Section 3. Modeling and software engineering

UDC 519.711.7:519.816

PROTECTION OF MULTILAYER NETWORK SYSTEMS FROM SUCCESSIVE, GROUP AND SYSTEM-WIDE TARGETED ATTACKS

Dr.Sci. O. Polishchuk ORCID: 0000-0002-0054-7159
Pidstryhach Institute for Applied Problems of Mechanics and Mathematics, National Academy of Sciences of Ukraine, Ukraine
E-mail: od_polishchuk@ukr.net

Abstract. *Structural and flow approaches to the vulnerability analysis of multilayer network systems (MLNS) from targeted attacks and non-target lesions of various origins are considered. Local and global structural and flow characteristics of monoflow multilayer system elements are determined to build scenarios of targeted attacks on the structure and operation process of MLNS and evaluation their consequences. In order to simplify the construction and improve the efficiency of such scenarios, the concepts of structural and flow aggregate-networks of monoflow MLNS are introduced, and the relationship between the importance indicators of their elements and corresponding indicators of multilayer system nodes is shown. The advantages of flow-based approach over structural ones have been demonstrated, both in the sense of analyzing the vulnerability of real MLNS and evaluation the consequences of negative influences of different nature.*

Keywords: *Complex network, network system, intersystem interactions, multilayer network system, flow model, aggregate-network, influence, betweenness, targeted attack, vulnerability.*

1. Introduction

Many internal and external negative influences can act on any real-world natural or man-made systems. Among such influences that can damage the system, we primarily highlight targeted attacks and its non-target lesions. A distinctive feature of targeted attacks is their intentionality and artificial nature (terrorist and hacker attacks, military aggression and financial and economic sanctions, etc.). In contrast to targeted attacks, non-target lesions can include various unintentional negative influences of natural or artificial origin (natural and man-made disasters, the spread of dangerous infectious diseases and so on). Such lesions can be local, group or system-wide and aimed at damaging both the structure and operation process of network systems (NS) and intersystem interactions. In paper [1], the typical scenarios of consecutive attacks on the structure and operation process of NS were considered and their connections with the development of countermeasures against the system non-target lesions were established. The usefulness of such scenarios lies in the fact that they, giving a picture of possible development of a certain type of

lesion, allow creating the most effective means of protection against it [2, 3]. In particular, the structural and flow NS models make it possible not only to build scenarios of the spread of negative influences of various origins, but also, compared to other system models, evaluate the level of local, group and system-wide losses resulting from the action of such influences during and after lesion [1]. The development of strategies for the protection of multilayer network systems, which describe the processes of intersystem interactions, is significantly complicated not only due to the increase of problem dimension, but also because the lesion of certain layer-system of such formation may not occur directly, for example, through a targeted attack on it, but consequentially as a result of attack on adjacent MLNS's layer [4, 5]. At the same time, lesions of various adjacent layers-systems can lead to different consequences (the influence of blocking the maritime and aviation layers of general transport system of Ukraine during Russian aggression on the railway and automobile layers is significantly different). Simultaneously, the quantity of local and global characteristics of MLNS elements, which describe the structural and functional features of not only internal, but also intersystem interactions, is increasing, and therefore, the amount of importance indicators of elements, which are used when building scenarios of targeted attacks on multilayer system, is increasing too [6]. The process of evaluation the consequences of MLNS lesions is also complicated, in particular, the negative influence of the directly damaged layers-systems on the adjacent ones [7, 8]. All these factors must be taken into account by the NS management systems, which are the part of man-made MLNS, for the effective organization of their protection and overcoming the consequences of various types of lesions. No large scale real-world complex system can protect or simultaneously restore all elements damaged by negative influences. Therefore, the calculation of objective importance indicators of nodes and edges of NS and MLNS plays a decisive role during the construction of effective scenarios of targeted attacks on them [9, 10]. Equally important is the value of these indicators for development the effective strategies for countering the spread of non-target lesions. The purpose of article is to determine on the basis of structural and flow models of intersystem interactions, the importance indicators of MLNS components and formation of effective scenarios of successive group and system-wide targeted attacks on multilayer network systems, as well as evaluation of consequences of separate system components lesions on different system layers and implementation of intersystem interactions in general.

2. A structural model of multilayer network system

The structural model of intersystem interactions is described by multilayer networks (MLNs) and displayed in the form [11]

$$G^M = \left(\bigcup_{m=1}^{M} G_m, \bigcup_{m,k=1,\, m \neq k}^{M} E_{mk} \right),$$

where $G_m = (V_m, E_m)$ determines the structure of mth network layer of MLN; V_m and E_m are the sets of nodes and edges of network G_m respectively; E_{mk} is the

set of connections between the nodes of V_m and V_k, $m \neq k$, $m,k = \overline{1,M}$, and M is the numder of MLN layers. The set $V^M = \bigcup_{m=1}^{M} V_m$ will be called the total set of MLN nodes, N^M – the number of elements of V^M.

Multilayer network G^M is fully described by an adjacency matrix

$$\mathbf{A}^M = \{\mathbf{A}^{km}\}_{m,k=1}^{M}, \quad (1)$$

in which the blocks $\mathbf{A}^{mm}$ determine the structure of intralayer and blocks $\mathbf{A}^{km}$, $m \neq k$, – interlayer interactions. Values a_{ij}^{km} =1 if the edge connected the nodes n_i^k and n_j^m exists, and a_{ij}^{km} =0, $i, j = \overline{1, N^M}$, $m,k = \overline{1,M}$, if such edge don't exist. Blocks $\mathbf{A}^{km} = \{a_{ij}^{km}\}_{i,j=1}^{N^M}$, $m,k = \overline{1,M}$, of matrix AM are determined for the total set of MLN nodes, i.e. the problem of coordination of node numbers is removed in case of their independent numbering for each layer. In this paper, we consider partially overlapped MLN [12], in which connections are possible only between nodes with the same numbers from the total set of nodes V^M (Figure 1).

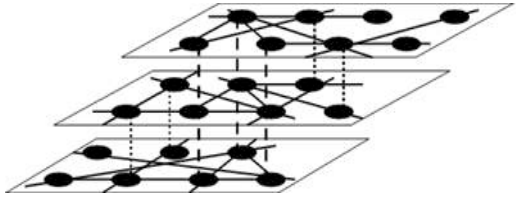


Figure 1. *An example of structure of partially overlapped three-layer MLN*

This means that each node can be an element of several systems and perform one function in them, but in different ways. Nodes through which interlayer interactions are carried out will be called MLNS transition points.

Multidimensional (multiflow) networks, which describe the structure of interactions between layers, each of which ensures the movement of specific type of flow different from other layers, are considered the most general case of MLN [13]. An example of two-dimensional network is a general transport system that ensures the movement of passenger and cargo flows [1]. A feature of such formations is the impossibility of flow transition from one layer to another (transformation of passengers into cargo and vice versa). Therefore, the characteristics of elements of multidimensional networks are usually described by vectors of these characteristics in each layer (degree, betweenness, closeness, eigenvector centralities and so on [14]). Scenarios of successive targeted attacks on the structure of such multilayer

networks are built using precisely these vectors of importance indicators of their elements [4, 6]. In the article [1] was proposed a method of decomposing multidimensional MLNS into monoflow multilayer systems, all layers of which ensure the movement of certain type of flow by different carriers or operator systems (movement of passengers or cargos through four-layer transport networks, which include railway, automobile, aviation and water system layers, respectively). The centrality of elements of monoflow MLNS can be determined not only for separate layers, but also for a multilayer network in general by constructing their aggregate-networks [15]. In addition to reducing the dimensionality of MLNS model by at least M times, the use of such structures makes it possible to solve a number of practically important problems of the theory of complex networks much more effectively [16].

2.1. A structural aggregate-network of multilayer system

The local characterictic ε_{ij} of the edge (n_i, n_j) in MLN, where n_i and n_j are the nodes from the total set of nodes V^M, which will be called its structural aggregate-weight, is the quantity of layers in which this edge is present. Structural aggregate-weight ε_{ii} of the MLN's node n_i is the quantity of layers of which it is a part, $i, j = \overline{1, N^M}$. For arbitrary multilayer network, the adjacency matrix $\mathbf{E} = \{\varepsilon_{ij}\}_{i,j=1}^{N^M}$ completely determines the weighted network (Figure 2), which will be called the structural aggregate-network of MLN. Since we are considering the case when interlayer connections are possible only between nodes with the same numbers of total set of MLNS nodes, the structure G_{ag}^M of this aggregate-network can be described in the form

$$G_{ag}^M = \left(V^M, E^M = \bigcup_{m=1}^M E_m\right) \qquad (2)$$

in which the set EM will be called the total set of MLN edges.

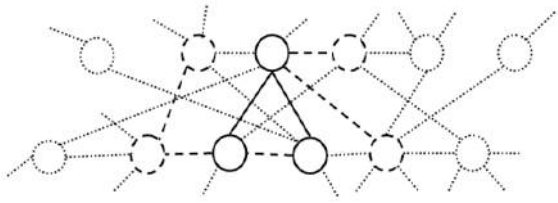


Figure 2. *Aggregate-network of reflected in fig. 1 three-layer MLN (____ – for* ε_{ij} *=3, _ _ _ – for* ε_{ij} *=2, ……. – for* ε_{ij} *=1,* $i, j = \overline{1, N^M}$ *)*

The elements of matrix E define integral structural characteristics of multilayer network nodes and edges. For multiflow multidimensional networks, the aggregate-weights of edges of weighted aggregate-network determine the quantity of interactions of various types between the nodes of such structures. For monoflow MLNs, the aggregate-weight of each edge reflects the number of possible carriers or operator systems that can ensure the movement of corresponding type of flow. Therefore, the input (output) aggregate-degree of each node of weighted aggregate-network of monoflow MLNS is equal to the sum of input (output) degrees of this node in all its layers.

2.2. Targeted attacks on multilayer systems

We will build a scenario of targeted attack on monoflow multilayer network, using as importance indicator of its nodes the centrality of generalized degree d_i of node n_i in the total set of nodes VM of aggregate-network, i.e.

$$d_i = \sum_{j=1,\ j\neq i}^{N^M} (\varepsilon_{ij} + \varepsilon_{ji}) + \varepsilon_{ii}, \quad i = \overline{1, N^M}. \tag{3}$$

This scenario consists of sequentially executing the following steps:

1) create the list of nodes of the set V^M in order of decreasing the values of their generalized degree centrality in aggregate-network;

2) delete the first node from created list;

3) if criterion of attack success is reached, then finish the execution of scenario, otherwise go to point 4;

4) since the structure of aggregate-network changes as a result of removal of node (and its connections), compile a new list of nodes of the set V^M that remained, in order of decreasing recalculated values of their generalized degree centrality, and proceed to point 2.

The criterion of attack success in this case can be division of MLN's aggregate-network into unconnected components, increase the average length of shortest path, etc. [9]. Likewise, similar scenarios can be developed for other types of structural centralities of aggregate-network nodes, including without recalculating the values of these centralities [17]. The last type scenarios are usually used when the system is unable to redistribute the functions of lesioned elements between those that remained undamaged. The main disadvantage of structural importance indicators of network system nodes is their ambiguity, because even D. Krackhardt, using example of fairly simple network, showed [18] that its node, which is important according to the value of one type centrality, may be unimportant according to the value of another type centrality. The most objective importance indicator of a node in MS's structure is its betweenness centrality [17], which is equal to the ratio of quantity of shortest paths passing through this node to the quantity of all shortest paths in the network [19]. However, the calculation of this indicator for networks that have billions of elements, is a rather difficult computational problem.

2.3. Evaluation of the lesion consequences

In paper [1], it was shown that the structural model of MLNS makes it possible to determine the integral and partially local losses of multilayer network during and after targeted attack or its non-target lesion. The criterion of attack success can be not only the quantity of directly damged (dd), but also quantity of consequentially injured (ci) by this attack MLN elements. The aggregate-network model allows us to identify such elements of multilayer network. Let us denote by $\Omega^s_{dd} = \{n_{i_1}, n_{i_2}, ..., n_{i_k}\}$ the set of directly damaged (that is destroyed, completely blocked), as a result of attack, nodes of aggregate-network, and through $U(n_{i_l}) = \{n^1_{i_l}, n^2_{i_l}, ..., n^{m_l}_{i_l}\}$ the set of nodes of this network adjacent to n_{i_l}, $l = \overline{1,k}$. Then the set $\Omega^s_{ci} = \bigcup_{l=1}^{k} U(n_{i_l})$ determines the group of consequentially injured by lesion of the set Ω^s_{dd} nodes of MLN's aggregate-network. It is obvious that the defeat of a certain group of nodes with the highest generalized degree, which is realized by the above scenario, will lead to maximization of the set of consequentially injured multilayer network nodes, which can serve as the main attack goal.

3. A flow model of multilayer network system

We will use the flow model proposed in the article [1] to determine the indicators of functional importance of monoflow MLNS's elements and build scenarios of successive targeted attacks on operation process of multilayer systems. This choice is explained by the fact that the majority of real-world systems are created precisely to ensure the movement of flows through the relevant networks (transport, financial, trade, energy, information, and so on) or the movement of flows directly ensures their vital activity (the movement of blood, lymph, neuroimpulses in a living organism, etc.). Stopping the movement of flows in such systems inevitably leads to the cessation of their existence. In general, by flow we mean a certain real positive function correlated to each edge of the network. Let us reflect the set of flows that pass through all edges of multilayer system in the form of flow adjacency matrix VM(t), the elements of which are determined by the volumes of flows that passed through the edges of MLN (1) for the period $[t-T, t]$ up to the current moment of time $t \geq T$:

$$\mathbf{V}^M(t) = \{V^{km}_{ij}(t)\}^{N}_{i,j=1}, {}^{M}_{k,m=1}, \quad V^{km}_{ij}(t) = \frac{\tilde{V}^{km}_{ij}(t)}{\max\limits_{s,g=\overline{1,M}} \max\limits_{l,p=\overline{1,N^M}} \{\tilde{V}^{sg}_{lp}(t)\}}, \quad V^{km}_{ij}(t) \in [0,1], \quad (4)$$

where $\tilde{V}^{km}_{ij}(t)$ is the volume of flows that passed through the edge (n^k_i, n^m_j) of multilayer network for the time period $[t-T, t]$, $i, j = \overline{1, N^M}$, $k, m = \overline{1, M}$,

$t \geq T > 0$. It is obvious that structure of matrix VM(t) completely coincides with the structure of matrix AM. The elements of MLNS flow adjacency matrix are determined on the basis of empirical data about movement of flows through MLNS edges. Currently, with the help of modern means of information extraction, such data can be easily obtained for many natural and the vast majority of man-made systems [20]. The matrix VM(t) similarly to AM also has a block structure, in which the diagonal blocks $\mathbf{V}^{mm}(t)$ describe the volumes of intralayer flows in the mth layer, and the off-diagonal blocks $\mathbf{V}^{km}(t)$, $m \neq k$, describe the volumes of flows between the mth and kth layers of MLNS, $m, k = \overline{1, M}$, $t \geq T > 0$.

3.1. Flow aggregate-network of monoflow multilayer system

Let's define the concept of a flow aggregate-network of monoflow partially overlapped MLNS. Since we are considering the case when interlayer connections are possible only between nodes with the same numbers in total set of MLNS nodes, the structure of such aggregate-network can also be described in the form (2). Then the adjacency matrix $\mathbf{F}(t) = \{ f_{ij}(t) \}_{i,j=1}^{N^M}$, the elements of which are calculated according to the formulas

$f_{ij}(t) = \sum_{m=1}^{M} V_{ij}^{mm}(t)$, $i \neq j$, $i, j = \overline{1, N^M}$, $f_{ii}(t) = \sum_{m,k=1,\, m \neq k}^{M} V_{ii}^{mk}(t)$, $i = \overline{1, N^M}$,

completely defines a dynamic (in the sense of dependence on time) weighted network, which will be called the flow aggregate-network of this MLNS. The elements of matrix F(t) determine the integral flow characteristics of the edges and transition points of multilayer system, namely, the off-diagonal elements of this matrix are equal to the total volumes of flows passing through the edge (n_i, n_j), and the diagonal elements are equal to the total volumes of flows passing through the transition point n_i of MLNS during the time period $[t - T, t]$, $t \geq T > 0$, where (n_i, n_j) are the edges from the total set of edges EM, and n_i and n_j, $i, j = \overline{1, N^M}$, are the nodes from the total set of nodes VM.

3.2. Local flow characteristics of multilayer network systems elements

Let's determine the most important local flow characteristics of the MLNS elements. By local we mean a characteristic that describes the properties of element itself or one or another aspect of its interaction with directly connected (adjacent) elements of the system. The local flow characteristic of the edge (n_i^k, n_j^m) is equal to corresponding element of the flow model (4), i.e., the volume of flows that passed through this edge during the time period $[t - T, t]$, $t \geq T$. The local flow

characteristic of edge (n_i, n_j) of the total set of edges EM is equal to the value of element $f_{ij}(t)$, $i \neq j$, and the transition points n_i – to the value of element $f_{ii}(t)$, of the flow adjacency matrix F(t), $t \geq T$. As mentioned above, during the study of monoflow MLNS properties, the flow characteristics can be determined for the set of interlayer interactions in general. Based on this, the parameters

$$\varsigma_{ij}^{in}(t) = \sum_{m=1}^{M} V_{ji}^{mm}(t) = f_{ji}(t),\ \varsigma_{ij}^{out}(t) = \sum_{m=1}^{M} V_{ij}^{mm}(t) = f_{ij}(t),\ i, j = \overline{1, N^M},$$

determine the input and output flow connection strength between nodes n_i and n_j of the total set of nodes VM, taking into account all ways of implementing this connection in different layers of MLNS. Then parameters

$$\varsigma_i^{in}(t) = \sum_{j=1}^{N^M} \varsigma_{ji}^{in}(t) = \sum_{j=1}^{N^M} \varsigma_{ji}^{in}(t) = \sum_{j=1}^{N^M} f_{ji}(t),\ \varsigma_i^{out}(t) = \sum_{j=1}^{N^M} \varsigma_{ji}^{out}(t) = \sum_{j=1}^{N^M} f_{ij}(t),$$

determine the input and output flow connection aggregate-strength of the node n_i, $i, j = \overline{1, N^M}$, with all adjacent nodes from the total set of MLNS nodes, respectively. Then the generalized flow aggregate-degree of node n_i in the process of intra- and intersystem interactions is determined by the formula

$$\psi_i(t) = \varsigma_i^{in}(t) + \varsigma_i^{out}(t) + \sum_{m=1}^{M}\sum_{k=1}^{M} V_{ii}^{mk}(t) = \sum_{j=1,\ j\neq i}^{N^M} (f_{ij}(t) + f_{ji}(t)) + f_{ii}(t),\ t \geq T,$$

and is a functional analogue of the concept of centrality by generalized degree d_i, $i = \overline{1, N^M}$, which is calculated by formula (3).

Analogously to the above scenario of sequential targeted attack on MLNS structure (see section 1.2) using as importance indicators of elements the generalized structural degree d_i, a scenario of attack on MLNS operation process is being built using the generalized flow aggregate-degree $\psi_i(t)$, $t \geq T$, $i = \overline{1, N^M}$. A significant advantage of this scenario, compared to the structural one, is the consideration of not only aggregate-network nodes destroyed or completely blocked as a result of the attack, but also those whose operation process was limited as a result of the corresponding negative influence. Thus, the flow approach makes it possible, even at the level of using local importance indicators of the elements, to more accurately determine both the results of targeted attack (the level of lesion of directly attacked nodes) and the consequences of this attack for consequentially injured adjacent nodes of MLNS. Moreover, structural and functional scenarios can be combined. In particular, if the first several nodes in the list of the most important in terms of generalized flow aggregate-degree have the same value of this indicator, then they can be additionally ordered according to the decreasing values of generalized structural degree of these nodes. However, as in the case of structural, the functional scenarios, which use as importance indicators the local characteristics of MLNS elements, among the consequentially injured only adjacent to directly

damaged nodes are taken into account. This situation is quite acceptable for assortative networks [21], in which connections between elements are generally limited to adjacent nodes, but not for disassortative ones, the structure of which has the majority of man-made NCs, the connections between elements of which are usually implemented by paths.

Another advantage of the flow-based approach compared to the structural ones is the possibility of prioritizing the recovery of damaged but not completely destroyed system elements. The list of recovery priorities in general may not coincide with the list of the most important MLNS nodes according to a certain centrality. In particular, the importance of object restoration can be determined by the formula

$$\gamma = \alpha\left(1-\beta_{after}/\beta_{before}\right)/\alpha_{\max}$$

in which α is the value of selected centrality for the damaged node, $\alpha_{\max}$ is the maximum value of this centrality for all system nodes, β_{after} is the average volume of flows in the node after damage, β_{before} is the average volume of flows in the node before the lesion. According to this formula, a more damaged node among less important ones may require priority restoration.

3.3. Global flow characteristics of multilayer network system elements

Let's determine the most important global flow characteristics of the MLNS elements. By global we mean the characteristics of system element which describe one or another aspect of its interaction with all other elements or the system at a whole [16].

3.3.1. Influence parameters of system noge

Denote by $V^{out}(t, n_i^m, n_j^l)$ the total volume of flows generated in the node n_i^m and directed for final acceptance at MLNS node n_j^l for the period $[t-T, t]$, $t \geq T$. Parameter $V^{out}(t, n_i^m, n_j^l)$ determines the real strength of influence of node n_i^m on node n_j^l of multilayer system for the duration period T, $i, j = \overline{1, N^M}$, $m, l = \overline{1, M}$. Denote by $R_i^{m,l,out}(t) = \{j_{i_1}^l, ..., j_{i_{L_i^{ml}(t)}}^l\}$ the set of numbers of all nodes of the lth MLNS layer, which are the final receivers of flows generated in the node n_i^m, $L_i^{ml}(t)$ is the quantity of elements of the set $R_i^{m,l,out}(t)$, which can also change during the period $[t-T, t]$, $t \geq T$. Parameter

$$\xi_i^{m,l,out}(t) = \sum_{j \in R_i^{m,l,out}(t)} V^{out}(t, n_i^m, n_j^l) \Big/ s(\mathbf{V}^M(t)), \quad \xi_i^{m,l,out}(t) \in [0,1], \qquad (5)$$

determines the strength of influence of node n_i^m, as a flow generator, on the lth layer-system in general, $t \geq T$, $i = \overline{1, N^M}$, $m, l = \overline{1, M}$. In formula (5), the value

$$s(\mathbf{V}^M(t)) = \sum_{m,k=1}^{M} \sum_{i,j=1}^{N^M} V_{ij}^{mk}(t),$$

as the sum of all elements of matrix $\mathbf{V}^M(t)$ is the global flow characteristic of MLNS, which is equal to the total volumes of flows that passed through the multilayer system during the period $[t-T, t], t \geq T$.

The power of influence of node n_i^m on the lth layer-system is determined by means of parameter

$$p_i^{m,l,out}(t) = L_i^{ml}(t) / N^M,\ p_i^{m,l,out} \in [0,1],$$

and the set $R_i^{m,l,out}(t)$ will be called the influence domain of node n_i^m on the lth MLNS layer-system. Parameters $\xi_i^{m,l,out}(t)$, $p_i^{m,l,out}(t)$, and $R_i^{m,l,out}(t)$ will be called the output influence parameters of the node n_i^m as generators of flows on the lth MLNS layer-system, $i = \overline{1, N^M}$, $m, l = \overline{1, M}$. Analogously to the output ones are determined parameters of the strength $\xi_i^{l,m,in}(t)$, power $p_i^{l,m,in}(t)$, and domain $R_i^{l,m,in}(t)$ of input influence, which will be called the input influence parameters of the lth MLNS layer-system on the node n_i^m, as final receiver of flows generated in the nodes of the lth layer. The values of input and output influence parameters of the node n_i^m on lth layer make it possible to quantitatively determine how the lesion of this node will influence on functioning of the lth MLNS layer, namely, how many, which elements of the lth layer and in which measure will be consequentially injured, $i = \overline{1, N^M}$, $m, l = \overline{1, M}$, $[t-T, t], t \geq T$.

The output strength of influence of the node n_i^m as generator of flows on MLNS at a whole during the time period $[t-T, t], t \geq T$, is calculated according to the formula

$$\xi_i^{m,out}(t) = \sum_{l=1}^{M} \xi_i^{m,l,out}(t) / M,\ \ \xi_i^{m,out}(t) \in [0,1], \qquad (6)$$

in which the value $\xi_i^{m,l,out}(t)$ is determined by the formula (5). Domain of output influence $R_i^{m,out}(t)$ of the node n_i^m on MLNS is defined by the ratio

$$R_i^{m,out}(t) = \bigcup_{l=1}^{M} R_i^{m,l,out}(t).$$

Then the power $p_i^{m,out}(t)$ of output influence of the node n_i^m on MLNS is equal to the ratio of quantity of elements of the set $R_i^{m,out}(t)$ to the value NM. Similarly to output ones, the strength $\xi_i^{m,in}(t)$, domain $R_i^{m,in}(t)$ and power $p_i^{m,in}(t)$ of the MLNS input influence on the node n_i^m, $i=\overline{1,N^M}$, $m=\overline{1,M}$, as the final receiver of flows, during the time period $[t-T,t], t \geq T$ are determined. Lesion of the node-generator of flows means the need to find a new source of supply for the final receivers, and the receiver node – to search for new markets for producers, which will lead to at least temporary difficulties in their functioning. The influence parameters of separate node of MLNS allow us to determine what quantitative losses this will lead to and how many elements and which elements of intra- and intersystem interactions will spread.

3.3.2. Betweenness parameters of system node

The next type of global flow characteristics of MLNS elements are their betweenness parameters [16], which determine the importance of a node or an edge of multilayer network system in ensuring the movement of transit flows during intra- and intersystem interactions. In order to shorten the presentation, we will focus on the determination of betweenness parameters of MLNS transition points, as the most important elements that ensure intersystem interactions in monoflow partially overlapped multilayer systems. Denote by $V_i^{ml}(t)$ the total volume of flows that passed through the transition point n_i^{ml} during period $[t-T,t], t \geq T$, $i=\overline{1,N^M}$, $m,l=\overline{1,M}$.

The value

$$\Phi_i^{ml}(t)=V_i^{ml}(t)/s(\mathbf{V}^M(t)),\ \Phi_i^{ml}(t)\in[0,1], \qquad (7)$$

which determines the specific weight in the system the flows passing through the transition point n_i^{ml} during time period $[t-T,t], t \geq T$, will be called the measure of betweenness of this transition point in the process of interaction of the lth and mth MLNS layers. The set M_i^{ml} of all nodes of lth and mth MLNS layers, which are generators and final receivers of flows transiting through the node n_i^{ml}, will be called the betweenness domain, and the ratio η_i^{ml} of the quantity of nodes in the set M_i^{ml} to the value NM is the betweenness power of transition point n_i^{ml}, $i=\overline{1,N^M}$, $m,l=\overline{1,M}$.

The betweenness parameters of transition point n_i^m in the process of intersystem interactions within the entire MLNS will be determined as follows. The measure of betweenness $\Phi_i^m(t)$ of transition point n_i^m in the entire multilayer system can be calculated using the formula

$$\Phi_i^m(t) = \sum_{l=1,\, l\neq m}^{M} \Phi_i^{ml}(t)/(M-1),\ \ \Phi_i^m(t) \in [0,1], \qquad (8)$$

in which the value $\Phi_i^{ml}(t)$ is calculated according to (7). The betweenness domain of transition point n_i^m in the entire MLNS is determined by the ratio

$$M_i^m(t) = \bigcup_{l=1,\, l\neq m}^{M} M_i^{ml}(t)\cdot$$

Then the power $N_i^m(t)$ of betweenness of transition point n_i^m in the MLNS at a whole is equal to the ratio of quantity of elements of the set $\mathrm{M}_i^m(t)$ to the value NM. Note, that for nodes that are not transition points of MLNS, the parameters of measure, domain and power of betweenness are determined according to the same principles. Similarly, it is possible to determine the parameters of measure $\Phi_{ij}^m(t)$, domain $M_{ij}^m(t)$, and power $N_{ij}^m(t)$ of betweenness for the edge (n_i^m, n_j^m) of MLNS mth layer, $i, j = \overline{1, N^M}$, $m = \overline{1, M}$, $[t-T, t], t \geq T$. The values of betweenness parameters of MLNS node n_i^m, $i = \overline{1, N^M}$, $m = \overline{1, M}$, allow us by means of quantitative measurement to determine how the lesion of this node will affect the provision of transit flows through the multilayer system and to what extent, how many and which elements will be consequentially injured.

3.3.3. Specific scenarios of targeted attacks

The importance of node n_i of the total set of MLNS nodes as generator, final receiver or flow transitor is calculated using formulas

$$\xi_i^{out}(t) = \sum_{m=1}^{M} \xi_i^{m,out}(t) \Big/ M, \qquad (9)$$

$$\xi_i^{in}(t) = \sum_{m=1}^{M} \xi_i^{m,in}(t)/M,\ \ \xi_i^{out}, \xi_i^{in}(t) \in [0,1], \qquad (10)$$

$$\Phi_i(t) = \sum_{m=1}^{M} \Phi_i^m(t)/M,\ \ \Phi_i(t) \in [0,1],\ i = \overline{1, N^M},\ [t-T, t],\ t \geq T, \qquad (11)$$

respectively. Domains of input $R_i^{in}(t)$, output $R_i^{out}(t)$ influence, and betweenness $M_i(t)$ of the node n_i in MLNS will determine by formulas

$$R_i^{in}(t) = \bigcup_{m=1}^{M} R_i^{m,in}(t),\ R_i^{out}(t) = \bigcup_{m=1}^{M} R_i^{m,out}(t),\ M_i(t) = \bigcup_{m=1}^{M} M_i^m(t),$$

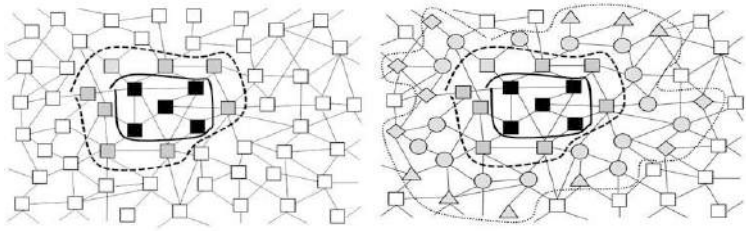
and the powers of input $p_i^{in}(t)$, output $p_i^{out}(t)$ influence, and betweenness $N_i(t)$ of the node n_i on MLNS at a whole as the ratio of quantity of elements of the sets $R_i^{in}(t)$, $R_i^{out}(t)$, and $M_i(t)$, $i=\overline{1,N^M}$, to the value NM respectively.

Depending on the purpose of attack, the targets of lesion can be nodes-generators, nodes – final receivers, nodes-transitors of flows or only transition points of MLNS. For each of these types of multilayer system elements, it is possible to build specific scenarios of targeted attacks, using as importance indicators of nodes the parameters of influence or betweenness, determined above by formulas (5), (6), (9), (10) or (7), (8), (11) respectively. For example, an embargo on energy carriers means blocking generator nodes (countries that extract and supply such carriers), a ban on the supply of high-tech products (microcircuits, modern computers or equipment) – blocking the final receivers of flows (countries or companies that use such products), blocking of transit nodes (prohibition of international air flights over the territory of Russia or crossing of the Bosphorus Strait by its military ships) – redirection of the flow traffic by other routes. Before carrying out an attack on generator (final receivers) or transit nodes, it is possible to identify domains of output (input) influence or domains of betweenness, which allow us to identify nodes that may be consequentially injured by the attack, as well as to quantify the possible level of their losses. It makes sense to carry out such actions before imposing sanctions against the aggressor country. Quantifying the losses of sanctioning party compared to the damage done to attacked system allows us to determine the feasibility of attack.

3.3.4. Aggregate-network and lesion consequences

It is obvious that the influence and betweenness parameters of MLNS nodes and edges are related to the influence and betweenness parameters of nodes and edges of its flow aggregate-network. Thus, the output strength of influence of node n_i of the general set VM in the aggregate-network is equal to the value $\xi_i^{out}(t)$ calculated by formula (9), the domain of output influence of this node is the projection of domain $R_i^{out}(t)$ onto the aggregate-network (2), and the power of output influence is equal to the ratio of quantity of elements of this projection to the value NM. The input strength of aggregate-network influence on a node n_i is equal to the value $\xi_i^{in}(t)$, which is calculated by formula (10), the domain of input influence of this node is the projection of domain $R_i^{in}(t)$ onto the aggregate-network (2), and the power of the input influence is equal to the ratio of quantity of elements of this projection to the value NM. The measure of betweenness of node n_i in the

aggregate-network is equal to the value $\Phi_i(t)$, which is calculated by formula (11), the domain of betweenness of this node is the projection of domain $M_i(t)$ onto the aggregate-network (2), and the power of betweenness is equal to the ratio of quantity of elements of this projection to the value NM.



a) b)

Figure 3. *Consequences of targeted attack, obtained on the basis of analysis of structural aggregate-network (a) and parameters of influence and betweenness of flow aggregate-network nodes (b)*

Figure 3 contains an example of lesions received by MLNS aggregate-network as a result of targeted attack. Here the black squares bounded by continuous curve indicate the directly damaged nodes, and dark gray squares bounded by a dashed curve indicate the consequentially injured nodes adjacent to the directly damaged ones obtained on the basis of structural approach, white squares indicate undamaged nodes (Figure 3 a). In Figure 3 b, the gray rhombuses, triangles, and circles bounded by a dotted curve indicate consequentially injured generator, final receivers, and transitor nodes obtained on the basis of flow approach, respectively. As follows from these figures, the domain of consequentially injured elements determined on the basis of flow approach can be much larger and more accurate in the sense of displaying the node type than the domain of adjacent to directly damaged nodes of the network system determined on the basis of the structural approach.

3.3.5. Parameters of interaction and comprehensive targeted attack scenario

Based on the input and output influence as well as betweenness parameters of the node n_i, we can determine the global indicators of interaction of this node with the MLNS at a whole, namely, the parameter $\Xi_i(t)$ of interaction strength of the node n_i with multilayer system, which is calculated according to the formula

$$\Xi_i(t) = (\xi_i^{out}(t) + \xi_i^{in}(t) + \Phi_i(t))/3, \quad t \geq T, \tag{12}$$

determines its overall role in multilayer system as generator, final receiver and flow transitor; the domain $\Omega_i(t)$ of interaction of the node n_i with MLNS is determined by the formula

$$\Theta_i(t) = R_i^{in}(t) \bigcup R_i^{out}(t) \bigcup M_i(t),$$

and the power of interaction of the node n_i with MLNS is equal to the ratio of quantity of elements of domain $\Omega_i(t)$, $t \geq T$, to the value NM. It is obvious that interaction parameters of MLNS nodes are related to the interaction parameters of its flow aggregate-network nodes. Thus, the strength of interaction of node n_i of the general set of nodes VM with the MLNS flow aggregate-network is equal to the value $\Xi_i(t)$, which is calculated according to formula (12), the domain of interaction of this node is the projection of domain $\Omega_i(t)$ onto the aggregate-network (2), and the power of interaction is equal to the ratio of quantity of elements of this projection to the value NM.

Let's build a scenario of consistent targeted attack on multilayer system, choosing as an importance indicator of node the strength of its interaction with MLNS flow aggregate-network. Such scenario, which achieves the comprehenciveness of attack on the functionally most important system nodes, will look like this:

1) compile a list of nodes of the set V^M in order of decreasing values of their strength of interaction with the flow aggregate-network;

2) delete the first node from the created list;

3) if the criterion of attack success is reached, then finish the execution of scenario, otherwise go to point 4;

4) since the operation process of flow aggregate-network changes as a result of removing a node (and its connections), compile a new list of nodes of the set V^M that remained in the order of decreasing recalculated values of their interaction strength with flow aggregate-network and proceed to point 2.

In this case, it is advisable to choose a reduction in the volume of flows $s(\mathbf{V}^M(t)), t \geq T > 0$, in MLNS by a certain predetermined value as the criterion for the attack success.

4. Simultaneous group attacks on the multilayer network systems

It is obvious that simultaneous group attacks or non-target system lesions are much more difficult than successive attacks on the most important MLNS elements, both from the point of view of its protection and overcoming the consequences [22]. In this section, for the sake of brevity, we will limit ourselves to the consideration of targeted attacks on the MLNS aggregate-network.

4.1. Classification of targetet group attacks

We divide simultaneous group negative influences into one-time (the attack on Pearl Harbor on December 7, 1941), repeated (the permanent bombing of London during the Second World War) and successive (attacks on transformer stations of the Ukrainian power system in 2022-2024). Repeated group attacks are carried out regularly at certain time intervals on the same system objects. Consecutive group attacks differ from repeated ones by changing the targets. A particular danger is that successful sequential group attacks can lead to system-wide MLNS lesions, for example, a prolonged blackout in the country. In the case of targeted attacks, this separation is often determined by the attacker's ability to launch subsequent massive attacks and the ability of attacked system to effectively defend and counter them. It is clear that each of above types of attack requires the development of specific type of scenarios for its most likely implementation. The simplest scenario of one-time group attack is obviously implemented by attempt to simultaneously defeat a group of the most important MLNS elements according to determined centrality. The repeated attack scenario is realized by attempt to damage a preselected and previously attacked, but not destroyed, group of multilayer system elements. A sequential group attack scenario involves the consecutive execution of following steps:

1) compile a list of groups of MLNS aggregate-network’s nodes in order of decreasing indicators of their importance in the system, selected according to a certain feature;

2) delete the first group from the created list;

3) if the criterion of attack success is reached, then complete the execution of scenario, otherwise go to point 4;

4) since the system structure changes as a result of removal of certain group of nodes (and their connections), compile a new list of groups in order of decreasing recalculated indicators of their importance in the MLNS aggregate-network and proceed to point 2.

If, during implementation of last scenario, a certain group of nodes contains too many elements that the attacker is unable to damage simultaneously, then such group is divided into the minimum quantity of connected subgroups available for such attacks. In addition, the execution of scenario may terminate when the attacking party has exhausted the resources to continue the attack. It follows from the above scenarios that the main way to increase their effectiveness is to choose the importance indicators of group in multilayer network system, the lesion of which will cause it the greatest damage. The most obvious way of such choice is to form a list of MLNS aggregate-network’s nodes in order of decreasing the values of their centrality of selected type and form a group from the first nodes of this list, the quantity of which is determined by the ability of attacker to carry out a simultaneous attack on them. The second method is based on the principle of nested hierarchy of the network system [23]. The method proposed by us consists in applying the concept of k-core of aggregate-network, as the largest subnet of source network, the centrality of which, according to the generalized structural degree of nodes, is at

least k >1 [24]. This method is based on the use of the most structurally important components of aggregate-network and obviously fits into the above scenario of successive group attacks. In particular, groups are initially selected for the maximum value k for a given aggregate-network, which is then sequentially reduced until the attack success criterion is met.

During the development of scenarios of simultaneous group attacks on MLNS operation process, as a functionally most important component of the flow aggregate-network, the concept of its flow λ-core [25] can be used, as the largest subsystem of source system, elements of flow adjacency matrix F(t) of which have values not less than $\lambda \in [0,1]$. It is obvious that the larger the value λ, the more important from a functional point of view aggregate-network's components are reflected by its λ-core. They can become one of the primary targets of simultaneous group attack. Similarly, as for aggregate-network elements, we can determine the parameters of influence and betweenness of its λ-core, which significantly deepens the analysis of network system lesions.

4.2. Optimization of targeted group attack scenarios

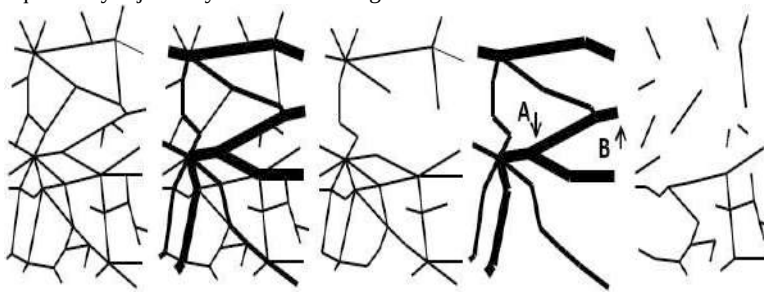
One of the ways to protect the system is to counterattack the intruder. It is clear that the organizers of such counterattacks also suffer considerable losses. That is, the problem of optimizing attack scenarios arises, namely, how to destroy or block the operation of minimum quantity of nodes of attacked party, to cause it the greatest possible lesion. A similar situation is observed during the development of scenarios for combating the spread of non-target lesions, for example, epidemics of dangerous infectious disease. In particular, how to minimize the volume of passenger flows through the transportation network by blocking the smallest possible quantity of nodes that ensure the movement of these flows. Obviously that it is advisable to take into account not only the magnitude of direct negative influence, but also the scale of mediated lesion consequences. Let's illustrate this on the example of railway layer of the country's general transport system. Above, for the construction of simultaneous group attacks scenarios, it was proposed to use the concepts of structural k- and flow λ-core of network system. We will show that the use of flow λ-cores compared to structural k-cores of MLNS aggregate-networks is significantly more effective when building scenarios of group targeted attacks, both from the point of view of possible lesion of the most functionally important aggregate-network's components, and for the purpose of optimizing these scenarios in terms of the quantity of attack targets. Let's consider the railway transport system (RTS) of the western region of Ukraine. In fig. 4a shows the structure of this system, and in fig. 4b – the same structure, but in the form of weighted network, which schematically reflects the volumes of cargo flows that passed through its edges during 2021 (the thickness of lines is proportional to the volumes of flows). Note that this real-world network contains 354 nodes in total, but in fig. 4a-b, only 29 nodes and 62 edges are displayed (transit nodes with structural degree 2 are not shown, and an edge is considered to be a line

that connects two nodes with degree greater than 2). In fig. 4c contains the structural 4-core of RTS, which includes 12 nodes and 35 edges, and in fig. 4d is its flow 0.8-core, which contains 4 nodes and 12 edges. One of the main disadvantages of k-cores compared to flow cores is the possibility of excluding functionally important components of the network system (path A-B in fig. 4d).

From fig. 4 also follows that the flow 0.8-core reflects a functionally more important subsystem of RTS and the target of group attack on it is a much smaller quantity of nodes than 4-core of corresponding structure.

Easy to see that in both cases, a successful attack on NS nodes selected with the help of k- and λ -cores will lead to actual termination of its operation process, as it divides RTS into unconnected components (fig. 4e), but in the second case, the goal of attack is achieved with significantly less efforts (three times in terms of quantity of nodes and edges). Thus, the flow-based approach allows us to build scenarios that are much more optimal from the point of view of attacking side's efforts than the structural one.

By analyzing the parameters of influence and betweenness of 0.8-core of given RTS fragment, it was established that all elements of this fragment will be consequentially injured by a successful targeted attack on it.



a) b) c) d) e)

Figure 4. *Examples of structure (a), operation process (b), structural 4-core (c), flow 0.8-core (d), and the addition to flow 0.8-core in the source structure (e) of railway transport system of the Western region of Ukraine*

5. Simultaneous system-wide attacks on the multilayer network systems

Typical scenarios of successive and simultaneous targeted attacks on layers-systems of MLNS are built according to the same principles as corresponding scenarios of such attacks on elements and subsystems of multilayer systems [1] with the difference that the objects of attack are the layers of MLNS. At the same time, the main way to improve the effectiveness of such attacks is the selection of structural and/or functional importance indicators of MLNS layer, the lesion of which will cause the greatest damage to it.

5.1. Structural and functional importance indicators of MLNS system-layers

The global structural characteristics of multilayer network, which is generated by monoflow partially overlapped MLNS, can be considered the quantity of layers-networks that are part of it, as well as the dimension, density, diameter, average length of the shortest path, the total quantity of transition points [11], etc. Let us define some general and most important for the vulnerability analysis of structure of intersystem interactions characteristics of the layers of partially overlapped MLNS, the values of which should be taken into account during construction the scenarios of system-wide targeted attacks:

1) the specific weight of the set of nodes of the m^{th} layer in the total set of nodes of partially overlapped MLN;

2) the specific weight of the set of edges of the m^{th} layer in the total set of edges of partially overlapped MLN;

3) input degree of the m^{th} layer, which is equal to quantity of transition point into this layer from all other layers of MLN or total quantity its input interlayer connections;

4) output degree of the m^{th} layer, which is equal to quantity of transition point from this layer into all other layers of MLN or total quantity its output interlayer connections;

5) the specific weight of transition points of the m^{th} layer in the set of all transition points of partially overlapped MLN, which is determined availability of interlayer interactions for this layer, $m = \overline{1, M}$.

Let us determine the most important flow characteristics of system-layers in monoflow partially overlapped multilayer network system, values of which should be taken into account during construction scenarios of system-wide targeted attacks:

1) the specific volumes of intralayers flows in the m^{th} MLNS layer;

2) the specific volumes of output flows of the m^{th} layer in MLNS, which reflect its role as a flow generator in multilayer system;

3) the specific volumes of input flows of the m^{th} layer in MLNS, which reflect its role as a final receiver of flows in multilayer system;

4) the specific volumes of transit flows that pass through the m^{th} MLNS layer, $m = \overline{1, M}$.

It should be noted that the layers with the highest values of structural and functional importance indicators listed above can become the primary targets of system-wide attacks on MLNS and be used during construction the scenarios for such attacks.

5.2. Influence and betweenness parameters of MLNS system-layers

Such global flow characteristics of nodes and subsystems of MLNS were determined in the monograph [16], as parameters of their output and input influence and betweenness.

These parameters make it possible to calculate the importance of separate components of multilayer system as generators, final receivers and transitors of flows and allow us to build effective scenarios of successive and simultaneous group targeted attacks on the process of intersystem interactions. However, for the formation of effective scenarios of successive system-wide attacks, it is advisable to determine the influence and betweenness parameters of separate MLNS system-layers.

5.2.1. Influence parameters of system-layers

Output strength of influence of the m^{th} layer as generator of flows on the l^{th} MLNS layer as final receiver of these flows during the time period $[t-T,t],\ t \geq T > 0$, is calculated by the formula

$$\xi^{m,l,out}(t) = \sum_{i=1}^{N^M} \xi_i^{m,l,out}(t) / N^M \ ,\ \xi^{m,l,out}(t) \in [0,1] , \qquad (13)$$

in which the parameter $\xi_i^{m,l,out}(t)$ is calculated by formula (5). Domain $R^{m,l,out}(t)$ of output influence of the m^{th} layer on the l^{th} MLNS layer are determined as union of influence domains of nodes-generators of flows of the m^{th} layer on the nodes – final receivers these flows in l^{th} MLNS layer, and the power of this influence $p^{m,l,out}(t)$ is equal to the ratio of guantity of elements of the domain $R^{m,l,out}(t)$ to value N^M. Similarly, the input strength $\xi^{m,l,in}(t)$, domain $R^{m,l,in}(t)$ and power $p^{m,l,in}(t)$ of influence of the m^{th} layer as final receiver of flows on the l^{th} MLNS layer as generator of these flows are determined during the time period $[t-T,t], t \geq T > 0$, $m \neq l$, $m,l = \overline{1,M}$.

The output strength of influence of the m^{th} layer as generator of flows on the MLNS at a whole is calculated by formula

$$\xi^{m,out}(t) = \sum_{l=1,\ l \neq m}^{M} \xi^{m,l,out}(t) / (M-1) \ ,\ \xi^{m,out}(t) \in [0,1] ,$$

in which the value $\xi^{m,l,out}(t)$ is calculated according to (13) and domain of this influence is determined by the ratio

$$R^{m,out}(t) = \bigcup_{l=1,\ l \neq m}^{M} R^{m,l,out}(t) .$$

The power of output influence of the m^{th} layer on the MLNS at a whole is equal to the ratio of quantity of element of domain $R^{m,out}(t)$ to the value N^M. Similarly, the input strength $\xi^{m,in}(t)$, domain $R^{m,in}(t)$ and power $p^{m,in}(t)$ of influence of

MLNS on the m^{th} layer as final receiver of flows are determined during the time period $[t-T,t], t \geq T > 0$. The values of parameters of input and output influence of the m^{th} layer on the l^{th} layer or MLNS at a whole make it possible to quantitatively determine how the lesion of this layer will impact on the operation process of the l^{th} layer and the multilayer network system in general, and how many components of intersystem interactions and to what extent will be affected, $m, l = \overline{1, M}$.

5.2.2. Betweenness parameters of system-layers

The next type of global flow characteristics of each MLNS layer is its betweenness parameters, which determine the importance of layer of the multilayer network system in ensuring the movement of transit flows during intersystem interactions [16]. The measure of betweenness of the m^{th} layer within the entire MLNS during period $[t-T,t], t \geq T > 0$, as transitor of flows is calculated using the formula

$$\Phi^m(t) = \sum_{i=1}^{N^M} \Phi_i^m(t) / N^M, \Phi^m(t) \in [0,1],$$

in which the value $\Phi_i^m(t)$ is calculated by formula (8). Betweenness domain of the m^{th} layer in MLNS at a whole is determined from the ratio

$$M^m(t) = \bigcup_{i=1}^{N^M} M_i^m(t),$$

in which $M_i^m(t)$ denotes the betweenness domain of i^{th} transition point of the m^{th} layer within the entire MLNS, and the betweenness power of the m^{th} layer $N^m(t)$ is equal to the ratio of quantity of elements of domain $M^m(t)$, $[t-T,t], t \geq T > 0$, to the value N^M. The values of betweenness parameters of the m^{th} layer in MLNS make it possible to quantitatively determine how the lesion of this layer will impact on the process of intersystem interactions in general and how many, which components of the multilayer system and to what extent will be affected, $m = \overline{1, M}$.

5.2.3. Interaction parameters of system-layers

Based on the parameters of input and output influence, as well as betweenness parameters of the m^{th} layer, we can form a global indicator of interaction of this layer with MLNS in general, namely, parameter $\Xi^m(t)$ of the strength of interaction of the m^{th} layer with multilayer system, which is calculated according to the formula

$$\Xi^m(t) = (\xi^{m,out}(t) + \xi^{m,in}(t) + \Phi^m(t))/3,$$

and defines its overall role in multilayer system as generator, final receiver and transitor of flows; domain $\Omega^m(t)$ of interaction of the m^{th} layer with MLNS is determined by the ratio

$$\Omega^m(t) = R^{m,out}(t) \bigcup R^{m,in}(t) \bigcup M^m(t),$$

and the power of interaction of the m^{th} layer with MLNS is equal to the ratio of quantity of elemens of the domain $\Omega^m(t)$, $m = \overline{1,M}$, $t \geq T > 0$, to the value N^M. It is clear that interaction parameters of a layer with multilayer network system define the importance of this layer in the process of intersystem interactions and make it possible to quantitatively determine how the lesion of this layer will impact on the MLNS opertion process and how many, exactly which of its components will be influenced and to what extent.

5.2.4. A comprehensive system-wide attack scenario

Let us build a scenario of successive targeted system-wide attack, choosing the parameter of strength of layer interaction with multilayer system as its importance indicator. Such scenario, which achieves the comprehensiveness of attack on the most functionally important MLNS layers, will look like this:

1) compile the list of MLNS layers in order of decreasing values of the parameters of strength of their interaction with multilayer system;

2) remove the first layer from the created list;

3) if the criterion of attack success is reached, then complete the execution of scenario, otherwise go to point 4;

4) since the structure and operation process of multilayer system changes due to removal of certain layer (and its interlayer connections), compile a new list of layers in the order of decreasing values of the parameters of strength of their interaction with MLNS and proceed to point 2.

In this case, it is advisable to choose the reduction of volumes of flows $s(\mathbf{V}^M(t)), t \geq T > 0$, in the multilayer system by a certain predetermined value as the attack success criterion. The next most difficult to ensure protection and overcome the consequences are simultaneous attacks on several or all MLNS layers, which include hybrid wars, sanctions against countries that pose a threat to world security, etc. The scenarios of such attacks are built according to the same principles as those of simultaneous group attacks, with the difference that the attack targets are not the most important by certain characteristics the groups of nodes, but the MLNS layers.

5.3. Targeted attacks and the scale of system lesions

The concept of system protection is closely related to determining the scale of targeted attack and its consequences. In particular, it is advisable to distinguish:

1) the scale of planned and executed attack, which is determined by the quantity of targets expected to be hit and the quantity of means used for this purpose, for example, 50 energy infrastructure facilities and 100 cruise and ballistic missiles aimed at them for destruction;

2) the scale of direct lesion, that is, the quantity of targets critically damaged or completely destroyed by the attack, for example, out of 100 launched missiles, 85 were shot down by anti-missile and anti-aircraft defenses, and the remaining 15 missiles hit 10 targets; in this case, the scale of direct damage is equal to the percentage of destroyed objects from the quantity of planned ones;

3) the scale of consequentially injured system elements, i.e. the quantity of objects that suffered certain damage as a result of the attack in addition to those directly damaged, for example, introduction of blackout schedules caused by consecutive and simultaneous attacks on the energy infrastructure of Ukraine.

The scale of direct damage is related to the quantitative indicator of system's protection, which is equal to the percentage of destroyed means of damage from all involved. The scale of mediated lesion, as well as the level of protection against it, is much more difficult to calculate, because it must take into account the disruption of structure and destabilization of the work of all system elements and even the moral and psychological damage caused as a result of the attack.

The concept of system sensitivity to the consequences of negative influence can be associated with the scale of indirect lesion, as the ratio of quantity of directly damaged to the quantity of consequentially injured elements. It is obvious that the closer the value of this indicator is to zero, the more sensitive the system is to negative influences, since a small quantity of directly damaged generates a large quantity of consequentially injured MLNS elements.

Therefore, it makes sense to believe that the nature of attack should be determined not only by the quantity of directly damaged MLNS elements, but also by the scale of indirect losses caused to the system, although it is much more difficult to ensure its protection against such lesion.

6. Conclusions

In 2019-2024, humanity faced two global challenges, the first of which (Covid-19 pandemic) is a vivid example of real-world system-wide non-target lesion, and the second is a targeted attack (attack of the russian federation on Ukraine) and the resulting threat of a global food, energy, and financial crisis and reverse comprehensive sanctions against the aggressor, the negative consequences of which affected almost all countries of the world.

Humanity proved to be unprepared for such challenges, but no less dangerous threats remain. Over the past half century, 67% of biological species known to man have disappeared, and over the past 20 years, the costs of combating climate disasters have increased 8 times. Currently, scientists know more than 20 viruses of dangerous infectious diseases, the mutations of which can lead to the spread of pandemics, much more catastrophic than Covid-19, the threat of global military and even intercivilizational conflicts is increasing, etc.

This confirms the relevance of studying the features of successive, group and system-wide lesions of complex network and multilayer network systems and developing methods of effective protection against them.

The concepts of structural and flow aggregate-networks of monoflow multilayer network system are introduced in the article in order to reduce the dimensionality of MLNS models and simplify the analysis of their vulnerability to heterogeneous negative influences.

The main local and global structural and flow characteristics of multilayer system and its aggregate-network elements are determined and the relationship between them is established.

These characteristics are chosen as importance indicators of MLNS nodes, with the help of which effective structural and functional scenarios of successive, group and system-wide targeted attacks on multilayer network systems are built. It is shown how, on the basis of various models of intersystem interactions, the domains of directly damaged and consequentially injured by the negative influence the system elements are determined. To optimize the scenarios of simultaneous group attacks, as a means of active protection against the source of negative influence, the application of flow cores of the MLNS's aggregate-network is proposed. The advantages of such scenarios over structural ones based on the concept of k-core of aggregate-network are shown on the example of real-world network system.

The advantages of flow-based approach for studying the vulnerability of intersystem interactions process and quantifying the level of losses caused to this process as a result of various negative influences are established. An objective evaluation of the scale of real or potential lesions allows us to develop strategies to protect not only separate MLNS elements, but also the system as a whole, and to prevent the consequences of local targeted attacks and non-target lesions from growing into group and system-wide ones.

The next steps of our research are the study of MLNS vulnerability damage not nodes, but connections of complex network and multilayer network systems, which are much easier to implement in practice and can lead to granulation of intra- and intersystem interactions.

7. References

[1] O. Polishchuk, M. Yadzhak, On the Vulnerability and Protection Strategies of Complex Network Systems and Intersystem Interactions, CEUR-WS 3538 (2023) 267-281.

[2] M. Bellingerio, D. Cassi, S. Vincenzi, Efficiency of attack strategies on complex model and real-world networks, Physica A: Statistical Mechanics and its Applications 414 (2014) 174-180. doi: 10.1016/j.physa.2014.06.079.

[3] Q. Nguyen et al, Conditional attack strategy for real-world complex networks, Physica A: Statistical Mechanics and its Applications 530 (2019) 12156. doi: 10.1016/j.physa.2019. 121561.

[4] M. Alonso et al, Cyber-physical vulnerability assessment in smart grids based on multilayer complex networks, Sensors 21 17 (2021) 5826.

[5] F. Zhou et al, Influence of interlink topology on multilayer network robustness, Sustainability 12(3) (2020) 1202. doi: 10.3390/su12031202.

[6] B. Fan et al, Critical nodes identification for vulnerability analysis of power communication networks, IET Communications 14 4 (2020) 703-713. doi: 10.1049/iet-com.2019.0179.
[7] J. Bluszcz, M. Valente, The Economic Costs of Hybrid Wars: The Case of Ukraine, Defence and Peace Economics 33 1 (2022) 1-25. doi: 10.1080/10242694.2020.1791616.
[8] N. Vindegaard, M.E. Benros, COVID-19 pandemic and mental health consequences: Sys-tematic review of the current evidence, Brain, Behavior, and Immunity 89 (2020) 531-542. doi: 10.1016/j.bbi.2020.05.048.
[9] J. Mariyam, D.S. Lekha, Need for a realistic measure of attack severity in centrality based node attack strategies, in Complex Networks and Their Applications XI, R.M. Benito et al, Eds. Springer: Cham (2022) 857-866.
[10] L. Glenn, Understanding the influence of all nodes in a network, Science Reports 5 (2015) 8665. doi: 10.1038/srep08665.
[11] S. Boccaletti et al, The structure and dynamics of multilayer networks, Physics Reports 544 1 (2014) 1-122. doi: 10.1016/j.physrep.2014.07.001.
[12] L.G. Alvarez-Zuzek et all, Dynamic vaccination in partially overlapped multiplex network, Physical Review E 99 (2019) 012302. doi: 10.1103/PhysRevE.99.012302.
[13] M. Berlingerio et all, Multidimensional networks: foundations of structural analysis, World Wide Web 16 (2013) 567–593. doi: 10.1007/s11280-012-0190-4.
[14] A. Saxena, S. Iyengar, Centrality measures in complex networks: A survey, arXiv: 2011. 07190, 2020.
[15] O. Polishchuk, Structural Cores and Problems of Vulnerability of Partially Overlapped Multilayer Networks, in Complex Networks and Their Applications XI, H. Cherifi et al, Eds. Springer: Cham (2023) 613-624. doi: 10.1007/978-3-031-21127-0_50.
[16] O.D. Polishchuk, M.S. Yadzhak, Models and methods of comprehensive research of complex network systems and intersystem interactions. Pidstryhach Institute for Applied Problems of Mechanics and Mathematics, National Academy of Sciences of Ukraine: Lviv, 2023.
[17] Yu. Holovach et al, Complex networks, Journal of physical studies 10 4 (2006) 247–289. doi: 10.30970/jps.10.247.
[18] D. Krackhardt, Assessing the political landscape: Structure, cognition, and power in organiza-tions, Administrative Science Quarterly 35 2 (1990) 342–369. doi: 10.2307/2393394.
[19] A. Barrett, M. Barthelemy, A. Vespignani, The architecture of complex weighted networks: Measurements and models, in Large Scale Structure and Dynamics of Complex Networks, G. Caldarelli, Eds. World Scientific: Singapore (2007) 67-92. doi: 10.1142/9789812771681_ 0005.
[20] A.-L. Barabasi, The architecture of complexity, IEEE Control Systems Magazine 27 4 (2007) 33-42. doi: 10.1109/MCS.2007.384127.
[21] R. Noldus, P. Van Mieghem, Assortativity in complex networks, Journal of Complex Networks 3 4 (2015) 507-542. doi: 10.1093/comnet/cnv005.

[22] O. Polishchuk, Protection of Complex Network Systems From Targeted Attacks and Non-Target Lesions, Qeios (2024) RH65Y6. doi: 10.32388/RH65Y6.
[23] O. Polishchuk, M. Yadzhak, Network structures and systems: III. Hierarchies and networks, System research and informational technologies 4 (2018) 82-95. doi: 0.20535/SRIT. 2308-8893.2018.4.07.
[24] S.N. Dorogovtsev, A.V. Goltsev, J.F.F. Mendes, K-core organization of complex networks. Physical review letters 96(4) (2006) 040601. doi: 10.1103/ PhysRevLett.96.040601.
[25] O. Polishchuk, M. Yadzhak, Network structures and systems: II. Cores of networks and multiplexes. System research and informational technologies 3 (2018) 38-51. doi: 10.20535/SRIT.2308-8893.2018.3.04.

ЗАХИСТ БАГАТОРІВНЕВИХ МЕРЕЖЕВИХ СИСТЕМ ВІД ПОСЛІДОВНИХ, ГРУПОВИХ І ЗАГАЛЬНОСИСТЕМНИХ ЦІЛЕСПРЯМОВАНИХ АТАК

Dr.Sci. О. Поліщук
Інститут прикладних проблем механіки і математики ім. Підстригача НАН України, Україна
E-mail: od_polishchuk@ukr.net

Анотація. *Розглянуто структурний і потоковий підходи до аналізу вразливості багаторівневих мережевих систем (MLNS) від цілеспрямованих атак і нецільових уражень різного походження. Визначено локальні та глобальні структурні та потокові характеристики елементів монопотокової багатошарової системи для побудови сценаріїв цілеспрямованих атак на структуру та процес функціонування MLNS та оцінки їх наслідків. Для спрощення побудови та підвищення ефективності таких сценаріїв введено поняття структурних і потокових агрегатів-мереж монопотокових МЛНС та показано зв'язок між показниками важливості їх елементів та відповідними показниками вузлів багатошарової системи. Продемонстровано переваги потокового підходу перед структурним, як у сенсі аналізу вразливості реальних MLNS, так і оцінки наслідків негативних впливів різного характеру.*

Ключові слова: *складна мережа, мережева система, міжсистемні взаємодії, багаторівнева мережева система, модель потоку, агрегатна мережа, вплив, міжсистемність, цілеспрямована атака, вразливість.*

UDC 621.391:519.2:519.7

NEW STATISTICAL CRITERIA FOR CHECKING INDEPENDENCE OF BIT RANDOM VARIABLES AND SEQUENCES

Dr.Sci. L. Kovalchuk ORCID: 0000-0003-2874-7950
G.E. Pukhov Institute for Modelling in Energy Engineering, Ukraine
E-mail: lusi.kovalchuk@gmail.com
Dr.Sci. A. Davydenko ORCID: 0000-0001-6466-1690
G.E. Pukhov Institute for Modelling in Energy Engineering, Ukraine
E-mail: davidenkoan@gmail.com
O. Bespalov 0000-0001-7126-6752
G.E. Pukhov Institute for Modelling in Energy Engineering, Ukraine
E-mail: alexb5dh@gmail.com

Abstract. *The paper proposes strictly justified two statistical criteria for checking the pairwise independence of bit sequences, which can be considered as realization of random variables from a certain set. The corresponding algorithms that implement the independence check according to these criteria are also developed and clearly step-by-step formulated. The resulting tools are very relevant for statistical verification of the cryptographic qualities of various cryptographic primitives, the functioning of which relates to the generation of random/pseudo-random sequences. Such crypto-primitives include not only random/pseudorandom sequence generators, but also stream ciphers, combined encryption algorithms (i.e., block algorithms in stream modes), etc. Using the proposed criteria allows checking the independence of the different parts of output sequences, or the independence of the output sequences from the sequences of internal states and input data. It should be noted that such independence is a necessary condition for the output sequence to be considered unpredictable. The article also considered comparative analysis of two given criteria and results of their applications to sequences of different types.*

Keywords: *statistical criterion, random variables, independence of random variables, random number generation, statistical checking of independence*

1. Introduction

When creating new cryptographic algorithms, in particular pseudorandom/random number generators (RNG/PRNG), or stream ciphers, the following requirements must be met [1, 2]: analytical justification of the algorithm's resistance to main known attacks; statistical investigations of the cryptographic properties of the algorithm's output gamma (often the properties of intermediate gammas of the algorithm are also investigated, such as the sequence of its internal states, or so).

At the same time, we can often see the lack of attention paid to an equally important issue concerning statistical investigations of the mutual independence of parts of the output sequence, as well as the independence of input data, intermediate gammas, sequences of internal states, and output sequences. Indeed, if, for example, the output sequence depends on the input sequence or on the intermediate gamma, the algorithm may be vulnerable to certain types of statistical attacks, or its output sequence may not have the property of unpredictability.

In this paper we propose two new criteria for checking independence of sequences, obtained from RNG/PRNG.

They are based on two different methods, though both of them use calculating the determinant of the correlation matrix. We also give detailed step-by-step description of two corresponding algorithms for independence checking. The first criterion is based on several statements, formulated and proved in this work below. It has such main advantages as significantly lower computational complexity and intuitive clarity of corresponding algorithm. The other advantage is that it needs no tables with probabilistic distributions, such as χ^2 or normal. Its disadvantage is some requirements to the length of the sequences to be analyzed, and the larger the number of sequences we check, the larger their length should be. The second criterion is based on the existing method, explained below. The algorithm, corresponding to the second criterion, is much more complicated and uses tables of χ-square distribution for large number of degrees of freedom. But its undeniable advantage is high accuracy, which testifies to a small value of second type error.

The paper is organized in the follow way. In the Introduction, we formulated the main task of our work and explained its importance for practical applications in cryptology. In the Section 1 we briefly described the state of the art in the considered topic and show the gap in existing results. In the Section 2 we formulate the problem and give main designations, which are necessary for further investigations. In the Section 3 we prove necessary statements to justify the first criterion. Then in the Section 4 we construct the first criterion of sequence independence, and describe corresponding Algorithm 1. The Section 5 gives the results of experiments with Algorithm 1 application. The Section 6 contains theoretical information, necessary for creation of the second criterion. Then in the Section 7 we construct the corresponding Algorithm 2 and describe the results of its applications. In the Section 8 we give conclusion and discussion of the obtained results.

2. Related work

A lot of statistical tools, like commonly used sets of statistical tests, exist for statistical investigations of the cryptographic properties of the output gamma of RNG/PRNG or stream ciphers [1, 2, 3]. But there is currently no standardized or at least generally accepted methodology for checking the statistical independence of sequences, which may be interpreted as independence of random variables with corresponding realizations. The relevance of creating a tool for checking the independence of input-output sequences is substantiated, for example, in [4],

although the number of scientific papers on this topic is much smaller than works on the development of new encryption algorithms or gamma generators [4, 5, 6, 7, 8], or work on estimation of the cryptographic qualities of generators and sequences [9, 10]. Paper [4] is quite close to this work, but its results are only experimental and were obtained for a different problem statement. In this paper, the author performs an experimental study of the independence of the statistical tests themselves rather than the independence of the sequences. Also, a similar problem concerning the dependencies between the bits of input/output sequences in stream ciphers is claimed in [11]. But the purpose of this work is almost the opposite of ours: it investigates how to define whether a stream cipher has sufficient diffusion property, i.e., whether each bit of the output sequence depends on many bits of the input sequence. An appropriate example of investigation of the independence of sequences of internal states for stream cipher was performed, for example, in [12]. For this purpose, the method proposed in [13] was used, which builds corresponding correlation matrix and calculates its determinant. This method is quite workable, but it has certain disadvantages. It is cumbersome and intuitively unclear, requires a large number of calculations, and requires the usage of tables of several distributions, which makes it not very convenient for practice. In addition, when performing statistical research, it is desirable to use several criteria rather than one – for example, the number of separate tests in the NIST STS [3] is 16 (15 in the updated one), and with all "subtests" it is more than 150.

3. Statement of the problem

In this section we introduce the main designations, which will be used below, and then conclude with formal description of the problem.

Let $m, n \in \mathbf{N}$, and

$$X = \left\{ X^{(i)},\ i \in \overline{1,n} \right\} \tag{1}$$

be a set of sequences of independent equally distributed random variables, $X^{(i)} = \left(x_1^{(i)}, ..., x_m^{(i)} \right)$ with $x_k^{(i)} \in \{0, 1\}$, and define $a^{(i)} = Ex_k^{(i)} = P\left(x_k^{(i)} = 1 \right)$,

$$\left(\sigma^{(i)} \right)^2 = Var\left(x_k^{(i)} \right). \tag{2}$$

In what follows, we consider the sequence $\left(x_1^{(i)}, ..., x_m^{(i)} \right)$ as a realization of a random variable $X^{(i)}$. Let's formulate a hypothesis H_0 as "the sequences in set (1) are pairwise independent". Alternative hypothesis H_1 is complex and is formulated as a "hypothesis H_0 is not true."

The purpose of this work is to build and justify a statistical criterion that will recognize the hypothesis H_0 with a given significance level α.

4. Designations and auxiliary statements needed for the first criterion and Algorithm 1

In this section, we will formulate and prove several statements necessary for solving the problem formulated in section 0.

Consider random variables

$$r_k^{(i,j)} = \frac{\left(x_k^{(i)} - a^{(i)}\right)\cdot\left(x_k^{(j)} - a^{(j)}\right)}{\sigma^{(i)}\cdot\sigma^{(j)}}, \quad i, j \in \overline{1,n}, \ k \in \overline{1,m}. \tag{3}$$

Proposition 1

Let the sequences $X^{(i)}$ and $X^{(i)}$ are independent. Then:

$E\left(r_k^{(i,j)}\right) = 0$, $Var\left(r_k^{(i,j)}\right) = 1$, $i, j \in \overline{1,n}$, $k \in \overline{1,m}$

if $a^{(i)} = \frac{1}{2}$, $i \in \overline{1,n}$, then $\left(r_k^{(i,i)}\right)^2 = 1$.

Proof:

1. If the sequences $X^{(i)}$ and $X^{(i)}$ are independent, then the random variables $x_k^{(i)}$ and $x_k^{(j)}$ are also independent, $k \in \overline{1,m}$. Therefore

$$E\left(r_k^{(i,j)}\right) = \frac{E\left(x_k^{(i)} - a^{(i)}\right)\cdot E\left(x_k^{(j)} - a^{(j)}\right)}{\sigma^{(i)}\cdot\sigma^{(j)}} = 0.$$

Next,

$$Var\left(r_k^{(i,j)}\right) = E\left(r_k^{(i,j)} - E\left(r_k^{(i,j)}\right)\right)^2 = E\left(r_k^{(i,j)}\right)^2 =$$

$$= E\frac{\left(x_k^{(i)} - a^{(i)}\right)^2\cdot\left(x_k^{(j)} - a^{(j)}\right)^2}{\left(\sigma^{(i)}\right)^2\cdot\left(\sigma^{(j)}\right)^2} = \frac{1}{\left(\sigma^{(i)}\right)^2\cdot\left(\sigma^{(j)}\right)^2}\cdot E\left(x_k^{(i)} - a^{(i)}\right)^2\cdot E\left(x_k^{(j)} - a^{(j)}\right)^2 =$$

$$= \frac{1}{\left(\sigma^{(i)}\right)^2\cdot\left(\sigma^{(j)}\right)^2}\cdot Var\left(x_k^{(i)}\right)\cdot Var\left(x_k^{(j)}\right) = 1.$$

2. If $a^{(i)} = \frac{1}{2}$, then $\left(r_k^{(i,i)}\right)^2 = \frac{\left(x_k^{(i)} - \frac{1}{2}\right)^2}{\left(\sigma^{(i)}\right)^2}$. Next, since $x_k^{(i)} \in \{0, 1\}$, then $x_k^{(i)} - \frac{1}{2} \in \left\{\pm\frac{1}{2}\right\}$, and $\left(x_k^{(i)} - \frac{1}{2}\right)^2 = \frac{1}{4}$.

In addition, $\left(x_k^{(i)}\right)^2 = x_k^{(i)}$, so

$$Var\left(x_k^{(i)}\right) = E\left(x_k^{(i)}\right)^2 - \left(E\left(x_k^{(i)}\right)\right)^2 = E\left(x_k^{(i)}\right)\left(1 - E\left(x_k^{(i)}\right)\right) = \frac{1}{4},$$

whence $\left(r_k^{(i,i)}\right)^2 = 1$.

The proposition is proved.

For arbitrary $n \in N$, define S_n the group of all permutations of length n, and also define $S_n^{(l)} \subset S_n$ as a subset of permutations that have exactly l fixed points.

Proposition 2

Let $n \in N$ and q be positive integer. Let matrix $A = \left(a_{ij}\right)_{i,j=1}^{n}$ be a square matrix such that:

1. $a_{ii} = 1$, $i \in \{1,...,n\}$;
2. $\forall i, j \in \{1,...,n\}$, $i \neq j$: $a_{ij} = a_{ji}$ and $|a_{ij}| \leq q$, $\forall i, j \in \{1,...,n\}$, $i \neq j$.

Then the following inequality holds:

$$\left|\det A - 1 + \sum_{1 \leq i < j \leq n} a_{ij}^2\right| \leq e^{n^2 q} - 1 - n^2 q - \frac{n^3 (n-1)}{2} \cdot q^2 . \quad (4)$$

Proof:

By the definition of the matrix determinant, and considering the notation introduced at the beginning of the section, we get

$$\det A = \sum_{\sigma \in S_n} \left\{(-1)^{\operatorname{sign}\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)}\right\} = \sum_{l=0}^{n} \sum_{\sigma \in S_n^{(l)}} \left\{(-1)^{\operatorname{sign}\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)}\right\}. \quad (5)$$

Next, we note that

$\left|S_n^{(n)}\right| = 1$, that is, the set $S_n^{(n)}$ consists of a single identical permutation

$\left|S_n^{(n-1)}\right| = 0$ since there are no permutations of the length n with $n-1$ fixed points

$\left|S_n^{(n-2)}\right| = C_n^2$, i.e., the number of elements of this set is equal to the number of ways in which the fixed points can be chosen, and the two remaining points are rearranged

in addition, note that this set coincides with the set of all possible permutations that are inversions of two elements; and $\left|S_n^{(l)}\right| \leq C_n^l \cdot (n-l)!$, in the general case.

To prove the last inequality, it is enough to choose fixed points in C_n^l ways, and all the other points can be rearranged in $(n-l)!$ ways; note that we get inequality

rather than equality, since the permutation of $n-l$ points may give us additional fixed points. That is, the value $\left|S_n^{(l)}\right| \le C_n^l \cdot (n-l)!$ determines the number of such permutations which have *not less than* $n-l$ fixed points.

Now consider separately each term on the right-hand side of (5) for $l \in \{0,...,n\}$:

if $l = n$ the corresponding term will be equal to 1

if $l = n-1$ the corresponding term will be equal to 0

if $l = n-2$ the corresponding term will be equal to $-\sum_{1\le i<j\le n} a_{ij}^{\ 2}$, since the inversion is an odd substitution, and taking into account the condition $a_{ij} = a_{ji}$.

Therefore, equality (5) can be rewritten as

$$\det A - 1 + \sum_{1\le i<j\le n} a_{ij}^{\ 2} = \sum_{l=0}^{n-3} \sum_{\sigma\in S_n^{(l)}} \left\{ (-1)^{\operatorname{sign}\sigma} \prod_{i=1}^{n} a_{i,\sigma(i)} \right\},$$

whence

$$\left| \det A - 1 + \sum_{1\le i<j\le n} a_{ij}^{\ 2} \right| \le \sum_{l=0}^{n-3} \sum_{\sigma\in S_n^{(l)}} \prod_{i=1}^{n} \left| a_{i,\sigma(i)} \right| . \qquad (6)$$

Next, it is easy to see that for $\sigma \in S_n^{(l)}$ there are exactly l units in the product $\prod_{i=1}^{n} \left| a_{i,\sigma(i)} \right|$ (the number of fixed points of the permutation is equal to the number of diagonal matrix elements in the product), then, according to the condition 2) of the Proposition 2, $\prod_{i=1}^{n} \left| a_{i,\sigma(i)} \right| \le q^{n-l}$.

Therefore, inequality (6) can be rewritten as

$$\left| \det A - 1 + \sum_{1\le i<j\le n} a_{ij}^{\ 2} \right| \le \sum_{l=0}^{n-3} C_n^l \cdot (n-l)! q^{n-l} \le$$

$$\le \sum_{l=3}^{n} C_n^l \cdot l! \cdot q^l \le \sum_{l=3}^{n} C_n^l \cdot l! \cdot q^l \le \sum_{l=3}^{n} C_n^l \cdot (n \cdot q)^l =$$

$$= \sum_{l=3}^{n} C_n^l \cdot (n \cdot q)^l = \sum_{l=0}^{n} C_n^l \cdot (n \cdot q)^l - 1 - n^2 \cdot q - \frac{n \cdot (n-1)}{2} \cdot (n \cdot q)^2 =$$

$$= (1 + n \cdot q)^n - 1 - n^2 \cdot q - \frac{n \cdot (n-1)}{2} \cdot (n \cdot q)^2 =$$

$$= \left(1 + \frac{n^2 \cdot q}{n}\right)^n - 1 - n^2 \cdot q - \frac{n \cdot (n-1)}{2} \cdot (n \cdot q)^2 \leq$$

$$\leq e^{n^2 \cdot q} - 1 - n^2 \cdot q - \frac{n^3 \cdot (n-1)}{2} \cdot q^2 ,$$

which proves equality (4).

Corollary 1

If, in the terms of Statement 2, set $q = \frac{\delta}{n^2}$, for some $\delta \in (0, 1)$, then from inequality (4) we obtain the inequality:

$$\left| \det A - 1 + \sum_{1 \leq i < j \leq n} a_{ij}^2 \right| \leq e^{\delta} - 1 - \delta - \frac{\delta^2 (n-1)}{2n} ,$$

that for sufficiently large n can be rewritten as

$$\left| \det A - 1 + \sum_{1 \leq i < j \leq n} a_{ij}^2 \right| \leq e^{\delta} - 1 - \delta - \frac{\delta^2}{2} .$$

In particular, when $\delta = 1$ we have:

$$\left| \det A - 1 + \sum_{1 \leq i < j \leq n} a_{ij}^2 \right| \leq e - 2.5 . \quad (7)$$

Define $R^{(i,j)}$ as

$$R^{(i,j)} = \frac{1}{m} \sum_{k=1}^{m} r_k^{(i,j)} . \quad (8)$$

Because $R^{(i,j)}$ is the average of independent random variables with zero expectation and unit variance, it is easy to see that

$$E\left(R^{(i,j)}\right) = 0 \quad \text{та} \quad Var\left(R^{(i,j)}\right) = \frac{1}{m} . \quad (9)$$

Below we prove two statements that allow us to construct the first statistical criterion. The first statement is based on Chebyshev inequality [14], the second one is based on Chernoff inequality [15]. Either of them can be used to construct the criterion, although, as it will be shown later, the probability estimate in the second one is always more accurate.

Proposition 3

Let $q > 0$ and $\varepsilon \in (0,1)$. Then, if the sequences $X^{(i)}$ and $X^{(j)}$, defined in (1), are independent, and $m > \frac{1}{\varepsilon \cdot q^2}$, then the next inequality holds:

$$P\left(\left|R^{(ij)}\right| > q\right) \le \varepsilon \quad . \tag{10}$$

The proof follows from Chebyshev inequality and the formulas (9).

Proposition 4

Let $q > 0$ and $\varepsilon \in (0,1)$. Then, if the sequences $X^{(i)}$ and $X^{(j)}$, defined in (1), are independent, and $m > \frac{\ln \frac{2}{\varepsilon}}{q^2}$, then the next inequality holds:

$$P\left(\left|R^{(ij)}\right| > q\right) \le \varepsilon \quad . \tag{11}$$

Proof:

According to Chernoff's inequality, $\forall a > 0$ the next inequality is true:

$$P\left(\left|\sum_{k=1}^{m} r_k^{(i,j)}\right| > a\right) \le 2e^{-\frac{a^2}{2 \cdot m}} \quad ,$$

whence

$$P\left(\frac{\left|\sum_{k=1}^{m} r_k^{(i,j)}\right|}{m} > \frac{a}{m}\right) = P\left(R^{(ij)} > \frac{a}{m}\right) \le 2e^{-\frac{a^2}{2 \cdot m}} \quad .$$

Set $q = \frac{a}{m}$; then the last inequality can be rewritten as

$$P\left(R^{(ij)} > q\right) \le 2e^{-q^2 \cdot m} \quad , \tag{12}$$

and for $m > \frac{\ln \frac{2}{\varepsilon}}{q^2}$ we get $2e^{-q^2 \cdot m} < \varepsilon$.

The Proposition is proved.

Note that from inequality $m > \frac{1}{\varepsilon \cdot q^2}$ we have $m > \frac{\ln \frac{2}{\varepsilon}}{q^2}$, so Proposition 4 gives a more accurate lower bound for the value of m.

5. Construction of a statistical criterion and algorithm for testing the hypothesis about sequence independence.

In this section we construct a new statistical criterion for checking the independence of sequences and develop the corresponding algorithm.

The basic idea behind the criterion can be described informally as follows. Let's set a certain level of significance α (probability of the 1st type error). If the sequences (1) are pairwise independent, then for sufficiently large m (chosen according to Propositions 3 or 4), we can construct correlation matrix $A = \left(R^{(ij)} \right)_{i,j=1}^{n}$, which elements (except diagonal elements) will be less than a certain value q with a probability not less than $1-\alpha$. The value q should be chosen in a such way that the right-hand side of (4) is sufficiently small – for example, $q = \frac{1}{n^2}$. Then calculate the elements of a matrix A and its determinant and check the inequality (4). Then, according to the choice of the sequence length, for pairwise independent sequences, the probability that inequality (4) does not hold does not exceed α.

Let us have access to sources of sequence generation (1) and can obtain sequences of sufficient length from them (the required length depends on the number of n sequences and the desired level of significance).

It should be noted that the algorithm below is focused on checking the independence of the sources that generate the corresponding sequences. This is the task that is most relevant for practical applications. But, with certain modifications, it can also be used to check the independence of given sequences.

Algorithm 1:

Input: level of significance α; number n of sequences to be tested.

1. Compute

$$q = \frac{1}{n^2} \quad \text{and} \quad m = \left\lceil \frac{\ln \frac{2}{\alpha}}{q^2} \right\rceil. \qquad (13)$$

2. Generate (from corresponding sources or one source, depending on the set task) n sequences $X^{(i)}$, $i \in \overline{1,n}$, of the lengths at least t. Note that there is no need to generate sequences if the task is to check the independence of sequences from a given set, but we need to be sure that their lengths are at least m.

3. Calculate the elements of a matrix $A = \left(R^{(ij)} \right)_{i,j=1}^{n}$ using formulas (3) and (8).

4. Compute $d = \det A$.

5. Compute

$$t = \left| \det A - 1 + \sum_{1 \le i < j \le n} a_{ij}^{\,2} \right|.$$

6. If $t \le e - 2.5 \approx 0.218$, then hypothesis H_0 accepted, otherwise rejected.

As we can see from the formula (13), the lower bound of the value m increases when the value α decreases, and increases when the value n increases. Note that if the level of significance α is not less than 0.001, and the number of sequences is not larger than 100, the lengths of sequences $m = 10^8$ is sufficient for testing.

To simplify the application of this algorithm, we propose the Table 1 with reference values of the lower bound of the sequences length m for different frequently used values of two other parameters, the level of significance α and the number of sequences n.

Table 1

The lower bound of parameter m

n / α	10	20	30	40	50
0.05	36 889	590 221	2 987 993	9 443 532	23 055 497
0.01	52 984	847 731	4 291 638	13 563 693	33 114 484
0.005	59 915	958 635	4 853 087	15 338 150	37 446 654
0.001	76 010	1 216 145	6 156 731	19 458 311	47 505 641

6. Experimental results of the Algorithm 1 application

Practical applications of the proposed criteria give the next results:

1) for sequences obtained from the certified generator (described in Appendix A in [16]), in 10 experimental sets of 100 sequences, hypothesis H_0 was accepted;

2) for sequences obtained using counter, in 10 experimental sets of 100 sequences, hypothesis H_0 was rejected 9 times and accepted 1 time;

3) for sequences obtained from file with extension avi, in 10 experimental sets of 100 sequences, hypothesis H_0 was rejected 10 times;

4) for set of 100 sequences, where 50 of them were generated from the certified generator, and other 50 were obtained from them, using some transformations, hypothesis H_0 was rejected 10 times out of 10.

We also applied this test to sets of sequences, which were created to be dependent. To describe these results, we introduce the term “experiment”. Under one experiment we will understand the following procedure:

- to generate definite number of sequences (for example, 20 sequences) of sufficient lengths (for example, 10^6 or larger), using some certified generator;
- to introduce dependences in chosen sequences (if necessary);
- to create correlation matrix and to evaluate the value t.

Below we describe the results obtained in such experiments.

1) **50% of sequences are dependent in half of the characters.** For 100 experiments, each with 20 sequences, we firstly got 20 sequences from certified PRNG, then chose 10 of them, and then each second element of these chosen sequences was changed in such a way, that introduce dependency on correspondent

elements of the first 10 sequences. For these 100 experiments, we got the next range of the value t: $t \in$ [1.55233507592079; 1.56300829290009]. It means that in all experiments the value lays deeply in critical region, so the criterion strongly recognized dependent sequences.

2) **50% of sequences are dependent in 33% of the characters.** For 100 experiments, each with 20 sequences, we firstly got 20 sequences from certified PRNG, then chose 10 of them, and then each third element of these chosen sequences was changed in such a way, that introduces dependency on correspondent elements of the first 10 sequences. For these 100 experiments, we got the next range of the value t: $t \in$ [0.416068024780703; 0.422475116952477]. It is still essentially larger than the edge of critical region (which is 0.218), but much closer to this edge than in previous experiments. These results show that the larger dependency between sequences, the deeply in critical region the statistics t takes values.

3) **25% of sequences are dependent in half of the characters.** For 100 experiments, each with 20 sequences, we firstly got 20 sequences from certified PRNG, then chose 5 of them, and then each second element of these chosen sequences was changed in such a way, that introduce dependency on correspondent elements of the first 5 sequences. For these 100 experiments, we got the next range of the value t: $t \in$ [0.484858277736367;- 0.489774359187274]. This range is closer to the edge of critical region, than in the first series of experiments, but further than in the second series of experiments. So the criterion can recognized dependency even in case when 75% of sequences are independent.

4) **Archive files.** For all 1000 experiments, $t \in$ [0.0000134436852349293; 124.401400670346], and in less than 1% cases the value of t is less than 0.218, or outside critical region.

7. Designations and auxiliary statements needed for the second criterion and Algorithm 2

Let us have a set ті of m bit sequences of the length l:

$$X^{(i)} = \left(x_1^{(i)}, ..., x_l^{(i)}\right), \quad i \in \overline{1,m}, \quad x_k^{(i)} \in \{0,1\}. \tag{14}$$

Define $a^{(i)} = \frac{1}{l}\sum_{k=1}^{l} x_k^{(i)}$ the sample average of these sequences. Usually, the sequences under investigations are obtained from RNG/PRNG; in such cases we may assume that $a^{(i)} = 0.5$. But the proposed criterion is still true for more general cases.

The criterion proposed below is based on the results, described in [13], and uses sample correlation matrix.

For sequences (14), define values

$$D_{ij} = \frac{1}{l-1}\sum_{k=1}^{l}\left(x_k{}^{(i)} - a^{(i)}\right)\left(x_k{}^{(j)} - a^{(j)}\right),$$

which are sample pairwise correlations, and build sample correlation matrix

$$\mathbf{R} = \left(R_{ij}\right)_{i,j=1}^{m}, \text{ де } R_{ij} = \frac{D_{ij}}{\sqrt{D_{ii}}\sqrt{D_{jj}}}, \; i, j \in \overline{1,m}.$$

Note that for pairwise independent random variables $X^{(i)}$, $i \in \overline{1,m}$, the matrix $\mathbf{R}$ is identity matrix I, so its determinant is equal to 1. But in our case this matrix was built as empirical, according to the results of statistical experiments. Then for creating the criterion for checking hypothesis H_0, that «sequences are pairwise independent» with given significance level α (for example, $\alpha = 10^{-3}$) we use the determinant $d = \det \mathbf{R}$ of the matrix $\mathbf{R}$.

The distribution low of the determinant d is rather complicated, but for large enough values of l we can use its asymptotical form: if sequences are pairwise independent, then

$$\mathbf{P}\left\{-n \cdot \ln d \le v\right\} = \mathbf{P}\left\{\chi_f{}^2 \le v\right\} + \frac{\gamma}{n^2}\left(\mathbf{P}\left\{\chi_{f+4}{}^2 \le v\right\} - \mathbf{P}\left\{\chi_f{}^2 \le v\right\}\right) + \mathrm{O}\left(n^{-3}\right), \quad (15)$$

where

$$f = \frac{m(m-1)}{2}, \quad n = l - \frac{2m+11}{6}, \quad \gamma = \frac{m(m-1)\left(2m^2 - 2m - 13\right)}{288},$$

and $\chi_f{}^2$ is a random variable with χ-square distribution with f degrees of freedom.

Note that for $l \ge 10^5$, $m \ge 100$ one can use approximation

$$\mathbf{P}\left\{-n \cdot \ln d \le v\right\} \approx \mathbf{P}\left\{\chi_f{}^2 \le v\right\}.$$

As we can see, in case of approximation of (15) we delete the positive values in its right part, so the obtained critical region corresponds to the large significance level then was set.

For hypothesis H_0, under given significance level $\alpha = 1 - P\left(-n \cdot \ln d \le v\right)$, the critical region is $-n \cdot \ln d > v$, or $d \in \left(0;\, e^{-\frac{v}{n}}\right)$.

After performing calculations according (15) or its approximation, we get critical region for parameter $d = \det \mathbf{R}$, which corresponds to given significance level α (for example, if $l = 10^6$, $m = 100$, then critical region for parameter d under given $\alpha = 10^{-3}$ is the region $d \in (0;\ 0.994754314)$; under given $\alpha = 10^{-2}$ the critical region isь $d \in (0;\ 0.994830549)$). The critical regions for these or other levels may be found using tables for χ-square distribution with f degree of freedoms (if $f \le 30$) or using normal approximation (if $f \ge 30$). Thus, for $f \to \infty$, χ-square distribution tends to normal distribution with parameters $N\left(f, \sqrt{2f}\right)$. It means that if $f \ge 30$ then one may use approximation

$$\mathbf{P}\left\{{\chi_f}^2 \le v\right\} = \mathbf{P}\left\{\frac{{\chi_f}^2 - f}{\sqrt{2f}} \le \frac{v - f}{\sqrt{2f}}\right\} \approx F\left(\frac{v - f}{\sqrt{2f}}\right),$$

where $F(x)$ is cumulative distribution function of normal distribution.

Table 2
Values of determinant of correlation matrix for 20 series of 100 sequences, each of the length of 10^6 bits, generated using certified PRNG

Series numbers	Values of determinant	Series numbers	Values of determinant
1	0.995169	11	0.995223
2	0.995150	12	0.995219
3	0.995164	13	0.995193
4	0.995178	14	0.995174
5	0.995172	15	0.995205
6	0.995182	16	0.995188
7	0.995163	17	0.995219
8	0.995168	18	0.995247
9	0.995166	19	0.995227
10	0.995201	20	0.995205
Sample average			0.99519065

In other words, for given significance level α, to calculate limit statistics v_α one needs, at first, to find corresponding limit statistics z_α, using tables of normal distribution, and after this to calculate limit statistics for ${\chi_f}^2$ as $v_\alpha = z_\alpha \cdot \sqrt{2f} + f$.

As example, in Table 2 the values of d are given, calculated for 20 series of 100 sequences, each of the length of 10^6 bits.

8. Description of Algorithm 2 and results of its application

Based on results described in Section 8, we can derive the next Algorithm 2.

Algorithm 2:

Input: level of significance α;
set of sequences (14).

1. For given sequences, calculate the values $a^{(i)} = \frac{1}{l}\sum_{k=1}^{l} x_k{}^{(i)}$.
2. For given sequences, calculate the values

$$D_{ij} = \frac{1}{l-1}\sum_{k=1}^{l}\left(x_k{}^{(i)} - a^{(i)}\right)\left(x_k{}^{(j)} - a^{(j)}\right).$$

3. Calculate the values $R_{ij} = \frac{D_{ij}}{\sqrt{D_{ii}}\sqrt{D_{jj}}}$, $i, j \in \overline{1,m}$.
4. Build correlation matrix $\mathbf{R} = \left(R_{ij}\right)_{i,j=1}^{m}$.
5. Calculate $d = \det \mathbf{R}$.
6. Calculate the next values:

$$f = \frac{m(m-1)}{2}, \quad n = l - \frac{2m+11}{6}, \quad \gamma = \frac{m(m-1)\left(2m^2-2m-13\right)}{288}.$$

7. If $f < 30$, then for given significance level α, using tables for $\chi_f{}^2$-distribution, find such value v_α, that $\mathbf{P}\left\{\chi_f{}^2 \geq v_\alpha\right\} = \alpha$.

Else, if $f \geq 30$, then do the next steps.

7.1. For given significance level α, using tables for normal distribution, find such value z_α, that $\Phi(z_\alpha) = 1-\alpha$ (if tables consists of values of cumulative distribution function for normal distribution $\Phi(x)$), or such z_α, that $\Phi_0(z_\alpha) = \frac{1}{2} - \alpha$ (if tables consists of values of function $\Phi_0(x) = \Phi(x) - \frac{1}{2}$).

Example: using tables of values of function $\Phi_0(x)$ in reference book of Bernshtain and Semendyaev, for $\alpha = 0.01$ we get $z_\alpha = 2.33$; for $\alpha = 0.001$ we get $z_\alpha = 3.1$.

7.2. Calculate the value ν_α as
$\nu_\alpha = z_\alpha \cdot \sqrt{2f} + f$.

8. If inequalityь $d \geq e^{-\frac{\nu_\alpha}{n}}$ holds, then hypothesis H_0 about pairwise independence is accepted, in the opposite case H_0 is rejected.

On the Figure 1 below we give the comparative results of application of Algorithm 2 to the sequences, obtained from the PRNG (black lines) and limit values of the value d, calculated according Algorithm 2 (red dotted lines), for different values of the sequences length l (from 200000 to 1000000 bits) and different numbers of sequences m (from 10 to 100). As we can see, the values of parameter d, calculated for given sequences, are very close to the edge of critical region. It means that this criterion is very accurate and has very small error of the second type (probability to accept H_0 in case when it is wrong).

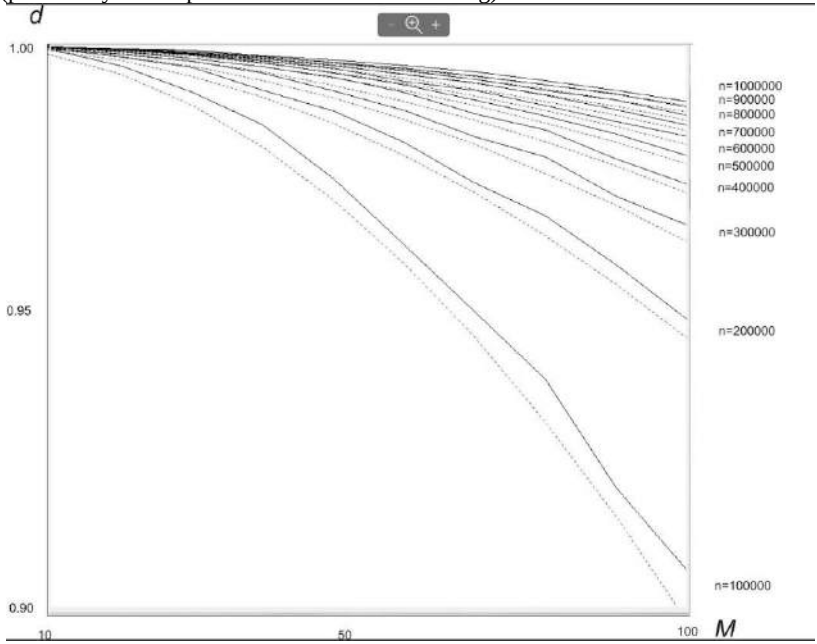


Figure 1. *Comparison of results of PRNG testing according Algorithm 2 and limit statistics for determinant of correlation matrix*

For these experiments, the stream cipher Krip [12] was used as PRNG. Note that this cipher was investigated using a lot of statistical tests, in particular NIST test suite [3], hence it may be considered as PRNG of high quality.

During investigations of statistical properties of RNG/PRNG, or stream ciphers, one may need to check the independence of different sequences, obtained on different steps of working of this generator. For example, it is necessary to check the independence of output gamma, internal state sequences, and input data, etc. And we need to be sure that all the parts of these sequences are pairwise independent.

In this case one may use Algorithms 1 or 2 for total set of the sequences, which consists of all sequences, obtained on different steps of generator's work. For example, we may take half of the sequences from inner states and half from output gamma.

If we cryptographic quality of stream cipher, then the set of sequences for checking should contain sequences created using different (randomly chosen) keys and initialization vectors. The number of sequences should be not less than 100, and their lengths should be not less than 10^6.

9. Conclusion and discussion

The proposed criteria for checking pairwise independence of random values/sequences are rather simple and convenient in practice. Their practical applications show that they work "in both directions": accept hypothesis H_0 for sets of independent sequences and reject H_0 in opposite case.

The experimental results show that the both proposed criteria reject hypothesis about independence in case when the number of sequences, which may be considered as independent, is up to 75%.

They may be used, along with other statistical tests, for checking statistical qualities of stream or combined ciphers and for random/pseudorandom number generators before adoption.

One of the tasks for the further investigations is more detailed comparative analysis of behavior of these two criteria for different types of sequences, obtained from sources with different nature, including stream cipher Krip [12]. It should be noted, that the first proposed criterion gives more rough estimation of the edges of critical region, if compare with the second one.

This leads to situation when the sequences with a slight dependency may be defined as independent. But in case when dependency is strong enough, it works well, as was shown in our experiments in Section 7. At the same time, the first criterion is much more easily to apply, it needs a small amount of calculations, and needs no tables of distribution functions for widely used distributions, like normal or hi-square distribution. From this point of view, we suggest to use the first criterion for primary analysis of RNG, or PRNG, or stream cipher.

If this criterion decides that sequences obtained from the source are independent, then we may continue to analyze the source more deeply, applying the second criterion and other tests. But in opposite case, when the source fails the first criterion, it means that this source can't be used in cryptographic applications and needs at least significant improvements.

The other interesting question is to find such modification of these criteria, which may be used not only for checking independence of sequences or corresponding random variables, but also for checking independence of statistical tests from the different test suits, like NIST or Diehard.

This problem is also of great importance, because excluding "redundant" tests from the suit allows to reduce significantly the time of application of the test suite. In the current form, the proposed criteria can't be used for this purpose, and it's a great challenge to invent suitable modification and to prove its relevance.

10. Funding and support

The research was conducted within the framework of project 2023.04/0020, "Development of Methods and Layout of the "DEMETRA" ARM for Regular and Periodic Control of the Functioning of Cryptographic Applications Using Statistical Methods," funded by the National Research Foundation of Ukraine (NRFU).

11. References

[1] Christof Paar, Jan Pelzl, "Stream Ciphers", Chapter 2 of "Understanding Cryptography, A Textbook for Students and Practitioners". Springer, 2009. URL: https://dosen.itats.ac.id/sitiagustini/wp-content/uploads/sites/78/2017/05/Understanding_Cryptography_Chptr_2-Stream_Ciphers.pdf.

[2] Matt J. B. Robshaw, Stream Ciphers Technical Report TR-701, version 2.0, RSA Laboratories, 1995 https://www.networkdls.com/Articles/tr-701.pdf.

[3] A Statistical Test Suite for Random and Pseudorandom Number Generators for Cryptographic Applications. NIST Special Publication 800-22, 1999. Rev. 1. – 131 p.

[4] Almaraz Luengo, E.; Román Villaizán, J. Cryptographically Secured Pseudo-Random Number Generators: Analysis and Testing with NIST Statistical Test Suite. Mathematics 2023, 11, 4812. URL: https://doi.org/10.3390/math11234812 (1 in the list).

[5] Suwais, K.; Almanasra, S. Strike: Stream Cipher Based on Stochastic Lightning Strike Behaviour. Appl. Sci. 2023, 13, 4669. URL: https://doi.org/10.3390/app13084669, https://www.mdpi.com/2076-3417/13/8/4669.

[6] Wu, S.-T. A Key-Based Multi-Mode Clock-Controlled Stream Cipher for Real-Time Secure Communications of IoT. Electronics 2023, 12, 1076. URL: https://doi.org/10.3390/electronics12051076.

[7] Melosik, M.; Galan, M.; Naumowicz, M.; Tylczyński, P.; Koziol, S. Cryptographically Secure PseudoRandom Bit Generator for Wearable Technology. Entropy 2023, 25, 976. URL: https://doi.org/10.3390/e25070976.
[8] Bikos, A.; Nastou, P.E.; Petroudis, G.; Stamatiou, Y.C. Random Number Generators: Principles and Applications. Cryptography 2023, 7, 54. URL: https://doi.org/10.3390/cryptography7040054.
[9] Piątkowski, J.; Szymoniak, S. Methodology of Testing the Security of Cryptographic Protocols Using the CMMTree Framework. Appl. Sci. 2023, 13, 12668. URL: https://doi.org/10.3390/app132312668.
[10] Crocetti, L.; Nannipieri, P.; Di Matteo, S.; Fanucci, L.; Saponara, S. Review of Methodologies and Metrics for Assessing the Quality of Random Number Generators. Electronics 2023, 12, 723. URL: https://doi.org/10.3390/electronics12030723.
[11] Madarro-Capó, E.J.; Ramos Piñón, E.C.; Sosa-Gómez, G.; Rojas, O. Practical Improvement in the Implementation of Two Avalanche Tests to Measure Statistical Independence in Stream Ciphers. Computation 2024, 12, 60. URL: https://doi.org/10.3390/computation12030060, https://www.mdpi.com/2079-3197/12/3/60.
[12] Kovalchuk L.V., Koriakov I.V., Alekseychuk A.N. Krip: high-speed hardware-oriented stream cipher based on a non-autonomous nonlinear shift register. 2023. Cybernetics and System Analysis Vol.59, P. 16–26. URL: https://doi.org/10.1007/s10559-023-00538-6 (date of access: 27.04.2024).
[13] Anderson, T.W. An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, New York. 1958. 500 p.
[14] Feller, W. (1968) An Introduction to Probability Theory and Its Applications, Vol. 1. 3rd Edition, John Wiley & Sons, New York.
[15] Michel Goemans. Chernoff bounds, and some applications. Lecture notes. 2015. URL: https://math.mit.edu/~goemans/18310S15/chernoff-notes.pdf.
[16] DSTU 9041:2020 Information technologies. Cryptographic protection of information. Short message encryption algorithm based on twisted elliptic Edwards curves.

НОВІ СТАТИСТИЧНІ КРИТЕРІЇ ДЛЯ ПЕРЕВІРКИ НЕЗАЛЕЖНОСТІ БІТНИХ ВИПАДКОВИХ ЗМІННИХ ТА ПОСЛІДОВНОСТЕЙ

Dr.Sci. Л. Ковальчук ORCID: 0000-0003-2874-7950
Г.Є. Пуховський інститут моделювання в енергетиці, Україна
E-mail: lusi.kovalchuk@gmail.com
Dr.Sci. А. Давиденко ORCID: 0000-0001-6466-1690
Г.Є. Пуховський інститут моделювання в енергетиці, Україна
E-mail: davidenkoan@gmail.com
О. Беспалов 0000-0001-7126-6752
Г.Є. Пуховський інститут моделювання в енергетиці, Україна
E-mail: alexb5dh@gmail.com

Анотація. *У статті запропоновано строго обґрунтовані два статистичні критерії для перевірки попарної незалежності бітових послідовностей, які можна розглядати як реалізацію випадкових величин із певної множини. Також розроблено та чітко покроково сформульовано відповідні алгоритми, які реалізують перевірку незалежності за цими критеріями. Отримані інструменти дуже актуальні для статистичної перевірки криптографічних якостей різних криптографічних примітивів, функціонування яких пов'язане з генерацією випадкових/псевдовипадкових послідовностей. Такі крипто-примітиви включають не лише генератори випадкових/псевдовипадкових послідовностей, але й потокові шифри, комбіновані алгоритми шифрування (тобто блокові алгоритми в потокових режимах) тощо. Використання запропонованих критеріїв дозволяє перевіряти незалежність різних частин вихідних послідовностей, або незалежність вихідних послідовностей від послідовностей внутрішніх станів і вхідних даних. Слід зазначити, що така незалежність є необхідною умовою для того, щоб вихідна послідовність вважалася непередбачуваною. У статті також розглянуто порівняльний аналіз двох наведених критеріїв та результати їх застосування до послідовностей різних типів.*

Ключові слова: *статистичний критерій, випадкові величини, незалежність випадкових величин, генерація випадкових чисел, статистична перевірка незалежності*