplest form in the case of a square matrix of weight coefficients. It is proposed to calculate the comparative estimates of the required calculation volume during rotation of square  $N \times N$  matrix and at pseudo-rotation of square  $N \times L$  matrix for asymptotic estimation of the tendency of increasing computational complexity at the deviation of the form of the matrix from the square one.

Obtaining (through computation or by computer simulation) of asymptotic estimates of computational complexity of the proposed method will make it possible to obtain potential characteristics of a control system. It offers good prospects of a systemic solution to the problem of optimal control of a telecommunication network. In this case, it is necessary to take into account that deriving expressions in a closed form

Table 1

and corresponding calculation expressions can be associated with cumbersome, although quite simple calculations. In turn, the computer simulation requires significant computational capacities. It seems that this problem can be solved in a quite powerful computational center, for example, in the computational center of a leading communication operator.

---

## 6. Conclusions

---

1. The problem of congestion detection based on the control of the quantitative value of the queue length is that the congestion detector can indicate only the existence of congestion by the connection, but not the location or the causes of congestion. In addition, an attempt to fight congestion by a simple increase in the buffer capacity does not lead to the solution of the problem, but vice versa, to bufferbloat and inadmissible increase in service delay.

2. As a result of the carried out research, a new method for optimal control of congestions of a telecommunication network with the use of a neural network as a system of monitoring and control was developed. The circuit of a multi-step prediction of the queue state was proposed and substantia-

ted. The apparatus of the general theory of sensitivity with the indirect feedback and control of message source activity was used for prediction and early detection of congestion. The results of this theory were used to construct a control system with indirect feedback that allows saving channel and computing resources.

3. The algorithm of neural network learning and formation of the network congestion feature was used. The fundamental difference of the obtained results is the development of the multi-layer dynamic neural network with a combination of explicit and hidden layers. Thanks to the choice of this architecture, firstly, the procedure of finding the optimal weight coefficients is simplified, and secondly, the process of neural network learning by the classic method of back error propagation is accelerated. As result of verification of theoretical results through computer simulation, the quantitative comparative estimates of efficiency (accuracy and required computing resource) of the developed method and the methods that existed before were obtained. It can be argued that with the perfect architecture of a neural network suitable for modeling the dynamics of a system, it is possible to receive rather satisfactory performance of a telecommunication system as a control object.

---

## References

1. Tanenbaum A. S., Wetherall D. J. *Computer Networks*. 5th ed. PrenticeHall, 2010. 960 p.
2. Stallings W. *Foundations of Modern Networking: SDN, NFV, QoE, IoT, and Cloud*. Pearson Education, Inc., OldTappan, New Jersey, 2016. 544 p.
3. Mao G. *Connectivity of Communication Networks*. Springer, 2017. 435 p. doi: <https://doi.org/10.1007/978-3-319-52989-9>
4. Analiz nagruzki na seti peredachi dannyh v sistemah kritichnogo primeneniya / Vinogradov N. A., Drovovozov V. I., Lesnaya N. N., Zembickaya A. S. // *Zviazok*. 2006. Issue 1 (61). P. 9–12.
5. Snarskiy A. A., Lande D. V. *Modelirovanie slozhnyh setey: ucheb. pos.* Kyiv: Inzhiniring, 2015. 212 p.
6. Bonaventure O. *Computer Networking: Principles, Protocols and Practices*. 2018. 272 p. URL: <https://www.computer-networking.info/2nd/cnp3bis.pdf>
7. Keshav S. *Congestion Control in Computer Networks*. University of California, 1991.
8. Kurose J. F., Ross K. W. *Computer Networking: A Top-Down Approach*. 7th ed. Pearson Education, Inc., 2017. 864 p.
9. Göransson P., Black C., Culver T. *Software Defined Networks: A Comprehensive Approach*. 2nd ed. Morgan Kaufmann, US, 2017. 436 p.
10. Korolkova A. V., Kulyabov D. S., Tchernoiyanov A. I. On the classification of RED algorithms // *Vestnik Rossiyskogo universiteta druzhby narodov. Seriya: Matematika, informatika, fizika*. 2009. Issue 3. P. 34–46.
11. Maximov V. V., Chmykhun S. O. Classification of algorithms of controlling networks congestions // *Naukovi zapysky Ukrainskoho naukovo-doslidnoho instytutu zviazku*. 2014. Issue 5 (33). P. 73–79.
12. Maxymov V. V., Chmykhun S. O. Research of the algorithm of controlling congestion TCP VENO // *Telekomunikatsiyni ta informatsiyni tekhnolohiyi*. 2015. Issue 4. P. 30–36.
13. Tomovich R., Vukobratovich M. *Obshchaya teoriya chuvstvitel'nosti*. Moscow: Sovetskoeradio, 1972. 240 p.
14. Shooman M. L. *Reliability of Computer Systems and Networks – Fault Tolerance, Analysis and Design*. JohnWiley & Sons, 2002. 546 p.
15. Toroshanko Ya. I. Management reliability of telecommunication network on the analysis of sensitivity of the complex systems // *Telekomunikatsiyni ta informatsiyni tekhnolohiyi*. 2016. Issue 3. P. 31–36.
16. Toroshanko Ya. I. Sensitivity analysis of systems of mass service on the base of model of adaptation and regulation of foreign traffic // *Herald of Khmelnytskyi national university*. 2016. Issue 6 (243). P. 171–175.
17. Overload Control for Signaling Congestion of Machine Type Communications in 3GPP Networks / Lu Z., Pan Q., Wang L., Wen X. // *PLOS ONE*. 2016. Vol. 11, Issue 12. P. e0167380. doi: <https://doi.org/10.1371/journal.pone.0167380>
18. Mel'nikov D. A. *Informacionnye processy v komp'yuternykh setyah*. Moscow, 1999. 256 p.
19. Gol'dshcheyn B. S., Sokolov N. A., Yanovskiy G. G. *Seti svyazi*. Sankt-Peterburg: «BHV – Peterburg», 2014. 400 p.
20. Tarhov D. A. *Neyrosetevye modeli i algoritmy*. Moscow: Izdatel'stvo «Radiotekhnika», 2014. 352 p.
21. Steklov V. K., Berkman L. N., Kilchyskiy Ye. V. *Optyimizatsiya ta modeliuвання prystroiv i system zviazku*. Kyiv: Tekhnika, 2004. 576 p.
22. Galushkin A. I. *Neyronnye seti: osnovy teorii*. Moscow: Goryachaya liniya – Telekom, 2010. 496 p.