Methods of Profiling the Behavior of Dynamic Objects of a Critically Important Information Infrastructure*

Sergei A. Petrenko[1][0000-0003-0644-1731], Alexander V. Olifirov[2][0000-0002-5288-2725], Krystina A. Makoveichuk[2][0000-0003-1258-0463], Nikolay N. Oleinikov[2][0000-0002-9348-9153]

[1]Innopolis University, Kazan, Russia
s.petrenko@rambler.ru
[2]V.I. Vernadsky Crimean Federal University, Yalta, Russia
alex.olifirov@gmail.com
christin2003@yandex.ru
oleinikov1@mail.ru

Abstract. According to ISO/IEC TR 18044: 2004, an incident means an undesirable or unexpected event (or a combination of such events) that could compromise the information interaction processes in a critically important infrastructure or threaten its information security and/or cyber resilience. Accordingly, the incident prediction means the identification process of vulnerable object interaction state of the critically important information infrastructure under the disturbances. According to the incident prediction results, it becomes possible to develop a profile of the profile of an observed object, containing information about the exploited vulnerability, the actions of the intruder and possible scenarios of a proactive counteraction against these attacking influences.

Keywords: inverse similarity theorem, dynamic control of correctness of calculation programs, correctness of computing processes.

1 Introduction

We propose a possible way of profiling the behavior of the key IT services and IT systems of a critically important information infrastructure under perturbation conditions. Here the dynamic profiles allow identifying the classes of the vulnerable states of the mentioned infrastructure. In this case, the recognition of the informative signs of the possible vulnerabilities is carried out in conditions of extremely large amounts of data monitoring. When selecting information, the dynamic weights of the recognition signs and the corresponding values of the profiling of the observed objects are determined; this can significantly reduce the response time to potential incidents and purposefully select the adequate measures to ensure the required cyber resilience [1, 4].

Thus, a new method is proposed for profiling the complex dynamic subsystems of critically important infrastructure under the incompleteness and competing information

on the state of the observed objects. This profiling method is based on the mathematical apparatus for iteratively diagnosing the potentially dangerous states of the complex dynamic systems using communication (Pr_1), behavioral (Pr_2) profiles, as well as profiles, providing the required cyber resilience (Pr_3) of observed objects [2, 5]. It is significant that the profiling method, mentioned above, makes it possible to model the potential behavior of an intruder, during the implementation of threats to resilience (security) and make decisions about the organization of the special scenarios to ensure the required cyber resilience and prevent serious incidents with the transfer of the critical information infrastructure to an irreversible catastrophic state.

2 The problem of profiling the objects' behavior of critical information infrastructure

Unlike the well-known cyber resilience approaches, the proposed profiling method is implemented both at the stages of the primary processing of the monitoring results of critical information infrastructure objects and at the stages of the analyzing and summarizing a heterogeneous information concerning the functioning processes of the observed infrastructure and its individual elements (devices and resources). At the first stage (analytical description of processes $Pr_1,\ldots, Pr_n$ of interaction of objects of critical information infrastructure G1) (Figure 1**Ошибка! Источник ссылки не найден.**) it is necessary to take into account the structural and functional characteristics of the observation objects, the composition and specificity of the system and application software, the characteristics of the operating system [3, 8]. This is necessary to form the sets of quantitative (B1) and qualitative (B2) signs, reflecting the options for the development of information technology impact situations on the objects of the critically important information infrastructure being protected.

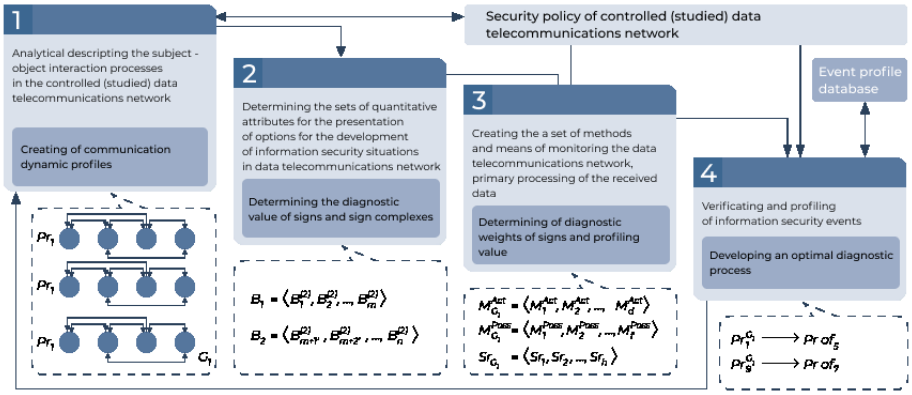


Fig. 1. Protected infrastructure profiling scheme

Based on the specifics and characteristics of disturbances in the functioning and composition of the feature set, at the third stage, a set of methods (active ($M_{G_1}^{Act}$) and / or

passive ($M_{G_1}^{Pass}$) and means (Sr_{G_1}) of monitoring the protected infrastructure G1 are formed. These methods and means should take into account the intruder impact type and their interconnection with a threat model of the protected infrastructure. At this stage, the degree of the interconnection between alternative groups of the negative sign impacts, the consequences (damage) of their manifestation are also determined, and a list of possible measures to ensure the required cyber resilience is developed. After the corresponding procedures of iterative diagnostics and primary processing of the obtained data are carried out, the intruder actions and the corresponding cyber resilience violation events are verified, and the profiles of the corresponding objects of the protected infrastructure are developed.

Thus, the effectiveness of ensuring the required cyber resilience of the protected infrastructure is ensured by diagnosing the potentially vulnerable states of the observed infrastructure, determining the type and criticality of vulnerability, and developing the plan of possible measures to ensure the required cyber resilience. The proposed approach of profiling the behavior of dynamic objects of the protected infrastructure required solving the problem of diagnosing complex dynamic cyber systems under the temporary observability absence of the corresponding interaction processes [6, 7].

Usually, a typical object of the protected infrastructure is a complex dynamic cyber system (both in structure and behavior), operating in the absence of temporal or partial observability of interaction with other infrastructure objects.

Here, the diagnosis task of the mentioned cyber systems is to determine the state of the object and the aggregate of monitored parameters, which can be used to judge the functional cyber resilience of the infrastructure object, i.e. to determine whether its current system configuration and application software is currently vulnerable, or whether the object has no distinguishable vulnerabilities. The desired solution involves the development of such diagnosis procedures, the content of which depends on the properties of the protected infrastructure, the priorities and diagnosis direction, as well as the conditions for its implementation.

Let some protected critically important information infrastructure S=P<B, L> (Figure 1 and Figure 2) be consisted of a set of objects B=<B1, B2, B3>, where $\mathbf{B_1} = \langle \mathbf{B_1^{(1)}}, \mathbf{B_2^{(1)}}, \dots, \mathbf{B_m^{(1)}} \rangle$ are many devices (routers) and web resources (servers), $\mathbf{B_2} = \langle \mathbf{B_{m+1}^{(2)}}, \mathbf{B_{m+2}^{(2)}}, \dots, \mathbf{B_n^{(2)}} \rangle$ - set of users (data sources) of the mentioned infrastructure, $\mathbf{B_3} = \langle \mathbf{B_1^{(3)}}, \mathbf{B_2^{(3)}}, \dots, \mathbf{B_h^{(3)}} \rangle$ - a set of an information, gathering and processing the means (nodal and network sensors of the cyber-attack detection system) associated with each other communication channels [12], represented by a connection matrix in the given units of measurement between points B1 and Bj(I, j =1,…n);

$\mathbf{L} = \left\| \begin{matrix} \mathbf{l_{11}}, \mathbf{l_{12}}, \dots, \mathbf{l_{1n}} \\ \mathbf{l_{21}}, \mathbf{l_{22}}, \dots, \mathbf{l_{2n}} \\ \mathbf{l_{n1}}, \mathbf{l_{n2}}, \dots, \mathbf{l_{nn}} \end{matrix} \right\|$ – the connection matrix between objects ($l_{ij} \geq 0$, with $i \neq j$, $l_{ii}=0$, j=1,…n).

Let the values of the monitoring data collection time (T_{col}), the recording time (T_0) (the action $\mathbf{d_1^{c6}}$) and the processing of the monitoring data be known, with the $\mathbf{d_1^{c6}} \in \mathbf{D_1^{c6}}$. The cyber attack detection systems allow receiving as a source of multiple

packet streams of the i-th node of the protected infrastructure ($b_1 \in N$) with intensities $\boldsymbol{\lambda} = \{\boldsymbol{\lambda}_{i_1}, \boldsymbol{\lambda}_{i_2}, \dots, \boldsymbol{\lambda}_{i_b}\}$ and generate a set of packets i infrastructure node ($b_1 \in N$) with $\boldsymbol{\nu}_i = \{\boldsymbol{\lambda}_{i_1}, \boldsymbol{\lambda}_{i_2}, \dots, \boldsymbol{\lambda}_{i_g}\}$.

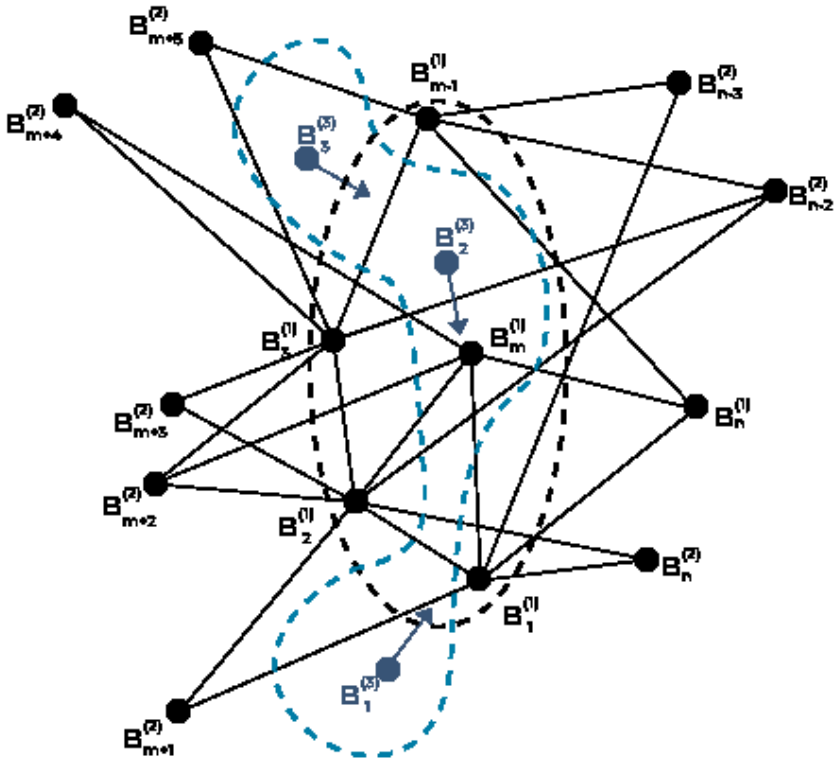


Fig. 2. Graphological representation of the protected infrastructure

In modern monitoring systems, subsystems of the active (based on scanning a network object according to the "request-response" principle with subsequent response processing) and passive (based on the analysis of network traffic parameters in the listening mode of the selected interface) data collection are implemented.

In general, the data processing system using active monitoring methods (i = 1, ..., m) can be represented by the seven arrays

$$\mathbf{B}_i^{(c)} = \left\{\mathbf{T}_i^{(1)}, \mathbf{Prs}_i^{(c)}, \mathbf{Prd}_i^{(c)}, \mathbf{V}_i^{(c)}, \mathbf{F}^{(c)}, \boldsymbol{\Phi}^{(c)}, \mathbf{T}_i^{(c)}\right\}, \quad (1)$$

where $\mathbf{T}_i^{(c)} = \left\{\mathbf{t}_{i_1}^{(c)}, \mathbf{t}_{i_2}^{(c)}, \dots, \mathbf{t}_{ir_1}^{(c)}\right\}$ – the set of t time values of the protected infrastructure object observation;

$\mathbf{Prs}_i^{(c)} = \left\{\mathbf{prs}_{i_1}^{(c)}, \mathbf{prs}_{i_2}^{(c)}, \dots, \mathbf{prs}_{iq_1}^{(c)}\right\}$, $\mathbf{q_1} \in \mathbf{N_q}$ is the set of parameter values (input signals) of scan sessions, conducted as regards the infrastructure object;

$\mathbf{Prd_i^{(c)}} = \left\{\mathbf{prd_{i_1}^{(c)}},\ \mathbf{prd_{i_2}^{(c)}}, \dots, \mathbf{prd_{ig_1}^{(c)}}\right\},\ \mathbf{g} \in \mathbf{N_a}$ is the set of values of passive traffic scanning (output signals) identifying the state of some infrastructure object;

$\mathbf{V_i} = \left\{\mathbf{V_{i_1}^{(c)}(t)},\ \mathbf{V_{i_2}^{(c)}(t)}, \dots, \mathbf{V_{id_1}^{(c)}(t)}\right\},\ \mathbf{d} \in \mathbf{N_a}$ is the statespace of the protected infrastructure object during monitoring;

$F^{(c)}$ - transition operator, reflecting the mechanism of changing the object state of the protected infrastructure under the action of internal and external cyber-attacks;

$\Phi^{(c)}$ is the output operator, describing the mechanism for generating the output signal as a response of the protected infrastructure object to internal and external disturbances;

$\mathbf{T_i^{(c)}} = \left\{\mathbf{T_{i_1}^{(c)}},\ \mathbf{T_{i_2}^{(c)}}, \dots, \mathbf{T_{ip_1}^{(c)}}\right\},\ \mathbf{p_1} \in \mathbf{N_a}$ is a set of the values, formed by the results of monitoring and establishing the truth values of passive scanning of the object of the protected infrastructure.

The structure of the process characterizing the dynamics of changes in the properties of devices and users of the protected infrastructure, when conducting the passive monitoring sessions $t \in [t_i, t_i + \triangle_i)$, $i=\overline{1, m}$), we will present in the form of a chain of mappings

$$R\langle\chi_{B^{(1)},B^{(2)}}(t), \chi_{B^{(3)}}(t)\rangle \to R\langle B_t^{(1)}, B_t^{(2)}, B_t^{(3)}\rangle,\ \ R\langle B_t^{(1)}, B_t^{(2)}, B_t^{(3)}\rangle \to$$
$$B_t^{(3)},\ \ R\langle B_t^{(1)}, B_t^{(2)}, B_t^{(3)}\rangle B_t^{(1)}, B_t^{(2)},$$
$$B_t^{(1)}, B_t^{(2)}\ ,\ R\langle\chi_{B^{(1)},B^{(2)}}(t),\ \chi_{B^{(3)}}(t)\rangle \to \chi_{B^{(1)},B^{(2)}}(t), B_t^{(3)} \to \chi_{B^{(3)}}(t),$$

where x(.)(t) - states of devices, users and controlled detection systems КА;

R<x(.), x(.)>$R\langle x_{\langle.\rangle}, x_{\langle.\rangle}\rangle$ – connections between states;

$R\langle B_t^{(1)}, B_t^{(2)}, B_t^{(3)}\rangle$ - connections between devices, users and sensors of the cyber-attack detection system, which change over time and characterize the above-mentioned process of monitoring the objects of the protected infrastructure.

Operators implement mappings:

$$F^{(c)}:\ T_i^{(c)} \times Prs_i^{(c)} \times V_{id_1}^{(c)}(t) \to V_i \tag{2}$$

$$\Phi^{(c)}:\ T_i^{(c)} \times Prs_i^{(c)} \times V_{id_1}^{(c)}(t) \to Prd_i^{(c)} \tag{3}$$

Every state of the protected infrastructure object V_i is characterized at each moment of time $t \in T$ by a set of variables $V_{id}^{(c)}, d \in N_a$, changing under the influence of cyber intruder attacks and the internal disturbances caused, for example, by component vulnerabilities of the system and/or application software.

Thus, with restrictions on the selected method of processing observations $u(t) \in Uadd$, on the intensity of the processed information flows ($\lambda 1 \le \lambda(t) \le \lambda 2$), on the amount of stored information about users and devices of the protected infrastructure ($V1 \le V(t) \le V2$), on the total time of collecting information about infrastructure users and devices ($\min\limits_{d_i \in D_{доп.}} \sum_{i=1}^{k} T_i(d_i^{сб.})$) need to find:

- Functional of state identification and control by the complex dynamic systems in the absence of time observability or partial observability of objects of the protected infrastructure ε:T×Prs×V→Prd, ϕ:Prd→T, φ:Prd→Tmon, k:T→Prd_{set}, γ:T→Tmon, i:Tmon→Prd_{set};
- Management law of the network (node) cyber-attack sensor, which would provide the total time spent on collecting the monitoring data of the protected infrastructure objects, not exceeding the directive value with restrictions on the acceptance region of management programs and a possible list of actions to ensure the required cyber resilience.

$$u^*(t) = \underset{\substack{u(t)\in\{U^{\partial}(t)\}\\ d_i\in\{D^{\partial}\}}}{arg}\left(\sum_{i=1}^{k} T\left(u(t), d_i^{\text{сб}}\right) \le T_{\Sigma}^{\partial}\right), \left\{\boldsymbol{U^{\partial}(t)}\right\} = \boldsymbol{u^{\partial}(t)} | (\boldsymbol{\lambda_1} \le \boldsymbol{\lambda(t)} \le \boldsymbol{\lambda_2}) \cap (\boldsymbol{V_1} \le \boldsymbol{V(t)} \le \boldsymbol{V_2}) \cap (\boldsymbol{N_1} \le \boldsymbol{N} \le \boldsymbol{N_2}). \quad (4)$$

In the secondary processing of monitoring data, the system for developing scenarios of proactively countering the cyber-attacks of the intruder and ensuring the required cyber resilience should assess the situation at $t=t_0$, determined by the dependencies between the states of the information sources and the sensors of the cyber-attack system. At the final time moment, the dependencies between the states become different, therefore the process of achieving the goal is described as a change in the dependencies

$$\boldsymbol{x_{B^{(1)},B^{(2)}}(t_0) R_{<\cdot>} x_{B^{(3)}}(t_0) x_{B^{(1)},B^{(2)}}(t_k) R_{<\cdot>} x_{B^{(3)}}(t_k)} \quad (5)$$

moreover, the logical entailment from the initial to the final state is associated with a set of possible informational actions.

The action list and sequence is determined by the logic of behavior $B^{(3)}$, its settings. In fact, $B^{(3)}$ performs the functions of a control unit that prepares some decision to ensure the required cyber resilience.

Working out a solution, it is necessary to consider all possible choices leading to the achievement of the goal $\boldsymbol{P}\left(\hat{\boldsymbol{\tau}}_{\text{req}} < \hat{\boldsymbol{\tau}} < \hat{\boldsymbol{\tau}}_{\text{enough}}\right) = \boldsymbol{P}_{\text{PV}}$, where $\hat{\boldsymbol{\tau}} = \hat{\boldsymbol{\tau}}_{\boldsymbol{p}} + \hat{\boldsymbol{\tau}}_{\text{pass}} + \hat{\boldsymbol{\tau}}_{\text{act}} + \hat{\boldsymbol{\tau}}_{\text{RV}}$, $\boldsymbol{p} \ge \hat{\boldsymbol{\tau}}_{\text{req}}$, and when deciding among the possible solutions it should be chosen the one most preferred choice.

Choosing the possible solutions and the actions behind them, it is necessary to choose such chains from them that satisfy the condition (5).

The emerging information situation at the protected infrastructure is fixed by a set of decision rules, reflecting the connections between the states $B^{(1)}$, $B^{(2)}$, $B^{(3)}$ with $t=t_k$. Thus, at the next stage of ensuring the required cyber resilience of the protected infrastructure, it is necessary to determine the observation parameters, based on the determining the diagnostic value of signs of a potentially vulnerable critically important information infrastructure.

3 Selection of observation parameters

In the technical diagnostics of the critically important information infrastructure, it is very important to describe the object in the system of signs that has a greater diagnostic

value. The use of the non-informative features not only turns out to be useless, but also reduces the efficiency of the diagnostic process itself, disturbing with recognition. We assume that the diagnostic sign value is determined by the information significance that is added by the sign into the observation object state system [9, 13].

Let there be a system Pr, which is in one of n possible states Pr_i(i=1,2,…,n).

Let us call this system - a system of profiles, and each of the states - a profile. Different states of the protected infrastructure at discrete instants of time are represented by a set of standards (profiles), while the choice of the number of profiles is determined by the study objectives. Recognition of the Pr system states is carried out by monitoring the system associated with it - the system of signs. We will call the survey result, expressed in one of two symbols or a binary number (0 and 1), a simple attribute.

From the point of information theory view, a simple feature can be considered as a system having one of two possible states. If k_j is a simple sign, then its two states will be denoted by k_j - the sign presence, $\overline{k}_j$ - the sign absence. A simple sign may indicate the presence or absence of the measured PST in a certain interval; it may also have a qualitative character (positive or negative test result, etc.) [11, 12].

The two-digit sign (m=2)) has two possible states. The states of the two-digit sign k_j are denoted by $\boldsymbol{k_{j_1}}$ and $\boldsymbol{k_{j_2}}$. Let, for example, the sign k_j be related to the measurement of PST x, for which two diagnostic intervals are established: x≤10 and x›0. Then $\boldsymbol{k_{j_1}}$ corresponds to x≤10, and $\boldsymbol{k_{j_2}}$ denotes x›10. These states are alternative because only one of them is realized.

It is obvious that the two-digit sign can be replaced by the simple sign k_j, putting $\boldsymbol{k_{j_1} = k_j, k_{j_2} = \overline{k}_j}$.

If the survey detect that the sign k_j has the value $\boldsymbol{k_{j_s}}$, for this object, then this value will be called the implementation of the sign k_j. Denoting it by $\boldsymbol{k_j^*}$, we will have $\boldsymbol{k_j^* = k_{j_s}}$.for the diagnosis Pr_i we take

$$\boldsymbol{Z_{Pr_i}(k_j^*) = Z_{Pr_i}(k_{j_s}) = log_2 \frac{P\left(\frac{Pr_i}{k_{j_s}}\right)}{P(Pr_i)}} \tag{6}$$

where $\boldsymbol{P\left(\frac{Pr_i}{k_{j_s}}\right)}$ – profile probability Pr_i provided that the sign k_j received the value $\boldsymbol{k_{j_s}}$; $P(Pr_i)$– is the prior profile probability.

The value $\boldsymbol{Z_{Pr_i}(k_{j_s})}$ was met in works on information theory under the name “information value”. From the point of view of information theory, the quantity $\boldsymbol{Z_{Pr_i}(k_{j_s})}$ is information on the state Pr_i, which the state of the sign $\boldsymbol{k_{j_s}}$ possesses. The diagnostic weight of a particular implementation of a sign does not yet give an idea of the diagnostic value of the examination for this sign. Thus, during a survey on a simple sign, it may turn out that its value does not have a diagnostic weight, whereas its absence is extremely important for establishing the profile of the object of the protected infrastructure.

We will consider the diagnostic survey value on the m-bit k_j sign for the profile Pr_i the information amount introduced by all implementations of the k_j sign to the profile Pr_i

$$\boldsymbol{Z_{Pr_i}(k_j) = \sum_{s=1}^{m} P\left(\frac{k_{js}}{Pr_i}\right) Z_{Pr_i}(k_{js})} \quad (7)$$

The diagnostic survey value takes into account all possible implementations of a sign and represents the amount expectation of information contributed by individual implementations. Since the value of $\boldsymbol{Z_{Pr_i}(k_j)}$ refers to only one profile Pr_i, we will call it the private diagnostic survey value based on k_j sign. $\boldsymbol{Z_{Pr_i}(k_j)}$ determines the independent diagnostic survey value. It is situation characteristic when the survey is conducted first or when the results of other surveys are unknown. Write $\boldsymbol{Z_{Pr_i}(k_j)}$ in a form convenient for further calculations

$$\boldsymbol{Z_{Pr_i}(k_j) = \sum_{s=1}^{m} P\left(\frac{k_{js}}{Pr_i}\right) log_2 \left[\frac{P\left(\frac{Pr_i}{k_{js}}\right)}{P(Pr_i)}\right]} \quad (8)$$

The generated attribute space allowed identifying and classifying the symptoms of the potentially vulnerable states of the protected infrastructure, determine the network traffic parameters used for communication and behavioral profiling of the protected infrastructure objects with atypical interaction macroparameters.

Let us further consider the procedure for determining the diagnostic sign weights of vulnerable states of the protected object infrastructure.

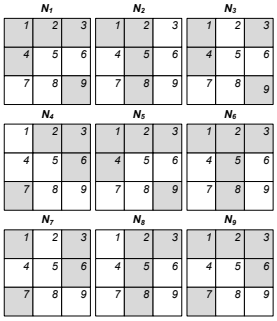
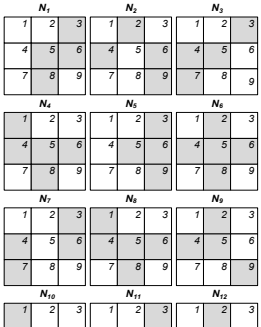
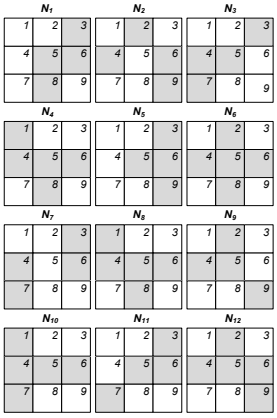
4 Reference behavior profiling

We will distinguish three different object states of the protected infrastructure (profiles), caused by the attacking effects of violators: Pr_1 is a profile, characterizing a vulnerable condition due to an unknown zero-day vulnerability; Pr_2 is a profile, characterizing the vulnerable state, due to the configuration of protection means; Pr_3 is a profile, characterizing a vulnerable condition, due to an impact on a known vulnerability. Profiling is carried out, according to the nine simple non-specific features: byte-frequency for TCP (k_1), byte-frequency for UDP (k_2), hash value (k_3), hash-value based on offset byte (k_4), based on the first **4** bytes repeated in packets (k_5), the hash value for pairs of the first **16** bytes of the first **4** packets (k_6), the length of the first four packets in one direction (k_7), the nibble number of the first packet from the server to the client (k_8), duplicate pairs of bytes (k_9) [14, 15].

For example, the functional state is diagnosed at **414** of the **450** network nodes of the protected infrastructure (having no known vulnerabilities), **10** of the **36** surveyed nodes that were attacked by the intruders, were in the first vulnerable state, **12** in the second, and **14** in the third [10, 16]. The results of profiling by the characteristics are shown in Table 1. Let us note that the first profile is characterized by the presence of at least two shaded squares (ones) in the first row and at least two white squares (zeros)

in the remaining rows, etc. The frequency of characteristic occurrence is taken as its probability.

Table 1. Statistical data of profiling infrastructure objects by a simple sign

Pr_i	Item No. N	k_i sign k_1	k_2	k_3	K_4	k_5	k_6	k_7	k_8	k_9	Geometric interpretation
Pr_1	1	1	1	1	1	0	0	0	0	1	
	2	1	1	0	0	1	0	0	1	0	
	3	1	0	1	1	0	0	0	0	1	
	4	0	1	1	0	0	1	1	0	0	
	5	1	1	1	1	0	0	0	0	1	
	6	1	1	1	0	1	0	0	1	0	
	7	1	1	0	1	0	0	0	0	1	
	8	1	0	1	0	0	1	1	0	0	
	9	0	1	1	0	0	1	1	0	0	
	10	1	1	1	0	0	1	1	0	0	
Pr_2	1	0	0	1	0	1	1	0	1	0	
	2	0	1	0	1	0	1	0	0	1	
	3	0	0	1	1	1	0	1	0	0	
	4	1	0	0	1	1	1	0	1	0	
	5	0	0	1	0	1	1	0	0	1	
	6	0	1	0	1	1	1	0	1	0	
	7	0	0	1	1	0	1	1	0	0	
	8	1	0	0	1	1	1	0	1	0	
	9	0	1	0	1	1	0	0	0	1	
	10	1	0	0	1	1	1	1	0	0	
	11	0	0	1	0	1	1	1	0	0	
	12	0	1	0	1	1	1	0	0	1	
Pr_3	1	1	0	0	0	1	0	1	1	0	
	2	0	1	0	0	0	1	0	1	1	
	3	1	0	0	1	0	0	1	1	1	
	4	0	0	1	0	1	0	1	0	1	
	5	1	0	0	0	0	1	1	1	0	
	6	0	1	0	0	1	0	1	1	1	
	7	1	0	0	1	0	0	0	1	1	
	8	0	0	1	0	1	0	1	1	1	
	9	0	1	0	0	0	1	1	0	1	
	10	0	0	1	1	0	0	1	1	1	
	11	0	0	1	1	0	0	1	1	1	
	12	0	1	0	0	1	0	0	1	1	
	13	0	0	1	1	0	0	1	1	0	
	14	1	0	0	0	0	1	1	0	1	

For example, for the first sign (presence of feature k_1, absence - $\overline{k_1}$): $P\left(\frac{k_1}{Pr_1}\right) = \frac{8}{10} = 0,8$; $P\left(\frac{k_1}{Pr_1}\right) = \frac{8}{10} = 0,80$; $P\left(\frac{k_1}{Pr_2}\right) = \frac{3}{12} = 0,25$; $P\left(\frac{k_1}{Pr_3}\right) = \frac{5}{14} = 0,357$; $P(k_1) = \frac{16}{36} = 0,444$. Then, we determine the independent diagnostic implementation weight of features using the expression (7) and the independent diagnostic survey value for equality (8).

The calculation results are shown in Table 2. For the Pr_1 profile, the survey by k_1, k_2, k_3 charcteristic is the most diagnostic; for the profile Pr_2 – by k_4, k_5, k_6 and for the profile Pr_3 – by k_7, k_8, k_9 signs. For the entire profile system, the diagnostic survey result values do not change a lot.

Table 2. Probabilities, diagnostic weights of implementation and diagnostic values of various signs

Feature k_j	Profile Pr_i												P(k_j)	$Z_{Pr}(k_j)$
	Pr_1				Pr_2				Pr_3					
	P(Pr_1)= 0,278				P(Pr_2) = 0,333				P(Pr_3) = 0,389					
	$P\left(\frac{k_j}{Pr_1}\right)$	$Z_{Pr_1}(k_j)$	$Z_{Pr_1}(\overline{k_j})$	$Z_{Pr_1}(k_j)$	$P\left(\frac{k_j}{Pr_2}\right)$	$Z_{Pr_2}(k_j)$	$Z_{Pr_2}(\overline{k_j})$	$Z_{Pr_2}(k_j)$	$P\left(\frac{k_j}{Pr_3}\right)$	$Z_{Pr_3}(k_j)$	$Z_{Pr_3}(\overline{k_j})$	$Z_{Pr_3}(k_j)$		
1	0,8	0,848	-1,475	0,383	0,25	-0,83	0,443	0,117	0,357	-0,315	0,21	0,023	0,444	0,154
2	0,8	0,848	-1,475	0,383	0,333	-0,415	0,263	0,037	0,286	-0,635	0,362	0,017	0,444	0,149
3	0,8	0,678	-1,322	0,278	0,417	-0,263	0,222	0,02	0,357	-0,486	0,363	0,059	0,5	0,107
…														
9	0,4	-0,4	0,346	0,047	0,333	-0,662	0,498	0,111	0,786	0,575	-1,141	0,208	0,528	0,141

Table 3 presents the condition of the diagnostic survey value after the surveying on the first characteristic. The table shows a significant change in the diagnostic survey value, depending on one or another implementation of the first sign.

Table 3. Conditional diagnostic survey values

Feature k_j	Profile Pr_i						$Z_{Pr}\left(\frac{k_j}{k_1}\right)$	$Z_{Pr}\left(\frac{k_j}{\overline{k_1}}\right)$
	Pr_1		Pr_2		Pr_3			
	$Z_{Pr_1}\left(\frac{k_j}{k_1}\right)$	$Z_{Pr_1}\left(\frac{k_j}{\overline{k_1}}\right)$	$Z_{Pr_2}\left(\frac{k_j}{k_1}\right)$	$Z_{Pr_2}\left(\frac{k_j}{\overline{k_1}}\right)$	$Z_{Pr_3}\left(\frac{k_j}{k_1}\right)$	$Z_{Pr_3}\left(\frac{k_j}{\overline{k_1}}\right)$		
2	0,42	1	0,678	0,009	0,678	0,009	0,606	0,284
3	0,42	0,737	0,678	0,006	0,678	0,006	0,606	0,209
4	0,011	0,863	0,83	0,136	0,077	0,041	0,31	0,301
…								
8	0,189	0	0,082	0,235	0,278	0,235	0,188	0,170

Thus, knowing the diagnostic survey value for the corresponding characteristic groups in the corresponding infrastructure, it is possible to conduct the selective monitoring, providing a significant reduction in the response time to potential incidents and ensuring the required cyber-resilience.

5 Procedure for iterative diagnosis

In the diagnostics tasks of the critically important information infrastructure, the selection of the most informative features for describing the object of the mentioned infrastructure and the subsequent construction of the diagnostic process is extremely important. In many cases, this is due both to the difficulty of obtaining the information itself (the node (network) number sensors of the cyber-attack detection systems, as a rule, is limited), and with the limited time of diagnostic survey under cyber-attacks. Imagine the process of diagnostic survey as follows [13, 15]. A system can be with a certain probability in one of the previously unknown states. If the prior probabilities of the states $P(Pr_i)$ can be obtained from a statistical data, then the system entropy is

$$H(Pr) = -\sum_{i=1}^{n} P(Pr_i)\, log_2\, P(Pr_i) \quad (9)$$

As a result of a full diagnostic survey of the complex of features K, the system state becomes known (for example, it turns out that the network object is in the state Pr1, then $P(Pr_1)=1$, $P(Pr_1)=0(i=2,\ldots n)$. After a complete diagnostic survey, the system entropy (uncertainty)

$$H(Pr/K)=0 \quad (10)$$

This information contained in the diagnostic survey, or the diagnostic survey value is

$$JPr(K)=ZPr(k)=H(Pr)-H(Pr/K)=H(Pr) \quad (11)$$

In fact, the condition (10) is far from being always fulfilled. In many cases, a recognition is statistical in nature and it is necessary to know that the probability of one of the states is quite high (for example, $P(Pr_1)=0,95$. For such situations, the residual system entropy $(Pr/K)\neq 0$.

In practical cases, the required diagnostic survey value is

$$ZPr(K)=\xi H(Pr) \quad (12)$$

where $\boldsymbol{\xi}$ is the survey completeness coefficient, $0<\xi<1$.

The coefficient $\boldsymbol{\xi}$ depends on the recognition reliability and for real diagnostic processes should be close to 1. If the prior probabilities of the system states are unknown, then one can always give an upper assessment for the system entropy $H(Pr)\leq \log_2 n$, where n is the number of the system states.

Under the (12) condition it follows that the amount of information that needs to be obtained during a diagnostic survey is given and it is required to make an optimal process for its accumulation.

When making a diagnostic process, it is necessary to take into account the difficulty of obtaining relevant information. Let us call the optimality coefficient of the diagnostic survey based on k_j for the profile Pr_i value is

$$\boldsymbol{\lambda_{ij}} = \frac{\boldsymbol{Z_{Pr_i}(k_j)}}{\boldsymbol{c_{ij}}} \quad (13)$$

where $\boldsymbol{Z_{Pr_i}(k_j)}$ is the diagnostic survey value based on $\boldsymbol{k_j}$ for the profile$\boldsymbol{Pr_i}$. In general, $\boldsymbol{Z_{Pr_i}(k_j)}$ is determined based on the results of previous surveys; c_{ij} is the coefficient of survey complexity based on $\boldsymbol{k_j}$ for the profile $\boldsymbol{Pr_i}$, it characterizes the laboriousness of the survey, its reliability, duration and other factors. It is assumed that c_{ij} does not depend on the previous surveys.

The optimality coefficient for the entire profile system is

$$\boldsymbol{\lambda_j} = \frac{\sum_{i=1}^{n} \boldsymbol{P(Pr_i)Z_{Pr_i}(k_j)}}{\sum_{i=1}^{n} \boldsymbol{P(Pr_i)c_{ij}}} = \frac{\boldsymbol{Z_{Pr_i}(k_j)}}{\boldsymbol{c_j}}. \quad (14)$$

When calculating λ_j, information is averaged and the survey complexity is carried out over all profiles. For survey of complex K of v signs, the optimality coefficient is

$$\boldsymbol{\lambda} = \frac{\boldsymbol{Z_{Pr}(K^{(v)})}}{\sum_{j=1}^{v} \boldsymbol{c_j}} \quad (15)$$

where ZPr(K(v)) is the diagnostic survey value of the complex of signs.

Thus, the optimality coefficient will be large if a smaller number of the individual surveys obtains the required diagnostic value. In the general case, an optimal diagnostic process should ensure that the maximum value of the optimality coefficient of the entire survey is obtained (conditions for the diagnostic survey optimality).

To describe the interaction (information transfer) between the objects of the protected infrastructure in time, dynamic communication profiles are used. The object profile of the protected infrastructure will be understood below as a formalized means of describing and displaying the characteristics of the infrastructure as a whole and its individual object in terms of the specification of rules (communication protocols, access to resources) and data exchange procedures at the corresponding observation interval.

The interaction features of the network nodes in a given observation interval are presented in three-dimensional space (Figure 3), where the start and end times of the

corresponding interaction processes are specified on the X-axis, the identified operating systems (OS) and applications installed on the network node are specified on the Y-axis, on the Z axis are the numbers used for TCP/UDP port interaction used by the corresponding applications. The communication profile (CP) of the network object is represented as

$$CP = Pr_1 = \langle Sft(Pt_k)_1^{Prt_i}, \dots, Sft(Pt_k)_n^{Prt_i} \rangle \quad (16)$$

where Sft is the software type (Operating system or application), Sft∈Os∪Apl, Pt - protocol, Prt- TCP/UDP port number, i=1,2,…65535; k,n∈N.

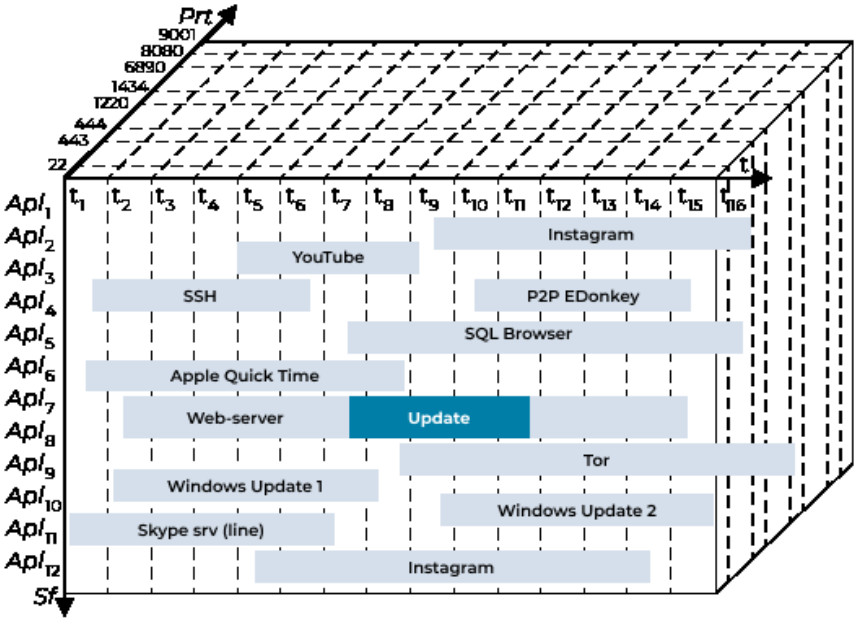


Fig. 3. Representation of the network object interaction

For example, the communication profile of a network object, shown in the diagram in Figure 3, has the following form:

$$\boldsymbol{Pr}_{NO_1}^{(1)} = \langle \boldsymbol{Apl}_1^{443}, \boldsymbol{Apl}_2^{443}, \boldsymbol{Apl}_3^{22}, \boldsymbol{Apl}_4^{6890}, \boldsymbol{Apl}_5^{1220}, \boldsymbol{Apl}_6^{8080}, \boldsymbol{Apl}_7^{4443}, \boldsymbol{Apl}_8^{443}, \boldsymbol{Apl}_9^{444}, \mathbf{Apl}_{10}^{9001}, \mathbf{Apl}_{11}^{1434}, \mathbf{Apl}_{12}^{443} \rangle.$$

Behavioral profile (BP)

$$BP=Pr2=<Nm1=<Vi, type, \xi k, Ds>> \quad (17)$$

where Nm – OS (application) name, V – network object identifier (application instance name); type - network object type (active or passive), type∈Act∪Psv; ξ - application version; D - a set of operations; I, k, n∈N.

For example, the behavioral profile of a network object represented in the diagram in Figure 3 has the following form:

$$BP_{NO_1}^{(1)} = \left\langle \begin{matrix} Apl_1 = < Instagram(IOS);6.0;act;chat>,..., \\ Apl_{11} = < Instagram(Android);5.1;act;chat> \end{matrix} \right\rangle \quad (18)$$

Protection profile (PP)

$$BP=Pr_3=<Nm_1=<\gamma I, \phi k, \psi>> \quad (19)$$

where γ – a security service name; ϕ - version; ψ - operation type (chat, file sharing (download), use of a web browser, download, file sharing (upload), IP-call); I, k, n∈N.

For example, the security profile (SP) of a network object represented in the diagram in Figure 3 has the following form:

$$SP_{NO_1}^{(1)} = \left\langle \begin{matrix} Apl_3 = \langle OpenSSH(sshd),\ 2,\ Kerberos\ v5\ auth \rangle,..., \\ Apl_8 = \langle MSCryptoAPI,\ 6.1,\ E2EE \rangle \end{matrix} \right\rangle \quad (20)$$

As an example, let us consider the detection of the certificate spoofing at one of the workplaces when accessing a web resource using the SSL/TLS protocols as a result of a passive monitoring. This situation has many alternatives in terms of the development of situations, related to the cyber-resilience violation of the protected infrastructure. If the destructive actions of the user were deliberate, this event (incident) can be associated with both previous incidents, and have a high probability of recurrence in the future (Table 4).

Table 4. Possible list of the preventive actions

No.	Network object	System or application software component	Exploited vulnerability/ vulnerable protocol or component	Action to prevent or respond to an incident
1.	Windows-hosts	Windows 2018	CVE-2016-3213/ NetBIOS, ISATAP	Installing security system updates MS16-063, MS16-077
2.	Client hosts	Web-browser	Internet Explorer, HTTP/HTTPS	Using Firefox, Opera, Chrome browsers with HPKP technology
			TLS	Mutual authentication when establishing a TLS connection
			HTTPS, TLS	Control of application software with access to the web browser
				Use of additional sources or databases of permitted keys and certificates
				Mutual client and server authentication
3.	Network traffic monitoring system	-	DNS	DNS name resolution
			SSL/TLS	SSL name resolution, maintenance of a registry of public server trusted key fingerprints

In addition, this incident poses a threat to the protected infrastructure from the intruder's point of view, gaining an access to the compromised node, as well as compromising other nodes or the entire infrastructure under study. At the first stage, based on the reverse data analysis, it will be necessary to verify the events (as well as their results) with the statistical characteristics are of interest in detecting cause-and-effect links between the user actions to determine his degree participation in the incident: certificate with the authentic issuer; certificate with fake issuer; certificate with valid expiration date; certificate with expired validity; certificate with original issuer, not expired; certificate with original issuer, expired; certificate with fake issuer, not expired; certificate with fake issuer, expired. According to the investigation results, the monitoring system forms a list of preventive (response) actions to the corresponding incident.

6 Conclusions

Further, a set of the qualitative features is formed, based on the results of the secondary processing of the monitoring results in the form of a decision tree, the interconnection degree between alternative feature groups, technical and economic consequences (damage) for the protected infrastructure and its assets during their manifestation is determined, and a set of possible actions is generated to localize the incident.

Thus, the proposed method of profiling the behavior of dynamic objects of a critically important information infrastructure allows selecting and putting into a practice (with scientific evidence) the corresponding organizational and technical measures to ensure the required cyber-resilience.

References

1. B. R. Shiller, 2014. “First-Degree Price Discrimination Using Big Data.” April 25, Brandeis University, Department of Economics Working Paper 58. [Electronic resource]. - Access mode: http://www.brandeis.edu/departments/economics/RePEc/brd/doc/Brandeis_WP58R.pdf
2. Beraud P., Cruz A., Hassell S. and Meadows S., "Using Cyber Maneuver to Improve Network Resiliency," in *MILCOM*, Baltimore, MD, 2011. DOI:10.1109/milcom.2011.6127449
3. Biryukov, D. N., Lomako, A. G. Approach to Building a Cyber Threat Prevention System. Problems of Information Security. Computer systems, Publishing house of Polytechnic University, vol. 2, pp. 13–19, St. Petersburg, Russia, 2013.
4. Bongard, M. M. The Problem of Recognition, Fizmatgiz, Moscow, Russia, 1967.
5. Bostick, T. P., Connelly, E. B., Lambert, J. H., & Linkov, I. (2018). Resilience Science, Policy and Investment for Civil Infrastructure. Reliability Engineering & System Safety *175*:19–23. DOI: 10.1016/j.ress.2018.02.025
6. Colbert, E. J., Kott, A., Knachel III, L., & Sullivan, D. T. (2017). *Modeling Cyber Physical War Gaming* (Technical Report No. ARL-TR-8079). US Army Research Laboratory, Aberdeen Proving Ground, United States.
7. Collier, Z. A., Linkov, I., DiMase, D., Walters, S., Tehranipoor, M., & Lambert, J. (2014a). Risk-Based Cybersecurity Standards: Policy Challenges and Opportunities. Computer 47:70–76. DOI: 10.1007/978-3-319-77492-3
8. D. J. Bodeau, "Analysis Through a Resilience Lens: Experiences and Lessons-Learned (PR 15-1309) (presentation)," in *5th Annual Secure and Resilient Cyber Architectures Invitational*, McLean, VA, 2015.
9. Dessavre D. G. and Ramirez-Marquez J. E., "Computational Techniques for the Approximation of Total System Resilience," in *Safety and Reliability of Complex Engineered Systems: ESREL 2015*, Zurich, Switzerland, 2015.
10. Dorofeev A.V., Markov A.S., Tsirlov V.L. Social Media in Identifying Threats to Ensure Safe Life in a Modern City, *Communications in Computer and Information Science*, 2016, vol. 674, pp. 441-449. DOI: 10.1007/978-3-319-49700-6_44.
11. Eisenberg, D. A., Linkov, I., Park, J., Bates, M., Fox-Lent, C., & Seager, T. (2014). Resilience metrics: Lessons from military doctrines. *Solutions, 5*(5), 76–87.

12. J. Park, T. P. Seager, P. S. Rao, M. Convertino and I. Linkov, "Integrating risk and resilience approaches to catastrophe management in engineering systems," *Risk Analysis,* vol. 33, no. 3, pp. 356-367, 2013. doi: 10.1111/j.1539-6924.2012.01885.x
13. Kelic, A., Collier, Z. A., Brown, C., Beyeler, W. E., Outkin, A. V., Vargas, V. N., Ehlen, M. A., Judson, C., Zaidi, A., Leung, B., & Linkov, I. (2013). Decision framework for evaluating the macroeconomic risks and policy impacts of cyber attacks. *Environment Systems & Decisions, 33*(4), 544–560. DOI: 10.1007/s10669-013-9479-9
14. Kotenko, I. V. Intellectual mechanisms of cybersecurity management. Proceedings of ISA RAS. Risk Manag. Safety, 41, pp. 74–103, Moscow, Russia, 2009.
15. Lomako, A. G., Petrenko, S. A., Petrenko, A. S. Realization of the immune system of the stable computations organization, In: Information systems and technologies in modelling and management, Materials of the All-Russian scientific and practical conference, pp. 255-259, Russia, 2017.
16. Patrick McDaniel and Ananthram Swami, The Cyber Security Collaborative Research Alliance:Unifying Detection, Agility, and Risk in Mission-Oriented Cyber Decision Making. CSIAC Journal, Army Research Laboratory (ARL) Cyber Science and Technology, 5(1), December, 2016.