

<https://doi.org/10.31891/csit-2026-3-19>

Vadim PAYUK

Khmelnyskyi National University
<https://orcid.org/0000-0002-9969-8239>
e-mail: vadypaiuk@gmail.com

Antonina KASHTALIAN

Doctor of Technical Sciences, Professor
Department of Computer Engineering and
Information Systems
Khmelnyskyi National University
<https://orcid.org/0000-0002-4925-9713>
e-mail: yantonina@ukr.net

Received: 14/06/2026

Accepted: 17/08/2026

Published: 30/09/2026

© Copyright
2026 by the author(s)



This is an Open Access article distributed
under the terms of the [Creative Commons
CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

UDC 004.75:004.49:004.3

SELECTION OF DATA IN SYSTEMS AND TOOLS FOR COMPUTER ATTACK DEFENSE TAKING INTO ACCOUNT THE HISTORICAL ASPECT OF THEIR USE

The article considers the problem of data selection for systems and means of countering computer attacks in corporate networks in the context of constant growth of information volumes, high dynamism of information flows and limited computing resources. It is shown that the effectiveness of the functioning of modern systems for countering computer attacks largely depends not only on the quality of threat detection algorithms, but also on the timely provision of decision-making processes with relevant, relevant and most valuable data. It is substantiated that the use of the entire accumulated array of information leads to an increase in the time of its processing, an increase in information redundancy and a decrease in the efficiency of response to computer attacks, while an excessive reduction in data volumes can lead to the loss of important information necessary to identify complex or recurring threats.

A method of data selection for systems and means of countering computer attacks is proposed, which is based on the complex use of the depth of the analysis horizon, self-learning mechanisms and previous experience of the system functioning. A feature of the method is the integrated evaluation of data according to a set of criteria, which allows taking into account their current relevance, predicted usefulness in different time horizons, the results of previous use, and the ability to adaptively change the selection parameters in accordance with changes in the information environment. In contrast to existing approaches, the proposed method provides an intelligent choice of data, taking into account the relationship between the historical effectiveness of information, its prospective significance and the results of the system's self-learning.

To implement the proposed method, an architecture of the memory subsystem of systems and means of countering computer attacks has been developed, which provides accumulation, classification, optimization, modification and adaptive selection of data in accordance with the needs of means of different levels. The proposed architecture allows maintaining a balance between completeness, duration of storage and relevance of information, reducing information redundancy, reducing access time to critical data and increasing the efficiency of information support of decision-making processes.

The practical significance of the results obtained lies in the possibility of using the proposed method in the development of intelligent memory subsystems of systems and means of countering computer attacks in corporate networks, which provides an increase in the accuracy of data selection, acceleration of information analysis processes, improvement of the adaptability of systems to changes in operating conditions and increase in the efficiency of detection and neutralization of computer attacks.

Prospects for further research are the development of a mathematical model for adaptive adjustment of the weights of the data selection method, the improvement of self-learning mechanisms using modern methods of artificial intelligence, and the experimental assessment of the effectiveness of the proposed method in real corporate networks under various scenarios of computer attacks.

Tags: corporate networks, computer attacks, data optimization, historical data, data selection, information redundancy, self-learning, depth of horizon for data

Problem statement

The rapid growth in the number of computer attacks, the complication of their structure and the constant increase in the volume of information generated in corporate networks necessitate the improvement of systems and means of countering computer attacks. The efficiency of such systems [1, 2] is determined not only by the quality of algorithms for detecting and responding to threats, but also by the timely provision of them with the necessary data. Under modern conditions, a significant

part of information accumulates over a long time, has different levels of relevance and practical value, which complicates its use in decision-making. As a result, there is a contradiction between the need to use the most complete information to improve the quality of counteraction to computer attacks and the limited capabilities of systems to store, process and quickly use it.

Existing approaches to the organization of memory subsystems are mainly focused on the accumulation and storage of information or on the optimization of its structure. However, they do not sufficiently take into account the need for intelligent selection of exactly those data that best meet the current conditions of the system functioning and ensure maximum effectiveness in countering computer attacks. At the same time, the use of all accumulated information leads to an increase in its processing time, while excessive reduction of data volumes can lead to the loss of important information necessary to detect complex or repetitive attacks.

This problem becomes especially relevant in the conditions of the dynamic information environment of corporate networks [3, 4], where the usefulness of data varies depending on the time, nature of attacks, features of the system functioning and the accumulated experience of its operation. Under such conditions, the choice of data should be carried out not only according to the criteria of their current relevance, but also taking into account the prospects for the use of information, the results of its previous application and the ability of the system to adapt to new operating conditions.

To solve this problem, it is advisable to use an integrated approach that combines the depth of the analysis horizon, self-learning mechanisms and previous experience of the system. The depth of the horizon allows you to assess the potential usefulness of information in the short, medium and long term, previous experience ensures that the actual effectiveness of data use in such situations is taken into account, and self-learning creates the possibility of automatic adjustment of selection criteria in accordance with changes in the information environment. The combination of these components ensures the formation of an adaptive data selection mechanism capable of maintaining high efficiency of the functioning of systems and means of countering computer attacks in conditions of uncertainty and constant changes in information flows.

The implementation of this approach requires the creation of a memory subsystem architecture that ensures the accumulation, classification, optimization, modification and use of data of all storage levels, as well as implements the proposed method of data selection based on horizon depth, self-learning and previous experience. This allows you to ensure a balance between the completeness, duration of storage and relevance of information, reduce information redundancy, increase the accuracy of data selection and speed up decision-making when countering computer attacks.

The article is aimed at developing a method for selecting data for systems and means of countering computer attacks in corporate networks, which is based on the complex use of the depth of the analysis horizon, self-learning mechanisms and previous experience and is implemented in the architecture of the adaptive memory subsystem to increase the efficiency of information support for the processes of countering computer attacks.

Analysis of recent research and publications

Choosing the depth of historical data is one of the key factors in building effective time series prediction models. The paper [5] shows that there is a compromise between the duration of the training sample and the forecasting horizon. The use of long historical periods (up to one year or more) makes it possible to reduce the uncertainty of the forecast and increase its statistical stability, while short time horizons better reflect the current state of the system. To increase the accuracy of forecasting, it is proposed to use adaptive distribution of historical data by forecasting horizons, statistical assessment of their consistency, methods for selecting informative features (Lasso, Mutual Information, Principal Component Analysis) and clustering of multidimensional time series. Such approaches make it possible to form representative training samples and reduce the impact of excess information on forecasting results [5].

The papers [6, 7] consider the problem of forecasting cyber events based on time series in the context of a constant increase in the number and complexity of cyberattacks. The authors note that modern cyber defense systems must use significant amounts of structured and unstructured data to detect threats in a timely manner. Particular attention is paid to the use of short time horizons and sliding windows, which ensure that models adapt to changes in attack characteristics and reduce the impact of concept drift. At the same time, it is emphasized that the prediction of cyber events based on time series has not yet been sufficiently researched compared to traditional methods of detecting attacks [7].

The study [8] shows that the effectiveness of cyber threat forecasting largely depends on the correct choice of historical data. To ensure the statistical relevance of information, it is proposed to use sliding time windows lasting from six to fourteen months, which allows taking into account current trends in the development of threats and at the same time avoiding the accumulation of outdated information. This approach provides an increase in forecasting accuracy and improves the adaptability of cyber defense systems to new types of attacks [8].

The paper [9] considers the consequences of information security breaches, which lead to the loss of personal, financial and medical data. The authors emphasize that timely detection of abnormal activity is one of the main means of preventing such incidents. At the same time, it is noted that the effectiveness of modern intrusion detection systems directly depends on the availability of high-quality historical data used to train machine learning models and build profiles of normal network behavior [9].

The paper [10] provides an overview of modern tools for analyzing network traffic and methods for forming training datasets for intrusion detection systems. The authors show that the development of machine learning models is largely determined by the quality and representativeness of historical data. The study proposes a classification of tools for collecting network traffic, considers the process of building data sets and analyzes their impact on the effectiveness of intrusion detection systems. Particular attention is paid to the need to create modern data sets that reflect real scenarios of network attacks [10].

The paper [11] investigates the application of big data technologies for building scalable intrusion detection systems in real time. The proposed architecture is based on the use of Apache Spark Structured Streaming, the MLlib library and the Microsoft Azure cloud infrastructure. A decision tree algorithm is used to classify network traffic, and the system is evaluated based on the MAWILab dataset. The results obtained confirm the effectiveness of using distributed platforms for the prompt processing of large data flows and timely detection of cyberattacks [11].

The study [12] considers the problem of forming realistic streaming datasets for network intrusion detection systems. The authors note that existing reference datasets do not always contain enough informative features for use in real networks. To eliminate this drawback, various methods of feature selection were applied, which made it possible to determine the minimum set of NetFlow parameters sufficient for effective training of machine learning algorithms. The need to accumulate the marked data directly in real information systems is also emphasized [12].

The paper [13] investigates the impact of class imbalance on the effectiveness of machine learning models in network intrusion detection problems. It is shown that the uneven distribution of training examples leads to a decrease in the accuracy of classification and worsens the interpretation of the results of the models. The authors evaluate the effectiveness of modern methods of data balancing and demonstrate their positive impact on both the quality of classification and the explainability of neural network decisions [13].

The paper [14] proposes a multimodal approach to building an intrusion detection system for Internet of Things networks. Apache Spark technologies, learning transfer methods, and convolutional neural networks are used to process large volumes of network traffic. A feature of the proposed solution is a combination of text and graphic features of network traffic, which allows to increase the accuracy of classification of attacks on CIC-IoT 2022, CIC-IoT 2023, and Edge-IIoT datasets [14].

The study [15] proposed a network attack detection system that combines big data and deep learning technologies. The architecture of the system is based on the use of a hybrid CNN-LSTM model, and SMOTE and Tomek Links methods are used to eliminate class imbalances. Data processing is implemented by PySpark tools, which provides scalability and high performance. Experimental results confirm the achievement of an accuracy of more than 99 % for the problems of binary and multiclass classification of network attacks [15].

The paper [16] performs a comprehensive analysis of modern data sets for intrusion detection systems in cyber-physical industrial systems and the industrial Internet of Things. The authors quantified 23 data sets according to the criteria of documentation completeness, realism, reproducibility, attack diversity, and class balance. The results obtained indicate the presence of significant gaps in the presentation of modern attack scenarios and industrial protocols, which must be taken into account when selecting training data for building highly effective intrusion detection systems [16].

Thus, historical data significantly affect the effectiveness of spacecraft and runway countermeasures systems in corporate networks, so the development of a method for selecting such data is an urgent scientific task.

Presentation of the main material

Architecture of Memory Subsystem Systems and Means of Countering Spacecraft in Corporate Networks Based on the Use of Historical Data for Decision-Making

Ensuring a balance in the duration of storage and use of data of systems and means of countering spacecraft of all seven levels, as well as problems regarding methodological, technical and semantic aspects, in order to resolve the contradiction regarding the need to combine the requirements for the completeness, duration and relevance of stored data with the limitations of computing resources, their distribution and adaptability of analysis systems, we will develop an architecture of the memory subsystem of systems and means of countering spacecraft in corporate networks. It should be the basis for the creation of means of all seven levels or a smaller number of them, depending on the existing systems and means of countering spacecraft in corporate networks. Such tools will be subsystems of memory systems and means of countering spacecraft with a focus on choosing the depth of the retrospective of the data used to counteract spacecraft in corporate networks, accumulating data according to the developed classes of different levels, optimizing outdated data and modifying them for the synthesis of new data. These tools will perform data processing tasks and select those that would most accurately ensure the functioning of systems and means of countering spacecraft at all seven levels in corporate networks.

The place of the memory subsystem of systems and means of countering spacecraft in corporate networks in the general architecture together with systems and means of countering spacecraft and data storage, including those that are contained directly in the systems and facilities themselves, is shown in Fig. 1. Such systems and means of countering spacecraft also include deceptive systems.

Let's set the list of tasks for storing, selecting and optimizing historical data of systems and means of countering computer attacks on corporate networks as follows: M_z^D

$$M_z^D = \{m_{z,1}^D, m_{z,2}^D, \dots, m_{z,n_{M_z^D}}^D\}, \quad (1)$$

where $m_{z,q}^D$ is the element of the set, which reflects one of the tasks of the means of the memory subsystem;
; – the number of tasks of the memory subsystem. $M_z^D q = 1, 2, \dots, n_{M_z^D}$

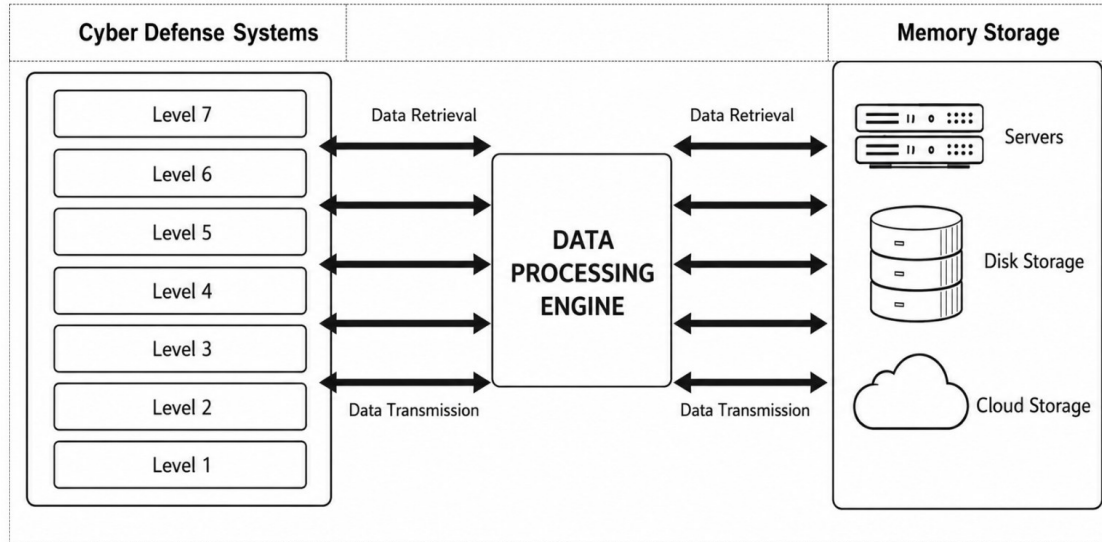


Fig. 1. The place of memory subsystem facilities in the general concept of means and systems for countering spacecraft and data storage

Memory subsystem tools can be separate independent objects in the context of their presentation in corporate networks, or they can be part of anti-spacecraft systems. In addition, part of the data will be stored not only separately, but also directly in the systems and anti-spacecraft tools themselves, so the memory subsystem tools can be partially or completely parts of them.

The main tasks of the memory subsystem are defined as follows:

- 1) – classification of the received data taking into account the data of deceptive systems; $m_{z,1}^D$
- 2) – storage of the received data in a distributed environment, taking into account their classification; $m_{z,2}^D$
- 3) – optimization of historical data in a distributed environment; $m_{z,3}^D$
- 4) – modification of data to ensure the adaptive functioning of systems and means of countering spacecraft; $m_{z,4}^D$
- 5) – evaluation of data according to various criteria; $m_{z,5}^D$
- 6) – data update; $m_{z,6}^D$
- 7) – selection of data for systems and means of countering spacecraft. $m_{z,7}^D$

The types of data processed by the memory subsystem tool are set as follows: M_d^D

$$M_d^D = \{m_{d,1}^D, m_{d,2}^D, \dots, m_{d,n_{M_d^D}}^D\}, \quad (2)$$

where $m_{d,q}^D$ is the element of the set, which reflects one data type; ; – the number of data types of the memory subsystem. $M_d^D q = 1, 2, \dots, n_{M_d^D}$

Data types will correspond to data classes. We divide the set M_d^D of data types into data classes, taking into account their purpose as follows:

- 1) – an element reflecting the data of systems and means of countering spacecraft that are available in them at the time of installation; $m_{d,1}^D$
- 2) – an element reflecting the accumulated data of systems and means of countering spacecraft during the time of their operation from the moment of installation; $m_{d,2}^D$
- 3) – an element reflecting the modified data of systems and means of countering spacecraft, which have been modified to ensure the property of adaptability in decision-making, i.e. they are artificial; $m_{d,3}^D$
- 4) – an element reflecting the updated data of systems and means of countering spacecraft that were received from outside corporate networks; $m_{d,4}^D$
- 5) – an element reflecting the raw data of systems and means of countering spacecraft and not assigned to any class for storage, but they were received in corporate networks; $m_{d,5}^D$
- 6) – an element reflecting false data of systems and means of countering spacecraft based on deceptive technologies that have been present in them since the beginning of their installation and operation; $m_{d,6}^D$
- 7) – an element reflecting modified data of systems and means of countering spacecraft based on deceptive technologies that were formed in them during their operation; $m_{d,7}^D$
- 8) – an element that reflects the data of systems and means of countering spacecraft based on deceptive

technologies that were obtained in them during their operation. $m_{d,8}^D$

According to this division (formula (2)) eight classes of data can be distinguished ($;$ – the number of tasks of the memory subsystem facilities) that should be processed by $K_{d,q}^D, q = 1, 2, \dots, n_{M_d^D} n_{M_d^D}$ the memory subsystem tool for their storage and optimization and selection of those that are necessary for the further functioning of the systems and means of countering spacecraft.

The main elements of the memory subsystem are set in a set M_{ez}^D as follows:

$$M_{ez}^D = \{m_{ez,1}^D, m_{ez,2}^D, \dots, m_{ez,n_{M_{ez}^D}}^D\}, \quad (3)$$

where $m_{ez,q}^D$ is the element of the set $;$ $;$ – the number of elements of the memory subsystem. $M_{ez,q} = 1, 2, \dots, n_{M_{ez}^D} n_{M_{ez}^D}$

Let's define the elements of the set M_{ez}^D as follows:

- 1) $m_{ez,1}^D$ – management of data on their placement in corporate networks;
- 2) $m_{ez,2}^D$ – search for data according to various criteria;
- 3) $m_{ez,3}^D$ – selection of data for systems and means of countering spacecraft;
- 4) $m_{ez,4}^D$ – modification of data for adaptive systems and means of countering spacecraft;
- 5) $m_{ez,5}^D$ – storage of data by classes in the distributed environment of corporate networks;
- 6) $m_{ez,6}^D$ – optimization of stored data, taking into account their relevance.

The management of data on their placement in corporate networks is highlighted as a separate element, since it is necessary to coordinate and carry out the general management of memory resources, free and filled places in memory storages, moving individual data classes to other places depending on the current state of memory and its volumes, accounting for the placement of sets and data classes, etc. These processes should be supported automatically in the memory subsystem, since the efficiency of the functioning of the means and systems of countering spacecraft of all seven levels will depend on them.

Since the duration of spacecraft is different, the means and systems of counteracting them can react in time depending on the levels to which they belong. But taking into account the specifics of the tools and systems themselves and their tasks, finding the necessary data should be ahead of the actions of attackers during their spacecraft, so that the relevance of data and, accordingly, reactions are not lost. Therefore, to ensure the search for the necessary data, it is necessary to implement various algorithms and methods in order to meet the requirements of the criteria for completeness, relevance and relevance of data.

For the means and systems of countering spacecraft, the choice of data for their decision-making is a difficult task, since the relevance of data in the context of the depth of the retrospective of their use remains a task that requires the provision of the necessary knowledge of the means of the memory subsystem, such tools must have the ability to self-learn. Since such selection of data for previous time periods of operation of systems and means of countering spacecraft will significantly affect the result, the ability to self-learn the memory subsystem tools based on the choice of data with a certain level of retrospective depth is necessary.

Data modification is necessary for the implementation of the property of adaptability in systems and means of countering spacecraft. Such tools and systems must respond adaptively to attacks, otherwise attackers will understand their behavior and will be able to bypass them. To ensure the property of adaptability, it is necessary to form artificial data and use them together with real data. But the formation of new artificial data requires the development of an element of the memory subsystem means, taking into account the requirements for countering CA data protection.

Storing data by class in a distributed environment of corporate networks involves ensuring that they are structured, typed and formatted to achieve typicality. An equally important aspect in this context is the means used for such storage, and the effectiveness in providing prompt access to data also depends on their organization.

In corporate networks, which may contain seven levels of computer attack countermeasures or less, the memory subsystem should be considered not just as a log store, but as an integrated multi-level information and analytical structure that provides accumulation, differentiated use and intelligent processing of historical data of different depths. The formation of such a subsystem involves the analysis of several models of hardware and software implementation, their comparison according to the criteria of performance, scalability, stability and security characteristics, as well as the subsequent synthesis of the optimal architecture for the conditions of the corporate environment. A centralized model, a distributed model, and a combined model can be used to provide data storage. These models are divided into those that support a hierarchical organization of memory or without division at the level of a hierarchy. In addition, the principle of adaptive depth of history can be used in the organization of memory.

A centralized model provides maximum completeness of event correlation and depth of retrospective analysis. It involves the use of hierarchical storage and is suitable for scenarios where a comprehensive analysis of the long history of incidents is crucial. However, the concentration of data streams creates potential bottlenecks and also increases redundancy and fault tolerance requirements. This model is optimal for environments with a powerful data center, but it is less efficient for geographically distributed structures.

The distributed model has a different logic for organizing memory compared to the centralized one. The short-term history is stored directly at the countermeasures layers (perimeter, network sensors, hosts), while the

aggregated or analytically enriched data is transmitted to the central core. This approach reduces response delays and increases the system's resilience to congestion or isolation of network segments. Such a model meets the requirements of real-time and scalability, but makes it difficult to ensure consistency of historical data and requires more complex mechanisms for synchronization, replication, and integrity control.

The combined architecture combines local infrastructure with cloud security services, expands the research field due to the almost unlimited scalability of archival storage and cloud analytics capabilities. RAM and short-term history remain within the corporate perimeter, while long-term data sets used for machine learning and behavioral analytics are moved to the cloud. It provides efficiency for large volumes of history, but requires careful consideration of regulatory restrictions, data sovereignty policies, and cryptographic channel protection.

The essence of the concept of hierarchical memory organization is to establish a different "memory horizon" for each of the seven levels of protection. The lower levels (perimeter, IDS/IPS) work with a short, high-frequency history, while the higher levels (analytics, attack signs, strategic management) operate with long-term arrays. This differentiation allows you to optimize the volume-speed-cost ratio, reducing the load on fast media and at the same time storing valuable historical data in compressed or aggregated form. The synthesis of hierarchical memory organization with the previous three models also makes it possible to implement adaptive data movement between storage tiers depending on their relevance and frequency of use.

Hardware accelerators can be used to process event streams and form signs of attacks in near real time. Such an addition to the model is advisable for heavy loads and significant volumes of telemetry. Also, memory can be organized as a separate protected domain with the implementation of the following procedures: microsegmentation; block-level encryption; access control; full access traceability; integration with systems.

The organization of memory itself can be carried out using intelligent methods for the formation of features, storage of behavioral profiles, adaptive reduction of historical data, auto-updating of signatures, etc.

None of the models provides optimal balance. Therefore, it is necessary to synthesize a submodel in the model of data storage tools for anti-spacecraft systems, which will be consistent with the seven-level division, combine distributed short-term storage at the edge, centralized or clustered data storage for deep analytics, hierarchical organization of media with automatic data lifecycle management, as well as the ability to integrate with cloud resources for long-term archives and ML analysis. Such a synthesis will ensure the adaptability of memory to different timescales of attacks from instantaneous incidents to slow, multi-year attacks. Let us define the following conceptual submodel of data storage organization for systems of seven levels as follows:

$$M_P^p = \langle R_{P,1}^p, R_{P,2}^p, M_{P,3}^p, I_{P,4}^p, A_{P,6}^p, A_{P,7}^p, R_{P,8}^p, S_{P,9}^p, V_{P,10}^p, A_{P,11}^p, D_{P,12}^p \rangle, \quad (4)$$

where is the number of the level of systems and means of countering spacecraft, for which the organization of data storage is represented by a submodel; – options for placing data storage facilities in corporate networks, taking into account the presence or absence of distribution; – the presence or absence of data migration or part thereof in corporate networks, if necessary, for example, during the reflection of the spacecraft; – the presence or absence of a hierarchy when storing data or part of it; – the presence or absence $R_{P,1}^p M_{P,2}^p R_{P,3}^p I_{P,4}^p A_{P,6}^p$ of automatic data lifecycle management of all data or part thereof; – the presence or absence of the function of archiving data that is not relevant at the current time of operation of anti-aircraft systems, for their long-term storage, for example, for several years; – the presence or absence of backup of all data or part thereof; – the presence or absence of backup of all data or part thereof; – selection of data taking into account the depth of the horizon based on intellectualization; – hardware and software in corporate networks that can be used to store data and means of their processing; – data. $A_{P,7}^p R_{P,8}^p S_{P,9}^p V_{P,10}^p A_{P,11}^p D_{P,12}^p$

Thus, a feature in the organization of data storage for systems of seven layers is taking into account the means and systems of countering spacecraft of seven layers, the distribution of data storage in corporate networks, the possibility of migration during spacecraft, the presence of a data hierarchy, and automatic management of the data life cycle all or part of the data, the availability of archiving and data backup, the selection of data taking into account the depth of the horizon based on intellectualization and the features of hardware and software for data storage.

Each level of means and systems for countering spacecraft has its own depth of the memory horizon, the volume of history in the context of the relevance of data for use. For example, for means and systems of seven levels, there may be such a duration that is given in Table. 1. In addition to taking into account history in the context of data relevance, it is also necessary to optimize data and modify to effectively counteract spacecraft that have an adaptive basis.

Table 1

Example of the volume and relevance of data for systems and means of countering spacecraft of seven levels

Level	Volume of history	Media Type
1	24–72 years	NVMe
2	7–14 days	SSD
3	30–90 days	SSD + NAS
4	6–12 months	NAS
5	1–3 years	Object Storage
6	Full story	Data Lake
7	Long-term	Archive

To ensure data storage for systems and means of countering spacecraft, we will define their architecture, in which it is necessary to combine automatic management of the life cycle of all data or part of them, features of hardware and software for data storage, intellectualization of the depth of the horizon for data, self-learning when selecting data for typical repetitive spacecraft. This should increase the efficiency and accuracy of data provision to countering spacecraft systems, and also to ensure the efficiency of their storage, taking into account the depth of the horizon.

We will define the conceptual architecture of the memory subsystem of systems and anti-aircraft tools in corporate networks based on the use of historical data for decision-making, taking into account the model of the process of formation of historical data of anti-aircraft systems, the model of data and time evaluation in the formation and storage of historical data of anti-aircraft systems, the model of data organization of anti-aircraft systems, the model of storing, selecting and optimizing historical data of systems and means of countering computer attacks on corporate networks, including data from deceptive systems, as well as taking into account subsystems that are based on models

$M_{zM,1}^D, M_{zM,2}^D, M_{zM,3}^D, M_{zM,4}^D, P_{P,1}, P_{P,2}, P_{P,3}$, yes:

$$A_p^{QS} = \begin{pmatrix} M_{zM,1}^D \\ M_{zM,2}^D \\ M_{zM,3}^D \\ M_{zM,4}^D \\ P_{P,1} \\ P_{P,2} \\ P_{P,3} \end{pmatrix}, \quad (5)$$

where $M_{zM,1}^D$ - models of the process of formation of historical data of spacecraft countermeasures systems; - a model for estimating data and time in the formation and storage of historical data of spacecraft countermeasures; $M_{zM,2}^D, M_{zM,3}^D$ - model of organization of these spacecraft countermeasures systems; $M_{zM,4}^D$ - a model for storing, selecting and optimizing historical data of systems and means of countering computer attacks on corporate networks; $P_{P,1}$ - a model for organizing data storage in systems and means of countering computer attacks, taking into account the historical aspect of their use; $P_{P,2}$ - a model of data optimization based on self-learning of memory subsystem tools; $P_{P,3}$ - a model of data selection taking into account horizon depth, self-learning and previous experience of use

We detail the models $M_{zM,1}^D, M_{zM,2}^D, M_{zM,3}^D, M_{zM,4}^D$ in the context of the input values regarding historical data and the relationships between them.

Let's set a model of the process of forming historical data of spacecraft countermeasures, including data from deceptive systems, as follows:

$$M_{zM,1}^D = (M_{DI}^1, M_{DI}^D, M_{DD}^r, M_{vp}^p, M_K^D, G_{zM,1}^D), \quad (6)$$

where M_{DI}^1 is the set of sources of information; M_{DI}^D - a set of vectors to ensure the storage of data from information sources of different levels; M_{DD}^r - a set of data sources for seven levels; M_{vp}^p - multiple influences on the preservation and relevance of data in various aspects of their use and ensuring access and relevance; M_K^D - set of data classes; $G_{zM,1}^D$ - a graph of relationships of set elements for the formation $M_{DI}^1, M_{DI}^D, M_{DD}^r, M_{vp}^p, M_K^D$ of historical data of anti-spacecraft systems, including data of deceptive systems.

The relationships between the elements of sets $M_{DI}^1, M_{DI}^D, M_{DD}^r, M_{vp}^p, M_K^D$ are depicted in Fig. 2, and for sets they are presented in more detail through subsets and classes. M_{DD}^r, M_K^D

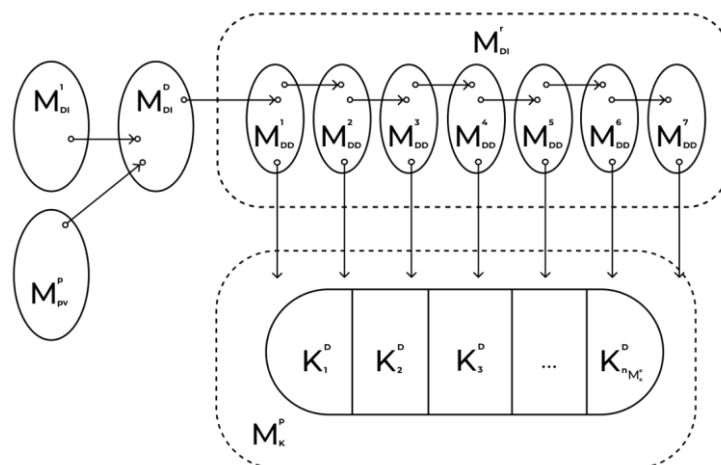


Fig. 2. Relationships between elements of sets $M_{DI}^1, M_{DI}^D, M_{DD}^r, M_{vp}^p, M_K^D$

According to the defined relationships between the sets $M_{DI}^1, M_{DI}^D, M_{DD}^r, M_{vp}^p, M_K^D$, we will set for the model (formula (5)) the process $M_{zM,1}^D$ of formation of historical data of the systems for counteracting the spacecraft graph $G_{zM,1}^D$, which is shown in Fig. 3.

Let's define a model for estimating data and time in the formation and storage of historical data of anti-aircraft systems, including data from deceptive systems, as follows:

$$M_{zM,2}^D = (M_{F_{vp,i}^p}, V_{vp,t}^p, F_{vp,t}^p, M_1^T, M_{v_{DI,m,n}^D}^T, G_{zM,2}^D), \quad (7)$$

where $M_{F_{vp,i}^p}$ is the set of functions for assessing the effects of elements of the set $M_{F_{vp,i}^p}$ of influences M_{vp}^p on the preservation and relevance of data in various aspects of their use and ensuring access and relevance; – evaluation functions; $F_{vp,i}^p$; $i = 1, 2, \dots, n_{M_{vp}^p}$; – the number of elements of the set of influences; – vector of values for assessing the impacts on the preservation and relevance of data in various aspects of their use and ensuring access and relevance; – vector $n_{M_{vp}^p} M_{vp}^p V_{vp,t}^p F_{vp,t}^p$ of influence functions $F_{vp,i}^p$ (; – the number of elements of the set of influences) $i = 1, 2, \dots, n_{M_{vp}^p} n_{M_{vp}^p} M_{vp}^p$ (formula (2.14)); t – the current time in which information about the data was received and the impact on them was evaluated using functions $F_{vp,i}^p$; M_1^T – a time matrix for these classes, taking into account the coordinate, which fixes the time of data placement in databases, vector (; – the number of vectors to ensure the storage of data from information sources of different levels) and the corresponding class (; – number of classes) from the plural; – $v_{DI,i}^D V_{DI,i}^D$; $i = 1, 2, \dots, N_{M_{DI}^D} N_{M_{DI}^D} K_i^D$ $i = 1, 2, \dots, n_{M_K^D} n_{M_K^D} M_K^D M_{v_{DI,m,n}^D}^T$ matrix of sequential assessment of impacts and relevant data from classes.

The graph for the model for estimating data and time in the formation and storage of historical data of anti-aircraft systems, including data from deceptive systems, is shown in Fig. 4. $G_{zM,2}^D M_{zM,2}^D$

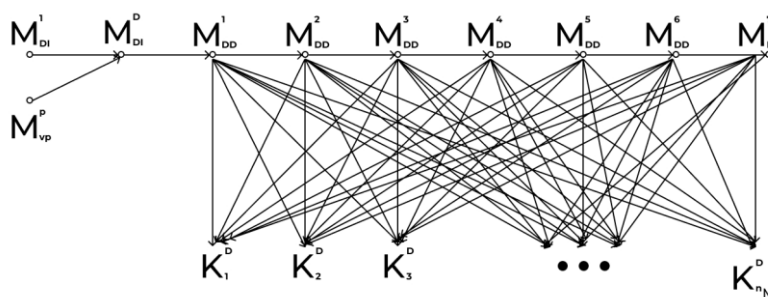


Fig. 3. Graph $G_{zM,1}^D$

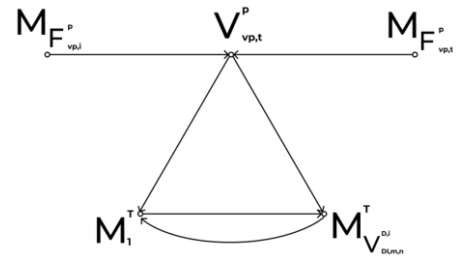


Fig. 4. Graph $G_{zM,2}^D$

Let's define the model for organizing data of anti-aircraft systems, including historical data and data of deceptive systems, as follows:

$$M_{zM,3}^D = (M_{D,r}^{OS}, M_{D,r}^{OS,D}, K_{D,r,i}^{OS,D}, G_{zM,3}^D), \quad (8)$$

where $M_{D,r}^{OS}$ is the set, which is a subset of all data and includes elements about data classes that are stored and can be used by deceptive systems of r level $M_{D,r}^{OS}$; – $M_{D,r}^{OS,D}$ a set of historical data of systems and means of all seven levels using deceptive technologies $M_{D,r}^{OS,D}$; – a $K_{D,r,i}^{OS,D}$ class of vectors of subclasses, and the first class without deceptive systems, and the second – with deceptive systems; $K_{D,r,i}^{OS,D}$ – the i th class of data of objects of deceptive systems; – the number of the level of means and systems of countering spacecraft; $i = 1, 2, \dots, 7$

The graph for the model of organization of data of anti-aircraft systems, including historical data and data of deceptive systems, is shown in Fig. 5. $G_{zM,3}^D M_{zM,3}^D$

Let's define a model for storing, selecting and optimizing historical data of systems and means of countering computer attacks on corporate networks, as follows:

$$M_{zM,4}^D = (M_z^D, M_d^D, M_{ez}^D, G_{zM,4}^D), \quad (9)$$

where M_z^D is the set of tasks of the means of storing, selecting and optimizing historical data of systems and means of countering computer attacks of corporate networks (formula (1)); M_d^D – the set of half of the data processed by the memory subsystem tool (formula (2)); M_{ez}^D is the set of basic elements of the memory subsystem (formula (2)).

The graph for the model $G_{zM,3}^D M_{zM,3}^D$ (formula (9)) of storing, selecting and optimizing historical data of systems and means of countering computer attacks of corporate networks is shown in Fig. 6.

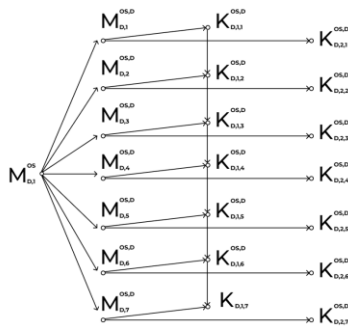


Fig. 5. Graph $G_{zM,3}^D$

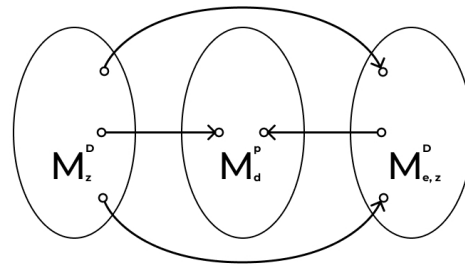


Fig. 6. Graph $G_{zM,4}^D$

Thus, the architecture of data storage facilities for systems and means of countering spacecraft has been improved, in which, unlike the known ones, a combination of automatic management of the life cycle of all or part of data, features of hardware and software tools for data storage, intellectualization of the depth of the horizon for data, self-learning in the selection of data for typical repetitive spacecraft, which is the basis for increasing efficiency and accuracy in providing data to systems and means of countering spacecraft, has been carried out, and also to ensure the efficiency of their storage, taking into account the depth of the horizon.

Storing data in the memory of facilities and systems of all seven levels is necessary to counteract computer attacks. At the same time, the accumulation of heterogeneous data creates problems in their volume, access, and unification of formats. Such data needs to be brought to consistent formats while maintaining connections and taking into account time relevance. The lack of information about the attackers' new activities or tools in the memory of the systems allows them to avoid detection, so they try to bypass or limit the use of existing data.

Spacecraft countermeasures systems use memory not only to detect spacecraft, but also to select further actions. This requires preserving the history of previous steps with an evaluation of their effectiveness for reuse. As the level of systems increases, the volume of historical data, the complexity of their processing, and the context of decision-making increase, while the dependence on signatures decreases. Therefore, it is necessary to improve the organization of memory and ways of accessing data.

Solving the contradictions between the volume of data, their relevance and adaptability of analysis is possible through the use of multilevel memory models. Such models combine the storage of data, attributes, and knowledge taking into account time, context, and the ability of systems to adapt.

The presentation of data in the form of vectors at different levels of the hierarchy is the basis of the memory subsystem, which provides the choice of the depth of the retrospective, the optimization of obsolete data and the synthesis of new ones. The improved architecture of data storage facilities provides for automatic management of the data life cycle, taking into account the features of storage facilities, adaptive horizon selection and self-learning, which is the basis for the development of methods for increasing the efficiency and accuracy of systems and means of countering space.

Method of data selection taking into account horizon depth, self-learning and previous experience of use

The use of horizon depth, self-learning mechanisms and previous experience in the selection of data is necessary due to the high dynamism of processes in the CM environment, in which the efficiency of the systems largely depends not only on the amount of information available, but also on the ability to correctly determine its relevance, practical value and prospects of use. Focusing only on the current state of data does not provide sufficient stability in decision-making, since information quickly loses relevance, and environmental conditions can change faster than the system has time to adapt to them. That is why there is a need to take into account the depth of the horizon, which allows you to assess not only the current usefulness of information, but also its potential significance in future scenarios of the system's functioning. This approach provides an opportunity to predict the effectiveness of data use in the short, medium and long term, which is especially important for adaptive and intelligent systems.

At the same time, focusing exclusively on the forecast horizon is not sufficient, since the system must take into account the accumulated experience of previous data use. Previous experience allows us to evaluate the real effectiveness of information based on the historical results of its application. Analysis of successful and unsuccessful scenarios makes it possible to identify patterns, identify consistently useful sources of information, and minimize the repetition of erroneous decisions. The use of previous experience increases the reliability of data selection, since the system gets the opportunity to rely not only on the theoretical assessment of relevance, but also on the actual results of functioning in similar conditions.

However, even the combination of the horizon of analysis and previous experience does not provide enough flexibility without self-learning mechanisms. Self-learning is necessary to constantly adjust the evaluation criteria, adapt the weighting factors and update the data selection models according to changes in the environment. In real conditions, the characteristics of information, the behavior of objects and the parameters of information flows can change so quickly that the accumulated experience partially loses its relevance. Self-learning allows the system to

identify new patterns, automatically take into account the results of its own work, and increase the accuracy of subsequent decisions without the need for a complete manual reconfiguration of the system.

The combination of horizon depth, previous experience, and self-study is not only possible, but also expedient, since these components complement each other. The depth of the horizon provides a strategic vision and orientation for future conditions, previous experience forms the basis for evaluating practical effectiveness, and self-learning provides adaptability and the ability for continuous improvement. Together, they form a multi-level data selection engine that combines forecasting, knowledge accumulation, and dynamic adaptation.

At the same time, certain contradictions may arise between these components. One of the main ones is the conflict between historical experience and new conditions of functioning. Data that has been effective in the past may lose relevance in the altered CM environment, while a system overly focused on previous experiences may continue to give it a high priority. Another contradiction is the contradiction between the long-term forecasting horizon and the need to quickly adapt to current changes. Long-term orientation can reduce the sensitivity of the system to operational changes in the environment, while over-adaptability of self-learning can lead to instability of models and loss of strategic coherence. There may also be a contradiction between the stability of accumulated knowledge and the constant updating of learning parameters, since too aggressive self-study can lead to the degradation of useful historical experience.

Despite these contradictions, they can be minimized by a balanced mechanism for integrating components. The use of adaptive weighting factors, multi-level data assessment, and learning stability control mechanisms allows for a trade-off between strategic sustainability and operational adaptability. As a result, the combination of horizon depth, self-learning and previous experience allows you to achieve an increase in the accuracy of data selection, a decrease in information redundancy, an improvement in adaptation to a dynamic environment, acceleration of decision-making, a decrease in the number of erroneous decisions and the accumulation of an effective knowledge base for the further functioning of the system. Such a combination of characteristic properties in the method will ensure the formation of an intelligent data selection mechanism capable of functioning effectively in conditions of uncertainty, changes in information flows and constant evolution of the environment.

The basic steps of the data selection method taking into account the depth of the horizon, self-learning and previous experience of use.

Step 1. Formation of a set of input data.

At this stage, the system collects all available data from various sources and forms a single information space.

Step 2. Definition of a data selection strategy and the formation of an optimization model.

At this stage, the system forms a single decision-making mechanism that transforms a set of heterogeneous data into an orderly system of evaluation and choice. The strategy defines not only the criteria for the importance of the data, but also the way they are compared, normalized and optimized in the face of limited resources.

2.1. Formation of a formal description of the incoming information data space

Let's define each element of the data with a feature vector as follows:

$$F_{DI}^{P,D} : V_{DI,i}^D \rightarrow V_{DI,i}^{P,D};$$

$$V_{DI,i}^{P,D} = (v_{DI,i,1}^{P,D}, v_{DI,i,2}^{P,D}, v_{DI,i,3}^{P,D}, v_{DI,i,4}^{P,D}, v_{DI,i,5}^{P,D}, v_{DI,i,6}^{P,D}, v_{DI,i,7}^{P,D}), \quad (10)$$

where $F_{DI}^{P,D}$ is the function of forming vectors of the input information space of data;

$V_{DI,i}^{P,D}$ – vector of the input information space of data; $-v_{DI,i,1}^{P,D}$ – relevance (horizon); $v_{DI,i,2}^{P,D}$ – the effectiveness of previous use; $v_{DI,i,3}^{P,D}$ – informativeness; $v_{DI,i,4}^{P,D}$ – redundancy; $v_{DI,i,5}^{P,D}$ – cost of processing; $v_{DI,i,6}^{P,D}$ – time characteristics; $v_{DI,i,7}^{P,D}$ – contextual correspondence; $M_{DI}^D = \{V_{DI,1}^D, V_{DI,2}^D, \dots, V_{DI,N_{M_{DI}}}^D\}$; $N_{M_{DI}}^D$ – the number of vectors to ensure the storage of data from information sources of different levels; $-i$ -th vector; $V_{DI,i}^D, i = 1, 2, \dots, N_{M_{DI}}^D$

2.2. Normalization and preparation of criteria.

Since all characteristics are of a different nature, they need to be brought to a single scale: unification of the range of values of values; elimination of dimensions; noise smoothing; filtering of anomalies.

2.3. Formation of an optimization strategy based on the optimization function.

The essence of this stage is that the system forms a formalized strategy of behavior, which simultaneously takes into account the usefulness, risks, resources and experience of previous use. The basic strategy determines the balance between utility and costs and we will set it as follows:

$$F_{DI}^{O,D}(V_{DI,i}^{P,D}) = \sum_{i_1=1}^{n_{i_1}} a_{DI,i,i_1}^{O,D} \cdot A_{DI,i,i_1}^{O,D} - \sum_{i_2=1}^{n_{i_2}} b_{DI,i,i_2}^{O,D} \cdot B_{DI,i,i_2}^{O,D}, \quad (11)$$

where $F_{DI}^{O,D}$ is the optimization function for the formation of selection strategies; $a_{DI,i,i_1}^{O,D}$ – coefficients for the values of local functions that accumulate the result; $-$ coefficients for the values of local functions that reduce the result; $-$ the number of local functions that accumulate the result; $-$ the number of local functions that reduce the result; $-$ the value of local functions that accumulate the result; $-$ the value of local functions that reduce the result; is the vector $b_{DI,i,i_2}^{O,D}, i_1 = 1, 2, \dots, n_{i_1}, i_2 = 1, 2, \dots, n_{i_2}, n_{i_1} n_{i_2} A_{DI,i,i_1}^{O,D} B_{DI,i,i_2}^{O,D} V_{DI,i}^{P,D}$ of the input information space of the data (formula (10)).

The utility in formula (11) is given by positive values of the sum, and the costs are given by negative values. We detail the functions in the formula (11).

The model of relevance (horizontal dynamics) depends on the time and scenario of use, and we will set it as follows:

$$A_{DI,i,1}^{O,D} = e^{-\mu \cdot t_i} \cdot H_{DI,1}^{O,D}(t_i), \quad (12)$$

where t_i is the duration of stay in the system of generated data; $-\mu$ the rate of aging of information; $H_{DI,1}^{O,D}$ – horizon function (operational / tactical / strategic mode).

A function $H_{DI,i,1}^{O,D}$ takes, for example, three values: $\{1\}$; $\{2\}$; $\{3\}$.

Data loses weight quickly if it is short-term, but can remain important in the strategic horizon.

The experience model (reinforcement assessment) is based on the probability of success in similar conditions.

Let's define it with a fugation like this:

$$A_{DI,i,2}^{O,D} = P_{DI,2}^{O,D}(V_{DI,i}^{P,D}, S_{DI,i}^{P,D}), \quad (13)$$

where $P_{DI,2}^{O,D}$ is the probability of success in similar conditions; $V_{DI,i}^{P,D}$ – vector of the input information space of data (formula (10)); $S_{DI,i}^{P,D}$ – the state of the CM environment.

Experience suggests that the data stores information about how it performed in similar situations and, then, the system adapts to the context, not just the story.

The model of informativeness provides a decrease in the uncertainty of the entropy of the system and we define it as a function as follows: $A_{DI,i,3}^{O,D}$

$$A_{DI,i,3}^{O,D} = P_{DI,3}^{O,D}(V_{DI,i}^{P,D}, S_{DI,i}^{P,D}) - P_{DI,3}^{O,D}(V_{DI,i}^{P,D}, S_{DI,i,p}^{P,D}), \quad (14)$$

where $P_{DI,3}^{O,D}$ is the probability of determining the state of the system; $V_{DI,i}^{P,D}$ – vector of the input information space of data (formula (10)); $S_{DI,i}^{P,D}$ – the state of the CM environment; $S_{DI,i,p}^{P,D}$ – the state of the CM environment after the event, i.e. after the state $S_{DI,i}^{P,D}$.

This model makes it possible to establish that the more data clarifies the situation in the system, the better its value.

2.4. Redundancy model (structural dependence)

Redundancy is defined as dependence on already available data:

$$B_{DI,i,1}^{O,D} = \sum_{r=1, r < i}^{M_{DI}^{D}} P_{DI,4}^{O,D}(V_{DI,r}^{P,D}, V_{DI,i}^{P,D}), \quad (15)$$

where $V_{DI,i}^{P,D}$ is the vector of the input information space of the data (formula (10)); $M_{DI}^{D} = \{V_{DI,1}^D, V_{DI,2}^D, \dots, V_{DI,N_{M_{DI}^D}}^D\}$; $N_{M_{DI}^D}$ – the number of vectors to ensure the storage of data from information sources of different levels; $-th$ vector; $-V_{DI,i}^D, ii = 1, 2, \dots, N_{M_{DI}^D}$ $P_{DI,4}^{O,D}$ correlation or similarity between the data of two different input vectors. $V_{DI,r}^{P,D}, V_{DI,i}^{P,D}$

Thus, if the data is repeated by others, then their value decreases, because the system seeks a variety of information.

The cost model, i.e. the resource function, $B_{DI,i,2}^{O,D}$ is set as follows:

$$B_{DI,i,2}^{O,D} = \sum_{r=1}^3 b_{DI,5,r}^{O,D} \cdot B_{DI,5,r}^{O,D}(V_{DI,i}^{P,D}), \quad (16)$$

where $V_{DI,i}^{P,D}$ is the vector of the input information space of the data (formula (3.27)); $B_{DI,5,1}^{O,D}$ – processing time; $B_{DI,5,2}^{O,D}$ – computational complexity; $B_{DI,5,3}^{O,D}$ – memory/resources; $r = 1, 2, 3$; $-b_{DI,5,r}^{O,D}$ the weight of resources.

Valuable data reduces the priority because the system incurs higher costs for its use. Thus, the determination of the data selection strategy will be carried out through the formation of an optimization model so that the system has numerical values for comparing data and can determine their priority.

Step 3. Analysis of the depth of the horizon.

At this stage, the system determines the usefulness of the data, taking into account the time and use case. The same data can have different values depending on the application horizon. Information that is critical for rapid response may be of little use in strategic analysis, and vice versa. Therefore, the system must not just evaluate the data, but determine the time of usefulness, significance and the possibility of using them in the current scenario.

The system evaluates the analysis of the depth of the data usage horizon in the context of future utility. To do this, we analyze the time characteristics of the data, the time of occurrence, the rate of aging, the frequency of renewal and the duration of the life cycle. In addition, we assess the dynamics of change in relevance, the rate of loss of value of information depending on environmental and time conditions. The system also determines the correspondence of the data to different horizons of use, i.e. operational, tactical or strategic, which allows you to adapt the choice of information to the type of task. Separately, we analyze the scenario of the functioning of the environment, in particular, the stability of conditions, the presence of critical events, and the likelihood of a change in the context. To improve accuracy, previous experience in using similar data, their effectiveness in similar situations and projected future usefulness are taken into account.

Step 4. Analysis of previous experience of use

The system refers to the history of data application and analyzes the usefulness of the data in previous uses. Let's use the results of step 15 "Formation of previous experience of the data used" on the method of organizing data

storage in systems and means of countering computer attacks, taking into account the historical aspect of their use. It uses methods and technologies based on Markov (formula (3.8)), Bayesian and statistical models of experience accumulation. To do this, we analyze the history of using information, the frequency of its involvement in the decision-making process, the level of success of the results obtained and the impact on the accuracy of the system's functioning. We evaluate the achievement of target indicators of systems and means of countering spacecraft, reducing uncertainty and improving the quality of forecasting with certain data. In addition, we will take into account errors that occurred during the use of data, the level of false positives, the stability of results in different scenarios, and the ability of data to adapt to changes in the environment. The system also analyzes the repeatability of successful use cases of information, the duration of its practical usefulness, and the dependence of effectiveness on the context of use. Based on the accumulated statistics, we form an assessment of the reliability and prospects of data, which allows us to determine the degree of trust in them in future cycles of use.

Step 5. Evaluating the information content of data

The system determines the content value of data, taking into account their ability to influence the quality of decision-making, the formation of new dependencies and the clarification of the current state of the environment. Unlike the previous stages, where time characteristics and experience of use were evaluated, here the focus is on the structural and semantic properties of information. The system analyzes the level of uniqueness of data, the degree of their connection with other elements of the information space, the presence of hidden patterns, as well as the contribution to improving the results of forecasting, classification or identification of system states.

Let's define the function of evaluating the information content of data as follows:

$$F_{DI,0}^{O,D}(V_{DI,i}^{P,D}) = \sum_{r=1}^3 u_{DI,r}^{O,D} \cdot C_{DI,r}^{O,D}(V_{DI,i}^{P,D}), \quad (17)$$

where $F_{DI,0}^{O,D}$ is the function of evaluating the information content of data; $u_{DI,r}^{O,D}$ – coefficients for the values of local functions; $C_{DI,r}^{O,D}$ – the value of local functions; $V_{DI,i}^{P,D}$ is the vector of the input information space of the data (formula (10)).

Local functions are set taking into account the following properties:

- 1) $C_{DI,1}^{O,D}$ – a function for determining semantic significance;
- 2) $C_{DI,2}^{O,D}$ – a function for determining the structural contribution;
- 3) $C_{DI,3}^{O,D}$ is a function for determining the increase in efficiency.

As a result, the system determines the level of meaningful usefulness of data, i.e. the ability to form new knowledge and improve the functioning of the system, which will increase the validity of further information selection.

Let's detail each of the local functions.

The local function is defined based on the use of cosine similarity. Cosine similarity in vector space makes it possible to measure the similarity of two objects, but not by comparing the words or symbols themselves, but by comparing their numerical semantic representations. Each text, piece of data, or context is transformed into a multidimensional vector by a special model. Such a vector carries hidden semantic information. Objects that are similar in meaning receive similar directions in space, even if they are formally different in words. After that, it is not the length of the vectors or their absolute values that is compared, but the angle between them. It is this angle that reflects the semantic similarity. If the two vectors point almost in the same direction, then this means that the corresponding data describe similar concepts or situations, even if the wording is different. If the vectors are perpendicular, then this means the absence of a meaningful connection. And if they are directed in opposite directions, this may mean a conceptual opposite or incompatibility of content. Thus, cosine similarity measures not syntactic similarity, but semantic proximity in a hidden semantic space where each coordinate reflects a certain abstract feature of meaning, learned by the model automatically from large amounts of data. This approach makes it possible to recognize similarities between data even when they do not have common words, but describe the same situation in the CM. Then, let's set the function like this:

$$C_{DI,1}^{O,D}(V_{DI,i}^{P,D}) = \frac{\overline{V_{DI,i}^{P,D}} \cdot \overline{V_{DI,j}^{P,D}}}{|\overline{V_{DI,i}^{P,D}}| \cdot |\overline{V_{DI,j}^{P,D}}|}, \quad (18)$$

where $V_{DI,i}^{P,D}$ is the vector of the input information space of the data (formula (3.27)); $K_{DI,i}^{P,D}$ – vector representation of the current context; $|\overline{V_{DI,i}^{P,D}}|$ – length (norm) of the data vector; $|\overline{K_{DI,i}^{P,D}}|$ – length (norm) of the context vector $K_{DI,i}^{P,D}$; $\overline{V_{DI,i}^{P,D}} \cdot \overline{V_{DI,j}^{P,D}}$ scalar product of vectors; $\overline{V_{DI,j}^{P,D}}$ is the vector of the input information space of the data (formula (3.27)), with which the vector is compared i .

The function for determining $C_{DI,2}^{O,D}$ the structural contribution of data $V_{DI,i}^{P,D}$ to the construction of information dependencies is defined as follows:

$$C_{DI,2}^{O,D}(V_{DI,i}^{P,D}) = \sum_{k=1}^{n_{DI,2}^{O,D}} u_{DI,2,k}^{O,D} \cdot C_{DI,2,k}^{O,D}(V_{DI,i}^{P,D}), \quad (19)$$

where $V_{DI,i}^{P,D}$ is the vector of the input information space of the data (formula (10)); $n_{DI,2}^{O,D}$ – the number of signs; $u_{DI,2,k}^{O,D}$ – weight of features; $C_{DI,2,k}^{O,D}(V_{DI,i}^{P,D})$ $k = 1, 2, \dots, n_{DI,2}^{O,D}$ – the meaning of the signs;

The function for determining $C_{DI,3}^{O,D}$ the efficiency gain is set as follows:

$$C_{DI,3}^{O,D}(V_{DI,i}^{P,D}) = C_{DI,3,n}^{O,D}(V_{DI,i}^{P,D}) - C_{DI,3,p}^{O,D}(V_{DI,i}^{P,D}), \quad (20)$$

where $V_{DI,i}^{P,D}$ is the vector of the input information space of the data (formula (3.27)); $C_{DI,3,p}^{O,D}(V_{DI,i}^{P,D})$ – the value of system quality without the use of data; $C_{DI,3,n}^{O,D}(V_{DI,i}^{P,D})$ – the value of system quality after data integration.

Step 6. Reduction of choice indicators (reduction of data space).

At this stage, the system performs a reduction in redundant information and a reduction in the dimensionality of the data space in order to increase the efficiency of further selection, reduce computational complexity and eliminate information noise. Not all features, parameters or data elements have the same effect on the outcome of decision-making. Some of the information can be duplicated, have a weak impact or create redundant dependencies, which complicates analysis and degrades the quality of optimization. Therefore, the system analyzes the data structure, the level of correlation between features, the degree of repeatability of information and the contribution of each parameter to the formation of the result. Based on this, the selection of the most informative components and the elimination of secondary or dependent features are carried out.

To reduce the selection indicators, we will apply a scheme for searching for data anomalies in classes. At the same time, it is mandatory to group close values to reduce the number of input data in the future at subsequent stages. The resulting scheme with profiles corresponds to the case of multivariate analysis, where several objects are considered according to many characteristics. and are based only on the results of observations. First, for the study, we build an observation matrix as follows:

$$X = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{1k} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2k} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ik} & \dots & x_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{s1} & x_{s2} & \dots & x_{sk} & \dots & x_{sn} \end{bmatrix}, \quad (21)$$

where n is the number of features observed on objects; s is the number of objects; – the number of manifestations of k – of that feature in i – of that object during the period of observations. x_{ik}

To normalize the features, let's perform the transformation as follows:

$$V_{ik} = \frac{x_{ik}}{\sum_{i=1}^s x_{ik}}, \quad (22)$$

and make the transition to matrix V like this:

$$V = \begin{bmatrix} V_{11} & V_{21} & \dots & V_{1k} & \dots & V_{1n} \\ V_{21} & V_{22} & \dots & V_{2k} & \dots & V_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ V_{i1} & V_{i2} & \dots & V_{ik} & \dots & V_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ V_{s1} & V_{s2} & \dots & V_{sk} & \dots & V_{sn} \end{bmatrix}, \quad (23)$$

Then it is possible to determine the isotonic index W, which will reflect the position of the S-th object according to the totality of all features:

$$W = \begin{bmatrix} W_1 \\ W_2 \\ \dots \\ W_i \\ \dots \\ W_s \end{bmatrix}, \quad (24)$$

where $W_i = \sum_{k=1}^n V_{ik}$

According to the matrix (23), we order objects by isotonic index, that is, by rank that characterizes the object by a set of features. We determine groups of objects by isotonic indicator. Next, we carry out a structural (isomorphic) ordering of objects. To do this, we form a matrix Z from matrix X according to formulas (22) and (23) as follows: W_i

$$Z_{ik} = \frac{x_{ik}}{\sum_{i=1}^s x_{ik}} / \frac{x_{ik}}{\sum_{i=1}^s x_{ik}}, \quad (25)$$

$$Z = \begin{bmatrix} Z_{11} & Z_{12} & \dots & Z_{1k} & \dots & Z_{1n} \\ Z_{21} & Z_{22} & \dots & Z_{2k} & \dots & Z_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Z_{i1} & Z_{i2} & \dots & Z_{ik} & \dots & Z_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Z_{s1} & Z_{s2} & \dots & Z_{sk} & \dots & Z_{sn} \end{bmatrix}, \quad (26)$$

Next, we plot the distance matrix in D using the formulas (3.42) and (3.43) as follows:

$$d_{im} = \left\{ \frac{1}{N} \sum_{k=1}^n (Z_{ik} - Z_{mk})^2 \right\}^{\frac{1}{2}}, \quad (27)$$

We can also use the average absolute difference in the values of the features, which we will define as follows:

$$d_{im} = \frac{1}{N} \sum_{k=1}^n |z_{ik} - z_{mk}|, \quad (28)$$

$$D = \begin{bmatrix} 0 & d_{12} & \dots & d_{1k} & \dots & d_{1s} \\ d_{21} & 0 & \dots & d_{2k} & \dots & d_{2s} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ d_{i1} & d_{i2} & \dots & 0 & \dots & d_{is} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ d_{s1} & d_{s2} & \dots & d_{sk} & \dots & 0 \end{bmatrix}. \quad (29)$$

That is, a quadratic matrix $S \times S$ is obtained, where the number of rows and the number of columns is equal to the number of objects. In this matrix, it will be possible to distinguish subsets for homogeneity. To do this, we will apply, for example, the Chekanovsky method.

Numerical values in the distance matrix are replaced with marks, for example, like this:

- × - distance $B < 0.4$;
- - distance $0.4 < B < 0.8$;
- - distance $B > 0.8$.

As a result, an unordered Chekanowski diagram will be obtained. To obtain an ordered chart, the rows and columns are rearranged so that these marks are as close to the main diagonal as possible. For example, let the following diagram be obtained for an 8×8 matrix (Table 3.1).

Table 2

Ordered Chekanovsky diagram

	1	2	3	4	5	6	7	8
1	×	×	×	×	○	○	○	·
2	×	×	×	×	○	○	·	·
3	×	×	×	×	○	○	·	·
4	×	×	×	×	·	·	○	·
5	○	○	○	·	·	·	○	○
6	○	○	○	·	·	×	·	·
7	·	·	·	○	○	·	×	·
8	·	·	·	·	○	·	·	×

From the diagram, objects 1, 2, 6, 7 are installed, which are grouped and have the smallest distance (group) from the main diagonal. The most distant objects belong to numbers 4, 8.

Thus, it is possible to divide into classes and, also, in the future, analyze deviations in classes to search for anomalies.

Step 7. Integral data ranking.

The system performs a generalized assessment of each data element by integrating the results of the previous stages of analysis. A single integral indicator is formed, which reflects the overall usefulness of the data, taking into account the temporal relevance, previous experience of use, semantic significance, resource costs and the results of data space reduction. Thus, the system moves from the analysis of individual characteristics to a comprehensive comparison of all elements of the information space.

The result should be a priority set of data that most fully corresponds to the current context, the goals of the system and the conditions of operation.

Step 8. Optimal data selection

The system performs the final selection of a subset of data that has the greatest integral value for the current decision-making process. The main goal is to form a compact set of information that provides maximum usefulness with minimal redundancy and permissible resource costs. After integral ranking, the system moves from evaluating individual elements to the problem of optimal selection of a set of data.

Step 9. Use of selected data

The selected set of data is used to perform the target tasks of the system, in particular, forecasting, analyzing the situation, classifying states, identifying patterns or supporting decision-making. The main goal is the practical application of an optimally formed set of information to obtain a result with maximum accuracy, stability and adaptability to the conditions of the CM environment.

Step 10. Self-learning system

We will self-study the system on the selection and optimization of integrated data sets. After using the formed set of information, the system analyzes the effectiveness of the result obtained, evaluates the contribution of individual data and their combinations to the quality of forecasting, analysis or decision-making, and accumulates experience of previous functioning. Based on this, the weights of the criteria, the parameters of the optimization function and the structure of the data selection are adjusted. The system gradually determines which combinations of information provide the best results, which data are redundant, and which are the most resistant to environmental changes. This allows you to adapt the selection process to new scenarios, increase the accuracy of decisions, reduce the impact of noise and form optimal integrated data sets for subsequent cycles of system operation.

Step 11. Updating the knowledge base

We will update the knowledge base of the system by accumulating information about the results of data selection, the efficiency of integrated populations and the peculiarities of functioning in different environmental scenarios. After the completion of the work cycle, the system stores assessments of the quality of decisions, parameters of the optimization function, successful combinations of data, the level of their informativeness, relevance and semantic significance. Additionally, we record errors, redundant structures and changes in the behavior of the environment. Based on the accumulated experience, we form a knowledge base that will be used for further self-learning, adaptation of selection criteria and forecasting the effectiveness of future decisions. This allows the system to gradually accumulate structured knowledge about effective strategies for working with data, reduce the time of adaptation to new conditions and increase the accuracy of the optimal choice of information in subsequent cycles of operation.

The developed method of data selection, taking into account the depth of the horizon, self-learning and previous experience of use, implements an integrated approach to the formation of an optimal information set for the implementation of the target tasks of the system. The essence of the method is a combination of the temporal relevance of data, their semantic significance, accumulated experience of use and adaptive optimization of the selection process. The method takes into account that the efficiency of the system depends not only on individual elements of information, but also on the structure of integrated data sets, their relationships and the ability to provide the most useful result in a specific environmental scenario.

Experiments

To confirm the effectiveness of the developed method for selecting historical data in systems and means of countering spacecraft in the CM, a series of experimental studies was carried out. The purpose of the experiment was to evaluate the impact of automated selection of historical data on the efficiency of CA detection compared to the traditional deterministic approach. For this purpose, two options for the functioning of the spacecraft counteraction system were studied:

1) the basic version, in which the choice of historical data is carried out deterministically according to pre-established rules without taking into account the current state of the system, self-learning mechanisms and the results of previous data use;

2) the proposed option, in which the selection of historical data is carried out automatically on the basis of the developed method, taking into account the optimal depth of the analysis horizon, the accumulated experience of the system functioning and self-learning mechanisms.

Experimental studies were performed on the same set of test data (Appendix D), which contained information security event logs, network traffic, and information about previous incidents. For each variant of the system, 50 independent simulation series were carried out under the same operating conditions. After the completion of each series, the probability of correct detection of spacecraft was determined by the formula: P_{vKA}

$$P_{vKA} = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (30)$$

where N_{TP} is the number of correctly detected spacecraft; N_{FN} – the number of missed spacecraft.

The results were compared according to the values of the probability of detecting attacks (formula (30)), the average decision-making time for the selection of historical data and the integral efficiency of the functioning of the spacecraft detection subsystem. The results are presented in Table 3.

Table 3

Results of the experiment with sequential addition of functionality to the spacecraft counteraction system

System variant	Automatic horizon selection	Self-study	Previous experience
Basic	Absent	Absent	Absent
Option 1	Available	Absent	Absent
Option 2	Available	Available	Absent
Option 3 (complete method)	Available	Available	Available

Then, on one graph, you can depict a gradual increase in the probability of detecting spacecraft. For example, the summary data in Table. 4 and the image of the graphs in Fig. 7.

Table 4

Results of experimental studies by average values

System variant	Average P_{vKA}
Basic system	0.846
Option 1	0.886
Option 2	0.928
Option 3 (complete method)	0.960

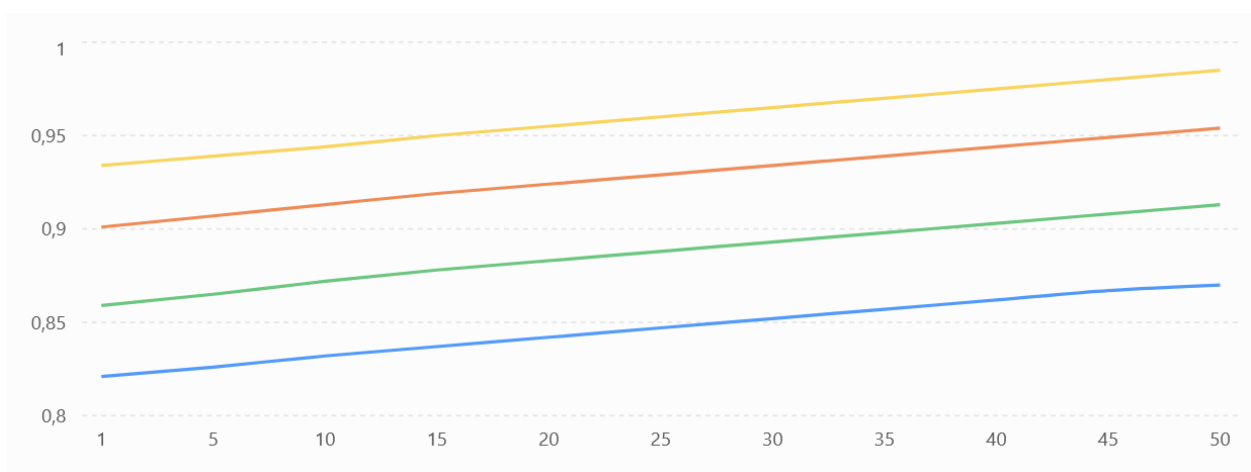


Fig. 7. Dependence of the probability of detecting spacecraft on the level of functionality of the system

Based on the results presented in Table, Table 4 and Fig. 7, it is found that each proposed improvement provides a stable increase in the probability of detecting spacecraft. Compared to the basic system, the automatic selection of the historical data horizon increases the average probability of detection by about 4 %, and taking into account previous experience using historical data allows you to achieve an average probability of detection of 0.96, which is about 13.5% higher than the baseline option.

The next step is to construct Table. 5 and Fig. 8 for the average time for making a decision on the choice of historical data. The graph according to the values is shown in Fig. 9.

Analysis of the results shown in Fig. 9, reflects a steady increase in the integral efficiency of the functioning of the spacecraft detection subsystem as the components of the proposed method for selecting historical data are sequentially added. The lowest values of integral efficiency are observed for the basic system, in which the choice of historical data is carried out according to deterministic rules without adaptation to the current conditions of functioning. The addition of an automatic horizon depth selection mechanism provides an increase in integral efficiency by generating more relevant historical data sets. Further use of self-learning mechanisms allows you to adapt the data selection process to the accumulated experience in detecting attacks, which further increases the effectiveness of the subsystem. The best results are achieved when using automatic horizon selection, self-learning mechanisms, and taking into account previous experience in using historical data. In this case, the integral efficiency reaches its maximum values.

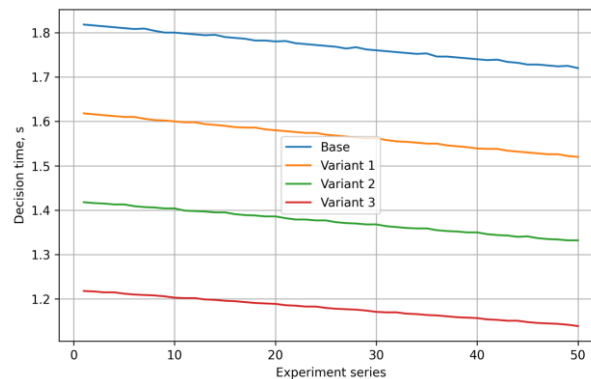


Fig. 8. Graph of the dependence of decision-making time on the number of the experiment series

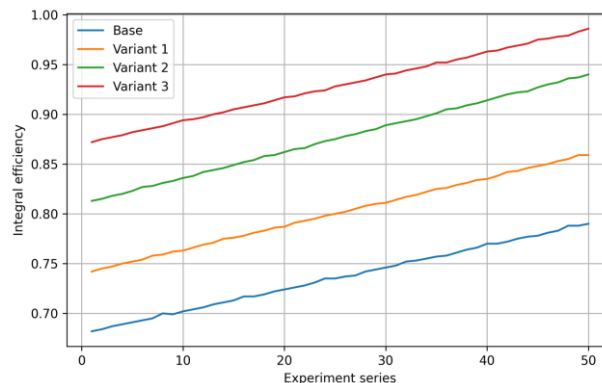


Fig. 9. Dependence of integral efficiency on the level of functionality of the system

For each variant of the system from Fig. 7, a linear approximation of the dependence of the probability of detecting a spacecraft on the number of the experiment series $k = 1$ is performed... 50. The results are presented in Table 5.

Table 5

Approximate functions for graphs from Fig. 9

System variant	Linear approximation
Basic system	$y = 0,8212 + 0,00102 \cdot k$
Option 1	$y = 0,8607 + 0,00106 \cdot k$
Option 2	$y = 0,9025 + 0,00104 \cdot k$
Option 3 (complete method)	$y = 0,9341 + 0,00103 \cdot k$

The proposed method of selecting historical data provides a statistically stable increase in the efficiency of detecting spacecraft compared to the deterministic approach of the basic system. Automatic selection of the depth of the historical data horizon increases the average probability of detection by about 4% relative to the baseline option. The addition of self-learning mechanisms provides an additional increase in detection accuracy by about 4.2%. Taking into account previous experience in the use of historical data allows you to achieve the highest integral efficiency and the shortest decision-making time for the selection of historical data. The comprehensive use of all the proposed improvements provides the best performance of the spacecraft detection subsystem and confirms the feasibility of implementing the developed method in the systems and means of countering spacecraft in the CM.

As a result of the experiments, it was found that the use of adaptive data selection taking into account the depth of the horizon, self-learning and previous experience provides an increase in the accuracy of attack detection compared to approaches that do not use these mechanisms. The results obtained confirm the effectiveness of the proposed method and the feasibility of its use in the systems of countering spacecraft in the CM.

Conclusions

The article considers the problem of data selection for systems and means of countering computer attacks in corporate networks in the context of constant growth of information volumes, high dynamism of information flows and limited computing resources. It is shown that the effectiveness of the functioning of modern systems for countering computer attacks largely depends not only on the quality of threat detection algorithms, but also on the timely provision of decision-making processes with relevant, relevant and most valuable data. It is substantiated that the use of the entire accumulated array of information leads to an increase in the time of its processing, an increase in information redundancy and a decrease in the efficiency of response to computer attacks, while an excessive reduction in data volumes can lead to the loss of important information necessary to identify complex or recurring threats.

A method of data selection for systems and means of countering computer attacks is proposed, which is based on the complex use of the depth of the analysis horizon, self-learning mechanisms and previous experience of the system functioning. A feature of the method is the integrated evaluation of data according to a set of criteria, which allows taking into account their current relevance, predicted usefulness in different time horizons, the results of previous use, and the ability to adaptively change the selection parameters in accordance with changes in the information environment. In contrast to existing approaches, the proposed method provides an intelligent choice of data, taking into account the relationship between the historical effectiveness of information, its prospective significance and the results of the system's self-learning.

To implement the proposed method, an architecture of the memory subsystem of systems and means of countering computer attacks has been developed, which provides accumulation, classification, optimization, modification and adaptive selection of data in accordance with the needs of means of different levels. The proposed architecture allows maintaining a balance between completeness, duration of storage and relevance of information, reducing information redundancy, reducing access time to critical data and increasing the efficiency of information support of decision-making processes.

The practical significance of the results obtained lies in the possibility of using the proposed method in the development of intelligent memory subsystems of systems and means of countering computer attacks in corporate networks, which provides an increase in the accuracy of data selection, acceleration of information analysis processes, improvement of the adaptability of systems to changes in operating conditions and increase in the efficiency of detection and neutralization of computer attacks.

Prospects for further research are the development of a mathematical model for adaptive adjustment of the weights of the data selection method, the improvement of self-learning mechanisms using modern methods of artificial intelligence, and the experimental assessment of the effectiveness of the proposed method in real corporate networks under various scenarios of computer attacks.

ADDITIONAL INFORMATION

AUTHOR CONTRIBUTIONS

Conceptualization, A.K.; methodology, V.P.; validation, A.K.; formal analysis, V.P.; investigation, V.P.; data curation, A.K.; writing-original draft preparation, V.P.; writing-review and editing, A.K.; visualization, V.P.;

supervision, A.K.; project administration, A.K. All authors have read and agreed to the published version of the manuscript.

DECLARATION ON THE USE OF GENERATIVE ARTIFICIAL INTELLIGENCE TOOLS

In preparing this work, the author used DeepL Translate and Grammarly for: grammar and spelling checks, paraphrasing, and rephrasing. After using these tools/services, the author reviewed and edited the content and takes full responsibility for the content of this publication.

REFERENCES

1. Kashtalian A., Ścisło Ł., Rucki R., Lysenko S., Sachenko A., Savenko B., Savenko O., Nicheporuk A. Control and Decision-Making in Deceptive Multi-Computer Systems Based on Previous Experience for Cybersecurity of Critical Infrastructure. *Applied Sciences*, 2025. Vol. 15(22), 12286. DOI: <https://doi.org/10.3390/app152212286>
2. Kashtalian A., Savenko O., Sachenko A. Agglomerative clustering of data collected by honeypots. *2021 11th IEEE international conference on intelligent data acquisition and advanced computing systems: technology and applications (IDAACS)*, Cracow, Poland, 22–25 September 2021. DOI: <https://doi.org/10.1109/idaacs53288.2021.9661027>
3. Kashtalian A., Lysenko S., Sachenko A., Savenko B., Savenko O., Nicheporuk A. Evaluation criteria of centralization options in the architecture of multicomputer systems with traps and baits. *Radioelectronic and Computer Systems*. 2025. Vol. 1. Pp. 264-297. DOI: <https://doi.org/10.32620/reks.2025.1.18>
4. Kashtalian A., Lysenko S., Kysil T., Sachenko A., Savenko O., Savenko B. Method and Rules for Determining the Next Centralization Option in Multicomputer System Architecture. *International Journal of Computing*. 2025. Vol. 24(1), 35-51. DOI: <https://doi.org/10.47839/ijc.24.1.3875>
5. Usta M., Turkmen H.I., Guvensan M.A. Distant and Recent Historical Data Fusion for Improving Short- and Medium-Term Traffic Forecasting. *Applied Sciences*. 2025; 15(24):13130. DOI: <https://doi.org/10.3390/app152413130>
6. del Pozo F.E. Jr., Kim C.K., Kim H.-G. Refining the Selection of Historical Period in Analog Ensemble Technique. *Energies*. 2023; 16(22):7630. DOI: <https://doi.org/10.3390/en16227630>
7. Ahmed Y., Azad M.A., Asyhari T. Rapid Forecasting of Cyber Events Using Machine Learning-Enabled Features. *Information*. 2024; 15(1):36. DOI: <https://doi.org/10.3390/info15010036>
8. Szczypiorski K. Cybersecurity and Data Science. *Electronics*. 2022; 11(15):2309. DOI: <https://doi.org/10.3390/electronics11152309>
9. Pimenta Rodrigues G.A., Marques Serrano A.L., Lopes Espiñeira Lemos A.N., Canedo E.D., Mendonça F.L.L.d., de Oliveira Albuquerque R., Sandoval Orozco A.L., García Villalba L.J. Understanding Data Breach from a Global Perspective: Incident Visualization and Data Protection Law Review. *Data*. 2024; 9(2):27. DOI: <https://doi.org/10.3390/data9020027>
10. Pinto D., Amorim I., Maia E., Praça I. A review on intrusion detection datasets: tools, processes, and features. *Computer Networks*. 2025, Volume 262, 111177, ISSN 1389-1286. DOI: <https://doi.org/10.1016/j.comnet.2025.111177>
11. Hafsa M., Jemili F. Comparative Study between Big Data Analysis Techniques in Intrusion Detection. *Big Data and Cognitive Computing*. 2019; 3(1):1. DOI: <https://doi.org/10.3390/bdcc3010001>
12. Komisarek M., Pawlicki M., Kozik R., Hołubowicz W., Choraś M. How to Effectively Collect and Process Network Data for Intrusion Detection? *Entropy*. 2021; 23(11):1532. DOI: <https://doi.org/10.3390/e23111532>
13. Szczepanski M., Komisarek M., Pawlicki M., Kozik R., Choraś M. The Proposition of Balanced and Explainable Surrogate Method for Network Intrusion Detection in Streamed Real Difficult Data. In: Wojtkiewicz, K., Treur, J., Pimenidis, E., Maleszka, M. (eds) *Advances in Computational Collective Intelligence. ICCCI 2021. Communications in Computer and Information Science*, vol 1463. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-88113-9_19
14. Ullah F., Turab A., Ullah S., Cacciagrano D., Zhao Y. Enhanced Network Intrusion Detection System for Internet of Things Security Using Multimodal Big Data Representation with Transfer Learning and Game Theory. *Sensors*. 2024; 24(13):4152. DOI: <https://doi.org/10.3390/s24134152>
15. Samed A.I., Murat Dener. STL-HDL: A new hybrid network intrusion detection system for imbalanced dataset on big data environment, *Computers & Security*, 2021, Volume 110, 102435, ISSN 0167-4048. DOI: <https://doi.org/10.1016/j.cose.2021.102435>
16. Termanini A, Bourdoucen H, Al-Abri D, Al Maashri A. Intrusion Detection Datasets for IIoT and ICS: A Taxonomic Review with a Decision-Aid Scoring Rubric. *Sensors*. 2026; 26(13):4099. DOI: <https://doi.org/10.3390/s26134099>

Вадим ПАЮК, Антоніна КАШТАЛЬЯН
Хмельницький національний університет

ВИБІР ДАНИХ В СИСТЕМАХ ТА ЗАСОБАХ ПРОТИДІЇ КОМП'ЮТЕРНИМ АТАКАМ З УРАХУВАННЯМ ІСТОРИЧНОГО АСПЕКТУ ЇХ ВИКОРИСТАННЯ

У статті розглянуто проблему вибору даних для систем та засобів протидії комп'ютерним атакам у корпоративних мережах в умовах постійного зростання обсягів інформації, високої динамічності інформаційних потоків та обмеженості обчислювальних ресурсів. Показано, що ефективність функціонування сучасних систем протидії комп'ютерним атакам значною мірою залежить не лише від якості алгоритмів виявлення загроз, а й від своєчасного забезпечення процесів прийняття рішень актуальними, релевантними та найбільш цінними даними. Обґрунтовано, що використання всього накопиченого масиву інформації призводить до зростання часу її оброблення, збільшення інформаційної надлишковості та зниження оперативності реагування на комп'ютерні атаки, тоді як надмірне скорочення обсягів даних може спричинити втрату важливої інформації, необхідної для виявлення складних або повторюваних загроз.

Запропоновано метод вибору даних для систем та засобів протидії комп'ютерним атакам, який базується на комплексному використанні глибини горизонту аналізу, механізмів самонавчання та попереднього досвіду функціонування системи. Особливістю методу є інтегроване оцінювання даних за сукупністю критеріїв, що дозволяє враховувати їхню поточну актуальність, прогнозовану корисність у різних часових горизонтах, результати попереднього використання та здатність адаптивно змінювати параметри вибору відповідно до змін інформаційного середовища. На відміну від існуючих підходів, запропонований метод забезпечує інтелектуальний вибір даних із урахуванням взаємозв'язку між історичною ефективністю інформації, її перспективною значущістю та результатами самонавчання системи.

Для реалізації запропонованого методу розроблено архітектуру підсистеми пам'яті систем та засобів протидії комп'ютерним атакам, яка забезпечує накопичення, класифікацію, оптимізацію, модифікацію та адаптивний вибір даних відповідно до потреб засобів різних рівнів. Запропонована архітектура дозволяє підтримувати баланс між повнотою, тривалістю зберігання та актуальністю інформації, зменшувати інформаційну надлишковість, скорочувати час доступу до критично важливих даних і підвищувати ефективність інформаційного забезпечення процесів прийняття рішень.

Практичне значення отриманих результатів полягає у можливості використання запропонованого методу під час розроблення інтелектуальних підсистем пам'яті систем та засобів протидії комп'ютерним атакам у корпоративних мережах, що забезпечує підвищення точності вибору даних, прискорення процесів аналізу інформації, покращення адаптивності систем до зміни умов функціонування та підвищення ефективності виявлення й нейтралізації комп'ютерних атак.

Перспективи подальших досліджень полягають у розробленні математичної моделі адаптивного налаштування вагових коефіцієнтів методу вибору даних, удосконаленні механізмів самонавчання з використанням сучасних методів штучного інтелекту та експериментальній оцінці ефективності запропонованого методу в реальних корпоративних мережах за різних сценаріїв реалізації комп'ютерних атак.

Ключові слова: корпоративні мережі, комп'ютерні атаки, оптимізація даних, історичні дані, вибір даних, інформаційна надлишковість, самонавчання, глибина горизонту для даних