DOI: 10.37943/AITU.2020.99.72.006

DOI: 10.37943/AITU.2020.99.72.006

O. Laptiev
Doctor of Technical Sciences, Professor Department of Information and Cybersecurity Systems
alaptev64@ukr.net, orcid.org/0000-0002-4194-402X
State University of Telecommunications, Ukraine

V. Savchenko
Doctor of Technical Sciences, Professor, Director of Cybersecurity Institute
savitan@ukr.net, orcid.org/0000-0002-3014-131X
State University of Telecommunication, Ukraine

Y. Kravchenko
Doctor of technical science, Professor, Head of the Department of Networking and Internet Technologies
kr34@ukr.net, orcid.org/0000-0002-0281-4396
Taras Shevchenko National University of Kyiv, Ukraine

O. Barabash
Doctor of Technical Sciences, Professor, Professor of the Department of Automation of Designing of Energy Processes and Systems
bar64@ukr.net, orcid.org/0000- 0003-1715-0761
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine

IMPROVING THE METHOD OF SEARCHING DIGITAL ILLEGAL MEANS OBTAINING INFORMATION BASED ON CLUSTER ANALYSIS

Abstract: In the article the possibilities of the multipositional technology of searching digital insertion devices are investigated based on clustering. Existing means of detecting radiation of digital illegal means obtaining information reception show that they are ineffective on the background of legal signals of a multiagent medium.

The constant improvement of digital illegal means obtaining information, masking their work under the signals of legal transmitters require the search for new approaches to the recognition and localization of these means. Prospects for the development of search technology today are associated with the creation of multi-position permanent detection and localization systems. However, the detection problem requires the recognition of harmful radiation on the set of statistics of signal parameters in the air by solving the problem of clustering.

The disadvantages of most classical clustering methods are the need for prior knowledge of the possible number of clusters and a sufficiently high interactiveness, which complicates their practical application, especially in real-time. At the same time, intelligent multi-agent methods are free from these shortcomings, although their application remains quite complicated.

The problem of recognizing the harmful signal against the background of similar legal signals is possible by using the method of a bee colony with direct communication between agents. In this case, the agents are individual elements of the multi-machine complex, which scan the ether at different points in space, then exchanging results with other agents, and, finally, come to a common conclusion about the nature of the signal.

Full-scale studies have been carried out that confirm the reliability of clusterization by 6 … 12% compared to the classic k-medium method.

Keywords: information protection, inbound device, multiagent system, clusterization.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031
ISSN (E): 2707-904X

Introduction

The development of modern hidden means of obtaining information requires constant improvement of methods and means of their detection. If a few years ago GSM-based devices were considered exotic, now their number and range are expanding almost daily on the black market. Identifying such systems is becoming an increasingly difficult task, as the methods and modes of their operation are also becoming more complicated. Radio availability for communication and data transmission is constantly increasing. Now almost the entire available radio frequency spectrum is used for the operation of various radio transmitters. It is possible to give an example of a typical institution where the inspection is carried out. Dozens of computers, DECT radio telephones, mobile phones of various standards (CDMA-2000, GSM-900/1800, 3G (UMTS), 4G (WiMax)), mobile communication amplifiers (in some buildings there are amplifiers of all standards), legal radio microphones, wireless headsets, Wi-Fi devices, various electronic readers of access control and management systems, wireless and wired security devices (which often have levels of spurious radiation, commensurate with the radiation of radio bugs), etc.

Analysis of literature data and setting research objectives

A wide range of radio scanning tools is used to detect hidden means of obtaining information (radio bugs) [1-2]. In particular, monitoring systems capable of scanning and storing panoramas of signal spectra are popular and inexpensive [3-4]. At the same time, such systems usually do not allow to solve the problem of analysis of digital legal communication channels due to the unsatisfactory quality of the radio path.

More promising are spectrum analyzers and measuring receivers such as Rohde & Schwarz [5]. They can solve the problem of finding means of hidden capture of information solely from the analysis of the spectrum of the radio. However, they are not able to analyze digital signals and perform the task of localizing the means of hidden recording of information.

Vector analyzers are used to study and demodulate signals of high-speed radio interfaces and signals with spectrum expansion [6-9]. However, these devices are designed to work with receivers and spectrum analyzers, i.e. they cannot independently perform search and localization tasks.

Thus, from the analysis we can conclude that nowadays there are no devices (devices, software packages) for the analysis of digital packets for solving the problem of searching of radio control. The task of finding a modern embedded device in a complex radio environment remains unsolved.

The main factors influencing the process of detecting hidden means of obtaining information are the short-term operation of the digital radio bug, operation in the ranges of “legal” transmitters (including GSM, Wi-Fi), the use of low-power broadband signals that are difficult to detect. The general essence of detection and localization of the embedded device is to capture and accumulate statistical information about the parameters of signals at different points of the object of study, analysis and clustering of accumulated information with subsequent decision on the presence of “unauthorized” radiation. The task of the collective work of agents is to reach an agreed conclusion on the classification of certain samples of the recorded signal as legal or illegal. This decision can be made based on conditional clustering in the field of signal parameters [10, 12].

Statistics on the electronic environment are collected by appropriate scanners and usually include arrays of the same type of data suitable for processing by Data Mining. The parameters of the signals are registered at different points in space and may include the duration of the signal T_c, the dynamic range D_c, the width of the spectrum ΔF_c, and others. Theoretically, to detect the radiation of the means of hidden obtaining of information, it is necessary to solve the problem of clustering on the set of statistics of the parameters of the signals in the air.

DOI: 10.37943/AITU.2020.99.72.006

Nowadays the term “cluster” does not have a precise definition and, therefore, there are many methods of clustering. The idea of combining similar objects into groups (clusters) is common to all methods [13, 16].

Various cluster analysis approaches can be used to solve the above-described problem of clustering signal samples. However, the main disadvantage of most methods is the need for prior knowledge of the possible number of clusters which complicates the application of these methods to identify means of hidden obtaining of information. Clustering based on the use of multi-agent methods does not require significant computational complexity and prior knowledge of clusters, however, it has another disadvantage - the ability to find the optimal solution.

In this paper it is proposed to solve the problem of cluster analysis based on the method of bee colony. The bee colony method is a heuristic iterative method of random search based on modeling the movement of bees. In this case, the relationship between software agents that simulate the behavior of bees is direct. Thus, the bee colony method is a multi-agent optimization method with a direct relationship between agents [14, 15].

Presenting main material

The work of the multiagent optimization method with direct communication between agents to perform clustering can be represented as the following algorithm:

1. Forming a search space with m cells. Cells are formed by dividing the radio frequency range into separate clusters corresponding to a certain type of radio transmitter (legal and illegal). Since agents are physically located at different points in space, the overall picture they will perceive will be somewhat different.

Agents are arranged in free cells randomly:

$$x_i^k = rand\left(m\right), i = \overline{1, m}\,, \tag{1}$$

where: x_i^k – the i-th coordinate of the location of the k-th agent in the search space;

$rand\ (m)$ – a random number selected in the range from 1 to m.

2. $t: = 1$ – set the iteration counter.
3. $i: = 1$ – set the coordinate of the agent.
4. $j: = 1$ – set the agent number.

5. If the j-th agent has not selected an object that it distributes to other agents, the j-th agent checks the neighboring cells of the space for the selection of the object for its distribution. If the j-th agent has already selected an object to distribute, go to step 6.

The j-th agent selects the object for distribution as follows:

$$o^j = \begin{cases} rand(o^l), \text{if } \left|o^l\right| = 2; \\ \left|o^l_{worst}\right|, \text{if } \left|o^l\right| > 2; \\ o^l, \text{if } \left|o^l\right| = 1, \end{cases} \tag{2}$$

where $|o^l|$ is the number of objects in cell l;

o^l – the set of objects that are in cell l;

$rand\ (o^l)$ – randomly selected object from the set o^l;

o^l_{worst} – the object with the worst conditions which is selected as follows:

$o^l_{worst} = \arg\max\ [\mathrm{D_n}\ (C^l,\ o^l_r)]$ where $\mathrm{D_n}\ (C^l,\ o^l_r)$ is the normalized difference between the r-th object of cell l and the center of this cell C^l. The center is defined as the average of each characteristic of all objects in cell l. The normalized difference is determined based on the

Scientific Journal of Astana IT University
ISSN (P): 2707-9031
ISSN (E): 2707-904X

distance $D_n (C^l, o^l_r)$ which is calculated according to the entered metric, for example, as the Euclidean distance:

$$D_n\left(C^l, o^l_r\right) = \left(\sum_{q=1}^{N}\left[C^l(q) - o^l_r(q)\right]^2\right)^{\frac{1}{2}}, \quad (3)$$

where:

$C^l(q)$, $o^l_r(q)$ – the value of the q-th characteristic of the object o^l_r and the center C^l, respectively.

If the agent has selected an object to distribute, it goes to cell l and takes the selected object for further distribution.

If the agent after examining all neighboring cells, does not select an object to distribute, he randomly moves to one of the neighboring cells.

6. If the j-th agent owns an object that propagates in the workspace, he studies the neighboring cells and decides where to duplicate the object he is propagating. If the j-th agent does not have an object to distribute, then the agent randomly moves to one of the neighboring points and proceeds to step 7.

If the cell being viewed by the agent does not contain any objects at all, the agent does nothing and views the next cell. If the cell contains only one object, then the agent with a probability of 0.5 duplicates the object that distributes:

$$\text{if} \quad rand > 0.5, \text{then} \, o^l_r = \left\{o^l_r, o^j\right\}, \quad (4)$$

where: $rand$ – a random number in the range [0,1].

If the cell contains more than one object, the following cases are possible:

6.1. The cell in question contains an object for which the conditions are worse than for the object that distributes the object. In this case, the agent performs the following actions:

a) An object for distribution is an object for which the conditions of stay in this cell are worse: $o^j = o^l_{worst}$.

b) The agent moves to this cell. Go to step 7.

6.2. The conditions in the cell for the object being propagated are better than in its original cell. In this case, the agent performs the following actions:

a) The object is duplicated in this cell: $o^l = \{o^l\}$.

b) The agent moves to this cell. Go to step 7.

6.3. If neither of the previous two cases occurs, the agent considers the next adjacent cell.

If, after considering all the neighboring cells, the agent has not moved to any of them, the agent proceeds to the neighboring cell, selected at random, and the transition to step 7 is performed.

7. $j := j + 1$ – switch to another agent.

8. $i := i + 1$ – change the coordinate of the agent.

9. If $i < N_{move}$, then go to step 4, otherwise – go to step 10.

10. Exchange of information between agents.

As a result of the exchange of information, some agents must inform others about the cells in which the objects distributed by the respective agents have a significant influence. Thus, agents are divided into two groups: agents that report information about the cell to which the object is distributed and agents that analyze information that is reported by other agents.

DOI: 10.37943/AITU.2020.99.72.006

Agents that inform other agents about the cell to which the object is distributed include the following agents:

1. Agents whose object is not further than the center of the cell $\Delta\left(D_n\left(C^l, o_r^l\right) < \Delta\right)$ provided that there are 3 or more objects in the cell. It is chosen experimentally and depends on the specific practical task. Half of such agents are randomly selected and they inform other agents about the corresponding cell.

2. Agents whose object belongs to a cell in which the object is unique $| o^l | = 1$. Half of such agents are also randomly selected to inform about the objects being distributed.

All agents that are not included in the group of agents that perform information are automatically included in the group of agents that analyze information from other agents.

After dividing into groups for each agent that analyzes the information, the distance between the object that it distributes and between the objects that distribute agents belonging to the informing group of agents is calculated. If the minimum of the resulting differences is less than ΔD, then the object that distributes the informed agent is duplicated in the cell with the object that distributes the corresponding informing agent.

11. Natural selection. Since one object can be in several cells at the same time, you need to select and leave each object in only one cell. To do this, you must perform the selection procedure. It is proposed to perform a rigid selection, according to which for each object it is necessary to consider how close it is to each of the centers of the cells $D(C^l, o_r^l)$, weighted by the normalized distance for the current cell. Therefore, it is necessary to leave the object in the cell in which the given weighted distance is the smallest

$$q = \arg\min_l \left[D\left(C^l, o_r^l\right) \cdot \left(1 - D\left(C^l, o_r^l\right)\right) \right], \forall l = \overline{1, m}, \tag{5}$$

where q is the cell in which you want to leave the object o_r.

12. $t := t + 1$ – go to the next iteration.

13. If $t < tmax$, then perform the transition to step 3, otherwise – go to step 14.

14. Calculate the end centers of clusters. Each individual cell is considered a cluster. Based on the objects in the cells, calculate the centers of the clusters:

$$x_i^c = \frac{1}{N^c} \cdot \sum_{j \in O^c} x_i^j . \tag{6}$$

15. The end.

In developing this method, some features are considered that provide a match for the optimal solution:

1. Direct communication between agents is ensured by the exchange of information between agents through which some agents can obtain information about search areas in which they were not and from which they are far away. Thus, a better study of the search space is achieved which has a positive effect on the convergence to the optimal solution.

2. The introduction of the natural selection procedure allows to exclude objects from clusters for which the location conditions are unsatisfactory. To do this, a measure is introduced that characterizes the conditions of the object in the cluster, as the distance of the object to the center of the cluster, weighted by the normalized distance considering both the absolute value of the distance and the relative impact of the object as a whole.

3. To better study the search space, it is suggested to perform step 6 several times which will allow each agent to study the area in which it is in more detail.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031
ISSN (E): 2707-904X

Experimental verification of results.

The proposed algorithm was implemented when searching for means of hidden removal of information in a room located in a multi-storey office building. A multi-position scanning complex formed of 3 Delta 4G complexes (2 ODA-4 antennas with a pie chart each) under general control is deployed for search. Two Delta complexes (4 spaced antennas) were in the middle of the room, another complex (2 spaced antennas) outside. An emulator of the embedded device in the GSM range was installed in the room. The surroundings of the premises are standard offices of companies during working hours.

An example of the algorithm is shown in Fig. 1. As you can see, the algorithm correctly determines the cluster of devices for hidden obtaining of information. At the same time, there is a certain percentage of errors in the work.

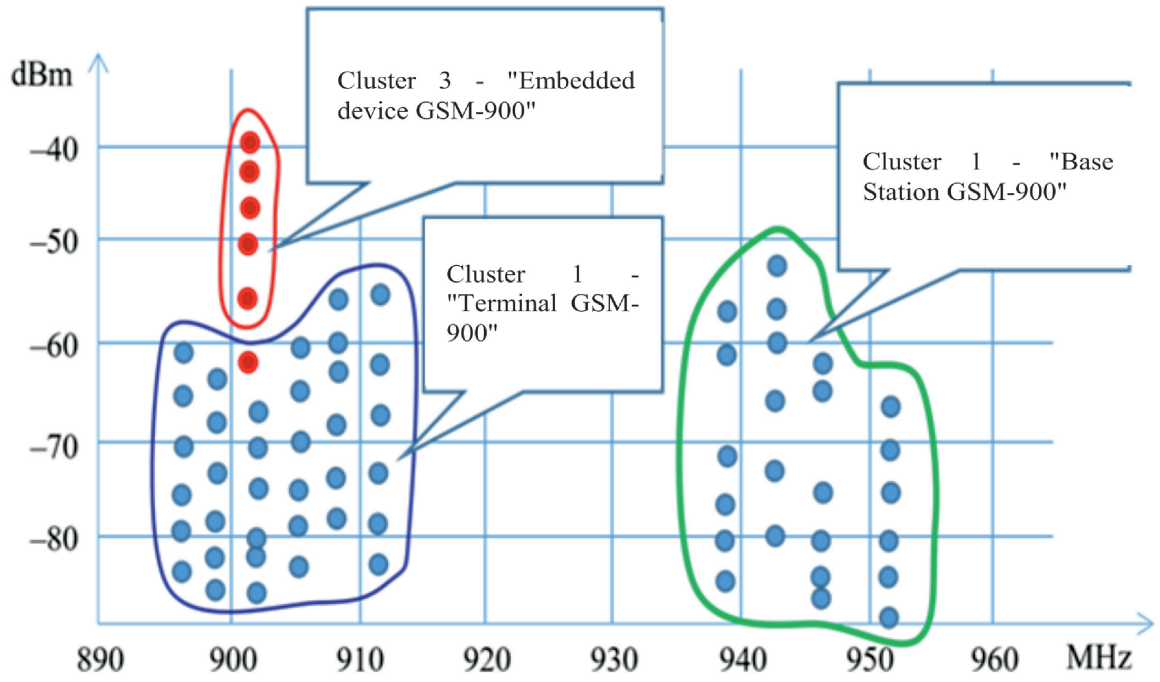

Fig. 1. An example of a multiagent clustering algorithm

To compare the accuracy of clustering, 150 experiments were performed during which the electronic situation in the middle and around the room was recorded by two parameters (operating frequency and signal strength) after which the results were processed by the known k-means method and the proposed multiagent method. In general, the method of k-means gives from 12 to 18% of errors in the classification of signal samples, while the multiagent method 6 - 8%. Thus, multi-agent clustering using direct communication between agents proves greater efficiency compared to classical methods. Another positive point is the lack of need for a priori assumptions about the number and nature of clusters.

Conclusion

The constant improvement of digital illegal means obtaining of information, masking their work under the signals of legal transmitters requires the search for new approaches to the recognition and localization of these means. Prospects for the development of search technology today are associated with the creation of multi-position permanent detection and localization systems. However, the detection problem requires the recognition of harmful

DOI: 10.37943/AITU.2020.99.72.006

radiation on the set of statistics of signal parameters in the air by solving the problem of clustering.

The disadvantages of most classical clustering methods are the need for prior knowledge of the possible number of clusters and a sufficiently high interactiveness which complicates their practical application, especially in real-time. At the same time, intelligent multi-agent methods are free from these shortcomings, although their application remains quite complicated.

The problem of recognizing the harmful signal against the background of similar legal signals is possible by using the method of a bee colony with direct communication between agents. In this case, the agents are individual elements of the multi-machine complex which scan the ether at different points in space, then exchanging results with other agents, and then come to a common conclusion about the nature of the signal.

The use of a distributed approach with joint processing of scan results based on information exchange between agents allows to increase the reliability of clustering 6 ... 12% compared to classical methods which is quite an acceptable result.

References

1. Yakovliev, A., and Lys, O. (2013). Special technical means of secret gathering of information. *Scientific works [Petro Mohyla Chornomorsk State University of the Kyiv-Mohyla Academy]*. Ser.: Computer technology. T. 229, No. 217, 39-43.
2. Digital direction finder "Rohde & Schwarz DDF0xE" / *Technology for special services, Bureau of Scientific and Technical Information*, founded in 1999. [Electronic resource] access mode: http://www.bnti.ru/des.asp?itm=4446&tbl =04.01.01.01.01.
3. Holembo, V., & Muliarevych, O. (2011). Modification of the goosebump colony method for the development of traveling salesman tasks by a group of autonomous agents. *Bulletin of the National University "Lvivska Politechnika"*. No. 717: Computer systems and systems, 24-30.
4. Danchuk, V. & Svatko, V. (2010). Optimization of the number of routs according to the graph in problems of logistic by the method of a modified goosebump algorithm. *Bulletin of the National Transport University*. No. 20, 109-114. - Access mode: http://nbuv.gov.ua/UJRN/Vntu_2010_20_21.
5. Khyzhniak, I., Makoveichuk, O., Khudov, R., Podlipaiev, V., Horban, H., & Khudov H. (2018). The method of swarm intelligence (piece bee colony (ABC)) of thematic segmentation of optical-electronic imaging. *Control systems, navigation and communication*, issue 2 (48), 91-96. DOI: 10.26906/SUNZ.2018.2.091
6. Davidović, T., Teodorović, D., & Selmic, M. (2015). Bee Colony Optimization – part I: The algorithm overview. *Yugoslav Journal of Operations Research*, 25, 33-56.
7. Wahid, A. Subhra, C., & Behera, D. (2015). Mohapatra Artificial Bee colony and its Application: An Overview. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 4(4), 1475–1480.
8. Korotin, S., Kravchenko, Y., Starkova, O., Herasymenko, K., & Mykolaichuk, R. (2019). Analytical determination of the parameters of the self-tuning circuit of the traffic control system on the limit of vibrational stability, *International Scientific-Practical Conference Problems of Infocommunications Science and Technology*, PIC S&T`– Proceedings, 471–476.
9. Rakushev, M., Permiakov, O., Tarasenko, S., Kovbasiuk, S., Kravchenko, Y., & Lavrinchuk, O. (2019). *Numerical Method of Integration on the Basis of Multidimensional Differential-Taylor Transformations*. In Proceedings of the IEEE International Scientific-Practical Conference Problems of Infocommunications Science and Technology, PIC S&T`2019. Proceedings, 675-678.
10. Kravchenko, Y., Leshchenko, O., Dakhno, N., Trush, O., & Makhovych O. *Evaluating the effectiveness of cloud services*. IEEE International Conference on Advanced Trends in Information Theory, ATIT`2019 – Proceedings, 120–124.
11. Savchenko, V., Ilin, O., Hnidenko, N., Tkachenko, O., Laptiev, O., & Lehominova, S. (2020). Detection of Slow DDoS Attacks based on User's Behavior Forecasting. *International Journal of Emerging Trends in Engineering Research (IJETER),* vol. 8. no. 5, – ISSN 2347–3983. 2019–2025.

Scientific Journal of Astana IT University
ISSN (P): 2707-9031
ISSN (E): 2707-904X

12. Savchenko, V., Syrotenko, A., Shchypanskyi, P., Matsko, O., & Laptiev, O. (2020). The Model of Localization Precision for Detection of Hidden Transmitters. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, no.4, ISSN: 2278–3075. 2114–2119.
13. Barabash, O., Laptiev, O., Tkachev, V., Maystrov, O., Krasikov, O., & Polovinkin, I. (2020). The Indirect method of obtaining Estimates of the Parameters of Radio Signals of hidden means of obtaining Information. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, Scopus, 8 (8), 4133–4139.
14. Stefurak, O., Tikhonov, Y., Laptiev, O., & Zozulya, S. (2020). Improving the stochastic model to identify threats of damage or unauthorized leakage. *Modern information protection: scientific and technical journal*. K.: DUT, no. 2 (42), 19-26.
15. Laptiev, O.A, Babenko, R.V, Pravdyvy, A.M, Zozulya, S.A, Stefurak, O.R. (2020). Improved methodology for selecting the sequence of priorities for servicing information flows. *Scientific and practical journal «Communication»*. K.: DUT, 4 (146), 45-49.
16. Oleksandr, L., Biehun, A., Hohoniants, S., Lisnevskyi, R., Pravdyvyi, A., & Lazarenko, S. (2020). Method of detecting signals of means of hidden obtaining of information on the basis of approximation of T-spectrum. The Intelligent Control System for infocommunication networks. *International Journal of Emerging Trends in Engineering Research (IJETER),* vol. 8, no. 10, 6835-6841.
17. Laptiev, O., Savchenko, V., Ablazov, I., Lisnevskyi, R., Kolos, O., & Hudyma, V. (2020). Method of detecting random signals based on determining the deviations of the main parameters of radio signals. *International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE)*, vol. 9, no. 5, 9204–9209.
18. Yevseiev, S., Korolyov, R., Tkachov, A., Laptiev, O., Opirskyy, I., & Soloviova, O. (2020). Modification of the algorithm (OFM) S-box, which provides increasing crypto resistance in the post-quantum period. *International Journal of Advanced Trends in Computer Science and Engineering (IJATCSE,)* vol. 9, no. 5, 8725-8729**.**
19. Laptev, A.A. Sachenko, V.A., Barabash, O.V. Sachenko, V.V., & Matsko A.I. (2019). The method of searching for digital means of illegal obtaining of information on the basis cluster analysis. *Magyar Tudományos Journal*. Budapest, Hungary, no. 31, 33–37
20. Laptev, A.A., Barabash, O.V., Savchenko, V.V., Savchenko, V.A., & Sobchuk, V.V. (2019). The method of searching for digital means of illegal reception of information in information systems in the working range of Wi-Fi. *International Journal of Advanced Research in Science, Engineering and Technology*. India, ISSN: 2350-0328, 6 (7), 10101–10105.
21. Hryshchuk, R., Korobiichuk, S., Ivanchenk, O., Roma, & A. Golishevsky. (2019). The Throughput of Technical Channels as an Indicator of Protection Discrete Sources from Information Leakage. *Computer Modeling and Intelligent Systems*, 2353, 523–532.
22. Korobiichuk, I., Ivanchenko, S., Havrylenko, O., Golishevsky, A., Hnatiuk, S., & Hryshchuk, R. (2019). *Protection of information from leakage by technical channels for sources with non range distribution of probability*. In CEUR Workshop Proceedings, 992-1003.
23. Hryshchuk, R., Korobiichuk, I., Horoshko, V., Khokhlacheva, Y. (2019). Microprocessor Means for Technical Diagnostics of Complex Systems. *Computer Modeling and Intelligent Systems*, vol. 2353, 1020–1029.