Selection of a Rational Composition of İnformation Protection Means Using a Genetic Algorithm

V. Lakhno, B. Akhmetov, O. Smirnov, V. Chubaievskyi, K. Khorolska, and B. Bebeshko

Abstract This article describes a modified genetic algorithm (MGA) for solving a multicriteria optimization problem for the selection and optimization of the information security means (ISM) quantity for sets located on the nodes of the informatization objects' (OBI) distributed computing system (DCS). Corresponding computational experiments were carried out, during which it was shown that MGA is distinguished by a sufficiently high efficiency. The time spent on solving the problem of the options evaluation for selecting and optimizing the placement of DSS sets along the DCS nodes for OBI, when using MGA, is approximately 16–25.5 times less in comparison with the indicators of the branch-and-bound method. The proposed approach of the MGA usage for solving the above written problem is characteristically exhibited by its integrated approach. In contrast to similar studies devoted to this problem, which, as a rule, consider only some aspects of information security (e.g., assessing the risks for OBI, comparing different information security systems, building maps of cyberthreats, etc.), the approach we are extending makes it possible to combine all areas of ISM selection in the process of the OBI information security (IS) contours optimization. The DSS module for solving the problem of selecting and optimizing

V. Lakhno
National University of Life and Environmental Sciences of Ukraine, Kiev, Ukraine
e-mail: lva964@nubip.edu.ua

B. Akhmetov
Yessenov University, Aktau, Kazakhstan
e-mail: berik.akhmetov@yu.edu.kz

O. Smirnov
Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine

V. Chubaievskyi · K. Khorolska (✉) · B. Bebeshko
Kyiv National University of Trade and Economics, Kyiv, Ukraine
e-mail: k.khorolska@knute.edu.ca

V. Chubaievskyi
e-mail: chubaievskyi_vi@knute.edu.ua

B. Bebeshko
e-mail: b.bebeshko@knute.edu.ua

G. Rajakumar et al. (eds.), *Intelligent Communication Technologies and Virtual Mobile Networks*, Lecture Notes on Data Engineering and Communications Technologies 131,
https://doi.org/10.1007/978-981-19-1844-5_2

the number of information security systems for the sets located on the nodes of the informatization objects' DCS was described.

Keywords Optimization · Genetic algorithm · Information security means · Informatization object · Distributed computing system

1 Introduction

Correctly substantiated selection of means and methods for protecting information (hereinafter ISS) for objects of informatization (hereinafter OBI, for example a company, a large enterprise or a network of government agencies), which have a distributed architecture of a computer network containing many nodes (which store information resources important for business processes), is important for the successful functioning of the analyzed protection object. This problem is especially relevant in the context of increasingly complex scenarios for carrying out cyber-attacks on OBI, as well as for the constantly changing of the threats topography to their information security.

Note that the selection of compatible hardware and software information security tools becomes more complicated as the range of products of this type offered on the market grows.

But the variety of information security systems offered on the market is not yet the main difficulty in the selection process. The problem of the correct selection of an information security system for a specific OBI is complicated by a number of specific factors. In particular, not everyone in the companies and enterprises management fully understands what exactly benefits a competent investment into information security system promises and how it will contribute to the profitability growth of the company's business processes as a whole [1].

Doubtless, when considering such a range of problems, one should take into account the presence of two mutually exclusive tendencies. The first one—is the desire to acquire an information security system, which will fully implement the standard range of tasks for ensuring OBI information security. And, the second one—is the concern about lowering costs for information security and the desire to make these costs pay off as soon as possible.

Moreover, in addition to each node of the OBI distributed computing system (hereinafter DCS), it is necessary to determine as accurately as possible the number of required ISS. Once, the lack of such an information security system will lead to an increase in the risks of the protected information resources loss or distortion. Therefore, it will entail both financial and reputational losses for the company. Meanwhile, redundant information security systems lead to unjustified investments into the OBI information security system, once the cost of modern hardware and software systems is quite high [2–6].

All of the above-stated predetermine the relevance of the studies, the results of which are presented in this article.

Research is focused, first of all, on the search for an effective algorithm that can help to quickly solve the problem of multicriteria optimization during the selection of the composition and optimization of the number of information security means (ISM). We believe that these information security means are collected in sets located on the nodes of the distributed computing system of the OBI. At the same time, information security means any hardware and technical means that can be used for this purpose (e.g., antivirus software, firewalls, access control means, intrusion detection systems, cryptographic protection means, etc.).

2 Literature Overview and Analysis

To date, the problem of justifying and optimizing the procedure for OBI information security system selection has received a lot of attention. The author in the scientific paper [7] proposes to optimize the ISS choice according to the criterion of "choice advantages."

The approaches outlined in studies [8, 9] propose the construction of an information security system based on the analytic hierarchy process (AHP) by Saati and the Pareto optimal set. Based on the results of such a calculation, the ISS security administrator can select solutions using expert knowledge.

Note that at the moment there are two most widespread approaches to determining the optimal option for company's information security development.

The first [9] of them is based on the compliance of the OBI information security level with the requirements of one of the IS standards verification.

The second approach is associated with the use of methods and models for the optimization of complex systems to determine the optimal option for an information security system development.

In scientific works [10–12], it is shown that the solution of these problems requires the inclusion of special optimization models in the procedures for the information security system construction. Such models make it possible to establish the relationship between the indicators of the final effect of the information security system functioning and the set of its parameters. It is this approach that can be used as the basis for information security optimization in the context of information and cyber confrontation. Thus, the problem of an optimal information security system development can be solved on the basis of a theoretical (systemic) approach. At the same time, the emphasis was made on a comprehensive consideration of the main factors affecting the effectiveness of the information security system for a particular OBI.

The expediency of using GA can be justified by the fact that the problem being solved belongs to multicriteria and multi-extreme problems [13, 14].

In studies [15, 16], the authors showed that GAs that can be used for solving multi-criteria optimization problems are variations of evolutionary search methods. However, the software implementation of the considered GA was not presented.

In scientific papers [17, 18], the authors analyzed the features of using GA in tasks related to the choice of equipment for OBI information protection systems. However, the solutions proposed in these papers are, in fact, a combination of standard greedy and GA.

Many researchers [12, 16, 18, 19] note that it is rather difficult to algorithmize the efficiency of ISS selection for OBI distributed computing systems (DCS). This is related to the description of the goal function. Such a function must be multi-parameter. It is due to the fact that it is influenced by a range of factors.

All of the above-written determines the relevance of our research.

3 The Objectives and Aim of the Research

The purpose of the research is to develop a modified genetic algorithm for solving a multi-criteria optimization problem for the selection and optimization of the number of information security tools for sets located on the nodes of the informatization objects' distributed computing system.

To achieve the goal of the research, the following tasks should be solved:

1. Modification of the classical genetic algorithm by applying the gray code to encode the number of DSS in the set and their position for the corresponding OBI IS circuit;
2. Software implementation and testing of the decision support system (based on MGA) module for the problem under consideration and its testing in comparison with classical methods for solving multicriteria optimization problems.

4 Models and Methods

4.1 Formulation of the Problem

It is required to find a solution to the optimization problem for the distribution of the information security system among the DCS nodes (see Fig. 1) while maximizing the information security metrics of the protected object (OBI) and minimizing the costs of their deployment and operation. At the same time, it is necessary to take into account the possibilities for the dynamic redistribution of the information security system among the DCS nodes on which the most valuable information resources are currently stored and at which the attack vector is aimed.

It is proposed to solve the process of optimizing the selection of information security systems for DCS nodes using a genetic algorithm (hereinafter GA). Once this class of problems belongs to multi-criteria and multi-extreme, the use of GA can provide a combination of acceptable quality of the solution and its efficiency.

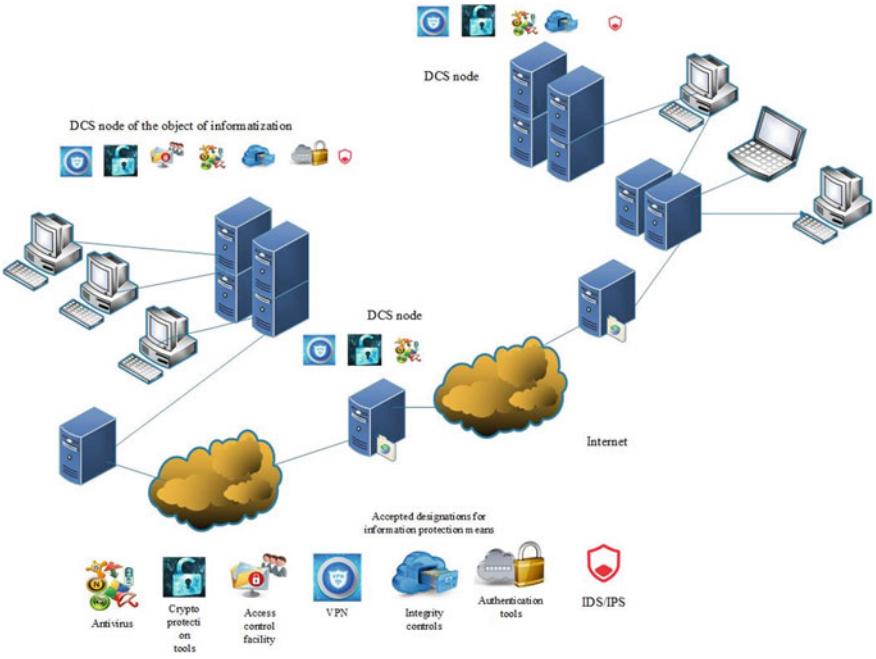

Fig. 1 Schematic Diagram

GA should combine two basic approaches for solving such problems. These are brute-force search and gradient search, respectively.

Note that it is rather difficult to set the goal function in the process of information security system selection by the nodes of the OBI DCS. It is due to the fact that such a goal function must be multivariable. It should be taken into account that stochastic factors also affect the functioning of the OBI information security circuits. After all, it is never known how the scenario of an attack on information resources located on a specific node of the OBI DCS will be implemented. And, such an attack will be launched.

Therefore, in the course of evaluating various options for equipping the circuits and nodes of the DCS with protection means, it is more expedient to initially perform mathematical modeling for each of the options used by the DSS for the corresponding node and OBI circuit. Only then should you consider how the chosen information security system will affect the OBI information security metrics.

4.2 Solution Example

The implementation of such an approach based on GA is described in this section. For the use of GA, data is required on the ISM sets for the DCS nodes. Moreover, this data must be presented in binary format. We believe that the number of information security systems of a certain type (antiviruses, firewalls, means of cryptographic protection, authentication, intrusion detection, etc.) can be encoded as a certain binary code with a certain number, see Fig. 2.

As an example, let us analyze any three adjacent rows in Table 1.

For example, consider rows 4 through 6 that are highlighted in the table with a colored fill. We will assume that, for example, row 4 (highlighted in blue), which assumes the use of five (5) ISM at the DCS node, is the best solution to the optimization problem. Such a solution is favorable for both versions of the code representation, both binary and gray code. Indeed, it is possible to implement a single operation which consists in replacing the fragment 0 by 1 in the penultimate bit. Similarly, one can illustrate the replacement of a single fragment for the next 5 lines.

The situation is more interesting for the sixth (6) line. To obtain the gray code, it is necessary to perform replacements in both the last and the penultimate bits. Accordingly, it is necessary to replace 1 with 0 in the penultimate right-hand bit and 0 with 1 in the last bit. Thus, the advantage of using gray codes in this problem will

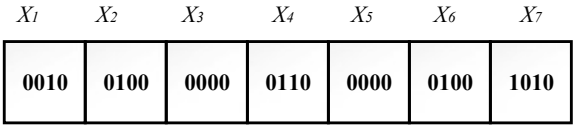

Fig. 2 Scheme of a conditional distributed computer network of an informatization object with the placement of information security tools sets on its nodes

Table 1 Example of coding by numbers and number of ISM for a DCS node

Number	The number of ISM for the considered DCS node	Binary code for the number	Gray Code
0	1	0000	0000
1	2	0001	0001
2	3	0010	0011
3	4	0011	0010
4	5	0100	0110
5	6	0101	0111
6	7	0110	0101
7	8	0111	0100
8	9	1000	1100
9	10	1001	1101

Table 2 Example of the formation of the initial GA population

ISM type	Placement node	Binary code	Gray code
Antivirus software	Servers (Ser), workstations (WSt), mobile clients (MobK), etc.	0000	0000
Firewalls	Ser, WSt, MobK, switches (Swit), routers (Rout)	0001	0001
Sandboxes, code analysis tools	Ser, WSt, MobK	0010	0011
VPN	Ser, WSt, MobK, Swit	0011	0010
Tools			
Cryptographic protection	Ser, WSt, MobK, Swit	0100	0110
User authentication and identification	Ser, WSt, MobK, Swit	0101	0111
Access control	Ser, third party hardware	0110	0101
Auditing and logging	Ser, WSt,	0111	0100
Code analysis	WSt, WSt, MobK	1000	1100
Systems			
Intrusion detection	Ser, Rout, Swit	1001	1101
Data backup	Ser	1010	1111

be that if the numbers differ from one another by only one (1), then their binary codes will differ only by one bit.

In the proposed GA, in the course of coding, numbers were first replaced with a binary code, which denote the number of ISM for the DCS node. And then, at the next stage, the resulting binary value was translated into the gray code.

Note that the type of ISM at the DCS node is adopted in accordance with Table 2. Once it is not advisable to use more than eight types of ISM on a typical DCS node (if it is not a critical computer system), then it is possible to limit oneself not to a 4-bit, but to a 3-bit coding.

A feature of using GA for solving the above problem is an integrated approach. In contrast to similar studies devoted to such issue, which, as a rule, consider only some aspects of information security (e.g., assessing the risks for OBI, comparing different information security systems, building maps of cyber threats, etc.), the approach we are extending makes it possible to combine all areas of information security systems selection during the process of the contours of information security optimization.

It is convenient to present the initial data for modeling using GA in such a tabular format, see Table 2.

The code for the initial sets of information security systems is generated randomly, for example due to the fact that we will consistently fill the bits in binary format. As described above, it is also quite convenient to use gray code for encoding.

The procedure for GA usage is presented in the form of an algorithm block diagram shown in Fig. 3.

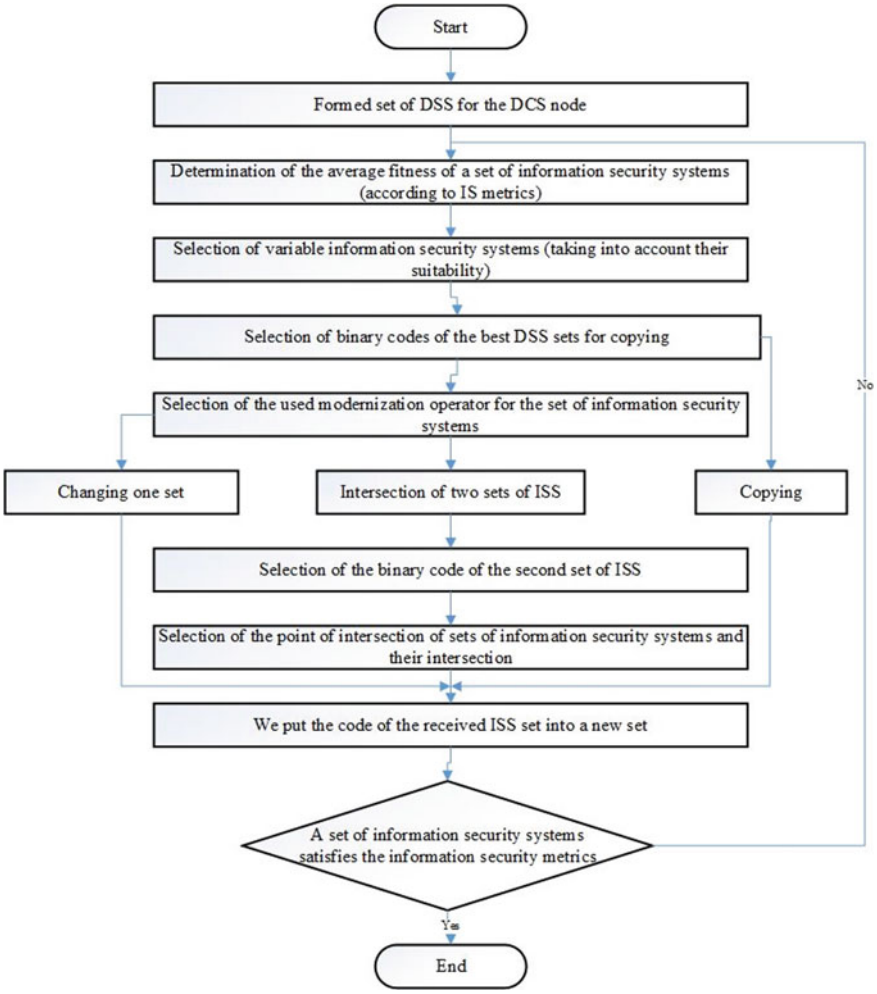

Fig. 3 Algorithm block diagram for MGA

Having decoded the values for the corresponding ISM in a set from binary to real, using Tables 1 and 2, it is possible to find the number of the minimum required ISM in each set. Further, by modeling the different-variant relations of the ISM in the set, we are able to find the fitness function of each solution. Then, depending on the value of this fitness function, we can arrange the sets of the corresponding codes. In particular, being guided by the average value of the fitness function, we can determine the probabilistic indicators of which of the ISM sets for the DCS node will have a fitness function with a large value and, accordingly, can further participate in the population change.

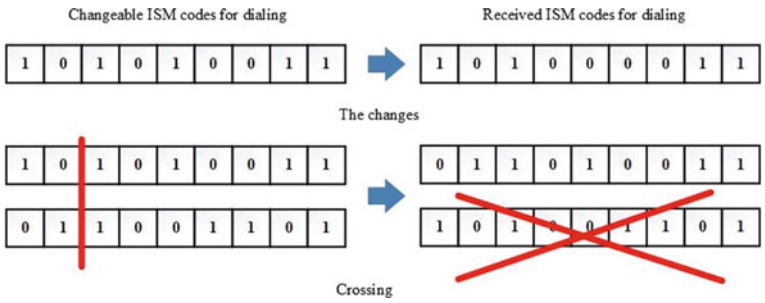

Fig. 4 Scheme of the MGA operator's work

Note that there are two options for each ISM set of the DCS node. The first one is to simply be copied into the next generation. The second is to be exposed to the modernization operator.

The work of the modernization operator is schematically shown in Fig. 4.

5 Software Implementation of the Model and Computational Experiment

The constant dynamic complication of decision-making procedures in the management of OBI information security necessitates the involvement of external funds to support decision-making related to the selection of information security systems along the DCS contours. In poorly structured subject areas, where there is no possibility of obtaining sufficient deterministic information for decision-making, expert support for decision-making is the only way to improve their quality. Once we are talking mainly about solving problems for high managerial levels, the "price" of wrong decisions at the present time may be too high, and it is constantly growing. Therefore, a clear presentation and processing of the data assessed by an expert in decision-making processes is one of the priority areas of relevant scientific research. And besides, critical questions related to these tasks require urgent solutions. It should also provide the ability for the provision of experts with the opportunity to clarify and correct their own previously entered estimates, in the process of further use of the DSS, for example, to solve such highly technical problems as making a decision on the need to select adequate protective measures and means to ensure the information security of OBI. In fact, it is all about the need to create a new type of DSS that could adapt to the level of the experts competence on a specific issue in the subject area.

Figure 5 represents a general view of the interface of the software implementation of the module for solving the problem on the basis of GA for the selection and optimization of the placement of ISM on the DCS nodes.

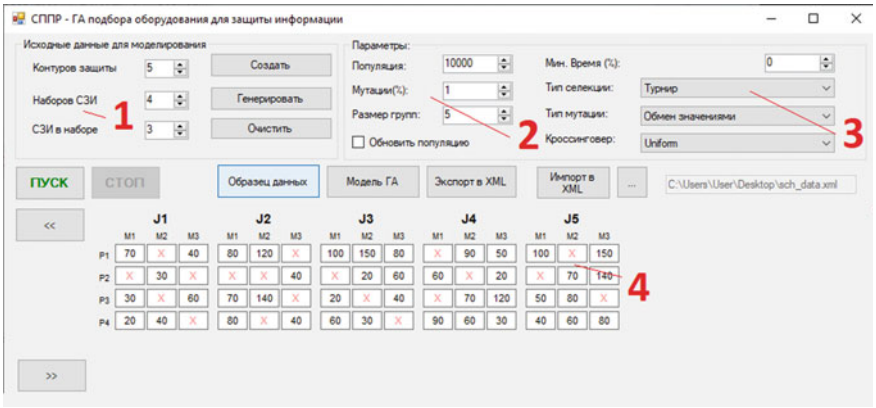

Fig. 5 General view of the interface of the developed decision support system for optimizing the selection and placement of information security systems on the DCS nodes

The numbers indicate the blocks of the window interface intended for the implementation of the following functions:

1. Initial data input unit (number of OBI IS contours, number of ISM sets, number of ISM in a set);
2. GA parameters (number of populations, percentage of mutations, etc.);
3. Additional GA parameters (type of selection, type of mutation, type of crossing over);
4. Table with set options (coded values) for OBI IS contours.

To verify the MGA adequacy, designed to solve the problem of selecting and optimizing the placement of ISM at the DCS nodes, and the DSS module described above in this paragraph of the article, the corresponding computational experiments were carried out, see Figs. 6 and 7.

Computational experiments were carried out for randomly generated variants of ISM sets.

To evaluate the proposed algorithm, test sets from 5 to 150 items were formed (ISM—information security means, ranging from antivirus software to intrusion detection systems) in the set. Was carried out 5 series of 50 experiments in a series. A total of 250 computational experiments were performed on a PC with an Intel i7 9750H processor (2.6–4.5 GHz).

Similar test suites were used for two other algorithms with which comparisons were made in terms of operating speed—selection and optimization of the placement of the ISM of the circuit IS OBI based on the branch-and-bound method [20, 21] and the "greedy algorithm" [22, 23].

To compare the performance of the MGA, three algorithms were chosen:

1. MGA: This algorithm was used as a basic one in the above-described DSS module for solving problems related to the selection and optimization of the placement of information security tools along the OBI contours.

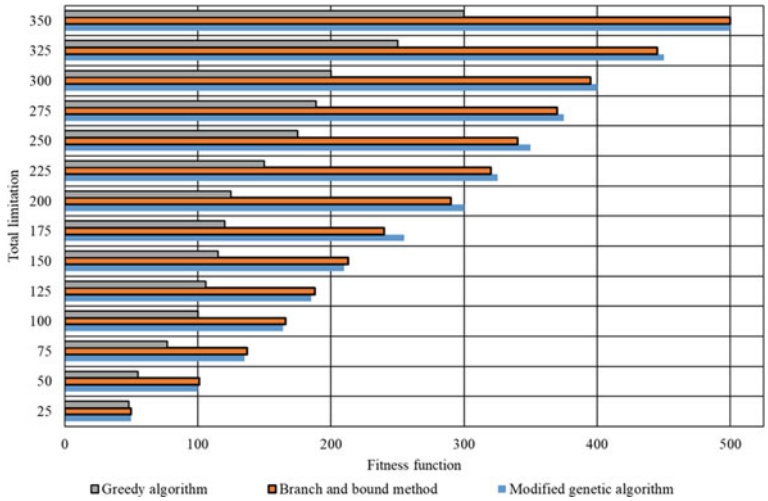

Fig. 6 Results of computational experiments comparing the effectiveness of algorithms used in the DSS module for the selection and optimization of the placement of information security tools along the information security contours of the informatization object

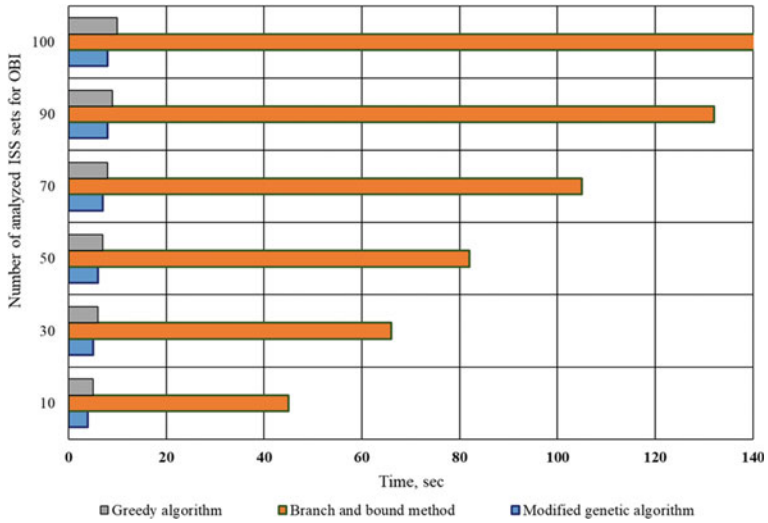

Fig. 7 Results of computational experiments comparing the running time of algorithms used to solve the problem

2. Selection and optimization of the ISM placement along the contours of the IS OBI based on the branch-and-bound method [20, 21];
3. "Greedy algorithm" [22, 23].

Obviously, by blocking poorly adapted ISM sets for the nodes of the DCS of the analyzed OBI, it is possible to increase the averaged fitness over the ISM set.

The work of the MGA will be terminated upon reaching the adaptation state by the ISM set. In such a state, the differences between the ISM sets in terms of fitness will be no more than 5%.

Intersection as a mechanism of variability will lose its attractiveness when crossing identical codes.

A simple change will facilitate the modification of the resulting ISM sets for the DCS node. It can be achieved by testing new points in the search space.

6 Discussion of the Results of the Computational Experiment

The results of computational experiments showed the following, see Figs. 6 and 7:

1. The branch-and-bound method and MGA demonstrate approximately the same efficiency in the course of solving the considered multicriteria optimization problem.
2. The maximum error was about 3.2–3.4%.
3. MGA is distinguished by a fairly high efficiency, as well as speed;
4. It was found that the time spent on solving the problem of evaluating the options for selecting and optimizing the placement of ISM sets on the OBI DCS nodes, when using MGA, is approximately 16–25.5 times less in comparison with the indicators of the branch-and-bound method. This circumstance allows in the future, while finalizing the DSS, to opt for this particular algorithm.

Certain disadvantages of using MGA include the fact that not all possible algorithms for solving the problem have been analyzed. In particular, the option of solving the specified problem by using other evolutionary algorithms was not considered. Also, a certain drawback of the study is the fact that the interface of the DSS module has not yet been localized for the English language.

7 Conclusions

A modified GA has been developed for solving a multi-criteria optimization problem for the selection and optimization of the number of information security tools for sets located on the nodes of a distributed computing system of an informatization object.

In the course of computational experiments, it was shown that the MGA is distinguished by a sufficiently high efficiency, as well as speed. The time spent on the solution of the problem of evaluating the options for selecting and optimizing the placement of ISM sets on the OBI DCS nodes, while using MGA, is approximately 16–25.5 times less in comparison with the indicators of the branch-and-bound method.

It is shown that the implementation of the modified GA allows to speed up the search for optimal options for placing information means for OBI. In addition, it is possible to solve the problem of redistributing resources in conditions of their limitedness. This advantage makes it possible not to perform a quick enumeration of various options for hardware and software information security means and their combinations for OBI, but also to subsequently combine the proposed MGA with the existing models and algorithms for optimizing the composition of OBI cybersecurity circuits. Potentially, such a combination of algorithms will make it possible to quickly rebuild the OBI defense in a dynamic confrontation with the attacking side.

The software implementation of the DSS module for solving the problem of selecting and optimizing the number of information security tools for the sets placed on the nodes of the DCS of the informatization object was developed.

Acknowledgements The research and article were carried out within the framework of the grant of the Republic of Kazakhstan registration number AP08855887-OT-20 "Development of an intelligent decision support system in the process of investing in cybersecurity systems."

References

1. Milov O, Yevseiev S, Aleksiyev V (2018) Development of structural models of stability of investment projects in cyber security. Ukrainian Sci J Inf Secur 24(3):181–194
2. Vijayakumar T (2019) Comparative study of capsule neural network in various applications. J Artif Intell 1(01):19–27
3. Pasumponpandian A (2020) Development of secure cloud based storage using the elgamal hyper elliptic curve cryptography with fuzzy logic based integer selection. J Soft Comput Paradigm 2(1):24–35
4. Samuel Manoharan J (2020) Population based metaheuristics algorithm for performance improvement of feed forward Neural Network. J Soft Comput Paradigm 2(1):36–46
5. Khorolska K, Lazorenko V, Bebeshko B, Desiatko A, Kharchenko O, Yaremych V (2022) Usage of clustering in decision support system. In: Intelligent sustainable systems. Lecture notes in networks and systems, vol 213. Springer, Singapore. https://doi.org/10.1007/978-981-16-2422-3_49
6. Bebeshko B, Khorolska K, Kotenko N, Kharchenko O, Zhyrova T (2021) Use of neural networks for predicting cyberattacks. Paper presented at the CEUR workshop proceedings, vol 2923, pp 213–223
7. Jerman-Blažič B (2008) An economic modelling approach to information security risk management. Int J Inf Manage 28(5):413–422
8. Smojver S (2011) Selection of information security risk management method using analytic hierarchy process (AHP). In: Central European conference on ınformation and ıntelligent systems. Faculty of Organization and Informatics Varazdin, p 119
9. Zhurin SI (2015) Comprehensiveness of response to internal cyber-threat and selection of methods to identify the insider. J ICT Res Appl 8(3):251–269

10. Trunova E, Voitsekhovska M (2019) The model of ınformation security culture level estimation of organization. In: Mathematical modeling and simulation of systems: Selected papers of 14th International scientific-practical conference vol. 1019, MODS, 2019 June 24–26, Chernihiv, Ukraine. Springer, p 249
11. Sushko OP (2018) Information security of power enterprises of North-Arctic region. J Phys: Conf Ser 1015(4):042058
12. Akhmetov BS, Akhmetov BB, Lakhno VA, Malyukov VP (2019) Adaptive model of mutual financial investment procedure control in cybersecurity systems of situational transport centers. In: News of the National Academy of Sciences of the Republic of Kazakhstan, Series of geology and technical sciences, vol 3(435), pp 159–172
13. Chiba Z, Abghour N, Moussaid K, El Omri A, Rida M (2019) New anomaly network ıntrusion detection system in cloud environment based on optimized back propagation neural network using ımproved genetic algorithm. Int J Commun Netw Inf Secur 11(1):61–84
14. Nozaki Y, Yoshikawa M (2019) Security evaluation of ring oscillator puf against genetic algorithm based modeling attack. In: International conference on ınnovative mobile and ınternet services in ubiquitous computing. Springer, Cham, pp 338–347
15. Dwivedi S, Vardhan M, Tripathi S (2020) Incorporating evolutionary computation for securing wireless network against cyberthreats. J Supercomput 1–38
16. Zhang F, Kodituwakku HADE, Hines JW, Coble J (2019) Multilayer data-driven cyber-attack detection system for industrial control systems based on network, system, and process data. IEEE Trans Ind Inf 15(7):4362–4369
17. Baroudi U, Bin-Yahya M, Alshammari M, Yaqoub U (2019) Ticket-based QoS routing optimization using genetic algorithm for WSN applications in smart grid. J Ambient Intell Humaniz Comput 10(4):1325–1338
18. Llansó T, McNeil M, Noteboom C (2019) Multi-criteria selection of capability-based cybersecurity solutions. In: Proceedings of the 52nd Hawaii International conference on system sciences, pp 7322–7330
19. Lakhno V, Akhmetov B, Adilzhanova S, Blozva A, Svitlana R, Dmytro R (2020) The use of a genetic algorithm in the problem of distribution of ınformation security organizational and financial resources ATIT 2020. In: Proceedings: 2020 2nd IEEE International conference on advanced trends in ınformation theory, No 9349310, pp 251–254
20. Lawler EL, Wood DE (1966) Branch-and-bound methods: a survey. Oper Res 14(4):699–719
21. Jabr RA (2013) Optimization of AC transmission system planning. IEEE Trans Power Syst 28(3):2779–2787
22. Tran V-K, Zhang H-S (2018) Optimal PMU placement using modified greedy algorithm. J Control Autom Electr Syst 29(1):99–109
23. Chen K, Song MX, He ZY, Zhang X (2013) Wind turbine positioning optimization of wind farm using greedy algorithm. J Renew Sustain Energy 5(2):023128