Decision support systems for information protection in the management of the information network

G I Korshunov[1,2], V A Lipatnikov[3] and A A Shevchenko[3]

[1] Innovation and Integrated quality systems dept. Saint-Petersburg State University of Airspace Instrumentation, 67, BolshayaMorskaya St., Sankt-Petersburg, 190000, Russia
[2] Ciberphisic systems and control high school, Peter the Great St.Petersburg Polytechnic University, 29, Polytechnicheskaya St., Saint-Petersburg, 195251, Russia
[3] S. M. BudjonnyMilitary Academy of the Signal Corps, TikhoretskiyProspekt, 3, Sankt-Petersburg, 194064, Russia kgi@pantes.ru

Abstract. The structure and algorithms of the decision support system for information network management for information protection based on intellectual analysis are proposed. Recognition of intrusions and analysis of the dynamics of the violator's actions are implemented on the basis of a tree classifier and Kohonen maps for neural networks. The system includes agents for intrusion detection and state forecasting, takes into account the dynamics and stochastic uncertainty of the main information protection processes in information network, provides monitoring, intrusion detection and forecasting of the information security state in intellectual protection processes.

1. Introduction

Intensive development of computer technology and industrial Internet has led to the unification of a huge number of heterogeneous networks. The transition to an information society sharply posed the problem of ensuring information security (IS) in the construction and management of distributed computing systems [1, 2]. An approach to the development and use of information security systems (IS) based on the allocation of an intellectual superstructure over traditional methods of protection and the construction of a single unified environment for the creation and support of the functioning of protection systems was consideredin [3]. A method for managing the information security of computer networks on the basis of a dedicated server with container virtualization is proposed, issues of intrusion detection and forecasting of the protection status of the information and computer network (ICN) of a system of distributed situational centers are considered. The main requirement for such systems is the ability to find anomalies and, accordingly, intrusion in real time. To effectively manage the IS, proactive protection means must ensure the collection of necessary information, analysis of security, monitoring of network conditions, detection of attacks, forecasting, counteracting their implementation, misleading the attacker [4]. Therefore, in order to increase the IS of ICN it is necessary to solve the actual tasks of classifying the classification when recognizing intrusions in intelligent ways of managing the IS system of distributed situational centers, analyzing the dynamics of the offender's actions, which include scenarios of external and internal intrusions. There is a contradiction between effective new means of information invasion and existing methods of protecting ICN. The purpose of the article is to increase the effectiveness of decision support in the management of ICN IS by analyzing the dynamics of the violator's actions.For this purpose, tasks are set to develop a method for supporting decision-making in the management of ICN with intrusion detection and forecasting the state of information security, identifying threats of intrusions using algorithms, and investigating the possibility of using methods of intellectualizing processes. Such a method is developed on the basis of a tree-like classifier and Kohonen maps.

2. Method of management of the ICN

The structural scheme of ICN in the problem of decision support for management based on intellectual analysis is shown in Figure 1. The ICN control should provide intrusion detection and prediction of the IS state when detecting intrusion threats using algorithms developed on the basis of a tree classifier and Kohonen maps [5, 6]. Management includes monitoring the situation, operational control, recognition of the sequence of actions of the offender, modeling the strategy of the violator's impact, the process of determining the situational parameters in a mutual conflict situation with a reliable prediction of the strategy of computer attacks. To solve the problems of ICN protection and monitoring, it is necessary not only to detect and block the actions of violators [7], but also to analyze computer attacks and distract infringers from information systems, by enticing infringers on false information systems and collecting information on the tactics of offenders, to carry out identification and exposure. Based on the analysis of the violator's activities, the weaknesses of the information protection system in the ICN aredetermined.

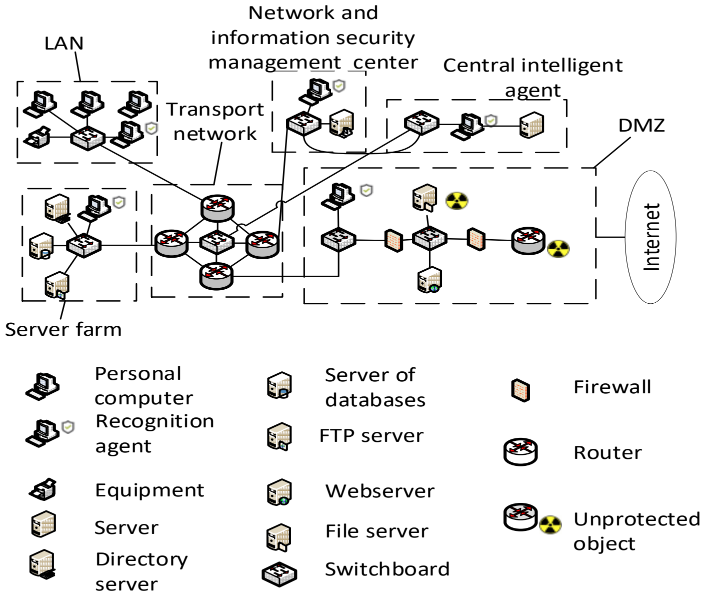

Figure 1. The structure of the ICN with the support of decision-making in the management based on intellectual analysis.

The functions of the ICN control algorithm include monitoring and the characteristics of digital streams assignmentwith data transfer protocols entering the ICN and the dedicated server, intrusion detection, selection and implementation of the protection method. The decision support algorithm for intelligent control is shown in Figure 2.

The algorithm implements:

- analysis of the current situation on the basis of data from a dedicated server with container virtualization;
- clustering of parameter values;
- processing of the received values;
- generation of a forecast based on the output values of the network;
- filtering of the received values and allocation of the target class that determines the predicted state of the ICN IS.

The methods and algorithms for decision support based on intellectual analysis are neededto prevent malware intrusions. The complexity of this problem is largely due to the incompleteness, inconsistency and variety of distribution laws in traffic flows. Possible scenarios detected by the agent intrusion detection system are [8 - 10]:

- monitoring and control of the attacker (actions to determine the network configuration, detection of hosts operating on the host services, the definition of the operating system, applications);
- Introduction to the system - the attacker's actions on hacking the host and introducing it into the system;
- Increase rights - intrusion attempts aimed at gaining elevated rights to access host objects;
- spread of the defeat on the host - illegitimate distribution of the attacker on host objects (directories, files, programs);
- distribution of the defeat over the network - the spread of the attacker over the protected ICN.

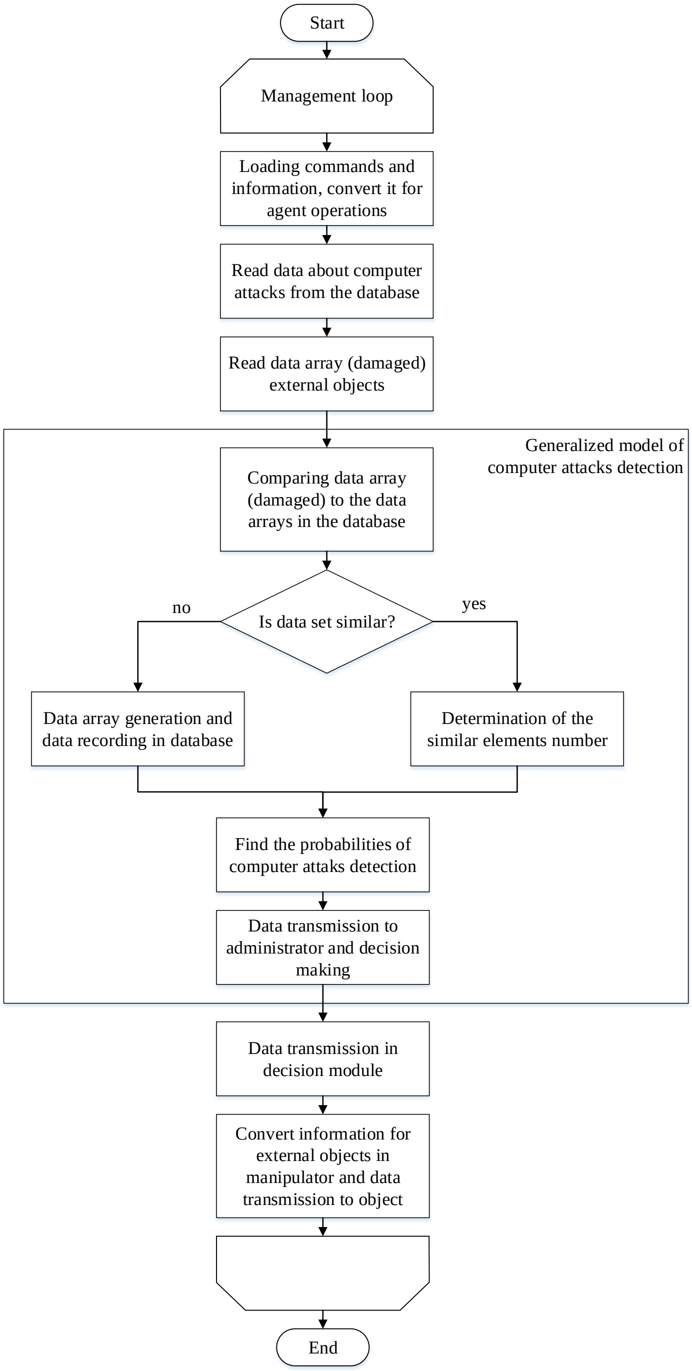
Start
Management loop
Loading commands and information, convert it for agent operations
Read data about computer attacks from the database
Read data array (damaged) external objects
Generalized model of computer attacks detection
Comparing data array (damaged) to the data arrays in the database
Is data set similar?
no
yes
Data array generation and data recording in database
Determination of the similar elements number
Find the probabilities of computer attaks detection
Data transmission to administrator and decision making
Data transmission in decision module
Convert information for external objects in manipulator and data transmission to object
End

Figure 2. Algorithm for decision support in management based on intellectual analysis.

3. The tree classifier

The Kohonen map is used for initial sorting and clustering of incoming values, it provides structuring of the raw data for a tree neural network. The tree classifier is intended for determining the degree of belonging of parameter values to a certain, early defined class that characterizes the current status of the ICN.

The functioning of the model realizes the clustering of values, the processing of the obtained values with the help of a neural tree classifier, the filtration of the obtained value and the allocation of the target class that determines the state of the ICN.

Consequently, having at the input of the considered model a certain set of knowledge of parameters, one can unambiguously interpret its output value as an estimate of the current state [11].

Identifying secondary signs of intrusion into information resources contains:

- identification of correlations between the signs;
- replenishment of the database of secondary invasion signs;
- clarification of the invasion scenarios;
- development of adequate deterrence and neutralization measures.

The advantage of the model for detecting abnormal deviations is the possibility of analyzing the dynamic processes of the ICN functioning and revealing new types of intrusion in them. The possibility of a priori recognition of anomalies by systematic scanning of the vulnerabilities is provided. The model ensures the implementation of:

- analysis of scenarios, ways of implementing and recognizing patterns of incursions of a potential violator;
- Identification of the status of ICN;
- detection, analysis and prevention of intrusions based on mathematically formalized and consistent logical rules [12];
- active counteraction to sources of intrusion based on the hierarchy analysis method;
- justification of requirements and proposals for developing ways of countering intrusions;
- substantiation of the test site for evaluation of models, methods and means of countering intrusions;
- assessing the effectiveness of anti-intrusion measures [12].

Neural network modeling of time series in multifactor forecasting consists in the formation of an structure that describes the behavior of the ICN at time points, and the prediction consists in predicting the future behavior of the system from the prehistory [13, 14]. The training of such neural networks is to adjust the weighting coefficients based on the change in the actual error of prediction at the iteration. Network training is performed separately for each time series, since an attempt to forecast a series on which the network was not trained will lead to an error result [15].

A formal description of the modification of the theory of attack trees - probabilistic attack trees is performed as follows. They describe ways to achieve targets by intruders and include the introduction of the composition "OAND" (Ordered AND), probability function $P(t)$, the ability to specify the time interval of the attacking action for each of the subtrees and the ability to indicate the level of complexity of the attacking action for the attacker. For a given quasi subnet, several attack trees are automatically generated [16]. Select all the targets for the attack (all nodes). For each goal, two attack trees are constructed: a tree of integrity attacks, accessibility, and a tree of attacks on confidentiality. The nodes from which it is possible to attack a specific target are counted. 3 trees are obtained for each target, and each node is provided with a probability function $P(t)$ for implementing a specific threat.

For internal attacks, the tree leaves are inside the network, and for external attacks, the leaves point to the root router (the network is attacked outside the perimeter). Each leaf node (elementary threat) of each attack tree is described by the function $P(t)$, the form of which depends on the mathematical model of the implementation of a specific threat. To calculate $P(t)$ for the subgoals and directly targets of attack trees, the corresponding formulas are given (depending on the composition: "AND", "OR", "OAND", their combinations, nesting, etc.), taking into account the time intervals and levels of attack actions. For each goal, three attack trees are built-integrity, accessibility and confidentiality-and a method for calculating the probabilities of achieving the root integrity goals $P_iS(t)$, privacy $P_iC(t)$, and availability $P_iA(t)$ in each tree is constructed. The Kohonen neural map (hereinafter NK) performs a partitioning of the original set of vectors into a set of classes combining vectors closest to the central vectors of classes in the sense of the proximity measure being used. NK, working on the principle of "the winner gets everything", gives out one of the reference vectors to the output, to which the input vector is closest.

Let there be a set of vectors *Y*such that:

$$Y = \{y^1, y^2, ..., y^k, ..., y^K\};\ k = \overline{1,K};$$
$$y^k = \{y_1^k, y_2^k, ..., y_q^k, ..., y_Q^K\};\ q = \overline{1,Q};$$
$$y_q^k \in \Re.$$

Here: y^k is a vector from the set of vectors *Y*;

K is the cardinality of the set *Y*;

*Q*is the dimension of the vectors making up the set *Y*.

HK is the two-dimensional network of neurons, each of which calculates the measure of proximity of the input vector to the corresponding reference vector:

$$W = \{w^{11}, w^{12}, ..., w^{ij}, ..., w^{IJ}\};\ i = \overline{1,I};\ j = \overline{1,J};$$
$$w^{ij} = \{w_1^{ij}, w_2^{ij}, ..., w_q^{ij}, ..., w_Q^{IJ}\};\ q = \overline{1,Q};$$
$$w_q^{ij} \in \Re.$$

Here:

*W*is the set of reference vectors that are the vectors of the weights of NK neurons;

I and *J* - the dimensions of the two-dimensional NK.

As a measure of the closeness of the vectors of the sets *Y* and *W*, we use the Euclidean distance, defined as:

$$d^{ij} = \left[\sum_{q=1}^{Q} \left(x_q - w_q^{ij}\right)^2\right]^{1/2}.$$

The problem of teaching the NK is reduced to Voronoi's partition [17] of the set *Y* - that is, to a partition that minimizes variance in the resulting classes relative to the corresponding reference vectors. On the output of the NK, there is a вектор $w^{i^r j^r}$ такой, что: $d^{i^r j^r} \equiv i = \overline{1,I}; j = \overline{1,J}$.

The learning algorithm for the NKs under consideration, called self-organizing Kohonen maps, consists of the following steps:

1. As reference vectors, zero vectors are assumed.

In some variations of the considered algorithm, random vectors or a vector can be taken as reference vectors, whose component values are the selective mathematical expectation of the corresponding components of the vectors of the space *Y*;

2. For each vector $y \in Y$, the following actions are performed:

- vector *y* is fed to the input of the neural map;
- The neural map defines the neuron - the winner (that is, the neuron whose reference vector is closest to the input vector);
- the values of the components of the reference vector of the winner neuron are corrected towards the values of the components of the input vector *y*.

The adjustment is as follows:

$$w_{\text{winner}} = w_{\text{winner}} + \eta \cdot \sigma \cdot (y - w_{\text{winner}}). \quad (1)$$

Here the coefficients ηandσ have the following meaning:

η is the coefficient determining the learning rate. At the first iterations of the learning process, η is assumed to be sufficiently small and increases to the last iteration of training to 1;

σ is the coefficient determining the degree of approximation of the reference vector to the input vector; $\sigma \in [0;1]$ and to adjust the values of the components of the reference vector of the winner neuron is assumed to be slightly less than or equal to 1.

The problem of the coefficients ηandσconsists in decreasing the probability of formation of "empty" neurons in the trained NK - that is, neurons that have been poorly involved in the learning process, as a result of which the neighborhoods of the reference vectors corresponding to these neurons turn out to be sparse. For spaces, the distribution of vectors in which has pronounced "points of attraction" (that is, for statistically inhomogeneous spaces), the appearance of such empty vectors appears to be a significant problem, causing a decrease in the effectiveness of the application of NK space to data.

The value of the components of the reference vectors of neurons adjacent to the winner neuron is corrected by the formula (1). For neurons - neighbors of the neuron - winner, the coefficient σ takes smaller values, compared to the value of this coefficient for the winner neuron. The coefficient σ is, therefore, a measure of the proximity of neurons in the NK. The simplest version of the Kohonen algorithm takes the value $\sigma = 1$ for the winning neuron and $\sigma = 0$ for other neurons, which seems computationally efficient, but effectively weak. In other versions of the Kohonen algorithm $\sigma \neq 0$ for a number of neurons located next to the neuron-winner.

Ideal, but computationally inefficient, the case - $\sigma \neq 0$ for all neurons NK and σ decreases with distance from the winner neuron according to the law chosen by the programmer. In the developed software complex, the coefficient σ is calculated as follows: $\sigma = \mu \cdot \left| i - i_{\text{winner}} \right| + \left| j - j_{\text{winner}} \right|$, here:

- i_{winner} and j_{winner} – coordinates of the neuron is the winner;
- *i* and *j* - coordinates of the corrected neuron;
- $0 < \mu < 1$ - some coefficient that regulates the degree of proximity of neurons to each other.

In particular, for the winner neuron we have: σ = μ_0 = 1.

3. A unit is added to the counter of the iteration of the training cycle and, if the counter does not reach the maximum value, go to step 2.

When using the NK, the following features of the algorithm for learning the map data for vector normalization should be considered.

The problem of vector normalization is the choice between the execution and non-fulfillment of the normalization of the vectors of the original space before learning and functioning of the NK. Normalization of the vectors of the source space leads to the fact that the NK learning algorithm is guaranteed to converge (that is, sooner or later it will end due to the fact that at the next iteration of the learning cycle, none of the vectors of the original space will change its class), but the collinear vectors will To be considered as equivalent without taking into account a possible significant difference of their modules. In the absence of normalization of the vectors of the original space, the calculation of the distances between the vectors will also take into account their moduli, but in this case the convergence of the Kohonen algorithm is not guaranteed. The choice between the fulfillment and non-fulfillment of the normalization of the vectors of the source space before the training and functioning of the NK is made based on existing practical considerations. In the developed software complex, the normalization of the input vectors of the NK is not performed.

The problem of "empty" neurons, as already mentioned, consists in the appearance of neurons in the NK that did not participate in the learning process, because of which the reference vectors corresponding to these neurons turn out to be in the sparse regions of the initial space of vectors. The cognitive function detection means provides the processes of cognition of the external environment, other agents, as well as its self-knowledge. Cognitive processes cover the agent's perception of the external environment, the formation of a generalized internal representation, an understanding of the principles of interaction and behavior, and training. At the same time, to the known properties of the artificial agent as activity, reactivity, autonomy and communication skills, important prediction capabilities are added [5, 6]. The algorithm of intelligent agent intrusion detection system contains real-time analytical tools and implements:

- collection of data on an attack by an attacker;
- obtaining for further comparison from the database of computer attacks trained by the system in advance, possible intrusions of the attacker;
- Preliminary classification of incursions based on the Kohonen map. This method is universal and depends only on the structure of the input data;

- comparison of the invasion pattern with known attacks and detection of the IS level;
- A condition for checking whether an intrusion algorithm exists in the agent database.

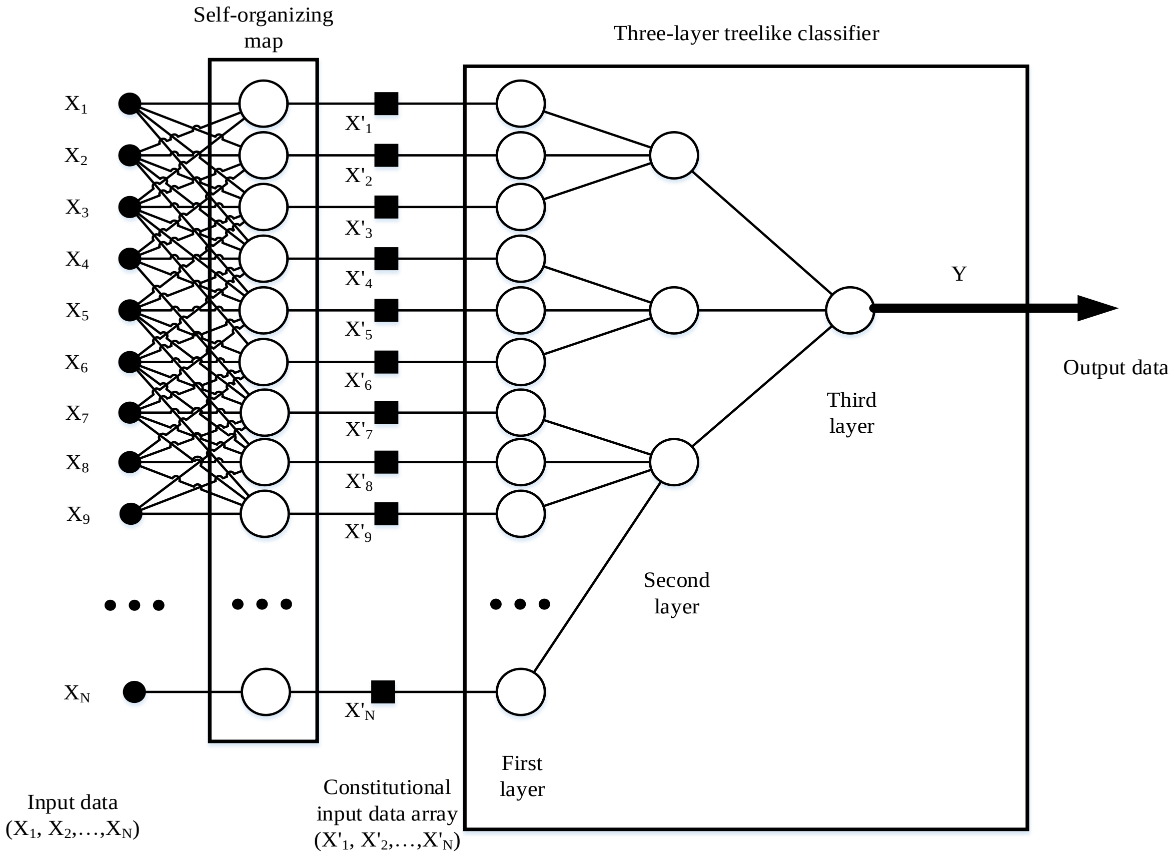

Figure 3. Generalized scheme of the model of abnormal abnormality detection.

The process of exposure by attackers to ICN in time can be viewed as a stream of random events (computer attacks) having a distribution density $w_A(t)$, and security measures taken by the ICN security administrator as a stream of random events with a distribution density $w_3(t)$ [4]. It is possible to determine the value of the probability of protection of the ICN$P_3(T)$ as

$$P_3(T) = \int_0^T w_3(\tau)\left[1 - \int_0^\tau w_A(t)\mathrm{d}t\right]\mathrm{d}\tau \qquad (2)$$

Provided that the streams of random events of the above processes have an exponential distribution, the intensities for the actions of the system administrator and the actions of the attacker have the following form λ=1/T_3 and λ=1/T_A, where T_3 is the average time required to implement the protection measures, and T_A– the average time required for the implementation of the computer attacks [4]. When the intensity expressions are substituted into formula (2), the value of $P_3(T)$ will be:

$$P_3(T) = (1 + T_3/T_A)^{-1} \cdot \{1 - \exp[-T \cdot (1 + T_3/T_A) \cdot T_3^{-1}]\} \qquad (3)$$

Taking into account that the process of information protection in ICN management using the proposed method is cyclical, the dependence of the probability of protection of the ICN$P_3(T)$ on the time calculated by the formula (3) is shown in Figure 4.

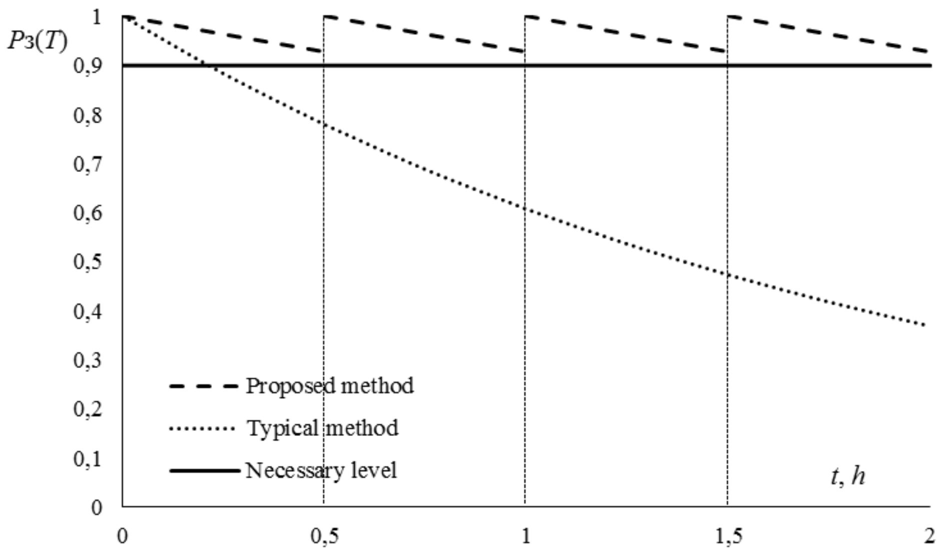

Figure 4. Graph of the dependence of the probability of protection of ITT on time, taking into account the cyclical nature of the information protection process in the management of ITTs.

Use of the proposed system allows to maintain the security of the ICN above the required valueof $P_3(T)$ within each iteration of the control cycle of the IS. When using a typical reactive approach to managing the IS at the time when the administrator has sufficient information about the computer attack to identify it and take measures to increase security, the security indicator will drop below the required value of $P_3(T)$.

The model and method of decision support for information protection during ICN management allow to significantly reduce the idle time of ICN work facilities due to the increase in the accuracy of the forecast of changes in the main IS indicators. The algorithm of the process by the intelligent agent intrusion detection system allows to track the attacker's actions in a timely manner, provides the possibility of a well-founded decision to take measures to prevent the implementation of security threats by predicting the state of the IS. The experimental results show that the algorithm has a fairly high efficiency.

4. Conclusion

The proposed decision support system is constructed using artificial intelligence algorithms and is designed to protect information during ICN management. The system includes agents for intrusion detection and state forecasting, and, unlike other systems, takes into account the dynamics and stochastic uncertainty of the main information protection processes in the ICN; provides monitoring, intrusion detection and predicting the state of intelligence in intelligent protection processes.

The model of detection of computer attacks based on intelligent NS is applied for preliminary classification of anomalies in ICN. It is based on the identification of the normal behavior of the system from the distribution function of receiving data packets (the execution of the specified operator commands), the training of an intelligent NS and the comparative analysis of events in the training sample. Anomalous deviation in the ICN is detected when the degree of confidence of the intellectual NS of its solution lies below a given threshold. The use of the model for the implementation of mechanisms for protecting information from computer attacks is preceded by the training of these networks to specified algorithms for normal functioning.

Combining approaches to detecting attacks and recognizing intrusions seems promising in the design of control systems. This allows you to combine the advantages of artificial intelligence and heuristic methods, as well as use the technology of large data and prediction of the state. This contributes to the provision of cognitive processes and operational analytical processing of events, the intelligent management of large amounts of information about the IS.

5. References

[1] ISO/IEC 27035-2:2016 Information technology - Security techniques - Information security incident management - Part 2: Guidelines to plan and prepare for incident response.

[2] GOST R ISO/MEHK 27005-2010. Informatsionnaya tekhnologiya. Metody i sredstva obespecheniya bezopasnosti. Menedzhment riska informatsionnoy bezopasnosti.[State Standard R ISO/IEC 27005-2010. Information Technology. Security techniques. Information security risk management].Available at: http: // docs.cntd.ru/document/gost-r-iso-mek-27005-2010 (accessed 14 August 2018).

[3] Kuznetsov I. A., Lipatnikov V. A., Shevchenko A. A. The way of Multifactor Management of the Security of the Information and Telecommunications Network of the Quality Management System of Enterprises of Integrated Structures. Voprosy radioehlektroniki [Questions of Radio Electronics], 2016, no. 6, pp. 23–28 (In Russian).

[4] Lipatnikov V. A., Shevchenko A. A., Yatskin A. D., Semenova E. G. Information Security Management of Integrated Structure Organization based on a Dedicated Server with Container Virtualization. Informatsionno-upravliaiushchie sistemy [Information and Control Systems], 2017, no. 4, pp 67–76 (In Russian). doi:10.15217/issn1684-8853.2017.4.67

[5] T. Kohonen, Self-OrganizingMaps (ThirdExtendedEdition), NewYork, 2001, p 501

[6] ManzhulaV. G., FedyashovD. S. Kohonen neural networks and fuzzy neural networks in data mining.Fundamental'nye issledovaniya [Fundamental research], 2011, no 4, pp 108-114(InRussian).

[7] Lukatskiy A. Obnaruzhenie atak [Attack Detection], Saint-Petersburg, BHV-Petersburg, 2008. p 304 (In Russian)

[8] Amoroso E.G. Intrusion Detection: An Introduction to Internet Surveillance, Correlation, Trace Back, Traps, and Response. Intrusion. Net Book, 1999.

[9] Moore A.P., Ellison R.J., Linger R.C. Attack Modeling for Information Security and Survivability // Technical Note CMU/SEI–2001–TN–001. Survivable Systems, 2001.

[10] Gorodetski V., Kotenko I. Attacks against Computer Network: Formal Grammar-based Framework and Simulation Tool // Lecture Notes in Computer Science, Vol. 2516, 2002. pp 219-238.

[11] Abraham S., Nair S. A predictive framework for cyber security analytics using attack graphs // International Journal of Computer Networks & Communications (IJCNC). 2015. vol. 7. no.1. pp 1–17.

[12] Karganov V. V., et al. Sposob zaschity informatsionno-vychislitel'noj seti ot nesanktsionirovannykh vozdejstvij [A Way to Protect the Information Network from Unauthorized Influences]. Patent RF, no. 2635256, 2016.

[13] Camacho E.F., Bordons C. Model predictive control. – London: SprinderVerlag, 2004. – 405 p.

[14] Ivo Batina. Model predictive control for stochastic systems by randomized algorithms - Eindhoven:TechnischeUniversiteit Eindhoven, 2004.

[15] E. Byres, J. Lowe. The myths and facts behind cyber security risk for industrial control systems. - In ISA Process Control Conference, 2003.

[16] Singhal A., Ou X. Security risk analysis of enterprise networks using probabilistic attack graphs // Network Security Metrics. 2017. pp 53–73.

[17] Preparata F. P., ShamosM.I.Computationalgeometry. An introduction. – M.: World, 1989 p 478