Genetic search of the optimal configuration of information security system using simulation modeling

N V Davidyuk¹, I M Kosmacheva¹, S V Belov² and I Yu Kuchin²

¹ Department of Information Security, Institute of Information Technologies and Communications, Federal State Budgetary Educational Institution of Higher Education "Astrakhan State Technical University", Tatishchev St., 16, 414056, Astrakhan, Russian Federation
² Institute of Information Technologies and Communications, Federal State Budgetary Educational Institution of Higher Education "Astrakhan State Technical University", Tatishchev St., 16, 414056, Astrakhan, Russian Federation

E-mail: davidyuknv@bk.ru

Abstract. The paper proposes the approach to develop a data protection system (information security system) that is optimal in effectiveness using evolutionary search methods. This approach is characterized by the ability to take into account the influence of random factors (staff qualifications, equipment failures, attack time on the protection system) when choosing a protection option and the possibility of adapting the protection system to changing environmental conditions. The development of an effective information security system using a genetic algorithm is possible on the basis of data on monitoring events in the system, data received from experts and during simulation modeling of the protection system. The research results are of applied nature and can be used in developments related to the design of information systems, decision support systems in the field of information security.

1. Introduction. Formulation of the problem

The development of an effective information security system (ISS) or data protection system, which is adequate for various potential destructive influences, is a rather time-consuming task. This is explained by the uncertainty of the functioning conditions of the protected information systems (IS), the behavior of ISS in non-standard and extreme situations, staff qualifications, their workload, and many other reasons [1].

At the same time, the number of unique cyber incidents continues to increase (average growth by more than 30%), and methods for realizing information security threats are constantly evolving. The development of IT technologies leads to the need to more frequently review current ISS models, adapt them to changing environmental conditions based on the methods of artificial intelligence, optimization, fuzzy logic, machine learning. Thus, the use of simulation and evolutionary modeling methods as a tool for choosing the optimal data protection solution is most preferable.

The development of tools for reading and recording settings and operating parameters of the entire IS and information security tools, in particular, creates the prerequisites for the development of security systems, the configuration of which would be optimized based on the processed information about events in the system in a dynamic mode.
As a result of the analysis of classical and heuristic optimization methods within the framework of the posed problem, it is proposed to use the genetic algorithm to find the optimal configuration of ISS [2], [3], [4], [5].

Genetic algorithms (GA) is a method for solving optimization problems, based on the biological principles of natural selection and evolution, simulating the “survival” of objects that are most adapted to environmental conditions (protection options, etc.).

In GA individuals included in the population are represented by chromosomes with sets of task parameters encoded in them, i.e. solutions that are otherwise called points in the search space. Chromosome is an object, a string, a sequence, an ordered sequence of genes. A gene (property, parameter, attribute, characteristic) is an atomic element of the genotype, in particular, the chromosome.

Let the configuration of the ISS (vector $h$, chromosome) store information about the inclusion or non-inclusion of a protective measure - settings of data protection tools (DPT). Vector $h$ has dimension $n$, $n$ is the number of protective measures. Their number can be different, possible options can be analyzed based on the classification of DPT. As a result, it is necessary to build a hierarchical tree of various settings for DPT that can be used in the system. The nodes of the second level of the tree are obtained on the basis of the classification of DPT in accordance with the requirements and regulatory guidelines. At the third level of the hierarchy are the most popular certified DPT. At the fourth level is a list of their settings, for example: control of the launch and activity of programs; registration of set events in the system; blocking the execution of malicious scripts on websites; device control / restriction of access to data storage and transmission devices; user identification and authentication; data encryption in cryptocontainers, etc. At the fifth level of the tree is specific DPT settings.

The parameters of protective measures are proposed to be encoded according to the scheme in Figure 1, for which, at first, it is necessary to determine the order of numbering settings.

| $h_j$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1     | 0 | 1 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2     | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 6 | 4 | 0 | 0 | 0 | 0 |
| 3     | 1 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 4 | 0 | 4 | 0 | 0 | 0 | 0 |

**Figure 1.** The example of the coding of the ISS configuration in the form of the chromosome

The numbers in the encoding presented are interpreted as follows:

- if the $i$-th setting in the $j$-th configuration is not activated or unavailable, the $i$-th component of the vector $h_j$ is 0;
- if the $i$-th setting is enabled in the $j$-th configuration, then the $i$-th component of the vector $h_j$ is equal to a certain number with which the number of DPT (or the combination of DPT) is encoded, where this setting is installed.

For example, let DPT (firewalls of various manufacturers, etc.) be used with the available settings (protection against DDoS attacks). Then the possible variants of DPT setting can be associated with the digital code like this:

- the setting is disabled -1;
- the setting is enabled on the firewall ViPNetCoordinator HW 1000 - 2;
- the setting is enabled on SecretNetStudio -3;
- the setting is enabled on the firewall DallasLock 8.0-K - 4.

To describe the relationships between many settings and many tools, as well as the relationships between settings and DPTs, we’ll use comparison matrices.
For example, let a matrix be given:

\[ A = [a_{ij}] , \]

where \( a_{ij} \) is the value reflecting the compatibility (incompatibility, conflict) of a pair of DPTs or their settings, \( i = 1, n \), \( j = 1, n \).

The values of \( a_{ij} \) can be not only binary. But it is possible to describe relations between settings to set matrices of pairwise comparisons (expertly or based on machine learning data), where the results of comparisons are determined in another way, for example:

\[
a_{ij} = \begin{cases} 
1, & \text{if } i - \text{th setting is compatible with the } j - \text{th one}, \\
0, & \text{if } i - \text{th setting is compatible with the } j - \text{th}, \\
0.3, & \text{if } i - \text{th setting is poorly compatible with the } j - \text{th}, \\
\ldots & 
\end{cases}
\]

Solution vectors (ISS configurations) should have the same dimensions for further unification of input data and their processing during genetic search of the “strongest” individuals.

2. The procedure of ISS optimal configuration search

The chromosome in GA will be associated with the previously defined vector \( h \), which is decision. To assess the suitability of the solution option, it is necessary to set the fitness function.

For the purpose of a more objective assessment of the effectiveness of the ISS configuration, it is possible to take into account not only the fact of the presence of the protective measure, or the recommended settings of the protective equipment, but other indicators, for example: C - price of the ISS configuration; V - weight coefficient of importance of the protected nodes (devices); S - average damage from downtime / data loss; T - idle time during failures of the device; K - the competence of specialists supporting the work of ISS; P is the probability of a successful attack to the protected node (its failure). In this way the fitness function \( F(h) \) is proposed to be calculated by the equation:

\[
F(h) = C + V \cdot T \cdot S \cdot K \cdot P
\]  

(1)

Then, the certain reference value of the fitness function \( F_E \) is set with a tolerance \( \beta \) taken into account in the optimization search for the best solution.

The fitness function can also take into account:

- the financial benefit of an attack to certain nodes of the organization’s network (priorities setting),
- costs, resources necessary for the development, modification and implementation of an attack scenario,
- damage, including: loss from lost profit, fines for failure to fulfill the terms of the contract for the continuous operation of information services.

Since attacks to an information system are often random in nature, it is not always possible to set the fitness function applicable for practical use, due to the immeasurability of some parameters [6], [7]. A possible solution for determining the values of the fitness function in this case is the use of simulation modeling, the results of which evaluate the parameters involved in the equation (1).

Then, let’s assume that the ISS is representable in the form of a queuing system, for which many threats \( U_i \) are defined.

ISS undergoes a certain set of failures or attacks \( A_i \): workstation overload, server overload (decreased performance), critical server load (denial of service), information distortion at the workstation, disconnection of some communication equipment (router, switch).

Many failures and attacks are characterized by the intensity vector \( \lambda = \{\lambda_1, \lambda_2, \ldots\} \), and the protection subsystems \( S_i \) by various operation characteristics \( \mu = \{\mu_1, \mu_2, \ldots\} \). Each threat can be associated with the following parameters: frequency of occurrence of the threat; probability of the threat; the coefficient
of destructiveness, expressing the degree of destructive effect of the threat on the asset(s); a set of assets or resources to which the threat is directed, etc.

It is possible to carry out modeling of various defense lines at the network level, user workstation, operation system, database management system, software, etc.

As a result, the simulation model was developed using GPSS World simulation software packages, which allows assessing the necessary system parameters for a given threat vector and taking into account environmental parameters. In the simulation model, several categories of DPTs were selected: firewall, intrusion detection system, antivirus protection tool, as well as various assets, such as servers.

The full procedure of searching and selection of an effective ISS configuration is presented in figure 2.

![Figure 2. The procedure of research of effective ISS configuration selection](image)

The work of the GA in the framework of this procedure, as well as the classical one, includes various operations on populations - selection, crossbreeding, mutation [8].

At the stage of formation of the initial population, the compatibility of the settings in each chromosome of the generated population is compared according to the matrix of the compatibility of the DPTs settings. The operation of chromosomes mutation is also performed when taking into account the DPTs settings compatibility matrix at a random location of the chromosome.

When using single-point crossing-over, one break point is selected inside the chromosome, in which two parent chromosomes are divided into two parts and exchanged between them. When crossing, it is impossible to “cut” the parent chromosomes in an arbitrary place, since this can lead to the fact that the new solution contains invalid DPT settings. All these restrictions were taken into account in the calculations.

As the criterion for stopping the configuration search algorithm, not only the reference value of the fitness function but also the certain number of iterations, after which further search is impractical, can be set.

In the course of work, by means of the created simulation model, the series of experiments was carried out using 9 vendors of DPTs according to the allocated 10 types of tools and 33 possible settings, allowing to identify the best of the configurations under the given restrictions. As a result of the experiment, the significant improvement in the suitability of the configurations compared with the reference value was revealed already at the second stage.
3. Conclusion

The advantages of the proposed approach to the development of an optimal (from the point of view of efficiency) configuration of ISS are following: automation of solving everyday tasks of information security specialists, reducing the probability of human factor errors, solving the problem of lack of qualified personnel.

During the work, the analysis of the use of evolutionary algorithms in information security was carried out, the structure and principles of heuristic optimization algorithms were studied, the classical GA was adapted to solve the problem of searching and choosing the effective configuration of ISS.

References

[1] Davidyuk N V, Davidyuk V V 2019 Determination of the required degree of data protection in automated control systems IOP Conference Series: Materials Science and Engineering 537(5) 052026

[2] Shreider M Iu, Borovskii A S, Tarasov A D 2017 Design of physical protection systems using a genetic algorithm J.Bulletin of Eurasian Sciences 4 (41)

[3] Davidyuk N V 2010 Development of decision support system for physical security of objects:dissertation for the degree of candidate of technical sciences (Astrakhan:Astrakhan State Technical University) p 209

[4] Davidyuk N V, Belov S V 2009 The procedure for the effective placement of detection tools at the facility using the genetic search method. J.Information and Security 12 4 559-68

[5] Vlasov A O 2014 Formation of a base of decisive rules for an attack detection system using a genetic algorithm.Information Space Security: Materials of the XII All-Russian Scientific and Practical Conf. of Students, Postgraduates and Young Scientist (Ekaterinburg) pp 126-33

[6] Kiselev D Ju, Kiselev JuV, Bibikov V V 2014 Simulation of information systems in the Arena package. (Samara: SamaraState Technical University) p 20

[7] Kurilov F M 2016 Modeling of information security systems. Application of graph theory. Technical Sciences: theory and practice: proceedings of the III International scientific conf. (Chita:ChitaState Technical University) pp 6-9

[8] Davidyuk N V, Belov S V 2010 Formation of the initial population in the genetic search procedure for the option of effective location of the detection means at the object of protection. J. Vestnik ASTU. Series: control, computer engineering and informatics, (Astrakhan: Astrakhan State Technical University) 1 114-8