

ХМЕЛЬЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МІНІСТЕРСТВА ОСВІТИ І НАУКИ УКРАЇНИ

Кваліфікаційна наукова праця
на правах рукопису

ПАЮК ВАДИМ ПАВЛОВИЧ

УДК 004.3:004.75:004.49:004.52:004.5

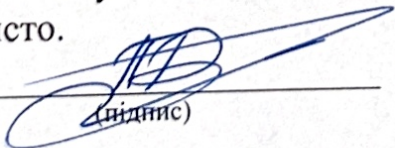
ДИСЕРТАЦІЯ

**МЕТОДИ ТА ЗАСОБИ ЗБЕРІГАННЯ, ВИБОРУ ТА ОПТИМІЗАЦІЇ ІСТОРИЧНИХ
ДАНИХ СИСТЕМ І ЗАСОБІВ ПРОТИДІЇ КОМП'ЮТЕРНИМ АТАКАМ В
КОРПОРАТИВНИХ МЕРЕЖАХ**

123 Комп'ютерна інженерія

12 Інформаційні технології

Подається на здобуття наукового ступеня доктора філософії
Дисертація містить результати власних досліджень. Подані до захисту наукові
положення є власним напрацюванням. Всі використані ідеї, наукові результати,
цитати супроводжуються належними посиланнями на їх авторів та джерела
опублікування. Всі частини тексту дисертації, під час написання яких
використовувались технології штучного інтелекту, перевірені та відредаговані
особисто.



(підпис)

В. П. Паюк

Наукові керівники:

Каштальян Антоніна Сергіївна, доктор технічних наук, доцент
Савенко Богдан Олегович, доктор філософії

Хмельницький – 2026

АНОТАЦІЯ

Паюк В.П. Методи та засоби зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії з галузі знань 12 Інформаційні технології за спеціальністю 123 Комп'ютерна інженерія. – Хмельницький національний університет, Хмельницький 2026.

У роботі подано результати дослідження, спрямованого на підвищення ефективності зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах шляхом удосконалення архітектури та методів організації зберігання, вибору та оптимізації історичних даних. Запропоновано удосконалену архітектуру таких систем, у якій синтезовано підходи до зберігання, оптимізації та вибору історичних даних для систем протидії комп'ютерним атакам та зловмисному програмному забезпеченню. Завдяки цьому досягається підвищення оперативності та точності при наданні даних системам протидії комп'ютерним атакам та зловмисному програмному забезпеченню, а також забезпечення ефективності їх зберігання з урахування глибини горизонту.

У дисертації здійснено обґрунтування актуальності наукової задачі, що полягає у підвищенні ефективності зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах шляхом удосконалення архітектури та методів організації зберігання, вибору та оптимізації історичних даних. Перспективним напрямком досліджень для підвищення ефективності протидії комп'ютерним атакам в корпоративних мережах визначено організацію зберігання, оптимізацію та вибір історичних даних з врахуванням глибини горизонту. В роботі удосконалено архітектуру засобів зберігання даних для систем і засобів протидії КА, розроблено метод організації зберігання даних у системах та засобах протидії комп'ютерним атакам, метод оптимізації даних на основі самонавчання засобів підсистеми пам'яті систем та засобів протидії КА та метод вибору даних, а також розроблено відповідний засіб зберігання даних для

систем і засобів протидії КА, здійснено постановку експериментів та проведені дослідження із ним.

Об'єктом дослідження є процес організації зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах

Предметом дослідження є методи організації зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах

Метою дисертаційного дослідження є підвищення ефективності зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах шляхом удосконалення архітектури та методів організації зберігання, вибору та оптимізації історичних даних.

Наукова новизна отриманих результатів полягає в наступному:

1) удосконалено архітектуру засобів зберігання даних для систем і засобів протидії комп'ютерним атакам, в якій на відміну відомих, здійснено поєднання автоматичного керування життєвим циклом всіх даних або їх частини, особливостей апаратно-програмних засобів для зберігання даних, інтелектуалізації щодо глибини горизонту для даних, самонавчання при виборі даних для типових повторюваних комп'ютерних атак, що дає змогу підвищити оперативність та точність при наданні даних системам протидії комп'ютерним атакам, а також забезпечити ефективність їх зберігання з урахування глибини горизонту;

2) вперше розроблено метод організації зберігання даних у системах та засобах протидії комп'ютерним атакам, особливістю якого є поділ даних на початкові, накопичувані та використані, а також врахування історичного аспекту їх застосування шляхом збереження результатів оцінювання ефективності використання даних із застосуванням марковських моделей, динамічних баєсівських мереж та статистичних методів, що забезпечує опис динаміки станів системи, врахування причинно-наслідкових залежностей і узагальнення накопиченого досвіду для підвищення ефективності повторного використання даних, адаптивності систем та автономності їх функціонування в умовах тривалої експлуатації;

3) розроблено новий метод оптимізації даних на основі самонавчання засобів підсистеми пам'яті систем та засобів протидії КА, який, на відміну від існуючих, забезпечує комплексну адаптивну оптимізацію початкових, накопичених, використаних та даних обманних систем з урахуванням критеріїв ресурсної ефективності, структури даних, шарового розміщення, апаратно-мережного розподілу та специфіки функціонування обманних систем, що дозволяє зменшити надлишковість інформаційних масивів, підвищити швидкодію доступу до даних, покращити інформативність інформаційного середовища, ефективність накопичення знань про дії зловмисника та адаптивність систем і засобів протидії КА до змін умов функціонування;

4) розроблено новий метод вибору даних, який, на відміну від відомих, що базуються на статичному оцінюванні ознак або фіксованих моделях ранжування, використовує інтегральну оптимізаційну функцію з урахуванням глибини горизонту, семантичної значущості та попереднього досвіду використання, доповнену механізмом самонавчання для адаптивного формування та корекції інтегрованих сукупностей даних, що дозволило підвищити точність відбору, зменшити інформаційну надлишковість і забезпечити динамічну адаптацію до змін середовища.

Практичне значення отриманих результатів. Розроблено архітектуру засобів зберігання даних для систем і засобів протидії КА, особливістю якої є поєднання автоматичного керування життєвим циклом даних, врахування особливостей апаратно-програмних засобів зберігання, інтелектуалізації глибини горизонту даних та механізмів самонавчання при виборі даних для типових повторюваних КА.

Розроблено метод організації зберігання даних у системах і засобах протидії КА, який забезпечує поділ даних на початкові, накопичувані та використані, а також врахування історичних аспектів їх застосування на основі оцінювання ефективності використання даних із використанням марковських моделей, динамічних баєсівських мереж і статистичних методів. Синтез методів оптимізації та самонавчання в архітектурі підсистем пам'яті систем і засобів протидії КА дав змогу реалізувати адаптивне керування процесами оптимізації початкових, накопичених, використаних даних та даних обманних систем. Це забезпечило зниження надлишковості

інформаційних масивів, підвищення швидкодії доступу до даних та покращення інформативності інформаційного середовища для підтримки процесів виявлення і протидії комп'ютерним атакам. Також розроблено метод вибору даних, який базується на використанні інтегральної оптимізаційної функції з урахуванням глибини горизонту, семантичної значущості та попереднього досвіду використання даних, доповнений механізмом самонавчання для адаптивного формування та коригування інтегрованих сукупностей даних. Це дозволило підвищити точність відбору даних, зменшити інформаційну надлишковість і забезпечити динамічну адаптацію систем і засобів протидії КА до змін умов функціонування.

За результатами проведених експериментальних досліджень встановлено, що розроблені архітектурні рішення та методи забезпечують ефективне функціонування засобів зберігання даних в умовах динамічної зміни параметрів інформаційного середовища, підвищують оперативність надання даних системам протидії КА, покращують точність вибору релевантної інформації та забезпечують підвищення загальної ефективності функціонування систем і засобів протидії КА.

Теоретичні та практичні результати дослідження впроваджені в ТОВ «ІТТ» (м. Хмельницький), ПП «Авіві» (м. Хмельницький), а також в освітньому процесі Хмельницького національного університету при викладанні дисциплін на кафедрі комп'ютерної інженерії та інформаційних систем для спеціальності F7 Комп'ютерна інженерія, зокрема в курсах «Безпека та захист комп'ютерних систем», «Моделювання та методи оптимізації в наукових та експериментальних дослідженнях», «Методології забезпечення якості, надійності, гарантоздатності та безпеки комп'ютерних систем та мереж».

У вступі представлено обґрунтування актуальності наукової задачі, що полягає у підвищенні ефективності зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах шляхом удосконалення архітектури та методів організації зберігання, вибору та оптимізації історичних даних. Перспективним напрямком досліджень для підвищення ефективності протидії комп'ютерним атакам в корпоративних мережах визначено організацію зберігання, оптимізацію та вибір історичних даних з врахуванням

глибини горизонту. Продemonстровано зв'язок тематики дослідження з напрямками наукових досліджень науковців за цим спрямуванням, представлено основні наукові результати роботи, їх практичне використання, перелік підприємств та установ, в яких впроваджено результат роботи.

У першому розділі проведено аналіз предметної області дослідження, існуючих систем протидії комп'ютерним атакам в корпоративних мережах, методів, що застосовуються для виявлення комп'ютерних атак та зловмисного програмного забезпечення, а також методів зберігання, оптимізації та вибору історичних даних з урахуванням їх впливу на показники виявлення комп'ютерних атак.

У другому розділі визначено область дослідження, розроблено моделі організації пам'яті з історичними даними щодо функціонування систем та засобів протидії КА, моделі обманних технологій в системах та засобах протидії КА, удосконалено архітектуру засобів підсистеми пам'яті систем та засобів протидії КА в корпоративних мережах на основі використання історичних даних для прийняття рішень. При розробленні архітектури засобів здійснено поєднання автоматичного керування життєвим циклом всіх даних або їх частини, особливостей апаратно-програмних засобів для зберігання даних, інтелектуалізації щодо глибини горизонту для даних, самонавчання при виборі даних для типових повторюваних комп'ютерних атак.

У третьому розділі представлено розроблені такі методи: організації зберігання даних у системах та засобах протидії комп'ютерним атакам, особливістю якого є поділ даних на початкові, накопичувані та використані, а також врахування історичного аспекту їх застосування шляхом збереження результатів оцінювання ефективності використання даних із застосуванням марковських моделей, динамічних баєсівських мереж та статистичних методів; метод оптимізації даних на основі самонавчання засобів підсистеми пам'яті систем та засобів протидії комп'ютерним атакам, який забезпечує комплексну адаптивну оптимізацію початкових, накопичених, використаних та даних обманних систем з урахуванням критеріїв ресурсної ефективності, структури даних, шарового розміщення, апаратно-мережного розподілу та специфіки функціонування обманних систем; метод вибору даних, який

використовує інтегральну оптимізаційну функцію з урахуванням глибини горизонту, семантичної значущості та попереднього досвіду використання, доповнену механізмом самонавчання для адаптивного формування та корекції інтегрованих сукупностей даних.

У четвертому розділі представлено методику оцінювання ефективності засобів зберігання, вибору та оптимізації історичних даних систем та засобів протидії комп'ютерним атакам. Також, представлено постановку та результати експерименту і засіб зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах.

У висновках представлено отримані наукові та практичні результати проведеного дослідження.

У додатках представлено наукові публікації, в яких відображено наукові результати роботи, акти впровадження результатів роботи, лістинг розробленого програмного забезпечення та деталізацію результатів експериментів.

Ключові слова: критична інформація, комп'ютерні системи, корпоративні мережі, обманні системи, комп'ютерні атаки, системи виявлення вторгнень, архітектура систем, оптимізація даних, графові моделі, марковські моделі, статистичні моделі.

ANNOTATION

Payuk V.P. Methods and means of storage, selection and optimization of historical data of systems and means of countering computer attacks in corporate networks. – Qualifying scientific work on manuscript rights.

Dissertation for obtaining the scientific degree of Doctor of Philosophy in the field of knowledge 12 Information technologies in the specialty 123 Computer engineering. – Khmelnytskyi National University, Khmelnytskyi 2026.

The work presents the results of research aimed at improving the efficiency of storage, selection and optimization of historical data of systems and means of countering computer attacks in corporate networks by improving the architecture and methods of organizing

storage, selection and optimization of historical data. An improved architecture of such systems is proposed, which synthesizes approaches to storage, optimization, and selection of historical data for systems to counter computer attacks and malicious software. Thanks to this, an increase in efficiency and accuracy is achieved when providing data systems against computer attacks and malicious software, as well as ensuring the effectiveness of their storage, taking into account the depth of the horizon.

The thesis substantiates the relevance of the scientific task, which consists in increasing the efficiency of storage, selection and optimization of historical data of systems and means of countering computer attacks in corporate networks by improving the architecture and methods of organizing storage, selection and optimization of historical data. The organization of storage, optimization and selection of historical data, taking into account the depth of the horizon, is defined as a promising direction of research to improve the effectiveness of combating computer attacks in corporate networks. The work improved the architecture of data storage facilities for anti-aircraft defense systems and means, developed a method of organizing data storage in systems and means of countering computer attacks, a method of data optimization based on self-learning of the memory subsystem of anti-aircraft defense systems and means, and a data selection method, as well as developed a suitable data storage facility for anti-aircraft defense systems and means, set up experiments and conducted research with it.

The object of the study is the process of organizing the storage, selection and optimization of historical data of systems and means of countering computer attacks in corporate networks

The subject of the research is methods of organizing storage, selection and optimization of historical data of systems and means of countering computer attacks in corporate networks

The aim of the dissertation research is to improve the efficiency of storage, selection and optimization of historical data of systems and means of combating computer attacks in corporate networks by improving the architecture and methods of organizing storage, selection and optimization of historical data.

The scientific novelty of the obtained results is as follows:

1) the architecture of data storage facilities for anti-aircraft systems and anti-aircraft vehicles has been improved, in which, in contrast to the known, a combination of automatic life cycle management of all or part of data, features of hardware and software for data storage, intellectualization regarding the depth of the horizon for data, self-learning when selecting data for typical repetitive anti-aircraft vehicles, which makes it possible to increase the efficiency and accuracy when providing data to anti-aircraft systems, as well as to ensure the efficiency of their storage taking into account the depth of the horizon;

2) for the first time, a method of organizing data storage in systems and means of countering computer attacks was developed, the feature of which is the division of data into initial, accumulated and used, as well as taking into account the historical aspect of their use by saving the results of evaluating the effectiveness of data use with the use of Markov models, dynamic Bayesian networks and statistical methods, which provides a description of the dynamics of system states, taking into account cause-and-effect relationships and generalizing accumulated experience to improve efficiency reuse of data, adaptability of systems and autonomy of their functioning in conditions of long-term operation;

3) a new data optimization method has been developed based on self-learning of the system memory subsystem and anti-aircraft countermeasures, which, unlike the existing ones, provides comprehensive adaptive optimization of initial, accumulated, used and data deception systems taking into account the criteria of resource efficiency, data structure, layer placement, hardware and network distribution and the specifics of the functioning of deception systems, which allows to reduce the redundancy of information arrays, increase the speed of data access, improve the informativeness of the information environment, the effectiveness of the accumulation of knowledge about the actions of the violator and the adaptability of systems and means of countermeasures to the changes in operating conditions;

4) a new method of data selection was developed, which, unlike the known ones based on static evaluation of features or fixed ranking models, uses an integral optimization function taking into account the depth of the horizon, semantic significance and previous experience of use, supplemented by a self-learning mechanism for adaptive formation and correction of integrated data sets, which made it possible to increase the accuracy of

selection, reduce information redundancy and ensure dynamic adaptation to changes environment.

Practical significance of the obtained results. An architecture of data storage facilities for anti-aircraft systems and means has been developed, the feature of which is a combination of automatic management of the life cycle of data, taking into account the peculiarities of hardware and software storage facilities, intellectualization of the depth of the data horizon, and self-learning mechanisms when selecting data for typical repetitive airfields.

A method of organizing data storage in anti-aircraft systems and means has been developed, which ensures the division of data into initial, accumulated and used, as well as taking into account the historical aspects of their application based on the evaluation of the effectiveness of data use using Markov models, dynamic Bayesian networks and statistical methods. The synthesis of optimization methods and self-learning in the architecture of system memory subsystems and anti-aircraft means made it possible to implement adaptive management of optimization processes of initial, accumulated, used data and data of deception systems. This ensured a reduction in the redundancy of information arrays, an increase in the speed of data access, and an improvement in the informativeness of the information environment to support the processes of detection and countermeasures against computer attacks. A data selection method was also developed, which is based on the use of an integral optimization function taking into account the depth of the horizon, semantic significance and previous experience of using data, complemented by a self-learning mechanism for adaptive formation and correction of integrated data sets. This made it possible to increase the accuracy of data selection, reduce information redundancy and ensure dynamic adaptation of systems and anti-aircraft countermeasures to changes in operating conditions.

According to the results of the experimental studies, it was found that the developed architectural solutions and methods ensure the effective functioning of data storage facilities in conditions of dynamic changes in the parameters of the information environment, increase the efficiency of providing data to anti-aircraft systems, improve the accuracy of selecting

relevant information, and ensure an increase in the overall efficiency of the functioning of anti-aircraft systems and means.

The theoretical and practical results of the research are implemented in LLC "ITT" (Khmelnyskyi), PE "Avivi" (Khmelnyskyi), as well as in the educational process of the Khmelnyskyi National University when teaching disciplines at the Department of Computer Engineering and Information Systems for the specialty F7 Computer Engineering, in particular in the courses "Security and protection of computer systems", "Modeling and optimization methods in scientific and experimental research", "Methodologies for ensuring the quality, reliability, warranty and security of computer systems and networks".

The introduction presents the justification of the relevance of the scientific task, which consists in increasing the efficiency of storage, selection and optimization of historical data of systems and means of countering computer attacks in corporate networks by improving the architecture and methods of organizing the storage, selection and optimization of historical data. The organization of storage, optimization and selection of historical data, taking into account the depth of the horizon, is defined as a promising direction of research to improve the effectiveness of combating computer attacks in corporate networks. The connection of the research topic with the directions of scientific research of scientists in this direction is demonstrated, the main scientific results of the work, their practical use, the list of enterprises and institutions in which the results of the work are implemented are presented.

In the first chapter, an analysis of the subject area of the study, existing systems for countering computer attacks in corporate networks, methods used to detect computer attacks and malicious software, as well as methods of storage, optimization and selection of historical data, taking into account their impact on the indicators of detection of computer attacks, is carried out.

In the second chapter, the research area is defined, memory organization models with historical data on the functioning of anti-aircraft systems and countermeasures, models of deceptive technologies in antiaircraft systems and countermeasures are developed, the architecture of the memory subsystem of systems and antiaircraft countermeasures in

corporate networks is improved based on the use of historical data for decision-making. When developing the architecture of the tools, a combination of automatic management of the life cycle of all or part of the data, features of hardware and software for data storage, intellectualization regarding the depth of the data horizon, self-learning when selecting data for typical repeated computer attacks was carried out.

The third section presents the following developed methods: the organization of data storage in systems and means of countering computer attacks, the feature of which is the division of data into initial, accumulated and used, as well as taking into account the historical aspect of their application by saving the results of evaluating the effectiveness of data use using Markov models, dynamic Bayesian networks and statistical methods; a method of data optimization based on self-learning of the system memory subsystem and means of countering computer attacks, which provides comprehensive adaptive optimization of the initial, accumulated, used and data of fraudulent systems, taking into account the criteria of resource efficiency, data structure, layer placement, hardware and network distribution, and the specifics of the functioning of fraudulent systems; a data selection method that uses an integral optimization function taking into account horizon depth, semantic significance, and previous usage experience, complemented by a self-learning mechanism for adaptively forming and correcting integrated data sets.

The fourth chapter presents the methodology for evaluating the effectiveness of means of storage, selection and optimization of historical system data and means of countering computer attacks. Also, the setting and results of the experiment and the means of storage, selection and optimization of historical data of systems and means of countering computer attacks in corporate networks are presented.

The scientific and practical results of the conducted research are presented in the conclusions.

The appendices present scientific publications that reflect the scientific results of the work, acts of implementation of the results of the work, a listing of the developed software and details of the results of the experiments.

Keywords: critical information, computer systems, corporate networks, deception systems, computer attacks, intrusion detection systems, system architecture, data optimization, graph models, Markov models, statistical models.

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Наукові праці, в яких опубліковані основні наукові результати дисертації

1. Паюк В.П., Савенко Б.О. Організація зберігання даних в системах та засобах протидії комп'ютерним атакам з урахуванням історичного досвіду їх використання. *Вимірювальна та обчислювальна техніка в технологічних процесах*. 2026. № 2. С. 321–336. DOI: <https://doi.org/10.31891/2219-9365-2026-86-39>

2. PAIUK V., KASHTALIAN A. SELECTION OF DATA IN SYSTEMS AND TOOLS FOR COMPUTER ATTACK DEFENSE TAKING INTO ACCOUNT THE HISTORICAL ASPECT OF THEIR USE. *Computer Systems and Information Technologies*. 2026. № 3. Pp. 186–204. DOI: <https://doi.org/10.31891/csit-2026-3-19>

3. Паюк В. П., Савенко Б.О. Оптимізація зберігання даних в системах та засобах протидії комп'ютерним атакам з урахуванням історичного аспекту їх використання. *Вимірювальна та обчислювальна техніка в технологічних процесах*. 2026. № 3. С. 217–236. DOI: <https://doi.org/10.31891/2219-9365-2026-87-26>

4. Паюк В. П., Гейдарова О. В., Косенков В. Д., Савенко О. С. Застосування багатфакторного аналізу з метою виявлення незадокументованих закладок програмного забезпечення в локальних комп'ютерних мережах. *Вісник Хмельницького національного університету. Технічні науки*. 2020. № 5. С. 280-285. DOI 10.31891/2307-5732-2020-289-5-280-285

Праці, які засвідчують апробацію матеріалів дисертації

5. Sierhieiev Y., Paiuk V., Sachenko A., Nicheporuk A., Kwiecien A. A graph-based vulnerability detection method. (2024) *CEUR Workshop Proceedings*, 3675, pp. 388-401. *The 5th International Workshop on Intelligent Information Technologies & Systems of Information Security (IntelITSIS-2024)* : CEUR-Workshop Proceedings. Vol. 3675.

(Khmelnyskyi, March 2024). Khmelnyskyi, 2024. Pp. 343-355. URL: <https://ceur-ws.org/Vol-3675/>

6. Sierhieiev Y., Paiuk V., Nicheporuk A., Kwiecien A., Huralnyk O. Detection and prediction of the vulnerabilities in software systems based on behavioral analysis with machine learning. *The 1th Proceedings of the 1st International Workshop on Intelligent & CyberPhysical Systems (ICyberPhyS 2024) Khmelnyskyi, Ukraine, June 28, 2024 : CEUR-Workshop Proceedings*. Vol. 3736. 2024. Pp. 239-254. URL: <https://ceur-ws.org/Vol-3736/>

7. Sierhieiev Y., Savenko O., Paiuk V., Drozd A. Effectiveness and improvement of Static Application Security Testing (SAST) in the context of SQL Injection vulnerabilities // *Proceedings of 2024 IEEE 14th International Conference on Dependable Systems, Services and Technologies (DeSSerT-2024, Athens, Greece, October 11-13, 2024)*. DOI: [10.1109/DESSERT65323.2024.11122171](https://doi.org/10.1109/DESSERT65323.2024.11122171)

8. Huralnyk O., Kwiecien A., Paiuk V., Klein O., Lyhun O. Method and Means of Critical Information Processing in Corporate Networks. *Intelitsis '25: The 6th International Workshop on Intelligent Information Technologies & Systems of Information Security*, Aprile 04, 2025, Khmelnyskyi, Ukraine. Pp. 314-328. URL: <https://ceur-ws.org/Vol-3963/paper25.pdf>

9. Stetsyuk Y., Stetsyuk M., Voznyi K., Paiuk V., Kvassay M. A model of a centralized security system, as an information technology for the synthesis of an OS architecture protected against the leakage of confidential information. *1st International Workshop on Advanced Applied Information Technologies: (AdvAIT-2024)*, Khmelnyskyi, Ukraine and Zilina, Slovakia, December 5, 2024. Vol. 3899. P. 224-233. URL: <https://ceur-ws.org/Vol-3899/paper20.pdf>

10. Paiuk V., Kosenkov V., Savenko O., Nicheporuk A., Geidarova O. Application of multifactor analysis for the purpose of detecting malicious software implants of the software in local computer networks. *2nd International Workshop on Intelligent Information Technologies and Systems of Information Security*, March 24–26, 2021. *Ceur Workshop Proceedings*. Vol. 2853. Pp. 517–526. URL: <https://ceur-ws.org/Vol-2853/paper46.pdf>

11. Nicheporuk A., Paiuk V., Savenko B., Savenko O., Geidarova O. Detecting Software Malicious Implant Based on Anomalies Research on Local Area Networks. *Ceur*

Workshop Proceedings, 1st International Workshop on Intelligent Information Technologies and Systems of Information Security, June 10-12, 2020. Khmelnytskyi, Ukraine 2020, Vol. 2623. Pp. 194–207. URL: <https://ceur-ws.org/Vol-2623/paper18.pdf>

Публікації, які додатково відображають наукові результати дисертації

12. Свідоцтво про реєстрацію авторського права на твір (програму) № 150528 Україна. Комп'ютерна програма «Програмне забезпечення зберігання, оптимізації і вибору даних в системах та засобах протидії комп'ютерним атакам з урахуванням історичного аспекту їх використання та глибини горизонту» («SECMON») / Нічепорук А. О., Паюк В.П., Савенко Б.О., Каштальян А.С. Дата реєстрації 03.09.2026.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ.....	18
ВСТУП.....	19
РОЗДІЛ 1 АНАЛІЗ МЕТОДІВ ТА ЗАСОБІВ ЗБЕРІГАННЯ, ВИБОРУ ТА ОПТИМІЗАЦІЇ ІСТОРИЧНИХ ДАНИХ СИСТЕМ І ЗАСОБІВ ПРОТИДІЇ КОМП'ЮТЕРНИМ АТАКАМ В КОРПОРАТИВНИХ МЕРЕЖАХ	30
1.1 Поняття про комп'ютерні атаки та вплив історичних даних на їх виявлення.	30
1.2 Методи та технології зберігання даних	37
1.3 Методи і засоби оптимізації даних систем та засобів протидії комп'ютерним атакам	43
1.4 Методи і засоби вибору даних з урахуванням характеристик та показників операційного середовища	47
1.5 Постановка задачі дослідження.....	51
1.6. Висновки до першого розділу.....	52
РОЗДІЛ 2 АРХІТЕКТУРА СИСТЕМ ПРОТИДІЇ КОМП'ЮТЕРНИМ АТАКАМ В КОРПОРАТИВНІ МЕРЕЖІ НА ОСНОВІ ОБМАННИХ ТЕХНОЛОГІЙ ТА ПАМ'ЯТІ З ІСТОРИЧНИМИ ДАНИМИ ЩОДО ФУНКЦІОНУВАННЯ.....	55
2.1 Визначення області дослідження	56
2.2 Моделі організації пам'яті з історичними даними щодо функціонування систем та засобів протидії КА в корпоративних мережах	67
2.3 Моделі обманних технологій в системах та засобах протидії КА в корпоративних мережах	80
2.4 Архітектура засобів підсистеми пам'яті систем та засобів протидії КА в корпоративних мережах на основі використання історичних даних для прийняття рішень	99
2.5 Висновки до другого розділу.....	113
РОЗДІЛ 3 МЕТОДИ ЗБЕРІГАННЯ, ВИБОРУ ТА ОПТИМІЗАЦІЇ ДАНИХ ДЛЯ СИСТЕМ ТА ЗАСОБІВ ПРОТИДІЇ КОМП'ЮТЕРНИМ АТАКАМ	114

3.1 Метод організації зберігання даних в системах та засобах протидії комп'ютерним атакам з урахуванням історичного аспекту їх використання ...	114
3.2 Метод оптимізації даних на основі самонавчання засобів підсистеми пам'яті	135
3.3 Метод вибору даних з урахуванням глибини горизонту, самонавчання та попереднього досвіду використання.....	158
3.4 Висновки до третього розділу.....	173
РОЗДІЛ 4 ЕФЕКТИВНІСТЬ, ЕКСПЕРИМЕНТИ ТА ЗАСІБ ЗБЕРІГАННЯ, ВИБОРУ ТА ОПТИМІЗАЦІЇ ІСТОРИЧНИХ ДАНИХ СИСТЕМ І ЗАСОБІВ ПРОТИДІЇ КОМП'ЮТЕРНИМ АТАКАМ В КОРПОРАТИВНИХ МЕРЕЖАХ.	174
4.1 Методика оцінювання ефективності засобів зберігання, вибору та оптимізації історичних даних систем та засобів протидії комп'ютерним атакам	174
4.2 Постановка та результати експерименту і засіб зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах	179
4.3 Висновки до четвертого розділу.....	192
ВИСНОВКИ.....	194
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ	196
ДОДАТКИ.....	210
ДОДАТОК А Список публікацій здобувача	210
ДОДАТОК Б Акти впровадження.....	213
ДОДАТОК В Дані експериментів.....	217
ДОДАТОК Г Дослідження вибірки даних	226
ДОДАТОК Д Програмний код.....	231

ПЕРЕЛІК СКОРОЧЕНЬ

ІС – інформаційна система

ІТ – інформаційні технології

КА – комп'ютерна атака

КД – критичні дані

КІ – критична інформація

КМ – корпоративна мережа

КС – комп'ютерна система

ОС – операційна система

ПЗ – програмне забезпечення

APT-атака (Advanced Persistent Threat) – розширена стійка загроза

IDS (Intrusion Detection System) – система виявлення вторгнень

NDR (Network Detection and Response) – виявлення та реагування на мережні загрози

NIDS (Network Intrusion Detection System) – це система виявлення мережних вторгнень

PCAP – архівування пакетів

SDN (Software-Defined Network) – програмно-конфігурована мережа

SIEM (Security Information and Event Management) – система управління інформацією та подіями безпеки

SOC (Security Operations Center) – центр операцій з безпеки

Threat Hunting – проактивний процес пошуку прихованих кіберзагроз або слідів компрометації, які вже знаходяться в мережі, але ще не виявлені стандартними системами захисту

XDR (Extended Detection and Response) – розширене виявлення та реагування

ВСТУП

Обґрунтування вибору теми дослідження. Зловмисники при здійсненні комп'ютерних атак (КА) на корпоративні мережі (КМ) стикаються з протидією їм різних апаратно-програмних та програмних систем, що відображено в класифікації індикаторів компрометації піраміди болю зловмисника. Долаючи перешкоди при проникненні в КМ зловмисники витрачають багато ресурсів та часу і в цьому процесі для них зростають складності, які пов'язані з проведенням КА. На кожному рівні піраміди болю зловмисника для забезпечення безпеки та захисту КМ використовуються різні апаратно-програмні та програмні засоби і системи. Всі вони залежать від даних щодо свого попереднього функціонування, які формують підсистему пам'яті з історичними даними. В залежності від етапу, на якому розміщуються засоби та системи протидії КА організація їх пам'яті з історичними даними є різною. Зловмисники розуміють ієрархію засобів та систем у співвіднесенні до рівнів піраміди болю зловмисника, а також особливо щодо організації в них підсистем пам'яті з історичними даними. Вважаючи елементи чи підсистеми пам'яті одними з основних джерел попереднього досвіду функціонування для протидії КА, зловмисники прагнуть обходити не стільки сенсори засобів та систем протидії КА, а саме пам'яті з історичними даними щодо функціонування таких засобів та систем. Тому, організація елементів чи підсистем пам'яті з історичними даними щодо функціонування потребує постійного удосконалення за різними рівнями та формами, зокрема і з використанням обманних технологій, які активно інтегруються в апаратно-програмні та програмні засоби і системи протидії КА на КМ, для заплутування зловмисників під час проведення ними КА. Це дало б змогу підвищити рівень кібербезпеки в КМ за рахунок змінюваної організації елементів чи підсистем пам'яті з історичними даними щодо попереднього функціонування систем і засобів та з використанням відповідних обманних технологій. Ще однією з проблем, які потребують вирішення в контексті зберігання даних, є проблема оптимізації таких даних, бо їх кількість накопичується швидко і частина з них рідко використовується.

В процесі функціонування систем та засобів протидії КА в КМ виникає потреба вибору даних з наявних або їх модифікація для використання при виникненні певних подій в КМ. Такий вибір має забезпечуватись з урахуванням глибини горизонту, що потребує оцінювання актуальності даних. Для вирішення таких проблем із зберіганням, оптимізацією та вибором даних необхідно розробляти нові підходи для забезпечення їх ефективного впливу на процес протидії КА і КМ.

Ученими, які проводять наукові дослідження з розроблення нових принципів, моделей та методів виявлення КА в КМ, є Бісвас Д. [1], Віола В. [2], Гнатюк С. [3], Занг Ю. [4], Карпінський М. [5], Коваленко А. [6], Корченко А. [7], Корченко О. [8], Летичевський О. [9], Мухін В. [10], Ров Н. [11], Саченко А. [12], Сохор Т. [13], Харченко В. [14], Хофер В. [15], Хохлачова Ю. [16], Фергюсен-Вальтер К. [17], Євсєєв С. [7], Юдін О. [18] та ін.

Протиріччя полягає в тому, що системи та засоби протидії КА в КМ повинні забезпечувати повноту, тривалість і історичність зберігання даних для ефективного аналізу та виявлення загроз, але водночас змушені обмежувати обсяги зберігання і тривалість використання даних через втрату їх актуальності, обмеженість обчислювальних ресурсів та необхідність адаптивної обробки, яка може змінювати або витіснити раніше накопичені дані.

Тому, важливою та актуальною науковою задачею є підвищення ефективності зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах за рахунок удосконалення архітектури та методів організації зберігання, вибору та оптимізації історичних даних.

Зазначена науково-прикладна задача відповідає предметній області Стандарту вищої освіти України зі спеціальності 123 – Комп'ютерна інженерія для третього (освітньо-наукового) рівня вищої освіти, зокрема, такому об'єкту вивчення та діяльності, як « - ... комп'ютерні системи, локальні, глобальні комп'ютерні мережі та мережа Інтернет, ..., IT-інфраструктури, методи та способи подання, отримання, зберігання, передавання, опрацювання та захисту в них інформації, ..., архітектура та організація їх функціонування, інтерфейси та протоколи взаємодії їх компонентів, ...;

- інформаційні процеси, технології, методи, способи, інструментальні засоби та системи для дослідження, ... й експлуатації комп'ютерів та комп'ютерних систем і мереж, ..., IT-інфраструктур, розроблення, верифікації та розгортання програмного забезпечення та систем у хмарних та інших середовищах, а також процедури та засоби підтримки та керування життєвим циклом, забезпечення якості, надійності та безпеки.».

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційне дослідження виконувалось у рамках науково-дослідної тематики Хмельницького національного університету: держбюджетної науково-дослідної теми №2Б-2024 «Система виявлення ЗПЗ та комп'ютерних атак в корпоративних мережах з використанням хибних об'єктів атак та пасток» (ДР № 0124U000980); держбюджетної науково-дослідної теми №1Б-2026 «Система забезпечення стійкості до витоку конфіденційної інформації в корпоративних мережах в умовах впливів комп'ютерних атак» (ДР № 0126U002082), в яких автор дисертації був виконавцем.

Мета і завдання дослідження.

Об'єктом дослідження є процес організації зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах

Предметом дослідження є методи організації зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах

Метою дисертаційного дослідження є підвищення ефективності зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах шляхом удосконалення архітектури та методів організації зберігання, вибору та оптимізації історичних даних.

Завдання дослідження сформульовані наступним чином:

1) провести аналіз типів комп'ютерних атак, методів та систем їх виявлення в корпоративних мережах, а також моделей, методів і засобів зберігання, оптимізації та вибору даних з урахуванням історичного контексту та глибини горизонту;

2) розробити моделі організації пам'яті з історичними даними щодо функціонування систем та засобів протидії КА в КМ, моделі обманних технологій в системах та засобах протидії КА в корпоративних мережах для їх використання при розробленні засобів зберігання, оптимізації та вибору даних для систем і засобів протидії КА;

3) розробити архітектуру засобів зберігання даних для систем і засобів протидії КА на сонвії поєднання автоматичного керування життєвим циклом всіх даних або їх частини, особливостей апаратно-програмних засобів для зберігання даних, інтелектуалізації щодо глибини горизонту для даних, самонавчання при виборі даних для типових повторюваних КА;

4) розробити метод організації зберігання даних у системах та засобах протидії комп'ютерним атакам з використанням поділу даних на початкові, накопичувані та використані, а також врахування історичного аспекту їх застосування шляхом збереження результатів оцінювання ефективності використання даних із застосуванням марковських моделей, динамічних баєсівських мереж та статистичних методів;

5) розробити метод оптимізації даних на основі самонавчання засобів підсистеми пам'яті систем та засобів протидії КА для здійснення комплексної адаптивної оптимізації початкових, накопичених, використаних та даних обманних систем з урахуванням критеріїв ресурсної ефективності, структури даних, шарового розміщення, апаратно-мережного розподілу та специфіки функціонування обманних систем;

6) розробити метод вибору даних на основі статичного оцінювання ознак або фіксованих моделей ранжування, інтегральної оптимізаційної функції з урахуванням глибини горизонту, семантичної значущості та попереднього досвіду використання, доповнену механізмом самонавчання для адаптивного формування та корекції інтегрованих сукупностей даних;

7) розробити засоби для зберігання, оптимізації та вибору даних для систем та засобів протидії КА і КМ і провести експериментальні дослідження таких засобів з метою покращення їх характеристик та виявлення КА.

Методи дослідження. Для вирішення поставлених задач використовуються основні положення теорії розподілених систем, теорії множин та методи оптимізації.

1. Теорії розподілених систем для визначення архітектури засобів зберігання, оптимізації та вибору даних систем та засобів протидії КА.

2. Теорії множин для визначення компонентів засобів зберігання, оптимізації та вибору даних систем та засобів протидії КА в КМ.

3. Методи оптимізації для розроблення підсистеми оптимізації даних систем та засобів протидії КА в КМ.

Обґрунтованість і достовірність наукових положень, висновків і рекомендацій. Наукові положення, висновки і рекомендації дисертації забезпечуються коректним використанням сучасного математичного апарату, методів теорії ймовірностей, математичної статистики, марковських моделей, динамічних баєсівських мереж, методів оптимізації та інтелектуального аналізу даних при розв'язанні задач організації зберігання, вибору та оптимізації даних у системах і засобах протидії комп'ютерним атакам. Достовірність отриманих результатів підтверджується обґрунтованістю удосконаленої архітектури засобів зберігання даних для систем і засобів протидії КА, яка забезпечує автоматичне керування життєвим циклом даних, врахування особливостей апаратно-програмного середовища, глибини горизонту даних та механізмів самонавчання для вибору інформації при типових повторюваних комп'ютерних атаках.

Отримані теоретичні результати узгоджуються з результатами експериментальних досліджень, що підтверджує адекватність запропонованих моделей, методів і архітектурних рішень, а також їх ефективність при практичному застосуванні у системах і засобах протидії комп'ютерним атакам.

Наукова новизна отриманих результатів полягає в наступному:

1) удосконалено архітектуру засобів зберігання даних для систем і засобів протидії КА, в якій на відміну відомих, здійснено поєднання автоматичного керування життєвим циклом всіх даних або їх частини, особливостей апаратно-програмних засобів для зберігання даних, інтелектуалізації щодо глибини горизонту для даних, самонавчання при виборі даних для типових повторюваних КА, що дає

змогу підвищити оперативність та точність при надані даних системам протидії КА, а також забезпечити ефективність їх зберігання з урахування глибини горизонту;

2) вперше розроблено метод організації зберігання даних у системах та засобах протидії комп'ютерним атакам, особливістю якого є поділ даних на початкові, накопичувані та використані, а також врахування історичного аспекту їх застосування шляхом збереження результатів оцінювання ефективності використання даних із застосуванням марковських моделей, динамічних баєсівських мереж та статистичних методів, що забезпечує опис динаміки станів системи, врахування причинно-наслідкових залежностей і узагальнення накопиченого досвіду для підвищення ефективності повторного використання даних, адаптивності систем та автономності їх функціонування в умовах тривалої експлуатації;

3) розроблено новий метод оптимізації даних на основі самонавчання засобів підсистеми пам'яті систем та засобів протидії КА, який, на відміну від існуючих, забезпечує комплексну адаптивну оптимізацію початкових, накопичених, використаних та даних обманних систем з урахуванням критеріїв ресурсної ефективності, структури даних, шарового розміщення, апаратно-мережного розподілу та специфіки функціонування обманних систем, що дозволяє зменшити надлишковість інформаційних масивів, підвищити швидкодію доступу до даних, покращити інформативність інформаційного середовища, ефективність накопичення знань про дії зловмисника та адаптивність систем і засобів протидії КА до змін умов функціонування;

4) розроблено новий метод вибору даних, який, на відміну від відомих, що базуються на статичному оцінюванні ознак або фіксованих моделях ранжування, використовує інтегральну оптимізаційну функцію з урахуванням глибини горизонту, семантичної значущості та попереднього досвіду використання, доповнену механізмом самонавчання для адаптивного формування та корекції інтегрованих сукупностей даних, що дозволило підвищити точність відбору, зменшити інформаційну надлишковість і забезпечити динамічну адаптацію до змін середовища.

Практичне значення отриманих результатів. Розроблено архітектуру засобів зберігання даних для систем і засобів протидії КА, особливістю якої є поєднання

автоматичного керування життєвим циклом даних, врахування особливостей апаратно-програмних засобів зберігання, інтелектуалізації глибини горизонту даних та механізмів самонавчання при виборі даних для типових повторюваних КА.

Розроблено метод організації зберігання даних у системах і засобах протидії КА, який забезпечує поділ даних на початкові, накопичувані та використані, а також врахування історичних аспектів їх застосування на основі оцінювання ефективності використання даних із використанням марковських моделей, динамічних баєсівських мереж і статистичних методів. Синтез методів оптимізації та самонавчання в архітектурі підсистем пам'яті систем і засобів протидії КА дав змогу реалізувати адаптивне керування процесами оптимізації початкових, накопичених, використаних даних та даних обманних систем. Це забезпечило зниження надлишковості інформаційних масивів, підвищення швидкодії доступу до даних та покращення інформативності інформаційного середовища для підтримки процесів виявлення і протидії комп'ютерним атакам. Також розроблено метод вибору даних, який базується на використанні інтегральної оптимізаційної функції з урахуванням глибини горизонту, семантичної значущості та попереднього досвіду використання даних, доповнений механізмом самонавчання для адаптивного формування та коригування інтегрованих сукупностей даних. Це дозволило підвищити точність відбору даних, зменшити інформаційну надлишковість і забезпечити динамічну адаптацію систем і засобів протидії КА до змін умов функціонування.

За результатами проведених експериментальних досліджень встановлено, що розроблені архітектурні рішення та методи забезпечують ефективне функціонування засобів зберігання даних в умовах динамічної зміни параметрів інформаційного середовища, підвищують оперативність надання даних системам протидії КА, покращують точність вибору релевантної інформації та забезпечують підвищення загальної ефективності функціонування систем і засобів протидії КА.

Теоретичні та практичні результати дослідження впроваджені в ТОВ «ІТТ» (акт від 30.06.2026, м. Хмельницький), ПП «Авіві» (довідка від 29.06.2026, м. Хмельницький), а також в освітньому процесі Хмельницького національного університету (акт від 26.06.2026) при викладанні дисциплін на кафедрі комп'ютерної

інженерії та інформаційних систем для здобувачів спеціальності F7 Комп'ютерна інженерія, зокрема в курсах «Безпека та захист комп'ютерних систем», «Моделювання та методи оптимізації в наукових та експериментальних дослідженнях», «Методології забезпечення якості, надійності, гарантоздатності та безпеки комп'ютерних систем та мереж».

Особистий внесок здобувача. Всі основні результати дисертаційного дослідження, які представлені до захисту, отримані автором особисто.

У роботах, які опубліковані у співавторстві, автору належать основні ідеї, теоретична та практична розробка положень, відображених у характеристиці наукової новизни отриманих результатів, а саме: [111] – розроблено моделі даних та метод організації зберігання даних в системах та засобах протидії КА з урахуванням історичного досвіду їх використання; [112] – розроблено архітектуру засобів підсистеми пам'яті систем та засобів протидії КА в корпоративних мережах на основі використання історичних даних для прийняття рішень та метод вибору даних з урахуванням глибини горизонту, самонавчання та попереднього досвіду використання; [113] – розроблено метод оптимізації даних на основі самонавчання засобів підсистеми пам'яті; [114, 120] – розроблено методику застосування багатofакторного аналізу з метою зменшення кількості показників в задачах виявлення ЗПЗ; [115] – здійснення порівняльного аналізу альтернативних подань програмного коду та формування експериментальних сценаріїв; [116] – здійснено вибір методів машинного навчання до задач аналізу програмного коду та систематизація характеристик програмних систем; [117] – здійснено постановку експерименту щодо вразливостей застосунків; [118] – розроблено алгоритм захисту критичної інформації в корпоративних мережах для реалізації засобів; [119] – розробка експериментів для перевірки роботи модуля шифрування для різних станів системи централізованого керування; [121] – розроблено метод виявлення вставок в програмний код на основі дослідження аномалій при його виконанні в комп'ютерних мережах; [122] – розроблено програмне забезпечення організації зберігання, оптимізації та вибору даних в системах та засобах протидії КА з урахуванням історичного досвіду їх використання.

У роботах, які опубліковані у співавторстві, співавторам належать такі результати: [111] – запропоновано концепцію організації зберігання даних в системах та засобах протидії КА з урахуванням історичного досвіду їх використання- (Савенко Б.); [112] – запропоновано концепцію архітектури засобів підсистеми пам'яті систем та засобів протидії КА в корпоративних мережах на основі використання історичних даних для прийняття рішень (Каштал'ян А.); [113] – запропоновано концепцію оптимізації даних на основі самонавчання засобів підсистеми пам'яті (Савенко Б.); [114, 121] – запропоновано підхід до застосування багатофакторного аналізу з метою зменшення кількості показників в задачах виявлення ЗПЗ (Гейдарова О. В., Косенков В. Д.), верифіковано застосування багатофакторного аналізу з метою зменшення кількості показників в задачах виявлення ЗПЗ (Савенко О.С.); [115] – розроблення підходу до побудови графових моделей коду для задачі виявлення вразливостей, визначення схеми перетворення таких моделей у вхідні дані для нейронних мереж і обґрунтування переваг графового представлення порівняно з лінійним (Сергєєв Є.), здійснено постановку завдання дослідження та участь у аналізі результатів (Саченко А.), інтерпретація результатів експериментів (Kwiecień A., Нічепорук А.); [116] – розроблення підходу до аналізу поведінкових характеристик програмних систем з використанням методів машинного навчання для виявлення та прогнозування вразливостей, а також адаптація отриманих положень для формування показників ризику (Сергєєв Є.), участь у зборі та систематизації даних для завдання дослідження (Нічепорук А., Гуральник О.), верифікація результатів дослідження (Kwiecień A.); [117] – здійснено формування вимог до сучасних статичних методів аналізу безпеки, розроблення підходів для підвищення ефективності SAST-засобів у контексті SQL-ін'єкцій, вдосконалення методу статичного тестування безпеки застосунків у контексті вразливостей SQL Injection (Сергєєв Є.), визначено стратегію тестування безпеки застосунків в корпоративних мережах (Дрозд А.), визначено концепцію здійснення тестування застосунків в корпоративних системах (Савенко О.); [118] – розроблено метод захисту критичної інформації в корпоративних мережах (Гуральник О.), здійснено верифікацію моделі подання критичної інформації (Kwiecień A.), здійснено постановку експерименту щодо надійності захисту

критичної інформації (Клейн О., Лигун О.); [119] – запропонована концепція перевірки ефективності централізованої системи (Стецюк М.), проведено аналіз множини механізмів захисту ОС (Возний К.), здійснено постановку експерименту для дослідження ефективності застосування централізованої системи безпеки ОС (Kvassay M.), розроблено модель централізованої системи безпеки, як інформаційної технології для синтезу архітектури ОС, захищеної від витоку конфіденційної інформації (Стецюк Ю.); [120] – розроблено показники для застосування багатофакторного аналізу ЗПЗ (Нічепорук А.); [121] – розроблено концепцію дослідження щодо виявлення вставок в програмний код на основі дослідження аномалій при його виконанні в комп'ютерних мережах (Савенко О., Нічепорук А.), здійснено верифікацію моделі вставок в програмний код на основі дослідження аномалій при його виконанні в комп'ютерних мережах (Каштальян А., Савенко Б.); [122] – здійснено постановку завдання щодо архітектури та завдань програмного забезпечення організації зберігання, оптимізації та вибору даних в системах та засобах протидії КА з урахуванням історичного досвіду їх використання (Савенко Б.), розроблено модулі програмного забезпечення організації зберігання, оптимізації та вибору даних в системах та засобах протидії КА з урахуванням історичного досвіду їх використання (Нічепорук А., Каштальян А.).

Апробація результатів дисертації. Апробацію основних положень, ідей, висновків дисертаційної роботи проведено на науковому семінарі кафедри комп'ютерної інженерії та інформаційних систем у Хмельницькому національному університеті. Наукові результати роботи доповідались на таких конференціях: 1nd International Workshop on Intelligent Information Technologies and Systems of Information Security (IntelITSIS, Khmelnytskyi, Ukraine, June 10-12, 2020); 2nd International Workshop on Intelligent Information Technologies and Systems of Information Security (IntelITSIS, Khmelnytskyi, Ukraine, March 24–26, 2021); 5th International Workshop on Intelligent Information Technologies & Systems of Information Security (IntelITSIS, Khmelnytskyi, Ukraine, March 28, 2024); 1st International Workshop on Intelligent & CyberPhysical Systems (ICyberPhyS, Khmelnytskyi, Ukraine, June 28, 2024); 1st International Workshop on Advanced Applied Information Technologies:

(AdvAIT, Khmelnytskyi, Ukraine and Zilina, Slovakia, December 5, 2024); 14th International Conference on Dependable Systems, Services and Technologies (IEEE, DeSSerT, Athens, Greece, October 11-13, 2024); 6th International Workshop on Intelligent Information Technologies & Systems of Information Security (IntelITSIS, Khmelnytskyi, Ukraine, April 04, 2025).

Публікації. За результатами проведених досліджень основні наукові результати опубліковано у чотирьох наукових статтях в трьох фахових наукових журналах України категорії Б [111-114]. Апробація засвідчена публікаціями 7 праць в матеріалах міжнародних конференцій [115-121], з яких 7 праць проіндексовано у наукометричній базі Scopus [115-121]. Опубліковано та отримано одне свідоцтво про реєстрацію авторського права на твір (програму) [122].

Структура та обсяг дисертації. Дисертаційна робота складається з анотації, змісту, переліку скорочень, вступу, чотирьох розділів, висновку, списку використаних джерел та чотирьох додатків. Повний обсяг роботи містить 245 сторінок друкованого тексту, з них анотація – на 12 стор., зміст – на 2 стор., перелік умовних скорочень – на 1 стор., основний текст – на 150 стор., список із 122 використаних джерел – на 14 стор., додатки – на 36 стор. Дисертація містить 12 рисунків та 10 таблиць.

РОЗДІЛ 1

АНАЛІЗ МЕТОДІВ ТА ЗАСОБІВ ЗБЕРІГАННЯ, ВИБОРУ ТА ОПТИМІЗАЦІЇ
ІСТОРИЧНИХ ДАНИХ СИСТЕМ І ЗАСОБІВ ПРОТИДІЇ КОМП'ЮТЕРНИМ
АТАКАМ В КОРПОРАТИВНИХ МЕРЕЖАХ

1.1 Поняття про комп'ютерні атаки та вплив історичних даних на їх виявлення

Корпоративні мережі, що широко експлуатуються підприємствами, організаціями та установами, все більше піддаються впливу КА. Такий вплив характеризується як [19] спрямовані (навмисні) дії в кіберпросторі, які здійснюються за допомогою засобів електронних комунікацій (включаючи інформаційно-комунікаційні технології, програмні, програмно-апаратні засоби, інші технічні та технологічні засоби і обладнання) та спрямовані на досягнення однієї або сукупності таких цілей: порушення конфіденційності, цілісності, доступності електронних інформаційних ресурсів, що обробляються (передаються, зберігаються) в комунікаційних та/або технологічних системах, отримання несанкціонованого доступу до таких ресурсів; порушення безпеки, сталого, надійного та штатного режиму функціонування комунікаційних та/або технологічних систем; використання комунікаційної системи, її ресурсів та засобів електронних комунікацій для здійснення кібератак на інші об'єкти кіберзахисту [19].

Збільшення інтенсивності та складності КА обумовлює необхідність удосконалення не лише засобів їх виявлення, а й інформаційного забезпечення процесів аналізу та підтримки прийняття рішень. Одним із найбільш інформативних джерел таких даних є історичні дані, накопичені в процесі функціонування систем кіберзахисту, що містять відомості про параметри мережного трафіку, журнали подій, ознаки попередніх атак, характеристики вразливостей та результати реагування на інцидент кібербезпеки, тобто події або ряд несприятливих подій ненавмисного характеру (природного, технічного, технологічного, помилкового, у тому числі внаслідок дії людського фактора) та/або таких, що мають ознаки можливої (потенційної) КА, які становлять загрозу безпеці систем електронних комунікацій,

систем управління технологічними процесами, створюють імовірність порушення штатного режиму функціонування таких систем (у тому числі зриву та/або блокування роботи системи, та/або несанкціонованого управління її ресурсами), ставлять під загрозу безпеку (захищеність) електронних інформаційних ресурсів; [19].

Використання історичних даних дає можливість виявляти приховані закономірності реалізації КА, визначати характерні послідовності дій порушника, прогнозувати розвиток атак та підвищувати достовірність їх ідентифікації. Разом із тим безперервне накопичення таких даних призводить до істотного збільшення їх обсягів, що ускладнює оперативне опрацювання інформації, збільшує час доступу до критично важливих відомостей і, як наслідок, негативно впливає на своєчасність виявлення КА. Методичні рекомендації щодо кіберзахисту автоматизованих систем управління технологічними процесами визначають необхідність використання результатів моніторингу та накопичених даних для своєчасного виявлення кіберінцидентів і підвищення ефективності функціонування засобів протидії КА [20].

Вирішення зазначеної проблеми потребує створення спеціалізованих засобів зберігання історичних даних, які забезпечують підтримання їх цілісності, конфіденційності, доступності та контрольованого використання відповідно до вимог законодавства України щодо захисту інформації [21] і міжнародних вимог до систем керування інформаційною безпекою [22]. Одночасно такі засоби повинні враховувати вимоги безпечного проєктування прикладних програм [23] та процесів керування інформацією про вразливості [24], оскільки саме історичні дані є основою для накопичення знань про реалізовані загрози та їх подальшого аналізу.

Разом із забезпеченням надійного зберігання особливого значення набуває оптимізація історичних даних. Використання повного масиву накопиченої інформації не забезпечує пропорційного підвищення ефективності виявлення КА, оскільки значна частина записів є дубльованою, застарілою або не містить ознак, релевантних поточній ситуації. Унаслідок цього збільшуються витрати часу на пошук необхідної інформації та зростає навантаження на обчислювальні ресурси. Тому засоби оптимізації повинні забезпечувати раціональну структурування, усунення інформаційної надлишковості та скорочення часу доступу до критично важливих

історичних даних із дотриманням принципів керування життєвим циклом інформації [25], планування та оцінювання якості програмних засобів [26], а також вимог до гарантування цілісності систем [27], гарантування процесів їх життєвого циклу [28] і формування сценаріїв гарантування [29].

Однак навіть оптимізовані історичні дані не забезпечують максимальної ефективності виявлення КА без їх цілеспрямованого вибору. У зв'язку з цим актуальним є розроблення засобів інтелектуального вибору історичних даних, які формують інформаційні вибірки з урахуванням типу КА, часової глибини аналізу, результатів попереднього функціонування системи та накопиченого досвіду протидії аналогічним загрозам. Комплексне поєднання засобів зберігання, оптимізації та адаптивного вибору історичних даних створює передумови для підвищення оперативності, достовірності та ефективності процесів виявлення КА в умовах постійного зростання інтенсивності кіберзагроз.

Для виявлення КА в КМ застосовують різноманітні засоби і системи. Розглянемо найбільш поширені системи і засоби виявлення КА (IDS/NIDS/NDR/XDR/SIEM), які використовуються для захисту корпоративних мереж.

Snort [30] є однією з найпоширеніших мережних систем виявлення КА, яка використовує сигнатурний аналіз пакетів для виявлення вторгнень, експлойтів і мережних аномалій. У КМ забезпечує контроль мережного трафіку в режимі реального часу. Історичні дані організовано у вигляді журналів подій та архівів спрацювань, що використовуються для ретроспективного аналізу інцидентів і налаштування правил виявлення.

Suricata [31] реалізує багатопотоковий аналіз мережного трафіку із застосуванням сигнатурного, протокольного та поведінкового аналізу. У КМ забезпечує виявлення складних КА навіть при високих швидкостях передавання даних. Історичні дані накопичуються у форматах Eve JSON та PCAP, що дозволяє виконувати детальний аналіз попередніх КА і цифрову криміналістику.

Zeek [32] орієнтований на поведінковий аналіз мережних взаємодій та виявлення КА за відхиленнями від типової активності користувачів і сервісів. У КМ

формує деталізовані журнали всіх мережних сеансів. Історичні дані структуровано за видами подій, що забезпечує ефективний ретроспективний аналіз і побудову сценаріїв розвитку КА.

Cisco Secure Network Analytics [33] здійснює виявлення КА шляхом аналізу мережних потоків та поведінкових характеристик вузлів КМ. Особливістю є виявлення прихованої активності та горизонтального поширення КА. Історичні дані накопичуються у централізованому сховищі мережної телеметрії для довгострокового аналізу.

Darktrace [34] використовує алгоритми штучного інтелекту для побудови моделі нормального функціонування КМ та виявлення КА за аномальними відхиленнями. Історичні дані безперервно накопичуються та використовуються для адаптивного навчання моделей, що дозволяє підвищувати точність виявлення нових і невідомих загроз.

Vectra AI [35] призначена для виявлення складних КА, зокрема компрометації облікових записів, lateral movement та командно-керуючих каналів. У КМ здійснює поведінковий аналіз взаємодії користувачів і вузлів. Історичні дані використовуються для побудови часових ліній КА та проведення проактивного пошуку прихованих кіберзагроз або слідів компрометації, які вже знаходяться в мережі, але ще не виявлені стандартними системами захисту.

Corelight [36] реалізує функції мережного виявлення КА шляхом інтеграції технологій Zeek, Suricata та глибокого аналізу мережного трафіку. У КМ забезпечує контроль усіх мережних сервісів і протоколів. Історичні дані представлені журналами подій, мережними метаданими та архівами PCAP для ретроспективного аналізу.

IBM QRadar [37] здійснює виявлення КА на основі кореляції журналів безпеки, мережних потоків, подій та зовнішньої розвідки загроз. У КМ забезпечує централізований моніторинг інцидентів. Історичні дані накопичуються у сховищі подій та використовуються для розслідування і повторного аналізу КА.

Microsoft Defender XDR [38] об'єднує телеметрію кінцевих пристроїв, мережі, електронної пошти та систем ідентифікації для комплексного виявлення КА. У КМ

використовує механізми кореляції та машинного навчання. Історичні дані організовано як централізовану телеметрію та часові шкали інцидентів.

Cortex XDR [39] забезпечує виявлення КА шляхом комплексного аналізу мережної і хостової телеметрії із застосуванням алгоритмів машинного навчання. У КМ дозволяє виявляти багатоступеневі КА. Історичні дані використовуються для реконструкції повного ланцюга компрометації та пошуку аналогічних інцидентів.

Trend Micro Vision One [430] реалізує централізоване виявлення КА шляхом кореляції подій із мережних пристроїв, серверів, робочих станцій та електронної пошти. Для КМ забезпечує наскрізний моніторинг безпеки. Історичні дані організовано у вигляді часової шкали розвитку інцидентів та архіву подій.

Logpoint NDR [41] виконує виявлення КА за допомогою аналізу мережного трафіку, правил кореляції, машинного навчання та інтеграції із SIEM. У КМ підтримує централізований моніторинг безпеки. Історичні дані містять журнали подій, мережні потоки та архіви пакетів для подальшого аналізу.

OPSWAT MetaDefender NDR [42] поєднує сигнатурний, поведінковий аналіз та зовнішню розвідку загроз для виявлення КА у КМ. Особливу увагу приділено моніторингу критичних сегментів інфраструктури. Історичні дані централізовано накопичуються та використовуються для аналізу повторюваних сценаріїв КА.

Secureworks Taegis XDR [43] здійснює виявлення КА шляхом аналізу мережних пакетів, журналів безпеки, поведінкових характеристик користувачів та індикаторів компрометації. У КМ забезпечує комплексний моніторинг інцидентів. Історичні дані підтримують довготривале зберігання для проведення проактивного пошуку прихованих кіберзагроз або слідів компрометації, які вже знаходяться в мережі, але ще не виявлені стандартними системами захисту, і для проведення цифрових розслідувань.

NETSCOUT Omnis Cyber Intelligence [44] призначена для виявлення КА шляхом аналізу мережних пакетів, потоків і метаданих у високошвидкісних КМ. Особливістю є можливість аналізу великих обсягів мережної телеметрії. Історичні дані організовано у вигляді довготривалих сховищ мережних потоків і архівування пакетів для детального ретроспективного аналізу.

Аналіз наведених систем і засобів протидії КА в КМ показує, що більшість сучасних засобів та систем реалізують схожу концепцію роботи з історичними даними: накопичення журналів подій; зберігання мережної телеметрії; архівування пакетів або їх фрагментів; побудову часових ліній розвитку інцидентів; використання історичних даних для навчання моделей машинного навчання; ретроспективний пошук; кореляцію нових подій із попередніми інцидентами; автоматичне формування баз знань щодо типових сценаріїв КА. Саме використання історичних даних сьогодні є однією з ключових характеристик сучасних NDR/XDR/SIEM-платформ, бо дозволяє не лише виявляти поточні КА, а й аналізувати їх розвиток, прогнозувати подальші дії зловмисника та підвищувати ефективність моделей виявлення.

Здійснимо порівняння (табл. 1.1) функціональних можливостей, які стосуються історичних даних. Виділимо для цього порівняння такі критерії: метод виявлення КА (сигнатурний, поведінковий, ML, кореляція тощо); робота в КМ (масштабованість, підтримка великих мереж); організація історичних даних (журнали, PCAP, телеметрія, централізоване сховище); ретроспективний аналіз; використання історичних даних для навчання або кореляції.

Таблиця 1.1 – Порівняння сучасних засобів виявлення КА

Засіб	Метод виявлення КА	Робота в КМ	Організація історичних даних	Ретроспективний аналіз та використання історичних даних
1	2	3	4	5
Snort [30]	Сигнатурний аналіз	+	Журнали подій	Аналіз журналів, без автоматичного навчання
Suricata [31]	Сигнатури, аналіз протоколів	+	Eve JSON, PCAP	Форензика, пошук по історичних даних
Zeek [32]	Поведінковий аналіз	+	Детальні журнали сесій	Ретроспективний аналіз мережної активності
Cisco Secure Network Analytics [33]	Поведінковий аналіз, NetFlow	++	Централізована телеметрія	Кореляція історичних потоків та аномалій

Кінець таблиці 1.1

1	2	3	4	5
Darktrace [34]	III, поведінковий аналіз	++	База поведінкової історії	Самонавчання на історичних даних
Vectra AI [35]	Поведінковий аналіз, AI	++	Історія мережної активності	Threat Hunting, часові лінії атак
Corelight Open NDR [36]	Zeek + Suricata + NDR	++	Метадані, журнали, PCAP	Глибокий ретроспективний аналіз
IBM QRadar NDR/SIEM [37]	Кореляція подій, SIEM	++	Централізоване сховище подій	Кореляція поточних та історичних інцидентів
Microsoft Defender XDR [38]	XDR, ML, кореляція	++	Телеметрія та історія інцидентів	Автоматичне розслідування за історичними даними
Palo Alto Cortex XDR [39]	ML, XDR	++	Централізована телеметрія	Реконструкція повного ланцюга атаки
Trend Micro Vision One [40]	XDR, кореляція	++	Єдина база подій	Аналіз часової шкали інцидентів
Logpoint NDR [41]	NDR, ML, SIEM	++	Журнали, мережні потоки	Threat Hunting за історичними даними
OPSWAT MetaDefender NDR [42]	Сигнатури, поведінковий аналіз	+	Централізовані журнали	Аналіз повторних атак та шаблонів
Secureworks Taegis XDR [43]	AI, XDR, поведінковий аналіз	++	Довготривале сховище телеметрії	Threat Hunting та цифрова криміналістика
NETSCOUT Omnis Cyber Intelligence [44]	Аналіз пакетів і потоків	++	Телеметрія, PCAP	Детальний ретроспективний аналіз мережних подій

Позначення в таблиці 1.1:

+ – підтримка функції;

++ – розвинена підтримка та спеціалізовані механізми.

Таким чином, сучасні методи [45; 46], засоби та системи протидії КА в КМ еволюціонували [47; 48] від традиційного сигнатурного аналізу до комплексних платформ класів NDR, XDR та SIEM, які використовують поведінковий аналіз [49,

50], машинне навчання та кореляцію подій. Майже всі сучасні рішення підтримують накопичення історичних даних, однак вони переважно використовуються для ретроспективного аналізу інцидентів, цифрової криміналістики або навчання моделей штучного інтелекту. Водночас у більшості розглянутих систем відсутні механізми адаптивного управління історичними даними, зокрема їх оптимізації, інтелектуального відбору за глибиною горизонту, життєвим циклом чи актуальністю для поточних процесів виявлення КА. Тому цей аспект є недостатньо дослідженим і обґрунтовує потребу в розробленні спеціалізованих методів організації та оптимізації історичних даних у системах та засобах протидії КА в КМ.

1.2 Методи та технології зберігання даних

Зберігання історичних даних систем та засобів протидії КА повинно базуватись на загальних стратегіях, принципах і методах зберігання даних, які застосовуються для різних предметних областей, але повинні враховувати специфіку, яка пов'язана із особливостями застосування та внутрішніми процесами таких систем та засобів.

Зростаючий масштаб і складність результатів інтелектуального аналізу даних, таких як часті набори елементів, правила асоціації, послідовності та підграфи, зробили ефективне пошук шаблонів критичним, але недостатньо вивченим завданням. У [51; 52] розглядаються стратегії організації, індексування та доступу, які забезпечують масштабований та швидкий пошук структурованих шаблонів. Було досліджено базові типи даних і вихідних шаблонів, загальні операції пошуку та різноманітність типів запитів, які зустрічаються на практиці. Розглянуто ключові структури індексування, включаючи дерева префіксів, інвертовані індекси, підходи на основі хешу та методи на основі растрових зображень, кожен з яких підходить для різних представлень шаблонів і робочого навантаження. Розглянуто стратегії оптимізації запитів, наголошуючи на обході з урахуванням індексу, кешуванні та механізмах ранжирування. Огляд в [51; 52] забезпечує комплексну основу для розробки систем пошуку шаблонів нового покоління, які є масштабованими, гнучкими та тісно інтегрованими в аналітичні робочі процеси. Пропонований аналіз

для всіх сфер застосування, включаючи кібербезпеку, де важливе значення має надійний пошук із можливістю інтерпретації.

У поточному сценарії універсальної доступності даних стикаються з надзвичайно складними проблемами, які пов'язані з інтеграцією та обробкою різноманітних наборів даних для задоволення своїх аналітичних потреб [53; 54]. У роботі аналізуються традиційні та інноваційні методи, що використовуються для зберігання та інтеграції даних, з особливим акцентом на їх вплив на масштабованість, узгодженість і взаємодію в аналітичній екосистемі. Зокрема, він вносить міжрівневу таксономію, що пов'язує механізми інтеграції, тобто відповідність схем, роздільна здатність об'єктів і семантичне збагачення, до укрупнених сховищ зберігання/запитів, разом із порівняльними таблицями та рисунками, які синтезують компроміси та важелі ефективності/керування. Завдяки рішенням відображення схем, які вирішують проблеми, спричинені структурною неоднорідністю, архітектурою зберігання даних, що варіюється від традиційних рішень до хмарних рішень, а також конвеєрної інтеграції ETL за допомогою об'єднаних процесорів запитів. В дослідженні приділено особливу увагу застосуванню керування метаданими, зосереджуючись на семантичному збагаченні за допомогою онтологій і керування лініями для забезпечення наскрізної відстежуваності та управління. Він також охоплює гарячі точки продуктивності та методи кешування, а також компроміси узгодженості, що виникають у розподілених системах.

В останні роки дослідження блокчейну привернули увагу з усього світу [55; 56]. Це децентралізована компетенція, яка розповсюджена та невизначена. Кілька країн і вчених вже успішно застосували блокчейн у багатьох сферах. Блокчейн необхідний у делікатних ситуаціях, оскільки він захищає дані та запобігає їх зміні чи підробці. Крім того, підвищений попит ринку на дані стимулює попит на масштабування даних у всіх галузях. Дослідники з багатьох країн з часом використовували блокчейн у різних секторах, таким чином приділяючи особливу увагу цьому домену блокчейну. Кожен дослідницький проект починається з глибоких знань про робочу область, а нова інформація про блокчейн досить розрізнена. У цьому дослідженні аналізується література про технологію блокчейн, наголошуючи на трьох ключових аспектах:

сховище блокчейну, масштабованість і доступність. Це критичні області в ширшому полі технології блокчейн. У цьому дослідженні використовуються CiteSpace і VOSviewer, щоб всебічно зрозуміти поточний стан досліджень у цих областях. Це інструменти бібліометричного аналізу, які зазвичай використовуються в академічних дослідженнях для вивчення моделей і зв'язків у науковій літературі. Таким чином, щоб візуалізувати спосіб зберігання даних із масштабованістю та доступністю, зберігаючи синхронізацію безпеки блокчейну, було проведено необхідні дослідження щодо зберігання, масштабованості та доступності даних у середовищі блокчейну. Кінцева мета полягала в сприянні розробці безпечних та ефективних рішень для зберігання даних у рамках технології блокчейн.

Пандемія COVID-19 підкреслила необхідність гнучких служб охорони здоров'я, які забезпечують надійний і безпечний обмін інформацією, але досягнення належного, приватного та безпечного обміну EMR залишається проблемою через різноманітні формати даних і фрагментовані записи в багатьох накопичувачах даних, що призводить до ускладненої координації між командами охорони здоров'я, можливих медичних помилок і затримок у догляді за пацієнтами. У той час як централізовані системи створюють ризики конфіденційності, а різноманітність форматів даних ускладнює взаємодію, технологія блокчейн пропонує ефективне рішення, забезпечуючи децентралізоване зберігання, гарантуючи цілісність даних, покращуючи контроль доступу, усуваючи посередників і підвищуючи ефективність охорони здоров'я. У роботі [57; 58] досліджено значення стандартів, проблеми безпеки та підходи на основі блокчейну для сприяння взаємодії та безпечного обміну даними.

Технологія блокчейн була успішно застосована в останні роки для сприяння незмінності, відстежуваності та автентичності раніше зібраних і збережених даних [59; 60]. Однак обсяг даних, що зберігаються в блокчейні, зазвичай обмежений з економічних і технологічних причин. Зокрема, блокчейн зазвичай зберігає лише відбиток даних, наприклад хеш даних, тоді як повна необроблена інформація зберігається поза ланцюгом. Зазвичай цього достатньо, щоб гарантувати незмінність і відстежуваність, але не підтримує іншу важливу властивість, тобто доступність

даних. Це особливо потрібно, коли традиційна централізована база даних вибирається для зберігання поза мережею. З цієї причини багато пропозицій намагаються правильно поєднати блокчейн із децентралізованим сховищем. Однак зберігання даних може спричинити деякі проблеми конфіденційності. У роботі запропоновано рішення, яке належним чином поєднує блокчейн і методи шифрування, щоб гарантувати незмінність, відстежуваність, доступність і конфіденційність даних.

Доступність Інтернету та його інтеграція з інтелектуальними технологіями віддали перевагу повсякденним предметам і речам і запропонували нові сфери, такі як Інтернет речей (IoT). IoT відноситься до концепції, коли розумні пристрої або речі підключаються та створюють мережу. Ця нова область страждає від обробки великих даних і проблем із безпекою [61; 62]. Існує потреба в розробці моделі аналізу даних з використанням нових технологій, архітектури та моделі безпеки 5G. Надійний обмін даними за наявності законних вузлів завжди є однією з проблем у цих мережах. Інфіковані вузли генерують неточну інформацію та порушують безпеку користувача. У роботі запропоновано модель аналізу даних і архітектура самоорганізації для мереж IoT, щоб зрозуміти різні рівні технології і процеси. У роботі також запропоновано модель безпеки, яка заснована на механізмі автентифікації, виявлення та прогнозування для мереж IoT. Запропонована модель підвищує безпеку та захищає мережу від DoS та DDoS атак.

Під час випробувань [63; 64], щоб відповідати вимогам послідовного отримання та зберігання кількох цілей, кількох систем і кількох типів даних, різні типи даних обробляються в потоки даних імпульсно-кодової модуляції (PCM) із використанням кодування PCM для зберігання. Відповідно до вимог зберігання в режимі реального часу потоків даних PCM з високою бітовою швидкістю розроблено систему зберігання великої ємності на основі Serial Advanced Technology Attachment 3.0 (SATA3.0). Система використовує програмовану вентиляну матрицю серії Kintex 7 (FPGA) як ядро керування, отримує потоки даних PCM через низьковольтний диференціальний інтерфейс низьковольтної диференціальної сигналізації (LVDS), зберігає отримані потоки даних PCM на диск mSATA через шину передачі SATA3.0 і передає збережені дані назад в головний комп'ютер через інтерфейс USB3.0 для

аналізу. Тим часом, щоб вирішити проблему експорту складних даних, система зберігання створює файлову систему FAT32 через програмне ядро MicroBlaze для оптимізації керування та роботи системи зберігання великої ємності.

Через проблеми із складним збором даних [65; 66], тривалим періодом реалізації проекту, складними даними, поганою безпекою, складною можливістю відстеження та взаємозв'язку даних, управління архівами більшості національної інфраструктури все ще перебуває в доінформаційній епосі. Для вирішення цих проблем у роботі запропоновано архітектуру зберігання даних для національної інфраструктури на основі блокчейну. Вибір даних, що зберігаються в ланцюжку, і дані кількох регіонів або полів моделюються спільно за допомогою федеративного навчання. Параметри та результати зберігаються в ланцюжку, а інформація кожного вузла спільно використовується для вирішення проблеми обміну даними.

У хмарних системах зберігання користувачі повинні мати можливість вимикати програму, коли вона не використовується, і перезапускати її з останнього узгодженого стану, коли потрібно [67; 68]. BlobSeer - це програма для зберігання даних, спеціально розроблена для розподілених систем, створена як альтернативне рішення для існуючої популярної системи зберігання з відкритим кодом розподіленої файлової системи Hadoop (HDFS). У хмарній моделі всі компоненти повинні зупинятися та перезапускатися з узгодженого стану, коли цього вимагає користувач. Одним із обмежень BlobSeer DFS є можливість втрати даних під час перезавантаження системи. Таким чином, важливо забезпечити послідовний стан запуску та зупинки компонентів BlobSeer під час використання в хмарному середовищі, щоб запобігти втраті будь-яких даних. У роботі досліджено можливість BlobSeer забезпечити систему розподіленого зберігання даних узгодженого стану з інтеграцією функції перезапуску контрольних точок. Щоб продемонструвати доступність узгодженого стану, було створено кластер із кількома машинами та розгорнуто сутності BlobSeer із функцією контрольних точок на різних машинах.

Конвеєри великих даних розроблені для обробки даних, що характеризуються однією або декількома з трьох особливостей великих даних, широко відомих як три V (об'єм, швидкість і різноманітність), за допомогою серії кроків (наприклад,

вилучення, трансформація та переміщення), створюючи основу для використання розширеної аналітики та методів ML/AI [69; 70]. Обчислювальний континуум (тобто хмара/туман/край) дозволяє отримати доступ до практично нескінченної кількості ресурсів, де конвеєри даних можуть виконуватися в масштабі. Однак реалізація конвеєрів даних у континуумі є складним завданням, яке потребує врахування обчислювальних ресурсів, каналів передачі даних, тригерів, методів передачі даних, інтеграції черг повідомлень тощо. Завдання стає ще більш складним, якщо зберігання даних розглядається як частина конвеєрів даних. Локальне сховище дороге, складне в обслуговуванні та пов'язане з кількома проблемами (наприклад, доступність даних, безпека даних і резервне копіювання). Використання хмарного сховища, тобто сховища як послуги (SaaS), замість локального сховища має потенціал для забезпечення більшої гнучкості з точки зору масштабованості, відмовостійкості та доступності. Запропоновано загальний підхід до інтеграції SaaS з конвеєрами даних, тобто обчислення на локальному сервері або в певній хмарі, але інтеграцію зі SaaS, і розроблено метод ранжирування доступних варіантів зберігання на основі п'яти ключових параметрів: вартість, близькість, продуктивність мережі, шифрування на стороні сервера та ваги/уподобання користувачів. Проведена оцінка демонструє ефективність запропонованого підходу з точки зору продуктивності передачі даних, корисності окремих параметрів і можливості динамічного вибору варіанта зберігання на основі чотирьох основних сценаріїв користувача.

Сума великих даних [71; 72], створених із різних джерел, значно збільшується з кожним днем до такої міри, що для традиційних методів зберігання стає складно зберігати цю велику кількість даних. З цієї причини більшість організацій вирішили використовувати сторонні хмарні сховища для зберігання даних. Останнім часом хмарне сховище прогресує, але все ще стикається з численними проблемами щодо безпеки та конфіденційності. У [71] розглядаються проблеми безпеки та конфіденційності великих даних, а також мінімальні вимоги, які мають забезпечувати майбутні рішення. Основна мета роботи полягає в розробленні нової технічної структури для контролю та управління ризиками безпеки та конфіденційності великих даних. Запропонований фреймворк використовує переваги технології

Blockchain для забезпечення безпечного зберігання великих даних шляхом керування їх метаданими та політиками та усунення зовнішніх сторін для підтримки безпеки та конфіденційності даних. Крім того, він використовує технологію мобільного агента, щоб скористатися перевагами, пов'язаними з продуктивністю системи в цілому.

Аналіз сучасних методів і засобів зберігання даних [51-74] засвідчив, що існуючі підходи переважно орієнтовані на універсальні інформаційні системи та не повною мірою враховують специфіку систем і засобів протидії КА в КМ. Незважаючи на розвиток хмарних технологій, розподілених сховищ, блокчейну та засобів інтеграції даних, залишаються невирішеними питання ефективного зберігання, актуалізації та використання історичних даних для підтримки процесів виявлення, прогнозування та протидії КА. Це обумовлює необхідність розроблення нових методів організації та зберігання історичних даних, адаптованих до особливостей функціонування систем і засобів протидії КА [74; 75], що забезпечуватимуть підвищення оперативності, достовірності та ефективності прийняття рішень.

1.3 Методи і засоби оптимізації даних систем та засобів протидії комп'ютерним атакам

Історичні дані систем та засобів протидії КА в процесі зберігання накопичуються і тому наявна потреба в здійсненні їх оптимізації. Цю оптимізацію повинні забезпечувати засоби, які підтримують загальну організацію сховища даних для систем та засобів протидії КА.

Для оптимізації даних використовуються різноманітні методи та технології, спрямовані на підвищення ефективності зберігання, оброблення та доступу до інформації. До найбільш поширених методів належать дедуплікація даних, компресія, індексування, кешування, архівування, реплікація, сегментація та кластеризація даних [76-78]. Застосування зазначених методів дає змогу зменшити обсяг інформації, що зберігається, скоротити час пошуку необхідних записів та підвищити продуктивність системи в цілому.

Значного поширення набули технології керування життєвим циклом даних, багаторівневого зберігання, інтелектуального керування даними, технології Data Lake, Data Warehouse, а також рішення, що базуються на штучному інтелекті та машинному навчанні для автоматичного аналізу та оптимізації інформаційних ресурсів. У сучасних системах та засобах протидії КА також використовуються технології потокового оброблення подій, розподіленого зберігання даних та горизонтального масштабування, що забезпечують роботу з великими обсягами телеметричної інформації в режимі реального часу.

Практична реалізація зазначених технологій здійснюється за допомогою сучасних систем керування базами даних, платформ розподіленого оброблення даних, засобів кешування, а також систем керування журналами подій та платформ класу SIEM [79; 80], які забезпечують централізоване накопичення, індексування та аналіз даних інформаційної безпеки.

Незважаючи на значний розвиток існуючих методів і технологій, більшість із них орієнтована переважно на підвищення продуктивності інформаційних систем загального призначення та не враховує специфіку функціонування систем і засобів протидії КА. Зокрема, недостатньо уваги приділяється адаптивному управлінню структурою сховища, інтелектуальному відбору даних з урахуванням глибини горизонту, накопиченого досвіду виявлення КА та необхідності оперативного надання критично важливої інформації підсистемам прийняття рішень. Це обумовлює актуальність розроблення спеціалізованих методів оптимізації даних, орієнтованих на забезпечення максимальної швидкодії та підвищення ефективності функціонування систем і засобів протидії КА [81-83].

Технологія блокчейн характеризується децентралізацією, безпекою та незмінністю даних, що забезпечило її широке застосування в різних сферах [84]. Однак зі зростанням обсягів даних суттєво збільшуються вимоги до ресурсів зберігання, що призводить до зростання вартості зберігання та зниження ефективності використання ресурсів. У зв'язку з цим оптимізація сховищ даних стала важливим напрямом досліджень. Основними підходами до оптимізації є скорочення

даних, шардинг, дедуплікація, стиснення та використання розподілених механізмів зберігання [84].

У роботі [85] досліджено задачу оптимізації зберігання даних у розподілених центрах обробки даних розумних мереж. Запропонований підхід базується на вдосконаленому алгоритмі Simulated Annealing у поєднанні з механізмами розподіленого зберігання на базі Hadoop. Метод дозволяє оптимізувати розподіл наборів даних між центрами обробки, підвищити ефективність використання обчислювальних ресурсів та скоротити витрати на передачу даних. Результати моделювання підтвердили підвищення ефективності та стабільності запропонованого підходу [85].

Великі дані мають значний потенціал для підвищення [86] ефективності різних галузей, однак дані, що зберігаються у різних системах, часто є несумісними, що призводить до їх ізолюваності та зниження загальної цінності. Ефективний обмін і спільне використання даних дозволяють формувати інтегровані екосистеми та підвищувати ефективність їх обробки. Однією з ключових проблем є необхідність оптимізації процесів зберігання, обробки та використання великих обсягів даних в умовах їх неоднорідності, обмеженості ресурсів і високих вимог до безпеки. Основними викликами залишаються інформаційна надлишковість, складність інтеграції даних та забезпечення ефективного доступу до релевантної інформації. Сучасні дослідження зосереджені на розробленні платформ і методів, що забезпечують оптимізацію обробки великих даних, підвищення ефективності їх спільного використання та зниження витрат ресурсів. Це створює передумови для побудови більш адаптивних, масштабованих і ефективних систем управління даними.

За останнє десятиліття кіберзлочинність суттєво еволюціонувала [87], охоплюючи КА на основі штучного інтелекту, глибокі фейки, атаки на 5G-мережі та ланцюги поставок. Це значно ускладнило забезпечення кібербезпеки та підвищило вимоги до систем виявлення й протидії кіберзагрозам. Важливим фактором підвищення ефективності кіберзахисту є аналіз історичних даних, який дозволяє виявляти закономірності, тенденції та еволюцію кіберзагроз. Дослідження багаторічних даних про кіберінциденти дає змогу оцінювати зміни у типах КА,

характеристиках зловмисників і масштабах завданих збитків. Результати таких досліджень підтверджують, що сучасні кіберзагрози стають дедалі складнішими, більш адаптивними та орієнтованими на отримання економічної вигоди. Це підкреслює необхідність удосконалення методів накопичення, обробки та оптимізації історичних даних у системах кіберзахисту для підвищення точності виявлення атак і ефективності протидії їм.

Кількість нових кіберуразливостей щороку зростає, що ускладнює своєчасне виявлення та усунення потенційних загроз. Водночас бази даних уразливостей часто мають обмеження щодо оперативного оновлення [88] та ефективного обміну інформацією про нові події безпеки. У таких умовах системи управління вразливістю відіграють важливу роль у виявленні слабких місць та підтримці процесів кіберзахисту. Проте їх ефективність значною мірою залежить від якості, повноти та актуальності даних про вразливості. Сучасні дослідження показують, що проблеми обробки, структурування та оптимізації баз даних уразливостей суттєво впливають на ефективність методів виявлення загроз. Це підтверджує необхідність розроблення більш ефективних методів організації та оптимізації даних для підвищення результативності систем кіберзахисту.

Сучасні технології зберігання даних [89-92] активно еволюціонують під впливом зростання обсягів великих даних, розвитку Інтернету речей, штучного інтелекту та машинного навчання. Традиційні підходи поступово доповнюються хмарними та периферійними обчисленнями, які забезпечують нові можливості для масштабованого та гнучкого зберігання даних. Дослідження показують, що хмарні технології значно підвищують ефективність управління даними завдяки масштабованості та гнучкості, тоді як периферійні обчислення дозволяють зменшити затримки доступу та навантаження на мережеву інфраструктуру. Особливу роль у розвитку сучасних систем зберігання даних відіграє інтеграція штучного інтелекту та машинного навчання, які дозволяють підвищити ефективність оптимізації даних, адаптивність систем та швидкість обробки інформації. Це підтверджує необхідність подальшого розвитку інтелектуальних, безпечних та ефективних технологій зберігання даних.

Згідно здійсненого аналізу встановлено, що існуючі методи, технології та засоби оптимізації даних забезпечують підвищення ефективності зберігання, оброблення й доступу до інформації, проте вони переважно орієнтовані на інформаційні системи загального призначення. Для систем та засобів протидії КА актуальним є розроблення спеціалізованих підходів, які враховують їх специфіку [93-95], забезпечують адаптивне управління історичними даними, інтелектуальний відбір релевантної інформації та скорочення часу прийняття рішень під час протидії КА.

1.4 Методи і засоби вибору даних з урахуванням характеристик та показників операційного середовища

При виборі даних в процесі здійснення КА і КМ для прийняття рішень системами та засобами протидії КА виникає потреба враховувати актуальність історичних даних, глибину горизонту, попередню ефективність їх використання тощо. Тому здійснення правильного вибору даних для систем та засобів протидії КА впливатиме на результат виявлення КА.

Ефективність такого вибору визначається здатністю системи ідентифікувати найбільш релевантні дані серед значних обсягів накопиченої інформації. Використання застарілих або малозначущих даних призводить до збільшення часу їх оброблення, зростання обчислювального навантаження та зниження точності виявлення КА. Водночас відбір інформації, що враховує часову актуальність, подібність до попередніх інцидентів, статистику результативності використання історичних даних та особливості поточної ситуації, забезпечує підвищення достовірності прийняття рішень. Особливого значення набуває адаптивний механізм вибору, який динамічно змінює склад даних залежно від типу КА, рівня її небезпечності та накопиченого досвіду функціонування системи. Такий підхід дозволяє скоротити час доступу до необхідної інформації, зменшити інформаційну надлишковість, підвищити точність виявлення КА і забезпечити більш ефективне функціонування систем та засобів протидії КА в умовах постійної еволюції кіберзагроз.

Вибір глибини історичних даних є одним із ключових чинників побудови ефективних моделей прогнозування часових рядів. У роботі [96] показано, що між тривалістю навчальної вибірки та горизонтом прогнозування існує компроміс. Використання тривалих історичних періодів (до одного року і більше) дає змогу зменшити невизначеність прогнозу та підвищити його статистичну стійкість, тоді як короткі часові горизонти краще відображають поточний стан системи. Для підвищення точності прогнозування пропонується використовувати адаптивний розподіл історичних даних за горизонтами прогнозування, статистичну оцінку їх узгодженості, методи відбору інформативних ознак та кластеризацію багатовимірних часових рядів. Такі підходи дозволяють формувати репрезентативні навчальні вибірки та зменшувати вплив надлишкової інформації на результати прогнозування [96].

У роботах [97; 98] розглянуто проблему прогнозування кіберподій на основі часових рядів в умовах постійного зростання кількості та складності КА. Автори відзначають, що сучасні системи та засоби протидії КА повинні використовувати значні обсяги структурованих і неструктурованих даних для своєчасного виявлення загроз. Особливу увагу приділено використанню коротких часових горизонтів і ковзних вікон, які забезпечують адаптацію моделей до змін характеристик атак та зменшують вплив дрейфу концепції. Водночас наголошується, що прогнозування кіберподій на основі часових рядів ще недостатньо досліджене порівняно з традиційними методами виявлення КА [99].

У дослідженні [100] показано, що ефективність прогнозування кіберзагроз значною мірою залежить від правильного вибору історичних даних. Для забезпечення статистичної релевантності інформації запропоновано використовувати ковзні часові вікна тривалістю від шести до чотирнадцяти місяців, що дозволяє враховувати актуальні тенденції розвитку загроз і одночасно уникати накопичення застарілої інформації. Такий підхід забезпечує підвищення точності прогнозування та покращує адаптивність систем кіберзахисту до нових типів атак [100].

У роботі [101] розглядаються наслідки порушень інформаційної безпеки, які призводять до втрати персональних, фінансових та медичних даних. Автори підкреслюють, що своєчасне виявлення аномальної активності є одним із головних

засобів запобігання таким інцидентам. Водночас зазначається, що ефективність сучасних систем виявлення вторгнень безпосередньо залежить від наявності якісних історичних даних, які використовуються для навчання моделей машинного навчання та побудови профілів нормальної поведінки мережі [101].

У роботі [102] наведено огляд сучасних інструментів аналізу мережного трафіку та методів формування навчальних наборів даних для систем виявлення вторгнень. Автори показують, що розвиток моделей машинного навчання значною мірою визначається якістю та репрезентативністю історичних даних. У дослідженні запропоновано класифікацію інструментів збору мережного трафіку, розглянуто процес побудови наборів даних та проаналізовано їх вплив на ефективність систем виявлення вторгнень. Особливу увагу приділено необхідності створення сучасних наборів даних, які відображають реальні сценарії мережних атак [102].

У роботі [103] досліджено застосування технологій великих даних для побудови масштабованих систем виявлення вторгнень у режимі реального часу. Запропонована архітектура базується на використанні Apache Spark Structured Streaming, бібліотеки MLlib та хмарної інфраструктури Microsoft Azure. Для класифікації мережного трафіку використовується алгоритм дерева рішень, а оцінювання системи проводиться на основі набору даних MAWILab. Отримані результати підтверджують ефективність використання розподілених платформ для оперативної обробки великих потоків даних та своєчасного виявлення кібератак [103].

У дослідженні [104] розглянуто проблему формування реалістичних потокових наборів даних для систем виявлення мережних вторгнень. Автори зазначають, що існуючі еталонні набори даних не завжди містять достатню кількість інформативних ознак для використання в реальних мережах. Для усунення цього недоліку застосовано різні методи відбору ознак, що дозволило визначити мінімальний набір параметрів NetFlow, достатній для ефективного навчання алгоритмів машинного навчання. Також підкреслено необхідність накопичення позначених даних безпосередньо в реальних інформаційних системах [104].

У роботі [105] досліджено вплив дисбалансу класів на ефективність моделей машинного навчання у задачах виявлення мережних вторгнень. Показано, що

нерівномірний розподіл навчальних прикладів призводить до зниження точності класифікації та погіршує інтерпретацію результатів роботи моделей. Автори оцінюють ефективність сучасних методів балансування даних і демонструють їх позитивний вплив як на якість класифікації, так і на пояснюваність рішень нейронних мереж [105].

У роботі [106] запропоновано мультимодальний підхід до побудови системи виявлення вторгнень для мереж Інтернету речей. Для обробки великих обсягів мережного трафіку використовуються технології Apache Spark, методи перенесення навчання та згорткові нейронні мережі. Особливістю запропонованого рішення є поєднання текстових і графічних ознак мережного трафіку, що дозволяє підвищити точність класифікації атак на наборах даних CIC-IoT 2022, CIC-IoT 2023 та Edge-IoT [106].

У дослідженні [107] запропоновано систему виявлення КА, що поєднує технології великих даних і глибокого навчання. Архітектура системи базується на використанні гібридної моделі CNN-LSTM, а для усунення дисбалансу класів застосовуються методи SMOTE та Tomek Links. Обробка даних реалізована засобами PySpark, що забезпечує масштабованість і високу продуктивність. Експериментальні результати підтверджують досягнення точності понад 99 % для задач бінарної та багатокласової класифікації мережних атак [107].

У роботі [108] виконано комплексний аналіз сучасних наборів даних для систем виявлення вторгнень у кіберфізичних промислових системах та промисловому Інтернеті речей. Автори здійснили кількісне оцінювання 23 наборів даних за критеріями повноти документації, реалістичності, відтворюваності, різноманітності атак і збалансованості класів. Отримані результати свідчать про наявність суттєвих прогалин у представленні сучасних сценаріїв атак та промислових протоколів, що необхідно враховувати під час вибору навчальних даних для побудови високоефективних систем виявлення вторгнень [108].

Згідно проведеного аналізу встановлено, що ефективність функціонування систем та засобів протидії КА значною мірою залежить від якості вибору історичних даних, які використовуються під час прийняття рішень. На відміну від традиційних

підходів, що базуються переважно на повноті накопиченої інформації, сучасні системи повинні здійснювати адаптивний відбір даних з урахуванням їх актуальності, глибини горизонту, попередньої результативності використання та відповідності поточній ситуації. Це дає змогу скоротити інформаційну надлишковість, зменшити час пошуку й оброблення необхідної інформації та підвищити оперативність прийняття рішень. Використання інтелектуальних механізмів вибору історичних даних забезпечує підвищення точності виявлення КА [109; 110], більш раціональне використання обчислювальних ресурсів і адаптацію системи до появи нових КА, що є необхідною умовою підвищення ефективності сучасних систем та засобів протидії КА.

1.5 Постановка задачі дослідження

Для розв'язання задачі підвищення ефективності зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах шляхом удосконалення архітектури та методів організації зберігання, вибору та оптимізації історичних даних, потрібно вирішити такі завдання:

1) провести аналіз типів комп'ютерних атак, методів та систем їх виявлення в корпоративних мережах, а також моделей, методів і засобів зберігання, оптимізації та вибору даних з урахуванням історичного контексту та глибини горизонту;

2) розробити моделі організації пам'яті з історичними даними щодо функціонування систем та засобів протидії КА в КМ, моделі обманних технологій в системах та засобах протидії КА в корпоративних мережах для їх використання при розробленні засобів зберігання, оптимізації та вибору даних для систем і засобів протидії КА;

3) розробити архітектуру засобів зберігання даних для систем і засобів протидії КА на основі поєднання автоматичного керування життєвим циклом всіх даних або їх частини, особливостей апаратно-програмних засобів для зберігання даних, інтелектуалізації щодо глибини горизонту для даних, самонавчання при виборі даних для типових повторюваних КА;

4) розробити метод організації зберігання даних у системах та засобах протидії комп'ютерним атакам з використанням поділу даних на початкові, накопичувані та використані, а також врахування історичного аспекту їх застосування шляхом збереження результатів оцінювання ефективності використання даних із застосуванням марковських моделей, динамічних баєсівських мереж та статистичних методів;

5) розробити метод оптимізації даних на основі самонавчання засобів підсистеми пам'яті систем та засобів протидії КА для здійснення комплексної адаптивної оптимізації початкових, накопичених, використаних та даних обманних систем з урахуванням критеріїв ресурсної ефективності, структури даних, шарового розміщення, апаратно-мережного розподілу та специфіки функціонування обманних систем;

6) розробити метод вибору даних на основі статичного оцінювання ознак або фіксованих моделей ранжування, інтегральної оптимізаційної функції з урахуванням глибини горизонту, семантичної значущості та попереднього досвіду використання, доповнену механізмом самонавчання для адаптивного формування та корекції інтегрованих сукупностей даних;

7) розробити засоби для зберігання, оптимізації та вибору даних для систем та засобів протидії КА і КМ і провести експериментальні дослідження таких засобів з метою покращення їх характеристик та виявлення КА.

1.6 Висновки до першого розділу

Згідно з проведеним аналізом існуючих систем і засобів протидії КА, архітектури засобів зберігання даних, методів організації, оптимізації та вибору інформації встановлено наявність проблем щодо забезпечення ефективного зберігання, обробки та використання даних у процесах виявлення і протидії КА. За результатами аналізу можна виокремити такі висновки.

1. Кількість КА постійно зростає, а їх структура стає більш складною, адаптивною та динамічною, що призводить до суттєвого збільшення обсягів даних

кібермоніторингу, які необхідно обробляти в реальному часі для забезпечення ефективного виявлення загроз.

2. Існуючі системи та засоби протидії КА генерують значні обсяги початкових, накопичених та використаних даних, однак у більшості випадків відсутні ефективні механізми їх структурованої організації, що призводить до накопичення інформаційної надлишковості та зниження оперативності доступу до критично важливої інформації.

3. Сучасні системи та засоби протидії КА є ефективними за умови своєчасного доступу до релевантних даних, проте збільшення обсягів інформаційних масивів ускладнює процеси їх зберігання, аналізу та використання при прийнятті рішень.

4. Існуючі підходи до організації зберігання даних у системах та засобах протидії КА переважно орієнтовані на статичні моделі управління інформацією та недостатньо враховують історію використання даних, причинно-наслідкові залежності між подіями та необхідність адаптивного накопичення досвіду функціонування систем.

5. Методи оптимізації даних, які використовуються у сучасних системах та засобах протидії КА, здебільшого спрямовані на скорочення обсягів даних або покращення швидкодії окремих компонентів, однак недостатньо враховують особливості структури даних, апаратно-програмного середовища та динамічних змін інформаційного навантаження.

6. Значною проблемою залишається вибір релевантних даних для аналізу, оскільки традиційні методи відбору інформації часто базуються на статичних правилах або фіксованих моделях ранжування, що знижує їх ефективність в умовах динамічної зміни характеристик КА.

7. Підвищення ефективності систем і засобів протидії КА значною мірою залежить не лише від алгоритмів виявлення загроз, а й від якості організації життєвого циклу даних, ефективності їх оптимізації та точності вибору релевантної інформації для підтримки процесів прийняття рішень.

8. Перспективним напрямом досліджень є розроблення адаптивних засобів зберігання даних та інтелектуальних методів їх організації, оптимізації й вибору, які

враховують глибину горизонту даних, механізми самонавчання, історичний досвід використання інформації та особливості функціонування систем.

Таким чином, актуальною науково-прикладною задачею є підвищення ефективності зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах за рахунок удосконалення архітектури та методів організації зберігання, вибору та оптимізації історичних даних.

Основні наукові результати першого розділу опубліковані в працях [111-122].

РОЗДІЛ 2

АРХІТЕКТУРА СИСТЕМ ПРОТИДІЇ КОМП'ЮТЕРНИМ АТАКАМ В
КОРПОРАТИВНІ МЕРЕЖІ НА ОСНОВІ ОБМАННИХ ТЕХНОЛОГІЙ ТА
ПАМ'ЯТІ З ІСТОРИЧНИМИ ДАНИМИ ЩОДО ФУНКЦІОНУВАННЯ

Зловмисники при здійсненні КА на корпоративні мережі стикаються з протидією їм різних апаратно-програмних та програмних систем, що відображено в класифікації індикаторів компрометації піраміди болю зловмисника. Долаючи перешкоди при проникненні в корпоративні мережі зловмисники витрачають багато ресурсів та часу і в цьому процесі для них зростають складності, які пов'язані з проведенням КА. На кожному рівні піраміди болю зловмисника для забезпечення безпеки та захисту корпоративних мереж використовуються різні апаратно-програмні та програмні засоби і системи. Всі вони залежать від даних щодо свого попереднього функціонування, які формують підсистему пам'яті з історичними даними. В залежності від етапу, на якому розміщуються засоби та системи протидії КА організація їх пам'яті з історичними даними є різною. Зловмисники розуміють ієрархію засобів та систем у співвіднесенні до рівнів піраміди болю зловмисника, а також особливо щодо організації в них підсистем пам'яті з історичними даними. Вважаючи елементи чи підсистеми пам'яті одними з основних джерел попереднього досвіду функціонування для протидії КА, зловмисники прагнуть обходити не стільки сенсори засобів та систем протидії КА, а саме пам'яті з історичними даними щодо функціонування таких засобів та систем. Тому, організація елементів чи підсистем пам'яті з історичними даними щодо функціонування потребує постійного удосконалення за різними рівнями та формами, зокрема і з використанням обманних технологій, які активно інтегруються в апаратно-програмні та програмні засоби і системи протидії КА на корпоративні мережі, для заплутування зловмисників під час проведення ними КА. Це дало б змогу підвищити рівень кібербезпеки в корпоративних мережах за рахунок змінюваної організації елементів чи підсистем пам'яті з історичними даними щодо попереднього функціонування систем і засобів та з використанням відповідних обманних технологій.

2.1 Визначення області дослідження

Системи протидії КА у корпоративні мережі містять апаратно-програмні чи програмні засоби, що захищають мережну інфраструктуру, дані та пристрої від несанкціонованого доступу, збоїв, ЗПЗ та кіберзагроз, забезпечуючи безперервність та конфіденційність бізнес-процесів і діючи як багаторівневі системи. Кожен рівень захисту має свої особливості, відповідні типові засоби і системи, а також спрямування щодо етапів КА і збереження даних про попередні події або бази сигнатур різного призначення тощо. Тобто, для кожного рівня систематизовано перелік потенційно зловмисних ситуацій та подій і засобів чи систем для їх опрацювання. Відповідно, на кожному рівні наявні певні типові методи щодо організації зберігання даних про попередній досвід функціонування систем та засобів при певних типах КА, включно з історичними даними щодо функціонування.

Організацію елементів чи підсистем пам'яті в засобах чи системах протидії КА розглянемо у співвіднесенні до такого поділу на рівні:

- 1) рівень джерел подій;
- 2) сенсорний рівень;
- 3) кореляційний рівень;
- 4) аналітичний рівень;
- 5) поведінковий рівень;
- 6) рівень розширеного виявлення;
- 7) стратегічний рівень.

На першому рівні, тобто рівні джерел подій, здійснюється фіксація первинних фактів діяльності такими засобами чи в таких засобах: мережні пристрої; сервери; комп'ютерні станції; операційні системи; прикладні сервіси; мережні протоколи. Задамо джерела інформації про події множиною джерел інформації так:

$$M_{DI}^1 = \{m_{DI,1}^1, m_{DI,2}^1, \dots, m_{DI,N_{M_{DI}^1}}^1\}, \quad (2.1)$$

де $m_{DI,i}^1$ - i -те джерело інформації про події в корпоративних мережах від систем та засобів першого рівня; $i = 1, 2, \dots, N_{M_{DI}^1}$; $N_{M_{DI}^1}$ - кількість джерел.

Отримані дані від джерел інформації різних рівнів задамо множиною векторів так:

$$M_{DI}^D = \{V_{DI,1}^D, V_{DI,2}^D, \dots, V_{DI,N_{M_{DI}^D}}^D\}, \quad (2.2)$$

де $N_{M_{DI}^D}$ – кількість векторів для забезпечення зберігання даних від джерел інформації різних рівнів; $V_{DI,i}^D$ – i -тий вектор; $i = 1, 2, \dots, N_{M_{DI}^D}$.

Вектори з формули (2.2) задамо загальними для всіх джерел інформації поданнями так:

$$V_{DI,i}^D = (v_{DI,i,1}^D, v_{DI,i,2}^D, \dots, v_{DI,i,N_{V_{DI,i}^D}}^D), \quad (2.3)$$

де $N_{V_{DI,i}^D}$ – кількість координат i -того вектору для забезпечення зберігання даних від джерел інформації різних рівнів; $v_{DI,i,j}^D$ – j -та координата i -того вектору; $i = 1, 2, \dots, N_{M_{DI}^D}$; $N_{M_{DI}^D}$ – кількість векторів для забезпечення зберігання даних від джерел інформації різних рівнів; $j = 1, 2, \dots, N_{V_{DI,i}^D}$.

Координатами векторів $V_{DI,i}^D$ (формула (2.3)) будуть дані для забезпечення зберігання даних від джерел інформації різних рівнів. Кількість координат може змінюватись. Спільна інформація у всіх векторах забезпечується однаковим призначенням перших відповідних координат в кожному векторі, зокрема:

- 1) $v_{DI,i,1}^D$ – номер типу пам'яті для зберігання даних;
- 2) $v_{DI,i,2}^D$ – номер бази зберігання даних;
- 3) $v_{DI,i,3}^D$ – дані місця зберігання даних;
- 4) $v_{DI,i,4}^D$ – номер джерела інформації;
- 5) $v_{DI,i,5}^D$ – дані про джерело інформації;
- 6) $v_{DI,i,6}^D$ – тип даних;
- 7) $v_{DI,i,7}^D$ – тип події, внаслідок якої отримано дані;
- 8) $v_{DI,i,8}^D$ – час появи даних;
- 9) $v_{DI,i,9}^D$ – час розміщення даних в базах даних;
- 10) $v_{DI,i,10}^D$ – номер рівня, на якому сформовано дані;

11) $v_{DI,i,11}^D$ – номери рівнів, на яких можуть бути використані дані;

12) $v_{DI,i,12}^D$ – наявність та тип ситуації, внаслідок якої відбулась подія;

13) $v_{DI,i,13}^D$ – інформація про вузли корпоративних мереж, в яких відбулась подія, внаслідок якої отримано дані;

14) $v_{DI,i,14}^D$ – інформація про кількість повторень події, тобто інформація про наявність даних про подію за попередній час функціонування систем;

15) $v_{DI,i,15}^D$ – інформація про результати реагування на подію, тобто інформація про наявність даних про реагування на подію за попередній час функціонування систем;

16) $v_{DI,i,16}^D$ – значення оцінки події з врахуванням її повторень та реагування на неї за попередній час функціонування систем;

17) $v_{DI,i,17}^D$ – значення оцінки події з врахуванням її ранжування між всіма наявними різними подіями, дані про яких збережено в пам'яті систем.

На рівні джерел подій, тобто на першому рівні, виявлення не відбувається, а лише здійснюється збір фактів та даних. Такі дані задамо множиною так:

$$M_{DD}^1 = \{m_{DD,1}^1, m_{DD,2}^1, \dots, m_{DD,N_{M_{DD}^1}}^1\}, \quad (2.4)$$

де $m_{DD,i}^1$ - i -те джерело даних в корпоративних мережах від систем та засобів першого рівня; $i = 1, 2, \dots, N_{M_{DD}^1}$; $N_{M_{DD}^1}$ – кількість джерел.

Елементами множини M_{DD}^1 (формула (2.4)) можуть бути такі:

1) $m_{DD,1}^1$ – мережні пакети;

2) $m_{DD,2}^1$ – системні логи;

3) $m_{DD,3}^1$ – журнали доступу;

4) $m_{DD,4}^1$ – результати телеметрії.

Кількість елементів множини M_{DD}^1 (формула (2.4)) може бути більше за чотири елементи.

На другому, тобто сенсорному рівні, здійснюють виявлення окремих подій, які є підозрілими. Для цього використовують міжмережні екрани, системи запобігання

та виявлення вторгнень, включно з мережними та хостовими варіантами їх встановлення.

На сенсорному рівні, тобто на другому рівні, дані задамо множиною так:

$$M_{DS}^2 = \{m_{DS,1}^2, m_{DS,2}^2, \dots, m_{DS,N_{M_{DS}^2}}^2\}, \quad (2.5)$$

де $m_{DS,i}^2$ - i -те джерело даних в корпоративних мережах від систем та засобів другого рівня; $i = 1, 2, \dots, N_{M_{DS}^2}$; $N_{M_{DS}^2}$ - кількість джерел.

Елементами множини M_{DS}^2 (формула (2.5)) можуть бути такі:

- 1) $m_{DS,1}^2$ - сталі сигнатури;
- 2) $m_{DS,2}^2$ - поведінкові сигнатури;
- 3) $m_{DS,3}^2$ - правила;
- 4) $m_{DS,4}^2$ - лічильники;
- 5) $m_{DS,5}^2$ - стани з'єднань.

Кількість елементів множини M_{DS}^2 (формула (2.5)) може бути більше за п'ять елементів. Для цього рівня характерною є реакція систем та засобів в реальному часі, а також виконання дій за невеликий проміжок часу протягом секунд чи хвилин з використанням історичних даних, які отримані за декілька хвилин чи секунд до проявів КА, тобто короткої історії щодо функціонування.

На третьому, кореляційному рівні здійснюється об'єднання подій у ширшому контексті з врахуванням часових вікон, кореляційних правил, кількості повторів подій, формуванні простих статистичних відомостей. Дані задамо множиною так:

$$M_{DK}^3 = \{m_{DK,1}^3, m_{DK,2}^3, \dots, m_{DK,N_{M_{DK}^3}}^3\}, \quad (2.6)$$

де M_{DK}^3 - i -те джерело даних в корпоративних мережах від систем та засобів третього рівня; $i = 1, 2, \dots, N_{M_{DK}^3}$; $N_{M_{DK}^3}$ - кількість джерел.

Елементами множини M_{DK}^3 (формула (2.6)) можуть бути такі:

- 1) $m_{DK,1}^3$ - відомості про часові вікна подій;
- 2) $m_{DK,2}^3$ - відомості про кореляційні правила;
- 3) $m_{DK,3}^3$ - відомості про кількість повторень усіх зафіксованих подій та час;

4) $m_{DK,4}^3$ – статистики для усіх подій.

Кількість елементів множини M_{DK}^3 (формула (2.6)) може бути більше за чотири елементів. Для цього рівня відносяться засоби проміжного рівня опрацювання даних, централізованого протоколювання та базові лог-корелятори.

Засоби та системи проміжного рівня опрацювання даних здійснюють фільтрацію та нормалізацію лог-даних перед їх потраплянням до основної SIEM-системи. Фільтрація дає змогу вилучити з подальшого розгляду нерелевантні події, що призводить до зменшення кількості та обсягів даних, що впливає на заповнення пам'яті. Здійснення нормалізації лог-даних з різних джерел до єдиного формату дозволяє здійснити їх підготовку до аналізу. Також, на цьому етапі дані можуть бути розділені на критичні, які потребують негайного опрацювання, та непередготовлені не критичні, які можуть бути опрацьовані пізніше, але потрібні будуть для використання після тривалого часу функціонування систем чи засобів.

Засоби та системи централізованого протоколювання або такі підсистеми забезпечують збір, поєднання та зберігання лог-даних в журналах з різних джерел в одному центральному сховищі лог-даних. Джерелами даних є комп'ютерні станції, сервери, застосунки, API-виклики тощо. Лог-дані індексуються для здійснення швидкого доступу в пам'яті.

Здійснення аналізу потоків логів у реальному часі для виявлення взаємозв'язків між окремими подіями забезпечують базові лог-корелятори. Вони шукають закономірності в послідовності подій, збирають лог-дані з різних джерел та пробують їх поєднати для встановлення закономірності, вилучають поодинокі події, фокусуються на критичних комбінаціях подій та генерують повідомлення про спрацювання визначеного правила кореляції. Тобто, лог-корелятор поєднує лог-дані для того, щоб зрозуміти суть ситуації в системах.

Таким чином, на третьому рівні забезпечується перехід від зберігання даних про події та їх опрацювання до зберігання даних про інциденти та їх опрацювання.

На четвертому, аналітичному рівні (SIEM) здійснюється виявлення складних, повільних і прихованих КА. Дані для засобів та систем цього рівня накопичуються та зберігаються тривалий час, що може тривати роками. Також, вони забезпечують

інцидент-орієнтоване виявлення, що полягає в реагуванні на конкретні ознаки вже розпочатої або завершеної КА. При цьому підході процес запускається після виявлення індикатору компрометації чи виявленні аномалії. Таким чином може бути здійснено керування подіями безпеки та виявлення у хостах. Дані для таких систем чи засобів задамо множиною так:

$$M_{DA}^4 = \{m_{DA,1}^4, m_{DA,2}^4, \dots, m_{DA,N_{M_{DA}^4}}^4\}, \quad (2.7)$$

де M_{DA}^4 - i -те джерело даних в корпоративних мережах від систем та засобів четвертого рівня; $i = 1, 2, \dots, N_{M_{DA}^4}$; $N_{M_{DA}^4}$ - кількість джерел.

Елементами множини M_{DA}^4 (формула (2.7)) можуть бути такі:

- 1) $m_{DA,1}^4$ - відомості про статистичний аналіз інцидентів та подій;
- 2) $m_{DA,2}^4$ - відомості про часові ряди для інцидентів та подій;
- 3) $m_{DA,3}^4$ - відомості про базові лінії зафіксованих подій та інцидентів;
- 4) $m_{DA,4}^4$ - ймовірнісні оцінки для усіх подій та інцидентів;
- 5) $m_{DA,5}^4$ - відомості про кореляції багатьох джерел зафіксованих подій та інцидентів.

Кількість елементів множини M_{DA}^4 (формула (2.7)) може бути більше за п'ять елементів.

На п'ятому, поведінковому рівні визначаються відхилення в поведінці, для яких зберігається в системах чи засобах тривала історія активності хостів і користувачів. Опрацювання історичних даних здійснюється з використанням багатовимірної статистики, метрик відстані, кластеризації та графів взаємодії. Засоби та системи цього рівня сфокусовані на виявленні загроз через аналіз аномальної поведінки. Також, їх можна поділити на ті, які здійснюють аналітику поведінки користувачів та сутностей, та ті, які спрямовані на виявлення і реагування на мережні загрози. Перший тип (UEBA) засобів та систем збирає дані про дії користувачів та сутностей, тобто пристроїв, серверів, застосунків у мережі. Як правило для їх функціонування використовується машинне навчання для створення профілю поведінки, при відхиленні від якого вони здійснюють відповідні заходи, включно з сповіщенням адміністраторів безпеки. Другий тип (NDR) засобів та систем безперервно аналізує

весь мережний трафік для виявлення підозрілої активності, яку можуть пропустити традиційні міжмережні екрани. Засоби та системи першого типу здійснюють пошук внутрішніх зловмисників, скомпрометовані облікові записи або аномальні дії адміністраторів, а другого типу - приховане переміщення зловмисників всередині мережі, витік даних, активність вірусів-вимагачів або зв'язок із командними серверами хакерів. Таким чином, перший тип призначений для збору і аналізу даних про користувачів, об'єкти і їхні дії, а другий тип призначений для дослідження того, що відбувається в мережних потоках даних. Для першого та другого типів засобів та систем потрібна ефективна організація пам'яті, оскільки необхідно забезпечити тривалу історію активності користувачів і хостів. Дані для такого типу систем чи засобів задамо множиною так:

$$M_{DP}^5 = \{m_{DP,1}^5, m_{DP,2}^5, \dots, m_{DP,N_{M_{DP}^5}}^5\}, \quad (2.8)$$

де M_{DP}^5 - i -те джерело даних в корпоративних мережах від систем та засобів i -го рівня; $i = 1, 2, \dots, N_{M_{DP}^5}$; $N_{M_{DP}^5}$ - кількість джерел.

Елементами множини M_{DP}^5 (формула (2.8)) можуть бути такі:

- 1) $m_{DP,1}^5$ - відомості про користувачів корпоративних інформаційних систем;
- 2) $m_{DP,2}^5$ - відомості про всіх працівників організації чи підприємства;
- 3) $m_{DP,3}^5$ - відомості про розподіл доступу;
- 4) $m_{DP,4}^5$ - відомості про сформовані профілі користувачів;
- 5) $m_{DP,5}^5$ - відомості про топологію корпоративних мереж;
- 6) $m_{DP,6}^5$ - відомості про розміщення конфіденційної інформації в корпоративних інформаційних системах;
- 7) $m_{DP,7}^5$ - відомості про довірені ресурси в корпоративних інформаційних системах;
- 8) $m_{DP,8}^5$ - відомості про трафік комп'ютерних мереж за період функціонування.

Кількість елементів множини M_{DP}^5 (формула (2.8)) може бути більшою, ніж конкретизовано елементів. Обидва типи таких систем є частиною концепції SOC (Security Operations Center) і часто інтегруються в SIEM-системи для повного

контролю безпеки, тому вони можуть і використовують як правило спільні дані, а тому організація та розподілення історичних даних для них потребують такої їх організації, щоб вони могли бути ефективно використані засобами та системами четвертого та п'ятого рівнів.

На шостому рівні, тобто рівні розширеного виявлення (XDR), здійснюється комплексне виявлення КА у різних доменах з використанням крос-доменної кореляції, ретроспективного аналізу, ризик-скорингу та побудови моделі захисту на основі етапів моделі kill chain, що вимагає формування історичних даних щодо попереднього досвіду функціонування. Дані для такого типу систем чи засобів задамо множиною так:

$$M_{DR}^6 = \{m_{DR,1}^6, m_{DR,2}^6, \dots, m_{DR,N_{M_{DR}^6}}^6\}, \quad (2.9)$$

де M_{DR}^6 - i -те джерело даних в корпоративних мережах від систем та засобів шостого рівня; $i = 1, 2, \dots, N_{M_{DR}^6}$; $N_{M_{DR}^6}$ – кількість джерел.

Елементами множини M_{DR}^6 (формула (2.9)) можуть бути такі:

- 1) $m_{DR,1}^6$ – відомості про КА за їх типами та для різних способів їх здійснення;
- 2) $m_{DR,2}^6$ – відомості про корпоративні інформаційні організації чи підприємства та їх ресурси;
- 3) $m_{DR,3}^6$ – відомості про засоби та системи протидії КА перших п'яти рівнів;
- 4) $m_{DR,4}^6$ – відомості про інциденти під час функціонування від початку встановлення систем та засобів;
- 5) $m_{DR,5}^6$ – відомості про рівні конфіденційної інформації корпоративних мереж та місця її розміщення;
- 6) $m_{DR,6}^6$ – відомості про довірені ресурси за периметром корпоративних мереж.

Кількість елементів множини M_{DR}^6 (формула (2.9)) може бути більшою, ніж конкретизовано елементів. Системи та засоби не тільки використовують і поєднують дані з попередніх п'яти рівнів, а також містять додаткові дані, які необхідні для крос-доменної кореляції, ретроспективного аналізу, ризик-скорингу та побудови моделі захисту на основі етапів моделі kill chain. Ці дані можуть оновлюватись в процесі тривалого функціонування систем та засобів цього рівня.

На цьому, стратегічному рівні (Threat Hunting / SOC) здійснюється пошук аналітиками безпеки приховані кіберзагрози, що оминули автоматичні системи захисту, щоб виявити зловмисників ще до того, як вони можуть завдати шкоди, на відміну від пасивного реагування. Тобто, на цьому рівні здійснюється пошук загроз, включно з потенційними, який виконується в центрі операцій з безпеки. Таким чином забезпечується виявлення складних, прихованих атак, наприклад повільних тривалих КА (APT), які використовують легітимні інструменти та маскуються під нормальну активність. При цьому здійснюється поєднання експертизи експертів, інтуїцію та знання про поведінку зловмисників з аналізом даних, наприклад журналів, мережного трафіку тощо. Далі висувуються гіпотези, перевіряються аномалії, аналізуються патерни поведінки, що в результаті формує нові правила виявлення.

Процес пошуку загроз на цьому рівні здійснюється згідно сформованих гіпотез, аналітичного процесу пошуку ознак компрометації та поведінкових патернів (TTP) зловмисника без очікування повідомлення про загрозу. Типовий цикл пошуку загроз без наявного повідомлення про загрозу містить такі етапи: формування гіпотези; збір даних; аналіз та кореляція; виявлення або спростування; створення нових правил та нових процесів для SIEM. На кожному з цих етапів використовуються дані, які були накопичені під час функціонування систем чи засобів. При формуванні гіпотез це дані про попередні інциденти. Етап збору даних використовує дані розподіленої телеметрії, топології та доменних імен, а також логів подій тощо. Етап аналізу та кореляції використовує дані для пошуку аномалій, здійснення поведінкового аналізу, про комунікацію між хостами, топологію мереж тощо. Для етапу виявлення або спростування гіпотези використовують дані індикаторів компрометації, дані про інциденти тощо. Якщо гіпотеза спростовується, то формуються нові знання і вони формують нові дані в пам'яті. На останньому етапі створюються нові правила та інформація про нові процеси для SIEM, які в подальшому зберігаються в системах або засобах. Взагалі сьомий рівень це рівень здійснення проактивного процесу, в якому аналітики, спираючись на гіпотези, Threat Intelligence та поведінковий аналіз, шукають приховані загрози в телеметрії, які не були виявлені автоматичними засобами, з метою зменшення часу перебування

зловмисника в середовищі. Наявність експерта на цьому етапі виявлення загроз обов'язкова, але він формує гіпотези, також, з урахуванням даних попереднього часу функціонування систем та засобів. Цей етап передбачає інтерпретації поведінки. Застосування систем на основі штучного інтелекту може зменшити шум у використовуваних даних, пришвидшити аналіз та вказати на підозрілі патерни, але не може розуміти мету зловмисників. Джерелами даних є дані попередніх рівнів та дані з пасивного збору телеметрії. Накопичені дані в пам'яті про події та інциденти є одними із ключових елементів в таких системах і засобах, але важливо як саме вони використовують.

Наприклад, чи усі накопичені дані треба аналізувати, щоб отримати певний результат. Може бути так, що достатньо невеликого відсотку від них, щоб мати той же результат. Тому, подання даних, їх організація в пам'яті та використання є важливою складовою частиною процесу підготовки результуючого рішення.

Дані для такого типу систем чи засобів задамо множиною так:

$$M_{DST}^7 = \{m_{DST,1}^7, m_{DST,2}^7, \dots, m_{DST,N_{M_{DST}^7}}^7\}, \quad (2.10)$$

де M_{DST}^7 - i -те джерело даних в корпоративних мережах від систем та засобів сьомого рівня; $i = 1, 2, \dots, N_{M_{DST}^7}$; $N_{M_{DST}^7}$ - кількість джерел.

Елементами множини M_{DST}^7 (формула (2.10)) можуть бути такі:

- 1) $m_{DST,1}^7$ - відомості про лог-дані всіх рівнів;
- 2) $m_{DST,2}^7$ - відомості про мережну інфраструктуру та ресурси;
- 3) $m_{DST,3}^7$ - відомості про систему доменних імен;
- 4) $m_{DST,4}^7$ - відомості про проксі-сервери;
- 5) $m_{DST,5}^7$ - дані з журналів аудиту;
- 6) $m_{DST,6}^7$ - відомості про інциденти;
- 7) $m_{DST,7}^7$ - відомості про наслідки інцидентів;
- 8) $m_{DST,8}^7$ - відомості про результати опрацювання інцидентів;
- 9) $m_{DST,9}^7$ - відомості про індикатори компрометації, включно з даними про хеші, домени, IP-адреси;

10) $m_{DST,10}^7$ – відомості про поведінкові сигнатури;

11) $m_{DST,11}^7$ – відомості про правила.

Кількість елементів множини M_{DST}^7 (формула (2.10)) може бути більшою, ніж конкретизовано елементів. Системи та засоби не тільки використовують і поєднують дані з попередніх шести рівнів, а також містять додаткові дані, які необхідні для пошуку загроз, включно з потенційними, який виконується в центрі операцій з безпеки та стосується загроз, що можуть виникнути. Ці дані використовуються для проведення аналізу експертами і можуть оновлюватись в процесі тривалого функціонування систем та засобів цього рівня.

На сьомому рівні наявність в пам'яті елементів множини M_{DST}^7 потрібна для здійснення аналітики. Для цього експерти з кібербезпеки беруть нову процедуру з матриці MITRE ATT&CK та здійснюють її застосування до наявних лог-даних за останні 90 днів. Таким способом можна знайти пропущену компрометацію. Для побудови базової лінії зв'язку подій чи групи індикаторів потрібні історичні дані, які як правило використовуються для пошуку аномалій і без яких такі пошуки неможливі. Зловмисники як правило не змінюють плани з проведення КА, а змінюють інструменти і засоби для досягнення мети, тому в системах і засобах мають зберігатись кроки КА та патерни КА для аналізу КА, які здійснювались раніше. Крім того, збереження даних необхідно для їх використання адміністраторами з кібербезпеки, які будуть замінювати своїх попередників. Тому, засоби та системи сьомого рівня містять дані в пам'яті, які використовують адміністратори з безпеки.

Таким чином, збереження даних в пам'яті засобів і систем всіх семи рівнів необхідно для протидії КА. Накопичення даних протягом всього часу функціонування систем та засобів протидії КА створює проблеми, які пов'язані з їх кількістю, організацією доступу до них та форматами і шаблонами подання, оскільки вони є різнотипними та різнорідними. Відповідно, такі дані потребують приведення до певних кількох форматів, а також потребують збереження встановлених між ними зв'язків. При використанні даних необхідним, також, є період часу, за який їх можна використовувати і їх актуальність в певні моменти часу. Фактично зловмисники можуть уникати виявлення своїх дій при проведенні КА першочергово уникаючи не

стілки проявів перед засобами чи системами, а відсутності відомостей про нові їх дії, інструменти, засоби, відомості про які відсутні в пам'яті таких систем. Тому, зловмисники намагаються уникати використання дій, дані про які містяться в пам'яті систем та засобів протидії КА, або здійснювати кроки для обмежень використання таких даних в пам'яті. Системи та засоби протидії КА в процесі свого функціонування використовують дані з пам'яті не тільки для протидії КА, але й для вибору своїх подальших кроків, зокрема і під час здійснення КА зловмисниками. Це потребує збереження попередніх кроків систем та засобів з подальшим оцінюванням їх доцільності та ефективності, щоб можна було такі кроки чи їх послідовності використати в подальшому при виборі наступних кроків систем та засобів. Для систем та засобів протидії КА всіх семи рівнів із зростанням рівнів від першого до сьомого рівнів історія даних в пам'яті стає довшою, для збереження, подання та доступу використовується складніший математичний апарат, здійснюється ширший контекст прийняття рішень і їх збереження, а також залежність від сигнатур стає меншою, тому враховуючи зростання складності у збереженні та використанні даних в пам'яті, а також актуальності їх для систем та засобів протидії КА і для прийняття ними рішень щодо подальших кроків, необхідним є покращення організації пам'яті з історичними даними щодо функціонування систем та засобів протидії КА в корпоративних мережах за рахунок розроблення нових моделей їх подання, організації та доступу до них, а також забезпечення збереження під час здійснення КА.

2.2 Моделі організації пам'яті з історичними даними щодо функціонування систем та засобів протидії КА в корпоративних мережах

Елементи чи підсистеми пам'яті в засобах чи системах протидії КА відповідно до поділу на рівні із зростанням рівнів, тобто для вищих рівнів порівняно з нижчими, містять більше даних про попередні події, зберігають дані триваліший час, а також дані для прийняття рішень, тобто загалом в них дані мають довшу історію, складніший математичний апарат їх опрацювання, ширший контекст прийняття рішень на їх основі та менша залежність від окремих сигнатур. Ефективність

функціонування систем та засобів протидії КА в корпоративних мережах суттєво залежить від організації збереження даних в пам'яті та доступу до них, бо процеси в мережах відбуваються швидко і актуальність опрацьовуваних даних повинна бути належною. Тобто використання наявних в поточний момент часу даних, які зберігались в системах та засобах на початку їх встановлення, накопичувались і зберігаються протягом певного часу, включно і про події в контексті КА, які задано за формулами (2.2) та (2.3), впливають на прийняття рішень системами і засобами протидії КА в корпоративних мережах.

Важливим для систем і засобів щодо даних є баланс в тривалості часу, тобто баланс короткотривалого та довготривалого зберігання даних і на цій основі збереження актуальності інформації в цих даних. Тривале зберігання даних може бути недоцільним через втрату актуальності, але може і бути доцільним. Крім того, довготривале накопичення і зберігання даних, включно з врахуванням розподілення пам'яті, призводить до формування їх великих обсягів, що породжує проблеми, які пов'язані з методологічними, технічними та семантичними аспектами. Використання адаптивних технологій в архітектурі систем та засобів протидії КА призводить до зміни парадигми щодо тривалого зберігання даних, оскільки адаптивність може передбачати модифікацію частини з них і, відповідно, втрату актуальності решти. Це формує внутрішні протиріччя у моделях організації пам'яті систем та засобів протидії КА в корпоративних мережах.

Таким чином, організація пам'яті з історичними даними в системах та засобах протидії КА в корпоративних мережах потребує вирішення протиріччя щодо необхідності поєднання вимог до повноти, тривалості та актуальності збережених даних з обмеженнями обчислювальних ресурсів, їх розподілення та адаптивністю систем аналізу.

Розглянемо особливості зберігання даних та їх вплив на ефективність систем та засобів протидії КА. Визначимо множину впливів M_{vp}^p на збереження та актуальність даних в різних аспектах їх використання та забезпечення доступу і актуальності так:

$$M_{vp}^p = \{m_{vp,1}^p, m_{vp,2}^p, \dots, m_{vp,n_{M_{vp}^p}}^p\}, \quad (2.11)$$

де $m_{vp,i}^p$ - i - тий елемент множини впливів M_{vp}^p , що означає певний конкретний вплив на ефективність систем та засобів протидії КА; $i = 1, 2, \dots, n_{M_{vp}^p}$; $n_{M_{vp}^p}$ - кількість елементів множини впливів M_{vp}^p .

Елементи множини впливів M_{vp}^p можуть бути такими:

- 1) $m_{vp,1}^p$ - обсяг дискових ресурсів для зберігання даних;
- 2) $m_{vp,2}^p$ - класифікація даних за приналежністю до завдань систем та засобів;
- 3) $m_{vp,3}^p$ - структурування та представлення класифікованих даних;
- 4) $m_{vp,4}^p$ - індексування даних в класах;
- 5) $m_{vp,5}^p$ - наявність різних методів пошуку даних за вимогами;
- 6) $m_{vp,6}^p$ - розподілення даних в мережі;
- 7) $m_{vp,7}^p$ - розміщення даних;
- 8) $m_{vp,8}^p$ - наявність різних методів опрацювання даних;
- 9) $m_{vp,9}^p$ - топології мережі;
- 10) $m_{vp,10}^p$ - розподіл ролей користувачів корпоративних мереж
- 11) $m_{vp,11}^p$ - технічна організація політик доступу в корпоративних мережах;
- 12) $m_{vp,12}^p$ - оновлення програмного забезпечення в комп'ютерних станціях;
- 13) $m_{vp,13}^p$ - доповнення / вилучення комп'ютерних станцій в корпоративних мережах;
- 14) $m_{vp,14}^p$ - заміна операційних систем в комп'ютерних станціях;
- 15) $m_{vp,15}^p$ - оновлення систем та засобів забезпечення безпеки і захисту корпоративних мереж та їх вузлів повністю або частково;
- 16) $m_{vp,16}^p$ - часткова заміна мережного обладнання корпоративних мереж;
- 17) $m_{vp,17}^p$ - оновлення баз сигнатур, баз правил;
- 18) $m_{vp,18}^p$ - відсутність оновлень баз сигнатур, баз правил протягом тривалого часу;

19) $m_{vp,19}^p$ – відсутність змін в топології та апаратно-програмних і програмних засобах корпоративних мереж, включно із засобами та системами забезпечення безпеки і захисту, протягом тривалого часу експлуатації корпоративних мереж;

20) $m_{vp,20}^p$ – автоматичне формування нових сигнатур для баз сигнатур засобами та системами забезпечення безпеки і захисту корпоративних мереж;

21) $m_{vp,21}^p$ – застосування методів протидії КА, які оперативно виконують завдання;

22) $m_{vp,22}^p$ – застосування методів протидії КА, які здійснюють аналіз ситуацій в корпоративних мережах і, відповідно, можуть повільно виконати завдання.

Кількість елементів множини впливів може бути більшою. Визначені двадцять два елементи охоплюють основні впливи на збереження та актуальність даних в різних аспектах їх використання та забезпечення доступу до них.

Обсяг дискових ресурсів для зберігання даних суттєво впливає на збереження даних, оскільки вони постійно накопичуються та потребують значного простору для розміщення. Дані, які зберігаються, мають різні типи, структуру, отримані з різних джерел та позиціонуються за різним призначенням для їх подальшого використання, тому для їх зберігання потрібно здійснювати класифікацію нових даних за приналежністю до завдань систем та засобів з подальшим їх розміщенням відповідно до приналежності визначеним класам. Для того, щоб класифіковані дані було зручно і швидко отримувати з місць їх розміщення потрібно забезпечити належне структурування та представлення. Нові дані повинні доповнювати наявні дані і, тому, їх узгодження щодо порядку повинно реалізовано через індексування даних у відповідних класах. Також, можуть варіанти з даними, які віднесені до певних класів, а в них надмірна або недостатня кількість параметрів порівняно з наявними даними в класах, що потребує додаткового їх приведення у відповідність, бо впливатиме на їх використання. В системах та засобах протидії КА наявні різні методи пошуку даних за вимогами, які впливають на організацію індексування та формати розміщення даних. При розміщенні даних в корпоративних мережах в зв'язку з потребою ефективного використання обсягів пам'яті комп'ютерних станцій та серверів вони

можуть бути розміщені за класами в різних вузлах мережі, а може бути частина класів подана в різних вузлах в межах даних кожного класу, що впливатиме на збереження та актуальність даних в різних аспектах їх використання та забезпечення доступу до них. Тобто дані певного класу можуть бути фізично розміщені в різних місцях. Обсяги даних визначають також швидкість опрацювання, а також ця швидкість залежить від наявності різних методів опрацювання даних, які реалізовані в системах і засобах протидії КА. Наявність різних методів опрацювання даних має таку особливість, що продуктивність частини з них знижується із збільшенням обсягів даних.

На збереження та доступ до даних впливають топології мережі, розподіл ролей користувачів корпоративних мереж, технічна організація політик доступу в корпоративних мережах, технічна організація політик доступу в корпоративних мережах, оновлення програмного забезпечення в комп'ютерних станціях, доповнення / вилучення комп'ютерних станцій в корпоративних мережах, заміна операційних систем в комп'ютерних станціях, оновлення систем та засобів забезпечення безпеки і захисту корпоративних мереж та їх вузлів повністю або частково, часткова заміна мережного обладнання корпоративних мереж, оновлення баз сигнатур, баз правил, відсутність оновлень баз сигнатур і баз правил протягом тривалого часу, відсутність змін в топології та апаратно-програмних і програмних засобах корпоративних мереж, включно із засобами та системами забезпечення безпеки і захисту, протягом тривалого часу експлуатації корпоративних мереж, автоматичне формування нових сигнатур для баз сигнатур засобами та системами забезпечення безпеки і захисту корпоративних мереж, застосування методів протидії КА, які оперативно виконують завдання, та застосування методів протидії КА, які здійснюють аналіз ситуацій в корпоративних мережах і, відповідно, можуть повільно виконати завдання.

Введемо множину функцій $M_{F_{vp,i}}^p$ для оцінювання впливів елементів множини впливів M_{vp}^p на збереження та актуальність даних в різних аспектах їх використання та забезпечення доступу і актуальності (формула (2.11)) так:

$$v_{vp,i}^p = F_{vp,i}^p(m_{vp,i}^p), \quad (2.12)$$

де $v_{vp,i}^p$ – значення функції оцінювання $F_{vp,i}^p$ для елемента $m_{vp,i}^p$; $m_{vp,i}^p$ – i – тий елемент множини впливів M_{vp}^p , що означає певний конкретний вплив на ефективність систем та засобів протидії КА; $i = 1, 2, \dots, n_{M_{vp}^p}$; $n_{M_{vp}^p}$ – кількість елементів множини впливів M_{vp}^p .

Величини $v_{vp,i}^p$ ($i = 1, 2, \dots, n_{M_{vp}^p}$; $n_{M_{vp}^p}$ – кількість елементів множини впливів M_{vp}^p) формують вектор значень оцінювання впливів на збереження та актуальність даних в різних аспектах їх використання та забезпечення доступу і актуальності, який задамо так:

$$V_{vp,t}^p = \left(t, v_{vp,1}^p, v_{vp,2}^p, \dots, v_{vp,n_{M_{vp}^p}}^p \right), \quad (2.13)$$

де t – поточний час, в який було отримано інформацію про дані і здійснено оцінювання впливу на них; $v_{vp,i}^p$ – значення функції оцінювання $F_{vp,i}^p$ для елемента $m_{vp,i}^p$; $i = 1, 2, \dots, n_{M_{vp}^p}$; $n_{M_{vp}^p}$ – кількість елементів множини впливів M_{vp}^p (формула (2.11)).

Також, задамо множину функцій для оцінювання впливів елементів множини впливів M_{vp}^p вектором її елементів, тобто функцій впливів $F_{vp,i}^p$ ($i = 1, 2, \dots, n_{M_{vp}^p}$; $n_{M_{vp}^p}$ – кількість елементів множини впливів M_{vp}^p) аналогічно до вектору значень оцінювання впливів (формула (2.13)) так:

$$F_{vp,t}^p = \left(t, F_{vp,1}^p, F_{vp,2}^p, \dots, F_{vp,n_{M_{vp}^p}}^p \right), \quad (2.14)$$

де t – поточний час, в який було отримано інформацію про дані і здійснено оцінювання впливу на них з використанням функцій $F_{vp,i}^p$; $F_{vp,i}^p$ – функція оцінювання елемента $m_{vp,i}^p$; $i = 1, 2, \dots, n_{M_{vp}^p}$; $n_{M_{vp}^p}$ – кількість елементів множини впливів M_{vp}^p (формула (2.11)).

Оцінювання за формулами (2.12) – (2.14) може бути здійснено за різними варіантами. Наприклад, можуть бути оцінені крайні останні значення елементів

множини впливів M_{vp}^p (формула (2.11)) або середні значення з усіх накопичених чи з певного проміжку часу тощо.

Таким чином, визначено множину впливів на збереження та актуальність даних в різних аспектах їх використання та забезпечення доступу і актуальності (формула (2.11)) та функції і вектор значень оцінювання таких впливів (формули (2.12) – (2.14)). Вони необхідні для забезпечення організації пам'яті з історичними даними при функціонуванні систем протидії КА в корпоративних мережах. Значення оцінок використовуватимуться для прийняття рішень щодо їх оптимізації, доповнення, вилучення, переміщення та опрацювання. Також, вони необхідні для безпосереднього врахування при здійсненні протидії КА наявними методами.

Деталізуємо функцію $F_{vp,1}^p$ оцінювання елементу $m_{vp,i}^p$ так:

$$v_{vp,1}^p = F_{vp,1}^p(m_{vp,1}^p) = \frac{\sum_{l=1}^{n_{vp,1}} d_{vp,1,l}}{\sum_{l=1}^{n_{z,vp,1}} d_{z,vp,1,l}}, \quad (2.15)$$

де $F_{vp,1}^p$ – функція оцінювання елементу $m_{vp,1}^p$; $d_{z,vp,1,l}$ – обсяг l – того дискового ресурсу для зберігання даних із загальної кількості таких ресурсів; $l = 1, 2, \dots, n_{z,vp,1}$; $n_{z,vp,1}$ – загальна кількість наявних окремих дискових ресурсів для зберігання даних; $d_{vp,1,l}$ – обсяг l – того дискового ресурсу для зберігання даних із визначених таких ресурсів та задіяних для зберігання; $l = 1, 2, \dots, n_{vp,1}$; $n_{vp,1}$ – кількість наявних задіяних окремих дискових ресурсів для зберігання даних.

Також, наприклад, деталізуємо функцію $F_{vp,6}^p$ оцінювання елементу $m_{vp,6}^p$, який відображає розподілення даних в мережі, так:

$$v_{vp,6}^p = F_{vp,6}^p(m_{vp,6}^p) = 1 - D_{l=1}^{n_{vp,6}} \frac{k_{vp,6,l}}{k_{z,vp,6,l}}, \quad (2.16)$$

де D – добуток; $l = 1, 2, \dots, n_{vp,6}$; $n_{vp,6}$ – загальна кількість класів даних, які потребують збереження і відрізняються між собою представленням даних, їх структуруванням та індексуванням; $F_{vp,6}^p$ – функція оцінювання елементу $m_{vp,6}^p$; $k_{z,vp,6,l}$ – загальна кількість окремих дискових ресурсів, в яких може бути розміщений l – тий клас даних; $k_{vp,6,l}$ – фактична кількість окремих дискових ресурсів, в яких

розміщено l – тий клас даних в поточний момент часу t (формула (2.13)); $m_{vp,6}^p$ – розподілення даних в мережі.

Крім визначення функції $F_{vp,6}^p$ за формулою (2.16), її можна деталізувати також із врахуванням кількості місць, в які розподілено кожен клас даних окремо, тоді таке значення буде знаходитись в знаменнику, а в чисельнику буде розміщено одиницю.

Частина класів даних може бути розміщена в однакових вузлах мережі, тобто кількість класів даних може бути більше, ніж кількість вузлів, в які вони розміщені. В загальному може бути багато різних варіантів щодо кількості класів даних та кількості вузлів корпоративних мереж, в яких можуть бути розміщені класи даних, зокрема: кожен клас даних розміщено в окремому вузлі і тільки один клас даних розміщено у окремому вузлі; всі класи даних розміщені в одному вузлі; всі класи даних розміщені в більше, ніж одному вузлі, але при цьому окремо певний клас даних не розміщується в декількох вузлах; всі класи даних розміщені в більше, ніж одному вузлі, але при цьому окремо певні класи даних розміщуються в декількох вузлах; всі класи даних розміщені в більше, ніж одному вузлі і всі класи даних обов'язково розміщуються в декількох вузлах тощо. З врахуванням таких особливостей деталізуємо функцію $F_{vp,7}^p$ оцінювання елементу $m_{vp,7}^p$, який відображає розміщення даних з урахуванням того, що частина класів даних може бути розміщена винятково в одному вузлі мережі і решта в різних або всі в одному чи всі в різних, так:

$$v_{vp,7}^p = F_{vp,7}^p(m_{vp,7}^p) = \frac{\sum_{l=1}^{n_{vp,6}} \frac{1}{n_{vp,7,l}}}{n_{vp,6}}, \quad (2.17)$$

де $n_{vp,6}$ – загальна кількість класів даних, які потребують збереження і відрізняються між собою представленням даних, їх структуруванням та індексуванням; $l = 1, 2, \dots, n_{vp,6}$; $F_{vp,7}^p$ – функція оцінювання елементу $m_{vp,7}^p$; $n_{vp,7,l}$ – загальна кількість окремих дискових ресурсів, в яких розміщено l - клас даних в поточний момент часу t (формула (2.13)); $m_{vp,7}^p$ – розміщення даних.

Решту функцій оцінювання $F_{vp,i}^p$ для елементу $m_{vp,i}^p$ ($i = 1, 2, \dots, n_{M_{vp}}^p$; $n_{M_{vp}}^p$ – кількість елементів множини впливів M_{vp}^p) можна деталізувати аналогічно.

Для забезпечення розмежування даних в системах та засобах протидії КА і їх використання на різних семи рівнях введемо та визначимо класи даних так:

$$M_K^D = \bigcup_{i=1}^{n_{M_K^D}} K_i^D, \quad (2.18)$$

де K_i^D – клас даних від джерел інформації різних рівнів (формули (2.2) та (2.3)); $i = 1, 2, \dots, n_{M_K^D}$; $n_{M_K^D}$ – кількість класів; M_K^D – множина класів даних.

Кожен клас даних містить унікальну для нього інформацію і має відповідне для неї структурування та індексування. А також, може бути використаний винятково на певному одному чи декількох рівнях або може бути використаний для всіх рівнів, включно з тим, що частково на певних рівня або повністю на усіх рівнях. Така вимога щодо класів даних актуалізує необхідність ефективної їх організації, збереження та актуалізації в пам'яті для засобів чи систем протидії КА.

Класи даних узгоджуються з векторами для забезпечення зберігання даних від джерел інформації різних рівнів з формули (2.2), які задамо загальними для всіх джерел інформації поданнями за формулою (2.3).

Формування класу даних від джерел інформації різних рівнів (формули (2.2) та (2.3)), які задано за формулою (2.18), задамо введенням координати $v_{DI,i,18}^D$ (формула (2.3)) у вектор $V_{DI,i}^D$ ($i = 1, 2, \dots, N_{M_{DI}^D}$; $N_{M_{DI}^D}$ – кількість векторів для забезпечення зберігання даних від джерел інформації різних рівнів), в які буде відображено структуру даних, що потребують збереження. Тоді, вектор $V_{DI,i}^D$ будемо включати до класу K_i^D ($i = 1, 2, \dots, n_{M_K^D}$; $n_{M_K^D}$ – кількість класів) з множини M_K^D (формула (2.18)), якщо сформована структура даних задана координатою $v_{DI,i,18}^D$ збіжна зі структурою даних класу. Якщо для сформованого вектору $V_{DI,i}^D$ координата $v_{DI,i,18}^D$ не буде відповідати жодному класу K_i^D , тоді такий вектор надсилаються адміністратору систему з відповідного змісту поясненням для прийняття ним рішення.

Множина функцій для оцінювання впливів елементів множини впливів M_{vp}^p на збереження та актуальність даних в різних аспектах їх використання та забезпечення доступу і актуальності (формула (2.11)), а також вектори, які задані за формулами (2.13) і (2.14), стосуються оцінювання в поточний момент часу t , тому введемо

матрицю оцінювання з урахуванням того, що доповнювані дані до класів будуть фіксуватись разом з часом їх отримання і формування. Це дасть змогу при виборі даних для використання в системах і засобах протидії КА здійснювати вибір за певні часові проміжки або враховувати інтервали доповнення до класів. Матрицю часу для даних класів задамо з урахуванням координати $v_{DI,i,9}^D$, що фіксує час розміщення даних в базах даних, вектору $V_{DI,i}^D$ ($i = 1, 2, \dots, N_{M_{DI}}^D$; $N_{M_{DI}}^D$ – кількість векторів для забезпечення зберігання даних від джерел інформації різних рівнів) і відповідного класу K_i^D ($i = 1, 2, \dots, n_{M_K}^D$; $n_{M_K}^D$ – кількість класів) з множини M_K^D (формула (2.18)) так:

$$M_1^T = \begin{pmatrix} K_1^D & v_{DI,1,1}^{D,1} & v_{DI,1,2}^{D,1} & \dots & v_{DI,1,j}^{D,1} & v_{DI,1,j+1}^{D,1} & \dots & v_{DI,1,18}^{D,1} \\ K_1^D & v_{DI,2,1}^{D,1} & v_{DI,2,2}^{D,1} & \dots & v_{DI,2,j}^{D,1} & v_{DI,2,j+1}^{D,1} & \dots & v_{DI,2,18}^{D,1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ K_1^D & v_{DI,n_{K_1}^D,1}^{D,1} & v_{DI,n_{K_1}^D,2}^{D,1} & \dots & v_{DI,n_{K_1}^D,j}^{D,1} & v_{DI,n_{K_1}^D,j+1}^{D,1} & \dots & v_{DI,n_{K_1}^D,18}^{D,1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ K_{n_{M_K}^D}^D & v_{DI,n_{K_{n_{M_K}^D}}^D,1}^{D,n_{M_K}^D} & v_{DI,n_{K_{n_{M_K}^D}}^D,2}^{D,n_{M_K}^D} & \dots & v_{DI,n_{K_{n_{M_K}^D}}^D,j}^{D,n_{M_K}^D} & v_{DI,n_{K_{n_{M_K}^D}}^D,j+1}^{D,n_{M_K}^D} & \dots & v_{DI,n_{K_{n_{M_K}^D}}^D,18}^{D,n_{M_K}^D} \end{pmatrix}, \quad (2.19)$$

де K_i^D – клас з множини M_K^D (формула (2.18)); $i = 1, 2, \dots, n_{M_K}^D$; $n_{M_K}^D$ – кількість класів; $s = 1, 2, \dots, n_{K_i^D}$; $n_{K_i^D}$ – кількість елементів i – того класу; $v_{DI,2,j}^{D,i}$ – j -та координата вектору, який відноситься до i -того класу K_i^D ; $j = 1, 2, \dots, 18$.

Таким чином, введемо компоненту $v_{DI,i,18}^D$ у вектори $V_{DI,i}^D$, якій призначимо номер класу K_i^D . Тоді, вектори $V_{DI,i}^D$ будуть класифікуватись згідно неї, а її значення буде визначено з урахуванням структурування даних та часу.

Для оцінювання даних згідно формули (2.19) важливим показником є дев'ята координата $v_{DI,i,9}^D$, яка відображає час фіксування даних. В межах певного одного класу цей час збільшується від першого зафіксованого в класі значення, тобто координат вектору, до останнього значення. Серед всіх цих даних можуть бути такі, що повторюються за основною групою показників, які подані координатами, і відрізняються лише частиною показників, включно з вісімнадцятим показником. Тобто, в класі можуть бути наявні підкласи, які описують повторювані події, що

зафіксовані в різний час. Тоді, при внесенні нових даних виникає потреба у здійсненні перевірки наявних даних певного класу з метою встановлення повторів, які відрізнятимуться за часом. Для встановлення зв'язків між такими даними внесемо у вектор $V_{DI,i}^D$ ($i = 1, 2, \dots, N_{M_{DI}}^D$; $N_{M_{DI}}^D$ – кількість векторів для забезпечення зберігання даних від джерел інформації різних рівнів) дев'ятнадцяту координату $v_{DI,i,19}^D$, яка буде містити номер підкласу для кожного даного. Якщо між всіма даними одного класу не буде встановлено повторів за групою параметрів, що відображені відповідними координатами, то кількість підкласів буде збіжна з кількістю даних класу. Компонента $v_{DI,i,19}^D$ вектору з певного класу буде відповідати за збіжність групи координат стосовно часу, який відображено координатою $v_{DI,i,9}^D$. Також, за потреби можуть бути введені додаткові координати після дев'ятнадцятої координати, які теж можуть відповідати за номери підкласів тих векторів, для яких обрані групи параметрів, що будуть використані в контексті дослідження їх зміни протягом певного часового періоду.

Оцінювання впливів на збереження та актуальність даних в різних аспектах їх використання та забезпечення доступу і актуальності будемо здійснювати в різних варіантах. Для здійснення оцінювання використаємо формулу (2.13). Також, оцінювання буде враховувати наявні дані у відповідних класах K_i^D , які задано з врахуванням формули (2.19). Така інформація необхідна в контексті актуальності даних та реалізації властивості адаптивності щодо зміни до використовуваних історичних даних через певний час.

Спочатку оцінимо крайні за часом значення і сформуємо відповідну для отриманих значень матрицю так, що кожен елемент даних з певного класу отримує і зберігає додатково результат оцінювання за формулою (2.13):

$$M_{v_{DI,m,n}^{D,i}}^T = \begin{pmatrix} K_i^D & v_{DI,m,n}^{D,i} & t_1^{v_{DI,m,n}^{D,i}} & v_{vp,1}^p & v_{vp,2}^p & \dots & v_{vp,n_{M_{vp}^p}}^p \\ K_i^D & v_{DI,m,n}^{D,i} & t_2^{v_{DI,m,n}^{D,i}} & v_{vp,1}^p & v_{vp,2}^p & \dots & v_{vp,n_{M_{vp}^p}}^p \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ K_i^D & v_{DI,m,n}^{D,i} & t_z^{v_{DI,m,n}^{D,i}} & v_{vp,1}^p & v_{vp,2}^p & \dots & v_{vp,n_{M_{vp}^p}}^p \end{pmatrix}, \quad (2.20)$$

де $t_q^{D,i}$ – час, в який було отримано інформацію про дані і здійснено оцінювання впливу на них; $q = 1, 2, \dots, z$; z – номер останнього часу, в який було отримано інформацію про дані і здійснено оцінювання впливу на них; $v_{vp,i}^p$ – значення функції оцінювання $F_{vp,i}^p$ для елементу $m_{vp,i}^p$; $n = 1, 2, \dots, n_{M_{vp}^p}$; $n_{M_{vp}^p}$ – кількість елементів множини впливів M_{vp}^p (формула (2.11)); m – номер елементу в класі; K_i^D – клас даних від джерел інформації різних рівнів (формули (2.2) та (2.3)); $i = 1, 2, \dots, n_{M_K^D}$; $n_{M_K^D}$ – кількість класів.

Матриця $M_{v_{DI,m,n}^{D,i}}^T$ (формула (2.20)) містить інформацію про послідовне оцінювання впливів та відповідних даних з класів, тобто накопичення даних супроводжується оцінюванням впливів. Тоді, така матриця буде містити статичні результати оцінювання для кожного елементу даних. Для здійснення аналізу та використання таких даних і значень оцінювання системи та засоби протидії КА повинні використовувати не тільки крайні останні значення, але й середні значення за певні проміжки часу або усі значення. Крім того, в засобах і системах, що відносяться до вищих з семи рівнів, необхідно встановлювати зв'язки між певною кількістю кроків, даних тощо і це вимагатиме реалізація властивості адаптивності. Тобто, в межах одного даного, для якого міститься різний час оцінювання впливів, необхідно забезпечити встановлення зв'язків в послідовності оцінювань з різним часом.

Таким чином, значення оцінювання даних будуть змінюватись з часом через зміни операційного середовища в корпоративних мережах формуючи певний зсув оцінок даних.

Розглянемо особливості даних і впливів в контексті різних часових термінів їх зберігання та використання. Перші вісім елементів ($m_{vp,1}^p - m_{vp,8}^p$) множини впливів M_{vp}^p відносяться до ресурсних обмежень, наступні три ($m_{vp,9}^p - m_{vp,11}^p$) – характеризують семантичну актуальність даних, одинадцять наступних елементів ($m_{vp,11}^p - m_{vp,22}^p$) – відносяться до зберігання та оновлення сигнатур різної природи та призначення.

Зі зростанням масштабів корпоративних мереж обсяг даних, що формується зростає та потребує систематизації, належного зберігання та оптимізації. Постійне накопичення первинних даних призводить до значного збільшення вимог до дискових ресурсів, ускладнення процесів індексації та пошуку, зниження продуктивності кореляційного аналізу. Тому, виникає протиріччя між потребою збереження повної історії подій та ускладненнями з її оперативного використання.

Історичні дані з часом втрачають контекстну релевантність унаслідок змін топології мережі, зміни ролей користувачів та політик доступу, оновлення програмного забезпечення та засобів захисту. Події, які були інтерпретовані як аномальні в певний часовий період, можуть бути нормальними для іншого стану системи. Це унеможлиблює пряме використання довготривало збережених даних без урахування часової динаміки контексту.

Первинні сигнатури (IP-адреси, доменні імена, хеші файлів, сигнатури мережних пакетів) характеризуються високою точністю, проте не тривалою довговічністю. Їх використання у довготривалій перспективі супроводжується швидкими застаріваннями та втратою актуальності, зростанням кількості хибнопозитивних спрацювань, конфліктами між історичними та актуальними наборами сигнатур. Це формує протиріччя між детермінованістю сигнатурного підходу та необхідністю адаптивності системи до нових типів атак.

Системи та засоби протидії КА активно використовують методи машинного навчання та поведінкового аналізу, але тривале зберігання історичних даних призводить до ефекту концептуального зсуву, за якого статистичні властивості даних змінюються з часом. Тому, виникає протиріччя між необхідністю збереження історичного досвіду для виявлення складних та тривалих атак і потребою адаптації моделей до поточного стану мережі та актуальних загроз. Довготривале зберігання історичних даних необхідне для розслідування інцидентів та аудиту безпеки. Системи чи засоби протидії КА у реальному часі потребують мінімальної затримки обробки та узагальнених ознак, що зумовлює протиріччя між архітектурами пам'яті, орієнтованими на швидкодію та архівними сховищами даних, які призначені для глибокого аналізу.

Довготривале зберігання даних трансформується з накопичення інформації у процес формування знань про загрози, бо в системах чи засобах протидії КА організація пам'яті повинна використовувати багаторівневі моделі зберігання, які передбачають розділення первинних неопрацьованих даних, ознак та знань, використання часових коефіцієнтів значущості для історичних даних, перехід від сигнатурної пам'яті до зберігання поведінкових шаблонів, ієрархію пам'яті відповідно до призначення. Довготривале зберігання історичних даних у системах чи засобах протидії КА є необхідною умовою підвищення рівня кіберзахисту.

Основні проблеми та протиріччя виникають на перетині обсягу даних, їх семантичної актуальності та адаптивності моделей аналізу. Тривале зберігання даних буде впливати на прийняття рішень, якщо буде використано адаптивну модель. Дані з різних епох будуть несумісні, бо події з попередніх періодів неможливо інтерпретувати однозначно, тобто вони втрачають контекст. Але для прийняття рішень вони можуть бути потрібні. Крім цього, потрібно встановити глибину повернення за часом до даних, щоб вони були корисними при прийнятті рішень.

Таким чином, ефективне вирішення протиріч щодо довготривалого зберігання даних, виникнення яких відбувається на перетині обсягу даних, їх семантичної актуальності та адаптивності моделей аналізу, можливе лише в межах багаторівневих моделей організації пам'яті, що поєднують зберігання даних, ознак та знань із урахуванням часової динаміки, контексту та адаптивності систем і засобів протидії КА.

2.3 Моделі обманних технологій в системах та засобах протидії КА в корпоративних мережах

Системи та засоби протидії КА в корпоративних мережах згідно з організацією елементів чи підсистем пам'яті поділено на сім рівнів з урахуванням ієрархії. Додатково на певних рівнях можуть бути застосовані засоби та системи, які розроблені із використанням обманних технологій. Вони доповнюють системи та засоби протидії КА та теж використовують дані щодо подій за попередній час їх

функціонування. Тому розглянемо моделі обманних систем та засобів, включно з приманками і пастками, і поєднані з системами та засобами певних рівнів в контексті зберігання для них даних за попередні часові періоди їх функціонування.

Моделі обманних технологій формуються на основі попереднього досвіду функціонування систем і засобів протидії комп'ютерним атакам, що акумулюється у вигляді журналів подій, інцидентів безпеки, сценаріїв атак, а також результатів реагування та розслідувань. Цей досвід є історичною основою для побудови, налаштування та еволюції таких систем. На відміну від класичних засобів виявлення, які здебільшого орієнтовані на пасивний аналіз подій, обманні технології реалізують активну модель захисту, у якій попередній досвід використовується для створення контрольованого середовища введення зловмисника в оману. Легітимний користувач не має причин взаємодіяти з обманними об'єктами, тому сигнал з них має надзвичайно низький рівень хибнопозитивних спрацювань.

Історичні дані відіграють ключову роль у формуванні моделей обманних технологій, зокрема: патерни попередніх атак (типи сканування, експлуатації, латеральне горизонтальне переміщення); характерні техніки, технології і процедури зловмисників згідно з MITRE ATT&CK; топологія реальної інфраструктури, яка раніше була об'єктом атак; поведінкові особливості користувачів і систем у нормальному та аномальному режимах. З урахуванням такої інформації, що формує досвід визначаються точки розміщення приманок і токенів, формується реалістичність обманних об'єктів, підвищується правдоподібність взаємодії зловмисника з фальшивими ресурсами. Ці системи та засоби створюють для зловмисників фальшиві активи (наприклад, сервери, сервіси, облікові записи), фальшиві дані (наприклад, API-ключі, токени), контрольоване «поле атаки», де будь-яка активність майже гарантовано є зловмисною. Такі фальшиві дані необхідно теж подавати в пам'яті систем та засобів протидії КА, а також видозмінювати. Це призводить до тих же проблем, що й при роботі з історичними даними. Крім того, використання в засобах чи системах обманних технологій теж передбачає використання даних, що відображають попередні досвід функціонування систем та засобів і прийняття рішень на його основі.

Розглянемо співвіднесення обманних технологій з рівнями ієрархії засобів протидії КА.

Системи та засоби першого рівня, тобто джерел подій, охоплюють всі первинні елементи, з яких формується телеметрія безпеки. Саме на цьому рівні виникають початкові події, що згодом аналізуються, корелюються та інтерпретуються вищими рівнями системи захисту. У контексті обманних технологій цей рівень є фундаментальним, оскільки забезпечує створення первинних обманних артефактів, взаємодія з якими ініціює подальший процес виявлення та реагування. На цьому рівні обманні моделі реалізуються базовими обманними об'єктами, що імітують реальні компоненти корпоративних мереж. До них віднесемо хости та сервери (реальні або віртуальні), мережні сервіси та порти, облікові записи користувачів і сервісів, файли, каталоги, конфігураційні елементи, облікові дані (паролі, ключі, токени). Значна частина з них вимагає підтримки певної кількості даних і функціонує як джерела подій, генеруючи журнали доступу, автентифікації, мережної взаємодії та системної активності у відповідь на дії зловмисника. Відтворення характеристик реальних ресурсів обманними системами з приманками та пастками на першому рівні полягає в їх максимальній правдоподібності, що включає іменування хостів і сервісів відповідно до прийнятих у системі шаблонів, типові конфігурації операційних систем і сервісів, реалістичну структури файлових систем, автентифікаційне операційне середовище.

Формування обманних моделей на першому рівні безпосередньо базується на попередньому досвіді функціонування систем і засобів протидії КА. Задамо дані цього рівня множиною, що буде підмножиною всіх даних, але буде включати елементи про класи даних, які зберігаються та можуть бути використані обманними системами першого рівня, так:

$$M_{D,1}^{OS} = \{m_{D,1,1}^{OS}, m_{D,1,2}^{OS}, \dots, m_{D,1,n_{M_{D,1}^{OS}}}^{OS}\}, \quad (2.21)$$

де $m_{D,1,i}^{OS}$ – i -тий елемент, що визначає клас даних для підтримки об'єктів обманних систем; $i = 1, 2, \dots, n_{M_{D,1}^{OS}}$; $n_{M_{D,1}^{OS}}$ – кількість типів даних.

Елементами множини $M_{D,1}^{OS}$ (формула (2.21)) для обманних технологій першого рівня визначимо такі:

- 1) $m_{D,1,1}^{OS}$ – елемент, що визначає клас історичних даних журналів подій (logs);
- 2) $m_{D,1,2}^{OS}$ – елемент, що визначає клас збережених конфігурацій реальних систем;
- 3) $m_{D,1,3}^{OS}$ – елемент, що визначає клас даних про раніше зафіксовані атаки та інциденти;
- 4) $m_{D,1,4}^{OS}$ – елемент, що визначає клас даних-результатів аналізу компрометацій.

Ці класи даних з множини $M_{D,1}^{OS}$ (формула (2.21)) використовуватимуться для вибору типів обманних об'єктів, визначення місць їх розміщення, налаштування параметрів взаємодії, формування подій, типових для реальної інфраструктури корпоративних мереж. Обманні об'єкти разом з даними є контекстно прив'язані до конкретної системи, що підвищує ефективність протидії КА.

Таким чином, засоби і системи першого рівня на основі обманних технологій формують первинний обманний простір, який створюється на основі історичного досвіду функціонування. При функціонуванні обманних технологій на першому рівні виділимо дві концептуально різних, але взаємопов'язані підмножини даних. До першої віднесемо історичні дані функціонування систем і засобів протидії, що використовують обманні технології, а до другої - дані, які згенеровані безпосередньо обманними об'єктами. Розмежування цих класів даних є принципово важливим для коректного зберігання, аналізу та подальшого використання у вищих рівнях ієрархії.

Задамо історичні дані систем та засобів всіх семи рівнів, що використовують обманні технології, так:

$$M_{D,r}^{OS,D} = \bigcup_{i=1}^2 K_{D,r,i}^{OS,D}, \quad (2.22)$$

де $K_{D,r,i}^{OS,D}$ – i -тий клас даних об'єктів обманних систем; $i = 1, 2$; r – номер рівня засобів і систем протидії КА; $r = 1, 2, \dots, 7$.

До першого класу $K_{D,1,1}^{OS,D}$ першого рівня віднесемо накопичені історичні дані реальних систем захисту, які не є результатом взаємодії із обманними об'єктами, але

визначають контекст їх побудови. Задамо клас $K_{D,1,1}^{OS,D}$ векторами підкласів, оскільки дані мають різну інформацію і відповідно структурування, так:

$$K_{D,1,1}^{OS,D} = \bigcup_{i=1}^{n_{M_{D,1,1}^{OS}}} K_{D,1,1,i}^{OS,D}, \quad (2.23)$$

де $K_{D,1,1,i}^{OS,D}$ – i -тий підклас даних об'єктів обманних систем класу $K_{D,1,1}^{OS,D}$; $i = 1, 2, \dots, n_{M_{D,1,1}^{OS}}$; $n_{M_{D,1,1}^{OS}}$ – кількість підкласів.

Елементи класу $K_{D,1,1}^{OS,D}$ з підкласів $K_{D,1,1,i}^{OS,D}$ такі:

- 1) $K_{D,1,1,1}^{OS,D}$ – підклас, що містить журнали подій реальних хостів і сервісів;
- 2) $K_{D,1,1,2}^{OS,D}$ – підклас, що містить конфігурації мережних сегментів та вузлів;
- 3) $K_{D,1,1,3}^{OS,D}$ – підклас, що містить історію змін інфраструктури корпоративних мереж;
- 4) $K_{D,1,1,4}^{OS,D}$ – підклас, що містить дані про попередні інциденти та атаки;
- 5) $K_{D,1,1,5}^{OS,D}$ – підклас, що містить шаблони нормальної поведінки користувачів і сервісів.

Такі дані зберігаються як контекстна база знань, що дозволяє адаптувати обманні моделі до еволюції інфраструктури корпоративних мереж та загроз.

До другого класу віднесемо дані, що безпосередньо згенеровані обманними об'єктами в процесі їх експлуатації, та задамо його так:

$$K_{D,1,2}^{OS,D} = \bigcup_{i=1}^{n_{M_{D,1,2}^{OS}}} K_{D,1,2,i}^{OS,D}, \quad (2.24)$$

де $K_{D,1,2,i}^{OS,D}$ – i -тий підклас даних об'єктів обманних систем класу $K_{D,1,2}^{OS,D}$; $i = 1, 2, \dots, n_{M_{D,1,2}^{OS}}$; $n_{M_{D,1,2}^{OS}}$ – кількість підкласів.

Визначимо елементи другого класу $K_{D,1,2}^{OS,D}$ його підкласами $K_{D,1,2,i}^{OS,D}$ так:

- 1) $K_{D,1,2,1}^{OS,D}$ – підклас, що містить дані про події доступу до приманок;
- 2) $K_{D,1,2,2}^{OS,D}$ – підклас, що містить дані про спроби автентифікації з використанням токенів-приманок;

3) $K_{D,1,2,3}^{OS,D}$ – підклас, що містить дані про команди та дії, які виконані зловмисником у обманному середовищі;

4) $K_{D,1,2,4}^{OS,D}$ – підклас, що містить дані про мережні з'єднання та трафік, які спрямовані на обманні об'єкти;

5) $K_{D,1,2,5}^{OS,D}$ – підклас, що містить дані про артефакти експлуатації та постексплуатаційної активності.

Потреба в даних другого підкласу суттєва, оскільки вони відображають зловмисну активність і можуть розглядатися як еталонні сигнали компрометації.

Для зберігання даних обох класів в часово-орієнтованому вигляді, а також з урахуванням того, що такий же підхід будемо застосовувати щодо даних всіх семи рівнів, задамо вектори з координатами так:

$$v_{D,r,i,j}^{OS,D} = \left(v_{D,r,i,j,1}^{OS,D}, v_{D,r,i,j,2}^{OS,D}, \dots, v_{D,r,i,j,n_{v_{D,r,i,j}^{OS,D}}}^{OS,D} \right), \quad (2.25)$$

де $v_{D,r,i,j}^{OS,D}$ - j -тий вектор; $s = 1, 2, \dots, n_{v_{D,r,i,j}^{OS,D}}$; $v_{D,r,i,j,s}^{OS,D}$ - s -та координата вектору $v_{D,r,i,j}^{OS,D}$; $n_{v_{D,r,i,j}^{OS,D}}$ – кількість координат вектору $v_{D,r,i,j}^{OS,D}$; j – змінна, що означає номер вектору $v_{D,r,i,j}^{OS,D}$; i - номер класу даних об'єктів обманних систем; $i = 1, 2$; r – номер рівня засобів і систем протидії КА; $r = 1, 2, \dots, 7$.

Координати вектору $v_{D,1,1,j}^{OS,D}$ визначимо так:

- 1) $v_{D,1,1,j,1}^{OS,D}$ – дані журналів подій реальних хостів і сервісів;
- 2) $v_{D,1,1,j,2}^{OS,D}$ – дані про конфігурації мережних сегментів та вузлів;
- 3) $v_{D,1,1,j,3}^{OS,D}$ – дані про історію змін інфраструктури корпоративних мереж;
- 4) $v_{D,1,1,j,4}^{OS,D}$ – дані про попередні інциденти та атаки;
- 5) $v_{D,1,1,j,5}^{OS,D}$ – дані про шаблони нормальної поведінки користувачів і сервісів.

Координати вектору $v_{D,1,2,j}^{OS,D}$ визначимо так:

- 1) $v_{D,1,2,j,1}^{OS,D}$ – дані про час взаємодії;
- 2) $v_{D,1,2,j,2}^{OS,D}$ – дані про тип обманного об'єкта;

3) $v_{D,1,2,j,3}^{OS,D}$ – дані про імітований ресурс;

4) $v_{D,1,2,j,4}^{OS,D}$ – дані про контекст інфраструктури корпоративних мереж на момент події;

5) $v_{D,1,2,j,5}^{OS,D}$ – дані про зв'язок з іншими подіями безпеки.

Так задані вектори $v_{D,1,2,j}^{OS,D}$ (j – змінна, що означає номер вектору $v_{D,1,i,j}^{OS,D}$) дають змогу формувати окремий часовий пласт подій, який доповнює традиційні журнали безпеки.

У межах першого рівня Event Sources історичні дані реальних систем і дані обманних об'єктів перебувають у циклічному взаємозв'язку, бо історичні дані використовуються для побудови обманних моделей, обманні об'єкти генерують нові дані, згенеровані дані накопичуються в історичному вимірі, на основі згенерованих даних відбувається подальше вдосконалення обманних моделей. Такий підхід забезпечує адаптивність і самонавчання систем протидії КА. На цьому рівні обманні моделі реалізуються у вигляді базових обманних об'єктів, які відтворюють характеристики реальних ресурсів, що сформовані на основі попередніх конфігурацій і журналів подій. Засобами першого рівня на основі обманних технологій є приманки низького, середнього та високого ступеня інтеграції, а також токени фейкових облікових даних (AD, LDAP, cloud), фейкових SSH-ключів, фейкових файлів з «чутливими» назвами, фейкових API-ключів.

Засоби та системи протидії КА на другому рівні забезпечують первинне аналітичне опрацювання подій, що надходять від джерел подій, з метою виявлення ознак атак за сигнатурним, евристичним або аномальним підходом. У поєднанні з обманними технологіями цей рівень переходить від ймовірнісного виявлення до детермінованого підтвердження зловмисної активності. При будь-якому зверненні до приманок виникає інцидент. Системи та засоби другого рівня, включно з обманними технологіями, залежать від історичних даних, бо потребують даних про атаковані раніше порти, скановані сервіси та використовувані протоколи. Використовуючи такі дані приманки налаштовуються з врахуванням, наприклад, актуальних технік,

технологій і процедур з матриці MITRE ATT&CK, а токени розміщуються в популярних для КА місцях корпоративних мереж.

До першого класу $K_{D,2,1}^{OS,D}$ віднесемо накопичені історичні дані систем та засобів виявлення та запобігання КА, які не є результатом взаємодії із обманними об'єктами, але визначають контекст їх побудови. Задамо клас $K_{D,2,1}^{OS,D}$ векторами підкласів, оскільки дані мають різну інформацію і відповідно структурування, так:

$$K_{D,2,1}^{OS,D} = \bigcup_{i=1}^{n_{M_{D,2,1}^{OS}}} K_{D,2,1,i}^{OS,D}, \quad (2.26)$$

де $K_{D,2,1,i}^{OS,D}$ – i -тий підклас даних об'єктів обманних систем класу $K_{D,2,1}^{OS,D}$; $i = 1, 2, \dots, n_{M_{D,2,1}^{OS}}$; $n_{M_{D,2,1}^{OS}}$ – кількість підкласів.

Елементи класу $K_{D,2,1}^{OS,D}$ з підкласів $K_{D,2,1,i}^{OS,D}$ такі:

- 1) $K_{D,2,1,1}^{OS,D}$ – підклас, що містить дані про сигнатури атак;
- 2) $K_{D,2,1,2}^{OS,D}$ – підклас, що містить дані про спрацювання правил;
- 3) $K_{D,2,1,3}^{OS,D}$ – підклас, що містить статистику хибних та істинних спрацювань;
- 4) $K_{D,2,1,4}^{OS,D}$ – підклас, що містить дані про історію мережних аномалій.

До другого класу віднесемо дані, що безпосередньо отримані обманними об'єктами в процесі їх мережної взаємодії, та задамо його так:

$$K_{D,2,2}^{OS,D} = \bigcup_{i=1}^{n_{M_{D,2,2}^{OS}}} K_{D,2,2,i}^{OS,D}, \quad (2.27)$$

де $K_{D,2,2,i}^{OS,D}$ – i -тий підклас даних об'єктів обманних систем класу $K_{D,2,2}^{OS,D}$; $i = 1, 2, \dots, n_{M_{D,2,2}^{OS}}$; $n_{M_{D,2,2}^{OS}}$ – кількість підкласів.

Визначимо елементи другого класу $K_{D,2,2}^{OS,D}$ його підкласами $K_{D,2,2,i}^{OS,D}$ так:

- 1) $K_{D,2,2,1}^{OS,D}$ – підклас, що містить дані про звернення до приманок;
- 2) $K_{D,2,2,2}^{OS,D}$ – підклас, що містить дані про взаємодію з пастками;
- 3) $K_{D,2,2,3}^{OS,D}$ – підклас, що містить дані про спроби експлуатації токенів;
- 4) $K_{D,2,2,4}^{OS,D}$ – підклас, що містить дані про трафік, який був спрямований на неіснуючі або фальшиві ресурси.

Таке роздільне, але узгоджене зберігання дозволяє аналізувати ефективність сигнатур, адаптувати окремі приманки під реальні вектори атак і зменшувати залежність від застарілих правил систем виявлення вторгнень.

Координати вектору $v_{D,2,1,j}^{OS,D}$ для другого рівня визначимо з урахуванням того, що дані систем та засобів другого рівня зберігаються як послідовності мережних подій, так:

- 1) $v_{D,2,1,j,1}^{OS,D}$ – дані про час;
- 2) $v_{D,2,1,j,2}^{OS,D}$ – дані про джерела та призначення;
- 3) $v_{D,2,1,j,3}^{OS,D}$ – дані про типу сигнатури або правила.

Координати вектору $v_{D,2,2,j}^{OS,D}$ для другого рівня визначимо з урахуванням того, що дані про обманні події зберігаються окремо, але з обов'язковим маркуванням, так:

- 1) $v_{D,2,2,j,1}^{OS,D}$ – дані про типу обманного об'єкта;
- 2) $v_{D,2,2,j,2}^{OS,D}$ – дані про рівні взаємодії (низький / середній / високий);
- 3) $v_{D,2,2,j,3}^{OS,D}$ – дані про відповідності конкретному сегменту мережі.

На другому рівні використовувані обманні технології перетворюють історичні мережні дані з допоміжного контексту на підтверджувальний фактор атаки, а збереження обманних даних у часовому вимірі дозволяє постійно коригувати як сигнатурні, так і обманні механізми.

На третьому рівні, тобто рівні кореляції, обманні технології використовують граничні значення кореляції, що дозволяє, спираючись на історичні ланцюги подій, реконструювати повний сценарій атаки та визначати її фазу. На цьому рівні відбувається об'єднання розрізнених подій у логічно пов'язані ланцюги, що відображають розвиток КА в часі. У обманних підходах цей рівень відіграє ключову роль у перетворенні окремих обманних подій на сценарії атак, тому обсяг даних для цього рівня є більшим. На цьому рівні обманні засоби дають високопріоритетний сигнал та використовуються як тригер для кореляції. Наприклад, результат можна отримати після таких переходів: сканування → бічний рух → приманка + пастка → підтверджена атака. Історичні дані для третього етапу є важливими, бо з їх

використанням будується ланцюг подій, обманні дії використовуються для всієї атаки та визначається етап моделі kill chain, на якому був зловмисник.

До першого класу $K_{D,3,1}^{OS,D}$ віднесемо накопичені історичні дані систем та засобів протидії КА, які не є результатом взаємодії із обманними об'єктами. Задамо клас $K_{D,3,1}^{OS,D}$ векторами підкласів, оскільки дані мають різну інформацію і відповідно структурування, так:

$$K_{D,3,1}^{OS,D} = \bigcup_{i=1}^{n_{M_{D,3,1}^{OS}}} K_{D,3,1,i}^{OS,D}, \quad (2.28)$$

де $K_{D,3,1,i}^{OS,D}$ – i -тий підклас даних об'єктів обманних систем класу $K_{D,3,1}^{OS,D}$; $i = 1, 2, \dots, n_{M_{D,3,1}^{OS}}$; $n_{M_{D,3,1}^{OS}}$ – кількість підкласів.

Елементи класу $K_{D,3,1}^{OS,D}$ з підкласів $K_{D,3,1,i}^{OS,D}$ такі:

- 1) $K_{D,3,1,1}^{OS,D}$ – підклас, що містить дані про раніше зафіксовані ланцюги атак;
- 2) $K_{D,3,1,2}^{OS,D}$ – підклас, що містить дані про типові послідовності подій;
- 3) $K_{D,3,1,3}^{OS,D}$ – підклас, що містить дані про часові інтервали між етапами атак.

До другого класу віднесемо дані, що безпосередньо отримані обманними об'єктами в процесі їх мережної взаємодії, та задамо його так:

$$K_{D,3,2}^{OS,D} = \bigcup_{i=1}^{n_{M_{D,3,2}^{OS}}} K_{D,3,2,i}^{OS,D}, \quad (2.29)$$

де $K_{D,3,2,i}^{OS,D}$ – i -тий підклас даних об'єктів обманних систем класу $K_{D,3,2}^{OS,D}$; $i = 1, 2, \dots, n_{M_{D,3,2}^{OS}}$; $n_{M_{D,3,2}^{OS}}$ – кількість підкласів.

Визначимо елементи другого класу $K_{D,3,2}^{OS,D}$ його підкласами $K_{D,3,2,i}^{OS,D}$ так:

- 1) $K_{D,3,2,1}^{OS,D}$ – підклас, що містить дані про перші звернення до приманок;
- 2) $K_{D,3,2,2}^{OS,D}$ – підклас, що містить дані про спроби використання токенів-приманок;
- 3) $K_{D,3,2,3}^{OS,D}$ – підклас, що містить дані про переходи з реального середовища в обманне середовище.

Координати вектору $v_{D,3,1,j}^{OS,D}$ визначимо з урахуванням того, що дані систем та засобів третього рівня зберігаються так:

- 1) $v_{D,3,1,j,1}^{OS,D}$ – дані про повні графи подій;
- 2) $v_{D,3,1,j,2}^{OS,D}$ – дані про зв'язки між реальними та обманними об'єктами;
- 3) $v_{D,3,1,j,3}^{OS,D}$ – дані про розвиток атак у динаміці.

Координати вектору $v_{D,3,2,j}^{OS,D}$ визначимо з урахуванням того, що обманні події виконують роль точок фіксації навколо яких окремо зберігаються дані, які задамо так:

- 1) $v_{D,3,2,j,1}^{OS,D}$ – дані побудованого ретроспективного аналізу;
- 2) $v_{D,3,2,j,2}^{OS,D}$ – дані про результат перевірки гіпотези щодо етапів моделі kill chain;
- 3) $v_{D,3,2,j,3}^{OS,D}$ – дані про підтверджений факт зловмисної активності.

На третьому рівні використовувані обманні технології забезпечують семантичну визначеність історичних даних, дозволяючи відокремити випадкові аномалії від цілеспрямованих атак і формувати довготривалі сценарії загроз.

Засоби та системи четвертого рівня є центральним елементом накопичення, зберігання та аналізу історичних даних безпеки. У контексті застосування додатково обманних технологій вони забезпечують довготривалу пам'ять для засобів і систем протидії КА. Це дає змогу виявляти повторювані КА, ідентифікувати тривалі приховані КА та підвищувати достовірність прийняття рішень.

До першого класу $K_{D,4,1}^{OS,D}$ віднесемо накопичені історичні дані систем та засобів протидії КА, які не є результатом взаємодії із обманними об'єктами. Задамо клас $K_{D,4,1}^{OS,D}$ векторами підкласів, оскільки дані мають різну інформацію і відповідно структурування, так:

$$K_{D,4,1}^{OS,D} = \bigcup_{i=1}^{n_{M_{D,4,1}^{OS}}} K_{D,4,1,i}^{OS,D}, \quad (2.30)$$

де $K_{D,4,1,i}^{OS,D}$ – i -тий підклас даних об'єктів обманних систем класу $K_{D,4,1}^{OS,D}$; $i = 1, 2, \dots, n_{M_{D,4,1}^{OS}}$; $n_{M_{D,4,1}^{OS}}$ – кількість підкласів.

Елементи класу $K_{D,4,1}^{OS,D}$ з підкласів $K_{D,4,1,i}^{OS,D}$ такі:

- 1) $K_{D,4,1,1}^{OS,D}$ – підклас, що містить дані про аналіз еволюції атак;
- 2) $K_{D,4,1,2}^{OS,D}$ – підклас, що містить дані про повторювані КА;
- 3) $K_{D,4,1,3}^{OS,D}$ – підклас, що містить дані про результати ретроспективного пошуку;
- 3) $K_{D,4,1,4}^{OS,D}$ – підклас, що містить дані про оцінювання доцільності і

результативності обманних механізмів.

До другого класу віднесемо дані, що безпосередньо отримані обманними об'єктами в процесі їх мережної взаємодії, та задамо його так:

$$K_{D,4,2}^{OS,D} = \bigcup_{i=1}^{n_{M_{D,4,2}}^{OS}} K_{D,4,2,i}^{OS,D}, \quad (2.31)$$

де $K_{D,4,2,i}^{OS,D}$ – i -тий підклас даних об'єктів обманних систем класу $K_{D,4,2}^{OS,D}$; $i = 1, 2, \dots, n_{M_{D,4,2}}^{OS}$; $n_{M_{D,4,2}}^{OS}$ – кількість підкласів.

Визначимо елементи другого класу $K_{D,4,2}^{OS,D}$ його підкласами $K_{D,4,2,i}^{OS,D}$ так:

- 1) $K_{D,4,2,1}^{OS,D}$ – підклас, що містить дані про тип пастки;
- 2) $K_{D,4,2,2}^{OS,D}$ – підклас, що містить дані про імітований актив;
- 3) $K_{D,4,2,3}^{OS,D}$ – підклас, що містить дані про контекст (користувач, хост, сегмент).

Координати вектору $v_{D,4,1,j}^{OS,D}$ визначимо з урахуванням того, що дані систем та засобів третього рівня зберігаються так:

- 1) $v_{D,4,1,j,1}^{OS,D}$ – дані про журнали подій;
- 2) $v_{D,4,1,j,2}^{OS,D}$ – дані про інциденти;
- 3) $v_{D,4,1,j,3}^{OS,D}$ – дані про результати реагування.

Координати вектору $v_{D,4,2,j}^{OS,D}$ визначимо з урахуванням того, що обманні події виконують роль точок фіксації навколо яких окремо зберігаються дані, які задамо так:

- 1) $v_{D,4,2,j,1}^{OS,D}$ – дані повних журналів взаємодії з обманними об'єктами;
- 2) $v_{D,4,2,j,2}^{OS,D}$ – дані про часові ряди інцидентів, які були пов'язані з обманними об'єктами;
- 3) $v_{D,4,2,j,3}^{OS,D}$ – дані статистики ефективності пасток.

На четвертому рівні використовувані обманні технології формують високодостовірний історичний шар подій, який є основою для аналітики, автоматизації реагування та стратегічного управління кіберзахистом.

Засоби та системи п'ятого рівня, тобто рівня UEBA / NDR, забезпечують поведінковий аналіз користувачів, сутностей і мережних потоків з метою виявлення відхилень від нормального функціонування. На цьому рівні історичні дані є обов'язковою передумовою, оскільки без них неможливе формування базових моделей поведінки. У поєднанні з обманними технологіями вони переходять від ймовірнісних аномалій до підтверджених поведінкових індикаторів атаки.

До першого класу $K_{D,5,1}^{OS,D}$ віднесемо накопичені історичні дані систем та засобів протидії КА, які не є результатом взаємодії із обманними об'єктами. Задамо клас $K_{D,5,1}^{OS,D}$ векторами підкласів, оскільки дані мають різну інформацію і відповідно структурування, так:

$$K_{D,5,1}^{OS,D} = \bigcup_{i=1}^{n_{M_{D,5,1}^{OS}}} K_{D,5,1,i}^{OS,D}, \quad (2.32)$$

де $K_{D,5,1,i}^{OS,D}$ – i -тий підклас даних об'єктів обманних систем класу $K_{D,5,1}^{OS,D}$; $i = 1, 2, \dots, n_{M_{D,5,1}^{OS}}$; $n_{M_{D,5,1}^{OS}}$ – кількість підкласів.

Елементи класу $K_{D,5,1}^{OS,D}$ з підкласів $K_{D,5,1,i}^{OS,D}$ такі:

- 1) $K_{D,5,1,1}^{OS,D}$ – підклас, що містить дані про часові ряди поведінки;
- 2) $K_{D,5,1,2}^{OS,D}$ – підклас, що містить дані про еволюцію профілів користувачів;
- 3) $K_{D,5,1,3}^{OS,D}$ – підклас, що містить дані про зміну ролей і контекстів.

До другого класу віднесемо дані, що безпосередньо отримані обманними об'єктами в процесі їх мережної взаємодії і стосуються еталонних маркерів компрометації, та задамо його так:

$$K_{D,5,2}^{OS,D} = \bigcup_{i=1}^{n_{M_{D,5,2}^{OS}}} K_{D,5,2,i}^{OS,D}, \quad (2.33)$$

де $K_{D,5,2,i}^{OS,D}$ – i -тий підклас даних об'єктів обманних систем класу $K_{D,5,2}^{OS,D}$; $i = 1, 2, \dots, n_{M_{D,5,2}^{OS}}$; $n_{M_{D,5,2}^{OS}}$ – кількість підкласів.

Визначимо елементи другого класу $K_{D,5,2}^{OS,D}$ його підкласами $K_{D,5,2,i}^{OS,D}$ так:

- 1) $K_{D,5,2,1}^{OS,D}$ – підклас, що містить дані, які підтверджують, що аномалія є атакою;
- 2) $K_{D,5,2,2}^{OS,D}$ – підклас, що містить дані для корекції поведінкових моделей;
- 3) $K_{D,5,2,3}^{OS,D}$ – підклас, що містить дані про еталонні маркери, які знижують рівень невизначеності під час аналізу.

Координати вектору $v_{D,5,1,j}^{OS,D}$ визначимо з урахуванням того, що дані систем та засобів третього рівня зберігаються так:

- 1) $v_{D,5,1,j,1}^{OS,D}$ – дані про профілі нормальної активності користувачів;
- 2) $v_{D,5,1,j,2}^{OS,D}$ – дані про типові шаблони мережного трафіку;
- 3) $v_{D,5,1,j,3}^{OS,D}$ – дані про часові та рольові характеристики доступу;
- 4) $v_{D,5,1,j,4}^{OS,D}$ – дані про відхилення від базових моделей;
- 5) $v_{D,5,1,j,5}^{OS,D}$ – дані про нетипові послідовності дій;
- 6) $v_{D,5,1,j,6}^{OS,D}$ – дані про аномальні мережні маршрути.

Координати вектору $v_{D,5,2,j}^{OS,D}$ визначимо з урахуванням того, що обманні події виконують роль точок фіксації навколо яких окремо зберігаються дані, які задамо так:

- 1) $v_{D,5,2,j,1}^{OS,D}$ – дані про доступ користувача або хоста до обманних об'єктів;
- 2) $v_{D,5,2,j,2}^{OS,D}$ – дані про використання обманних токенів;
- 3) $v_{D,5,2,j,3}^{OS,D}$ – дані про мережні взаємодії з приманками і пастками.

На п'ятому рівні використовувані обманні технології надають історичним поведінковим моделям об'єктивну базу, перетворюючи абстрактні аномалії на доказову базу зловмисної активності.

Засоби та системи шостого рівня інтегрують дані з кінцевих точок, мережі, результатів ідентифікації, хмари та обманних систем, забезпечуючи єдиний простір виявлення та реагування. На цьому рівні обманні технології використовуються як індикатори для автоматизованих дій реагування. Засоби та системи цього рівня використовують історію атак, ланцюги подій та реакції, які вже спрацювали

ефективно. А обманні системи забезпечують зменшення часу реагування та дозволяють реагувати агресивніше, бо впевненість в діях максимальна.

До першого класу $K_{D,6,1}^{OS,D}$ віднесемо накопичені історичні дані систем та засобів протидії КА, які не є результатом взаємодії із обманними об'єктами. Задамо клас $K_{D,6,1}^{OS,D}$ векторами підкласів, оскільки дані мають різну інформацію і відповідно структурування, так:

$$K_{D,6,1}^{OS,D} = \bigcup_{i=1}^{n_{M_{D,6,1}^{OS}}} K_{D,6,1,i}^{OS,D}, \quad (2.34)$$

де $K_{D,6,1,i}^{OS,D}$ – i -тий підклас даних об'єктів обманних систем класу $K_{D,6,1}^{OS,D}$; $i = 1, 2, \dots, n_{M_{D,6,1}^{OS}}$; $n_{M_{D,6,1}^{OS}}$ – кількість підкласів.

Елементи класу $K_{D,6,1}^{OS,D}$ з підкласів $K_{D,6,1,i}^{OS,D}$ такі:

- 1) $K_{D,6,1,1}^{OS,D}$ – підклас, що містить дані про сценарії атак;
- 2) $K_{D,6,1,2}^{OS,D}$ – підклас, що містить дані про ефективності реагування;
- 3) $K_{D,6,1,3}^{OS,D}$ – підклас, що містить дані про час до виявлення і реагування.

До другого класу віднесемо дані, що безпосередньо отримані обманними об'єктами в процесі їх мережної взаємодії, та задамо його так:

$$K_{D,6,2}^{OS,D} = \bigcup_{i=1}^{n_{M_{D,6,2}^{OS}}} K_{D,6,2,i}^{OS,D}, \quad (2.35)$$

де $K_{D,6,2,i}^{OS,D}$ – i -тий підклас даних об'єктів обманних систем класу $K_{D,6,2}^{OS,D}$; $i = 1, 2, \dots, n_{M_{D,6,2}^{OS}}$; $n_{M_{D,6,2}^{OS}}$ – кількість підкласів.

Визначимо елементи другого класу $K_{D,6,2}^{OS,D}$ його підкласами $K_{D,6,2,i}^{OS,D}$ так:

- 1) $K_{D,6,2,1}^{OS,D}$ – підклас, що містить дані про найбільш ефективні сценарії реагування;
- 2) $K_{D,6,2,2}^{OS,D}$ – підклас, що містить дані про жорсткіші автоматизовані дії;
- 3) $K_{D,6,2,3}^{OS,D}$ – підклас, що містить дані про легітимні об'єкти для зменшення ризику їх блокування.

Координати вектору $v_{D,6,1,j}^{OS,D}$ визначимо з урахуванням того, що дані систем та засобів третього рівня зберігаються так:

- 1) $v_{D,6,1,j,1}^{OS,D}$ – дані про корельовані інциденти;
- 2) $v_{D,6,1,j,2}^{OS,D}$ – дані про ланцюги атак;
- 3) $v_{D,6,1,j,3}^{OS,D}$ – дані про результати попередніх реакцій;
- 4) $v_{D,6,1,j,4}^{OS,D}$ – дані про підтверджені інциденти;
- 5) $v_{D,6,1,j,5}^{OS,D}$ – дані про контекстні сигнали з різних доменів.

Координати вектору $v_{D,6,2,j}^{OS,D}$ визначимо з урахуванням того, що обманні події виконують роль точок фіксації навколо яких окремо зберігаються дані, які задамо так:

- 1) $v_{D,6,2,j,1}^{OS,D}$ – дані про події взаємодії з обманними об'єктами;
- 2) $v_{D,6,2,j,2}^{OS,D}$ – дані про переходи зловмисника між доменами;
- 3) $v_{D,6,2,j,3}^{OS,D}$ – дані про повторні спроби експлуатації.

На шостому рівні використовувані обманні технології виконують роль каталізатора автоматизованого реагування, а їх історичні дані забезпечують адаптацію та оптимізацію рішень у часі.

Засоби та системи сьомого рівня є верхнім рівнем ієрархії та орієнтовані на аналітичну, дослідницьку й стратегічну діяльність. На цьому рівні історичні дані розглядаються не лише як джерело реагування, а як знання про еволюцію загроз. Засоби і системи використовують обманні технології для виявлення прихованих АРТ, перевірки гіпотез та тестування ефективності контролів.

До першого класу $K_{D,7,1}^{OS,D}$ віднесемо накопичені історичні дані систем та засобів протидії КА, які не є результатом взаємодії із обманними об'єктами. Задамо клас $K_{D,7,1}^{OS,D}$ векторами підкласів, оскільки дані мають різну інформацію і відповідно структурування, так:

$$K_{D,7,1}^{OS,D} = \bigcup_{i=1}^{n_{M_{D,7,1}^{OS,D}}} K_{D,7,1,i}^{OS,D}, \quad (2.36)$$

де $K_{D,7,1,i}^{OS,D}$ – i -тий підклас даних об'єктів обманних систем класу $K_{D,7,1}^{OS,D}$; $i = 1, 2, \dots, n_{M_{D,7,1}^{OS}}$; $n_{M_{D,7,1}^{OS}}$ – кількість підкласів.

Елементи класу $K_{D,7,1}^{OS,D}$ з підкласів $K_{D,7,1,i}^{OS,D}$ такі:

- 1) $K_{D,7,1,1}^{OS,D}$ – підклас, що містить дані про місяці або роки активності;
- 2) $K_{D,7,1,2}^{OS,D}$ – підклас, що містить дані про порівняння подій та інцидентів;
- 3) $K_{D,7,1,3}^{OS,D}$ – підклас, що містить дані про атрибути та базис для перевірки гіпотез;
- 4) $K_{D,7,1,4}^{OS,D}$ – підклас, що містить дані про засіб виявлення раніше непомічених атак.

До другого класу віднесемо дані, що безпосередньо отримані обманними об'єктами в процесі їх мережної взаємодії і стосуються еталонних маркерів компрометації, та задамо його так:

$$K_{D,7,2}^{OS,D} = \bigcup_{i=1}^{n_{M_{D,7,2}^{OS}}} K_{D,7,2,i}^{OS,D}, \quad (2.37)$$

де $K_{D,7,2,i}^{OS,D}$ – i -тий підклас даних об'єктів обманних систем класу $K_{D,7,2}^{OS,D}$; $i = 1, 2, \dots, n_{M_{D,7,2}^{OS}}$; $n_{M_{D,7,2}^{OS}}$ – кількість підкласів.

Визначимо елементи другого класу $K_{D,7,2}^{OS,D}$ його підкласами $K_{D,7,2,i}^{OS,D}$ так:

- 1) $K_{D,7,2,1}^{OS,D}$ – підклас, що містить дані про точки входу для threat hunting;
- 2) $K_{D,7,2,2}^{OS,D}$ – підклас, що містить дані про зміну тактик зловмисників;
- 3) $K_{D,7,2,3}^{OS,D}$ – підклас, що містить дані про повільні або «тихі» атаки.

Координати вектору $v_{D,7,1,j}^{OS,D}$ визначимо з урахуванням того, що дані систем та засобів третього рівня зберігаються так:

- 1) $v_{D,7,1,j,1}^{OS,D}$ – дані про інциденти за місяці й роки;
- 2) $v_{D,7,1,j,2}^{OS,D}$ – дані про кампанії атак;
- 3) $v_{D,7,1,j,3}^{OS,D}$ – дані про результати розслідувань.

Координати вектору $v_{D,7,2,j}^{OS,D}$ визначимо з урахуванням того, що обманні події виконують роль точок фіксації навколо яких окремо зберігаються дані, які задамо так:

- 1) $v_{D,7,2,j,1}^{OS,D}$ – дані про повні журнали взаємодії з приманками і пастками;
- 2) $v_{D,7,2,j,2}^{OS,D}$ – дані про повторювані шаблони поведінки зловмисників;
- 3) $v_{D,7,2,j,3}^{OS,D}$ – дані про сліди тривалої присутності (APT).

На цьому рівні використовувані обманні технології формують високодостовірний історичний пласт знань, який підтримує стратегічне управління кібербезпекою та розвиток систем протидії КА.

Таким чином, на всіх семи рівнях ієрархії обманні технології формують окремий, високонадійний історичний контур даних, який доповнює класичні журнали безпеки, підвищує точність аналітичних моделей та забезпечує адаптивність системи захисту в часі. Моделі обманних технологій є адаптивними та ієрархічно інтегрованими, а їх ефективність безпосередньо залежить від обсягу й якості історичних даних, які накопичені засобами протидії КА.

Перетини між сімома рівнями ієрархії у контексті протидії КА формують єдиний безперервний обманний контур даних, у якому обманні технології забезпечують достовірність, історичні дані – спадкоємність, аналітичні рівні – трансформацію даних у знання. Саме ці перетини роблять обманні технології інтеграційним елементом усієї ієрархії, а не ізольованим засобом захисту. Агрегація перетинів рівнів ієрархії, обманних технологій та історичних даних подана в табл. 2.1.

У межах семи рівнів одна обманна подія одночасно належить кільком класам даних, змінюючи свою семантичну роль так: подія → індикатор → якір кореляції → інцидент → поведінковий маркер → тригер реагування → аналітичне знання. Таким чином, перетини рівнів є точками трансформації даних, обманні дані забезпечують достовірність на кожному перетині, історичний вимір гарантує спадкоємність і адаптивність системи. Перетини між рівнями ієрархії формують єдиний обманний контур даних, у якому обманні технології виступають не окремим засобом захисту, а інтеграційним механізмом, що поєднує події, інциденти, поведінкові моделі та аналітичні знання в часовому вимірі.

Таблиця 2.1 - Перетин рівнів ієрархії, обманних технологій та історичних даних

Перетин рівнів	Спільні (перехресні) класи даних	Роль обманних даних у перетині	Результат перетину
1 ↔ 2	Первинні журнали подій; мережні події доступу; обманні події	Перетворення сирогої події на підтверджений індикатор атаки	Перехід від факту взаємодії до факту вторгнення
2 ↔ 3	Підтверджені спрацювання; часові послідовності подій; обманні події як «якорі»	Фіксація ключових точок сценарію атаки	Перехід від окремих індикаторів до сценаріїв атак
3 ↔ 4	Корельовані ланцюги атак; збагачені інциденти; історичні обманні інциденти	Закріплення сценаріїв у довготривалій пам'яті	Перехід від оперативного аналізу до історичного знання
4 ↔ 5	Історичні журнали; поведінкові профілі; обманні маркери компрометації	Валідація та навчання поведінкових моделей	Перехід від історії подій до поведінкової інтерпретації
5 ↔ 6	Підтверджені поведінкові інциденти; мультидоменні контексти; обманні події високої довіри	Тригер для автоматизованого реагування	Перехід від аналітики до дії
6 ↔ 7	Історія інцидентів і реакцій; сценарії атак; аналітичні обманні артефакти	Точки входу для розслідувань і гіпотез	Перехід від тактичного реагування до стратегічного знання

Розроблені подання даних, які зберігаються та використовуються на різних рівнях ієрархії, векторами є основою для розроблення підсистеми пам'яті систем та засобів протидії КА з орієнтацією на вибір глибини ретроспективи використовуваних даних для здійснення протидії КА в корпоративних мережах, накопичення даних згідно розроблених класів різних рівнів, оптимізації застарілих даних та їх модифікації для синтезу нових даних.

2.4 Архітектура засобів підсистеми пам'яті систем та засобів протидії КА в корпоративних мережах на основі використання історичних даних для прийняття рішень

Забезпечення балансу в тривалості зберігання і використання даних систем та засобів протидії КА всіх семи рівнів, а також проблеми щодо методологічних, технічних та семантичних аспектів, для вирішення протиріччя щодо необхідності поєднання вимог до повноти, тривалості та актуальності збережених даних з обмеженнями обчислювальних ресурсів, їх розподілення та адаптивністю систем аналізу розробимо архітектуру підсистеми пам'яті систем та засобів протидії КА в корпоративних мережах. Вона повинна бути основою для створення засобів всіх семи рівнів або меншої кількості з них в залежності від наявних систем та засобів протидії КА в корпоративних мережах. Такі засоби будуть підсистемами пам'яті систем та засобів протидії КА з орієнтацією на вибір глибини ретроспективи використовуваних даних для здійснення протидії КА в корпоративних мережах, накопичення даних згідно розроблених класів різних рівнів, оптимізації застарілих даних та їх модифікації для синтезу нових даних. Ці засоби будуть виконувати завдання з опрацювання даних та вибір з них таких, які б найбільш точніше забезпечували функціонування систем та засобів протидії КА всіх семи рівнів в корпоративних мережах.

Місце засобів підсистеми пам'яті систем та засобів протидії КА в корпоративних мережах в загальній архітектурі разом із системами і засобами протидії КА та сховищами зберігання даних, включно і тими, які містяться безпосередньо в самих системах і засобах зображено на рис. 2.1. До таких систем та засобів протидії КА належать також обманні системи.

Задамо множиною M_Z^D перелік завдань засобів зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам корпоративних мереж так:

$$M_Z^D = \{m_{z,1}^D, m_{z,2}^D, \dots, m_{z,n_{M_Z^D}}^D\}, \quad (2.38)$$

де $m_{z,q}^D$ – елемент множини M_z^D , який відображає одне із завдань засобів підсистеми пам'яті; $q = 1, 2, \dots, n_{M_z^D}$; $n_{M_z^D}$ – кількість завдань засобів підсистеми пам'яті.

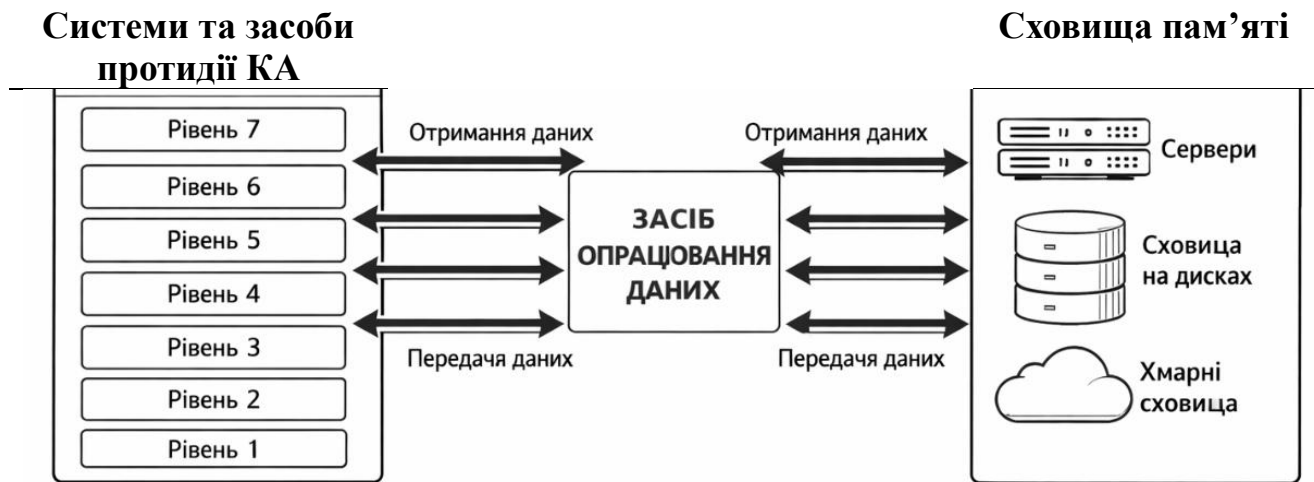


Рисунок 2.1 – Місце засобів підсистеми пам'яті в загальній концепції засобів та систем протидії КА і сховищах даних

Засоби підсистеми пам'яті можуть бути окремими незалежними об'єктами в контексті їх подання в корпоративних мережах, а можуть бути частиною систем протидії КА. Крім того, частина даних буде зберігатись не тільки окремо, але й безпосередньо в самих системах і засобах протидії КА, тому засоби підсистеми пам'яті можуть бути частково або повністю їх частинами.

Основні завдання засобів підсистеми пам'яті визначимо так:

- 1) $m_{z,1}^D$ – класифікація отриманих даних з урахуванням даних обманних систем;
- 2) $m_{z,2}^D$ – зберігання отриманих даних в розподіленому середовищі з урахуванням їх класифікації;
- 3) $m_{z,3}^D$ – оптимізація історичних даних в розподіленому середовищі;
- 4) $m_{z,4}^D$ – модифікація даних для забезпечення адаптивного функціонування систем та засобів протидії КА;
- 5) $m_{z,5}^D$ – оцінювання даних за різними критеріями;
- 6) $m_{z,6}^D$ – оновлення даних;
- 7) $m_{z,7}^D$ – вибір даних для систем та засобів протидії КА.

Типи даних, які опрацьовує засіб підсистеми пам'яті, задамо множиною M_d^D так:

$$M_d^D = \{m_{d,1}^D, m_{d,2}^D, \dots, m_{d,n_{M_d^D}}^D\}, \quad (2.39)$$

де $m_{d,q}^D$ – елемент множини M_d , який відображає один тип даних; $q = 1, 2, \dots, n_{M_d^D}$; $n_{M_d^D}$ – кількість типів даних засобів підсистеми пам'яті.

Типи даних будуть відповідати класам даних. Множину M_d^D типів даних поділимо на класи даних з урахуванням їх призначення так:

- 1) $m_{d,1}^D$ – елемент, що відображає дані систем та засобів протидії КА, які наявні в них в момент встановлення;
- 2) $m_{d,2}^D$ – елемент, що відображає накопичені дані систем та засобів протидії КА протягом часу їх функціонування від моменту початку встановлення;
- 3) $m_{d,3}^D$ – елемент, що відображає модифіковані дані систем та засобів протидії КА, які були змінені для забезпечення властивості адаптивності при прийнятті рішень, тобто вони є штучними;
- 4) $m_{d,4}^D$ – елемент, що відображає оновлені дані систем та засобів протидії КА, які були отримані ззовні корпоративних мереж;
- 5) $m_{d,5}^D$ – елемент, що відображає неопрацьовані дані систем та засобів протидії КА та не віднесені до жодного класу для зберігання, але вони були отримані в корпоративних мережах;
- 6) $m_{d,6}^D$ – елемент, що відображає фальшиві дані систем та засобів протидії КА на основі обманних технологій, які були наявні в них з початку їх встановлення та функціонування;
- 7) $m_{d,7}^D$ – елемент, що відображає модифіковані дані систем та засобів протидії КА на основі обманних технологій, які були сформовані в них під час їх функціонування;
- 8) $m_{d,8}^D$ – елемент, що відображає дані систем та засобів протидії КА на основі обманних технологій, які були отримані в них під час їх функціонування.

Згідно такого поділу (формула (2.39)) можна виділити вісім класів даних $K_{d,q}^D$ ($q = 1, 2, \dots, n_{M_d}^D$; $n_{M_d}^D$ – кількість завдань засобів підсистеми пам'яті), які повинен опрацьовувати засіб підсистеми пам'яті для їх зберігання та оптимізації і вибору з них тих, що необхідні для подальшого функціонування систем та засобів протидії КА.

Основні елементи засобів підсистеми пам'яті задамо множиною M_{ez}^D так:

$$M_{ez}^D = \{m_{ez,1}^D, m_{ez,2}^D, \dots, m_{ez,n_{M_{ez}}^D}^D\}, \quad (2.41)$$

де $m_{ez,q}^D$ – елемент множини M_{ez}^D ; $q = 1, 2, \dots, n_{M_{ez}}^D$; $n_{M_{ez}}^D$ – кількість елементів засобів підсистеми пам'яті.

Задамо елементи множини M_{ez}^D так:

- 1) $m_{ez,1}^D$ – керування даними щодо їх розміщення в корпоративних мережах;
- 2) $m_{ez,2}^D$ – пошук даних за різними критеріями;
- 3) $m_{ez,3}^D$ – вибір даних для систем та засобів протидії КА;
- 4) $m_{ez,4}^D$ – модифікація даних для адаптивних систем та засобів протидії КА;
- 5) $m_{ez,5}^D$ – зберігання даних за класами в розподіленому середовищі корпоративних мереж;
- 6) $m_{ez,6}^D$ – оптимізація даних, які зберігаються, з урахуванням їх актуальності.

Керування даними щодо їх розміщення в корпоративних мережах виділено окремим елементом, оскільки потрібно координувати та здійснювати загальний менеджмент ресурсів пам'яті, вільного та заповненого місць в сховищах пам'яті, переміщення окремих класів даних в інші місця в залежності від поточного стану пам'яті і її обсягів, облік розміщення множини та класів даних тощо. Ці процеси повинні підтримуватись в засобах підсистеми пам'яті автоматично, оскільки від них буде залежати ефективність функціонування засобів і систем протидії КА усіх семи рівнів.

Оскільки тривалість КА є різною, то засоби і системи протидії їм можуть реагувати за часом в залежності від рівнів, до яких вони відносяться. Але враховуючи специфіку самих засобів і систем та їх завдання, знаходження необхідних даних повинно випереджати дії зловмисників під час проведення ними КА, щоб не

втратилась актуальність даних і відповідно реакцій. Тому, для забезпечення пошуку необхідних даних потрібно реалізовувати різні алгоритми та методи, щоб задовольняти вимогам критеріїв до повноти, актуальності та релевантності даних.

Для засобів та систем протидії КА вибір даних для прийняття ними рішень є складним завданням, оскільки актуальність даних в контексті глибини ретроспективи їх використання залишається завданням, яке потребує забезпечення необхідними знаннями засобів підсистеми пам'яті, тобто такі засоби повинні мати можливості до самонавчання. Оскільки такий вибір даних за попередні часові періоди функціонування систем та засобів протидії КА суттєво впливатиме на результат, то спроможність до самонавчання засобів підсистеми пам'яті на основі вибору даних з певним рівнем глибини ретроспективи є необхідним.

Модифікація даних необхідна для реалізації властивості адаптивності в системах та засобах протидії КА. Такі засоби і системи повинні адаптивно реагувати на атаки, інакше зловмисники зрозуміють їх поведінку і зможуть їх обходити. Для забезпечення властивості адаптивності потрібно формувати штучні дані і використовувати їх разом з реальними даними. Але формування нових штучних даних вимагає розроблення елементу засобів підсистеми пам'яті з урахуванням вимог до протидії КА щодо даних.

Зберігання даних за класами в розподіленому середовищі корпоративних мереж передбачає забезпечення їх структурування, типізацію та форматування для досягнення типовості. Не менш важливим аспектом в цьому контексті є засоби, які використовуються для такого зберігання, і від їх організації також залежить ефективність при забезпеченні оперативного доступу до даних.

У корпоративних мережах, які можуть містити сім рівнів засобів протидії комп'ютерним атакам або менше, підсистема пам'яті повинна розглядатися не просто як сховище логів, а як інтегрована багаторівнева інформаційно-аналітична структура, яка забезпечує накопичення, диференційоване використання та інтелектуальну обробку історичних даних різної глибини. Формування такої підсистеми передбачає аналіз декількох моделей апаратно-програмної реалізації, їх порівняння за критеріями продуктивності, масштабованості, стійкості та безпекових характеристик, а також

подальший синтез оптимальної архітектури для умов корпоративного середовища. Для забезпечення зберігання даних можуть бути використані централізована модель, розподілена модель та комбінована модель. Ці моделі поділяються на такі, що підтримують ієрархічну організацію пам'яті або без поділу на рівні ієрархії. Крім того, може бути використано при організації пам'яті принцип адаптивної глибини історії.

Централізована модель забезпечує максимальну повноту кореляції подій та глибину ретроспективного аналізу. Вона передбачає використання ієрархічного зберігання і підходить для сценаріїв, де вирішальне значення має комплексний аналіз багаторічної історії інцидентів. Однак концентрація потоків даних створює потенційні вузькі місця, а також підвищує вимоги до резервування та відмовостійкості. Така модель є оптимальною для середовищ із потужним центром обробки даних, але вона менш ефективна для географічно розподілених структур.

Розподілена модель має іншу логіку організації пам'яті порівняно з централізованою. Короткострокова історія зберігається безпосередньо на рівнях протидії (периметр, мережні сенсори, хости), тоді як агреговані або аналітично збагачені дані передаються до центрального ядра. Такий підхід зменшує затримки реагування та підвищує стійкість системи до перевантажень або ізоляції сегментів мережі. Така модель відповідає вимогам реального часу та масштабованості, але ускладнює забезпечення узгодженості історичних даних і потребує складніших механізмів синхронізації, реплікації та контролю цілісності.

Комбінована архітектура поєднує локальну інфраструктуру з хмарними сервісами безпеки, розширює дослідницьке поле за рахунок практично необмеженої масштабованості архівного зберігання та можливостей хмарної аналітики. Оперативна пам'ять та короткострокова історія залишаються в межах корпоративного периметра, тоді як довгострокові масиви даних, що використовуються для машинного навчання та поведінкової аналітики, переміщуються до хмари. Вона забезпечує ефективність при великих обсягах історії, однак вимагає ретельного врахування регуляторних обмежень, політик суверенності даних і криптографічного захисту каналів.

Суть концепції ієрархічної організації пам'яті полягає у встановленні різного “горизонту пам'яті” для кожного з семи рівнів захисту. Нижчі рівні (периметр, IDS/IPS) працюють з короткою, високочастотною історією, тоді як вищі рівні (аналітика, формування ознак атак, стратегічне управління) оперують довготривалими масивами. Така диференціація дозволяє оптимізувати співвідношення “обсяг – швидкість – вартість”, зменшуючи навантаження на швидкі носії та водночас зберігаючи цінні ретроспективні дані в компресованому або агрегованому вигляді. Синтез ієрархічної організації пам'яті з попередніми трьома моделями дає змогу також реалізувати адаптивне переміщення даних між рівнями зберігання залежно від їх актуальності та частоти використання.

Для обробки потоків подій і формування ознак атак у режимі близькому до реального часу можуть бути використані апаратні прискорювачі. Таке доповнення моделі є доцільним при великих навантаженнях і значних обсягах телеметрії. Також пам'ять може бути організована окремим захищеним доменом з реалізацією таких процедур: мікросегментація; шифрування на рівні блоків; контроль доступу; повна трасованість доступу; інтеграція з системами.

Організація безпосередньо пам'яті може бути здійснена з використанням інтелектуальних методів для формування ознак, зберігання поведінкових профілів, адаптивного скорочення історичних даних, автооновлення сигнатур тощо.

Жодна з моделей не забезпечує оптимального балансу. Тому, необхідно синтезувати в моделі засобів зберігання даних для систем протидії КА підмодель, яка буде узгоджуватись з семирівневим поділом, поєднувати розподілене короткострокове зберігання на периферії, централізоване або кластеризоване зберігання даних для глибокої аналітики, ієрархічну організацію носіїв із автоматичним керуванням життєвим циклом даних, а також можливість інтеграції з хмарними ресурсами для довгострокових архівів та ML-аналізу. Такий синтез дозволить забезпечити адаптивність пам'яті до різних часових масштабів атак від миттєвих інцидентів до повільних, багаторічних атак. Задамо таку концептуальну підмодель організації зберігання даних для систем семи рівнів так:

$$M_P^P = \langle R_{P,1}^P, R_{P,2}^P, M_{P,3}^P, I_{P,4}^P, A_{P,6}^P, A_{P,7}^P, R_{P,8}^P, S_{P,9}^P, V_{P,10}^P, A_{P,11}^P, D_{P,12}^P \rangle, \quad (2.42)$$

де $R_{P,1}^P$ – номер рівня систем та засобів протидії КА, для яких організація зберігання даних подана підмоделлю M_P^P ; $R_{P,2}^P$ – варіанти розміщення засобів збереження даних в корпоративних мережах з урахування наявності або відсутності розподілення; $M_{P,3}^P$ – наявність або відсутність міграції даних або їх частини в корпоративних мережах за потреби, наприклад під час відбиття КА; $I_{P,4}^P$ – наявність або відсутність ієрархії при збереженні даних або їх частини; $A_{P,6}^P$ – наявність або відсутність автоматичного керування життєвим циклом даних всіх даних або їх частини; $A_{P,7}^P$ – наявність або відсутність функції архівування даних, які не актуальні в поточний час функціонування систем протидії КА, для їх тривалого зберігання, наприклад, напротязі декількох років; $R_{P,8}^P$ – наявність або відсутність резервування всіх даних або їх частини; $S_{P,9}^P$ – наявність або відсутність резервування всіх даних або їх частини; $V_{P,10}^P$ – вибір даних з урахуванням глибини горизонту на основі інтелектуалізації; $A_{P,11}^P$ – апаратно-програмні засоби в корпоративних мережах, які можуть бути використані для зберігання даних та засобів їх опрацювання; $D_{P,12}^P$ – дані.

Таким чином, особливістю в організації зберігання даних для систем семи рівнів є врахування засобів та систем протидії КА семи рівнів, розподілення сховищ з даними в корпоративних мережах, можливість міграції під час КА, наявність ієрархії даних, автоматичне керування життєвим циклом даних всіх даних або їх частини, наявність архівування та резервування даних, вибір даних з урахуванням глибини горизонту на основі інтелектуалізації та особливості апаратно-програмних засобів для зберігання даних.

Кожен рівень засобів та систем протидії КА має свою глибину горизонту пам'яті, тобто обсяг історії в контексті актуальності даних для використання. Наприклад, для засобів та систем семи рівнів може бути така тривалість, яку подано в табл. 2.2.

Крім врахування історії в контексті актуальності даних потрібно також здійснювати оптимізацію даних та модифікацію для ефективної протидії КА, які мають адаптивну основу.

Таблиця 2.2 – Приклад обсягу та актуальності даних для систем та засобів протидії КА семи рівнів

Рівень	Обсяг історії	Тип носія
1	24–72 год	NVMe
2	7–14 днів	SSD
3	30–90 днів	SSD + NAS
4	6–12 міс	NAS
5	1–3 роки	Object Storage
6	повна історія	Data Lake
7	довгостроково	Архів

Для забезпечення зберігання даних для систем і засобів протидії КА задамо їх архітектуру, в якій необхідно поєднати автоматичне керування життєвим циклом всіх даних або їх частини, особливості апаратно-програмних засобів для зберігання даних, інтелектуалізацію щодо глибини горизонту для даних, самонавчання при виборі даних для типових повторюваних КА. Це повинно підвищити оперативність та точність при надані даних системам протидії КА, а також забезпечити ефективність їх зберігання з урахування глибини горизонту.

Задамо концептуальну архітектуру засобів підсистеми пам'яті систем та засобів протидії КА в корпоративних мережах на основі використання історичних даних для прийняття рішень з урахуванням моделі $M_{zM,1}^D$ процесу формування історичних даних систем протидії КА, моделі $M_{zM,2}^D$ оцінювання даних та часу при формуванні та зберіганні історичних даних систем протидії КА, моделі $M_{zM,3}^D$ організації даних систем протидії КА, моделі $M_{zM,4}^D$ зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам корпоративних мереж, включно з даними обманних систем, а також з врахуванням підсистем, які базуються на моделях $P_{P,1}, P_{P,2}, P_{P,3}$, так:

$$A_P^{OS} = \begin{pmatrix} M_{zM,1}^D \\ M_{zM,2}^D \\ M_{zM,3}^D \\ M_{zM,4}^D \\ P_{P,1} \\ P_{P,2} \\ P_{P,3} \end{pmatrix}, \quad (2.43)$$

де $M_{zM,1}^D$ - моделі процесу формування історичних даних систем протидії КА; $M_{zM,2}^D$ - модель оцінювання даних та часу при формуванні та зберіганні історичних даних систем протидії КА; $M_{zM,3}^D$ - модель організації даних систем протидії КА; $M_{zM,4}^D$ - модель зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам корпоративних мереж; $P_{P,1}$ - модель організації зберігання даних в системах та засобах протидії комп'ютерним атакам з урахуванням історичного аспекту їх використання; $P_{P,2}$ - модель оптимізації даних на основі самонавчання засобів підсистеми пам'яті; $P_{P,3}$ - модель вибору даних з урахуванням глибини горизонту, самонавчання та попереднього досвіду використання

Деталізуємо моделі $M_{zM,1}^D$, $M_{zM,2}^D$, $M_{zM,3}^D$, $M_{zM,4}^D$ в контексті введених величин щодо історичних даних та зв'язків між ними.

Задамо модель процесу формування історичних даних систем протидії КА, включно з даними обманних систем, так:

$$M_{zM,1}^D = (M_{DI}^1, M_{DI}^D, M_{DD}^r, M_{vp}^p, M_K^D, G_{zM,1}^D), \quad (2.44)$$

де M_{DI}^1 - множина джерел інформації (формула (2.1)); M_{DI}^D - множина векторів для забезпечення зберігання даних від джерел інформації різних рівнів (формула (2.2)); M_{DD}^r - множина джерел даних для семи рівнів (формули (2.4)-(2.10)); M_{vp}^p - множина впливів на збереження та актуальність даних в різних аспектах їх використання та забезпечення доступу і актуальності (формула (2.11)); M_K^D - множина класів даних (формула (2.18)); $G_{zM,1}^D$ - граф зв'язків елементів множин $M_{DI}^1, M_{DI}^D, M_{DD}^r, M_{vp}^p, M_K^D$ для формування історичних даних систем протидії КА, включно з даними обманних систем.

Зв'язки між елементами множин $M_{DI}^1, M_{DI}^D, M_{DD}^r, M_{vp}^p, M_K^D$ визначені згідно формул (2.1), (2.2), (2.4)-(2.11)), (2.18) і зобразимо їх на рис. 2.2, причому для множин M_{DD}^r, M_K^D здійснено їх детальніше подання через підмножини та класи.

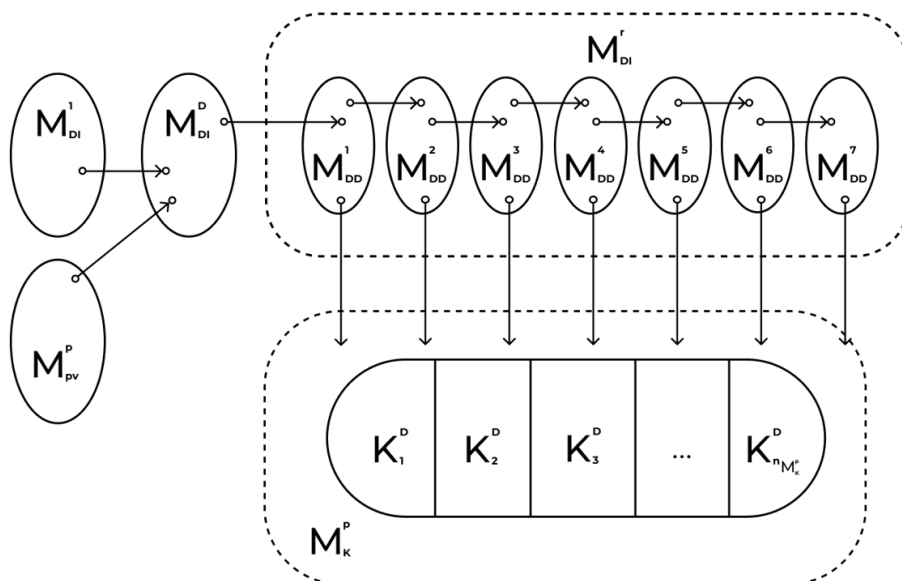


Рисунок 2.2 – Зв'язки між елементами множин $M_{DI}^1, M_{DI}^D, M_{DD}^r, M_{vp}^p, M_K^D$

Згідно визначених зв'язків між множинами $M_{DI}^1, M_{DI}^D, M_{DD}^r, M_{vp}^p, M_K^D$ задамо для моделі $M_{zM,1}^D$ (формула (2.43)) процесу формування історичних даних систем протидії КА граф $G_{zM,1}^D$, який зображено на рис. 2.3.

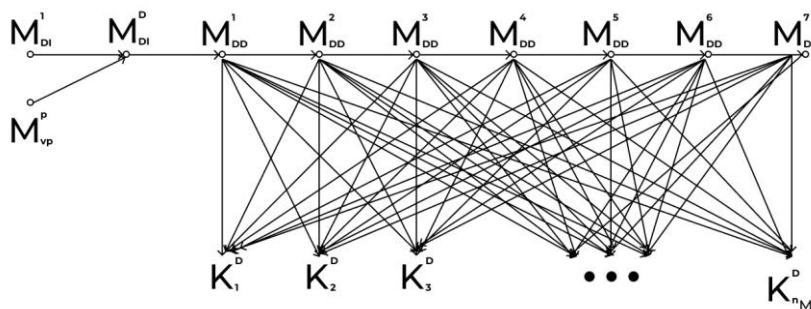


Рисунок 2.3 – Граф $G_{zM,1}^D$

Задамо модель оцінювання даних та часу при формуванні та зберіганні історичних даних систем протидії КА, включно з даними обманих систем, так:

$$M_{zM,2}^D = \left(M_{F_{vp,i}^p}, V_{vp,t}^p, F_{vp,t}^p, M_1^T, M_{v_{DI,m,n}^T}^T, G_{zM,2}^D \right), \quad (2.45)$$

де $M_{F_{vp,i}}^p$ – множина функцій $M_{F_{vp,i}}^p$ (формула (2.12)) для оцінювання впливів елементів множини впливів M_{vp}^p на збереження та актуальність даних в різних аспектах їх використання та забезпечення доступу і актуальності; $F_{vp,i}^p$ – функції оцінювання; ; $i = 1, 2, \dots, n_{M_{vp}^p}$; $n_{M_{vp}^p}$ – кількість елементів множини впливів M_{vp}^p ; $V_{vp,t}^p$ – вектор значень (формула (2.13)) оцінювання впливів на збереження та актуальність даних в різних аспектах їх використання та забезпечення доступу і актуальності; $F_{vp,t}^p$ – вектор функцій впливів $F_{vp,i}^p$ ($i = 1, 2, \dots, n_{M_{vp}^p}$; $n_{M_{vp}^p}$ – кількість елементів множини впливів M_{vp}^p) (формула (2.14)); t – поточний час, в який було отримано інформацію про дані і здійснено оцінювання впливу на них з використанням функцій $F_{vp,i}^p$ (формула (2.12)); M_1^T – матриця часу (формула (2.19)) для даних класів з урахуванням координати $v_{DI,i,9}^D$, що фіксує час розміщення даних в базах даних, вектору $V_{DI,i}^D$ ($i = 1, 2, \dots, N_{M_{DI}^D}$; $N_{M_{DI}^D}$ – кількість векторів для забезпечення зберігання даних від джерел інформації різних рівнів) і відповідного класу K_i^D ($i = 1, 2, \dots, n_{M_K^D}$; $n_{M_K^D}$ – кількість класів) з множини M_K^D (формула (2.18)); $M_{v_{DI,m,n}^D,i}^T$ – матриця (формула (2.20)) послідовного оцінювання впливів та відповідних даних з класів.

Граф $G_{zM,2}^D$ для моделі $M_{zM,2}^D$ (формула (2.45)) оцінювання даних та часу при формуванні та зберіганні історичних даних систем протидії КА, включно з даними обманних систем, зображено на рис. 2.4.

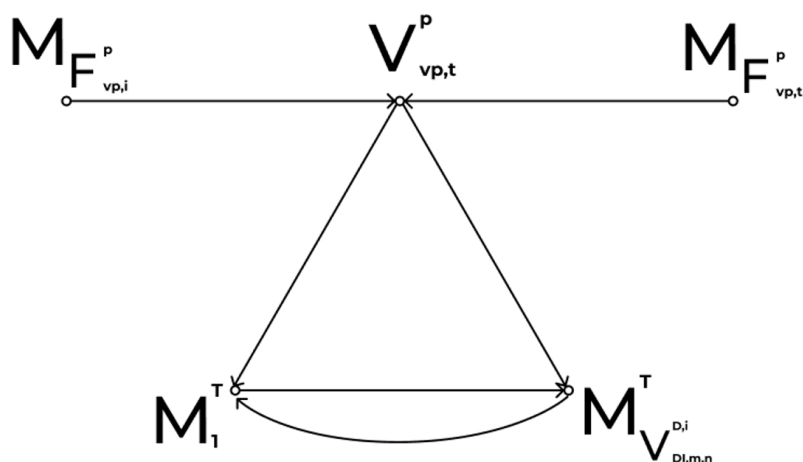


Рисунок 2.4 – Граф $G_{zM,2}^D$

Задамо модель організації даних систем протидії КА, включно з історичними даними та даними обманних систем, так:

$$M_{zM,3}^D = (M_{D,r}^{OS}, M_{D,r}^{OS,D}, K_{D,r,i}^{OS,D}, G_{zM,3}^D), \quad (2.46)$$

де $M_{D,r}^{OS}$ – множина (формула (2.21)), яка є підмножиною всіх даних та включає елементи про класи даних, які зберігаються та можуть бути використані обманними системами r рівня $M_{D,r}^{OS}$; $M_{D,r}^{OS,D}$ – множина історичних даних систем та засобів всіх семи рівнів (формула (2.22)), що використовують обманні технології $M_{D,r}^{OS,D}$; $K_{D,r,i}^{OS,D}$ – клас векторів підкласів (формули (2.23), (2.24), (2.26)-(2.37)), причому перший клас без обманних систем, а другий – з обманними системами; $K_{D,r,i}^{OS,D}$ – i -тий клас даних об'єктів обманних систем; $i = 1, 2$; r – номер рівня засобів і систем протидії КА; $r = 1, 2, \dots, 7$.

Граф $G_{zM,3}^D$ для моделі $M_{zM,3}^D$ (формула (2.46)) організації даних систем протидії КА, включно з історичними даними та даними обманних систем, зображено на рис. 2.5.

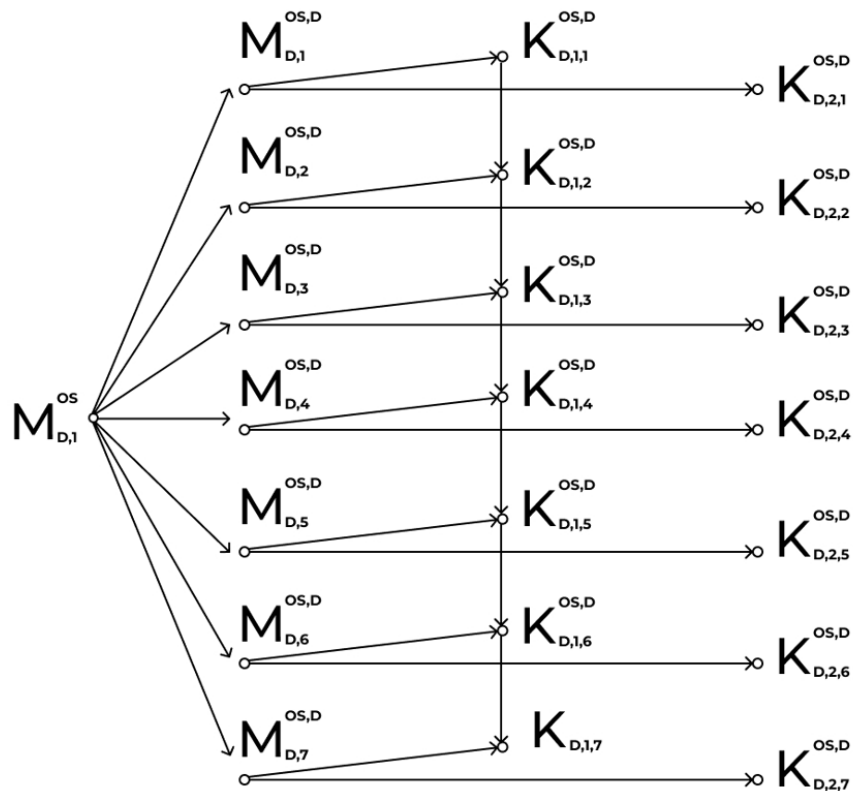


Рисунок 2.5 – Граф $G_{zM,3}^D$

Задамо модель зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам корпоративних мереж, так:

$$M_{zM,4}^D = (M_z^D, M_d^D, M_{ez}^D, G_{zM,4}^D), \quad (2.47)$$

де M_z^D – множина переліку завдань засобів зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам корпоративних мереж (формула (2.38)); M_d^D – множина типів даних, які опрацьовує засіб підсистеми пам'яті (формула (2.39)); M_{ez}^D – множина основних елементів засобів підсистеми пам'яті (формула (2.39)).

Граф $G_{zM,3}^D$ для моделі $M_{zM,3}^D$ (формула (2.47)) зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам корпоративних мереж зображено на рис. 2.6.

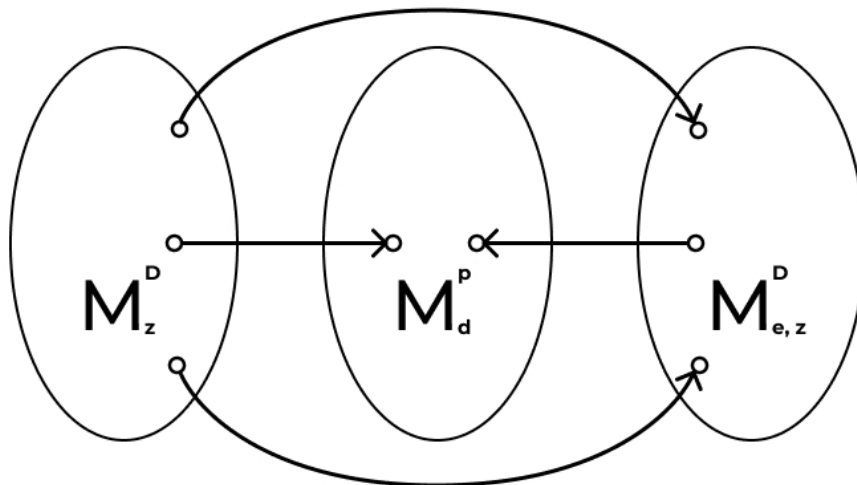


Рисунок 2.6 – Граф $G_{zM,4}^D$

Таким чином, удосконалено архітектуру засобів зберігання даних для систем і засобів протидії КА, в якій на відміну відомих, здійснено поєднання автоматичного керування життєвим циклом всіх даних або їх частини, особливостей апаратно-програмних засобів для зберігання даних, інтелектуалізації щодо глибини горизонту для даних, самонавчання при виборі даних для типових повторюваних КА, що є основою для підвищення оперативності та точності при наданні даних системам і засобам протидії КА, а також забезпечити ефективність їх зберігання з урахування глибини горизонту.

2.5 Висновки до другого розділу

Збереження даних у пам'яті засобів і систем усіх семи рівнів є необхідним для протидії комп'ютерним атакам. Водночас накопичення різномірних даних створює проблеми їх обсягу, доступу та уніфікації форматів. Такі дані потребують приведення до узгоджених форматів із збереженням зв'язків і врахуванням часової актуальності. Відсутність інформації про нові дії чи інструменти зломисників у пам'яті систем дає їм змогу уникати виявлення, тому вони намагаються обходити або обмежувати використання наявних даних.

Системи протидії КА використовують пам'ять не лише для виявлення КА, а й для вибору подальших дій. Це потребує збереження історії попередніх кроків із оцінюванням їх ефективності для повторного використання. Із підвищенням рівня систем зростає обсяг історичних даних, складність їх обробки та контекст прийняття рішень, водночас зменшується залежність від сигнатур. Тому потребує вдосконалення організації пам'яті та способів доступу до даних.

Розв'язання суперечностей між обсягом даних, їх актуальністю та адаптивністю аналізу можливе через застосування багаторівневих моделей пам'яті. Такі моделі поєднують зберігання даних, ознак і знань із урахуванням часу, контексту та здатності систем до адаптації.

Подання даних у вигляді векторів на різних рівнях ієрархії є основою підсистеми пам'яті, що забезпечує вибір глибини ретроспективи, оптимізацію застарілих даних і синтез нових. Удосконалена архітектура засобів зберігання даних передбачає автоматичне керування життєвим циклом даних, урахування особливостей засобів зберігання, адаптивний вибір горизонту та самонавчання, що є основою для розроблення методів підвищення оперативності і точності роботи систем та засобів протидії КА.

Основні наукові результати другого розділу опубліковані в працях [111-114; 119].

РОЗДІЛ 3

МЕТОДИ ЗБЕРІГАННЯ, ВИБОРУ ТА ОПТИМІЗАЦІЇ ДАНИХ
ДЛЯ СИСТЕМ ТА ЗАСОБІВ ПРОТИДІЇ КОМП'ЮТЕРНИМ АТАКАМ

Усунення протиріччя між необхідністю забезпечення максимально повного, довготривалого та історично репрезентативного зберігання даних і потребою обмеження їх обсягів та часу зберігання з метою підтримання актуальності, ефективності обробки та відповідності ресурсним і адаптивним можливостям систем потребує розроблення нових та удосконалення наявних методів опрацювання історичних даних. Основою для розроблення та удосконалення таких методів є удосконалена архітектура засобів зберігання даних, в якій передбачено автоматичне керування життєвим циклом даних, урахування особливостей засобів зберігання, адаптивний вибір горизонту та самонавчання. Таке поєднання архітектури засобів зберігання даних та методів забезпечення зберігання, вибору та оптимізації історичних даних повинно підвищити ефективність зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах.

3.1 Метод організації зберігання даних в системах та засобах протидії комп'ютерним атакам з урахуванням історичного аспекту їх використання

Системи та засоби протидії КА всіх семи рівнів по різному використовують дані, які зберігаються для їх функціонування, включно з даними, що накопичуються в процесі їх тривалої експлуатації. Отримані дані від джерел інформації різних рівнів формують базу множини векторів (формула (2.2)) з урахуванням формату векторів згідно із заданою в формулі (2.3). Також для даних різних рівнів тривалість збереження даних є різною, але в контексті застосування засобів всіх семи рівнів є потреба в збереженні всіх наявних та отримуваних даних. Крім того, частина даних могла бути використана при прийнятті рішень, що дає змогу оцінити ефективність такого застосування і, відповідно, при подальшому збереженні таких даних є потреба

в доповненні їх результатами оцінювання. Такі результати можуть теж накопичуватись з часом.

Таким чином, розділимо формування даних для систем та засобів протидії КА в залежності від їх джерел формування та використання так:

- 1) початкові дані;
- 2) накопичувані дані;
- 3) використані дані.

Початкові дані наявні в системах чи засобах протидії КА безпосередньо з початку їх встановлення в КМ. До накопичуваних даних віднесемо ті дані, які отримуються системами та засобами протидії КА в процесі опрацювання інцидентів в КМ. До використаних даних віднесемо ті дані, які сформовано системою чи засобами протидії КА на основі використання початкових та/або накопичених даних. При цьому вони містять результат оцінювання їх використання. Якщо їх використано один раз, то вони формуються як такі, що вже мають оцінку результату і є цілісними, тобто новими даними. При цьому всі вони взяті з певною глибиною горизонту початкових та накопичених даних. Якщо їх використано більше одного разу, то вони залишаються такими ж як були сформовані першого разу і містять додатково час використання та результат їх оцінювання для кожного випадку використання.

У процесі функціонування систем та засобів протидії КА в КМ формується значний обсяг різнорідних даних, які відрізняються за джерелами походження, часом актуальності та характером використання. Основною проблемою є не лише накопичення цих даних, а й забезпечення їх ефективного повторного використання з урахуванням історії обробки та результатів попередніх оцінювань. Крім того, важливою вимогою є автономність функціонування систем. Навіть у разі недоступності спеціалізованого засобу зберігання даних, системи та засоби протидії КА повинні продовжувати роботу без втрати базової функціональності.

Розглянемо кроки методу організації зберігання даних в системах та засобах протидії комп'ютерним атакам з урахуванням історичного аспекту їх використання.

Крок 1. Структуризація даних за джерелами формування з урахуванням 7-рівневої архітектури систем протидії КА.

У КМ системи та засоби протидії КА побудовані як багаторівнева ієрархічна структура з семи функціональних рівнів, кожен з яких генерує, обробляє або використовує дані різного типу. Тому на першому етапі необхідно виконати структуризацію даних одночасно за джерелами формування і за рівнями системи, що дозволяє встановити повну ієрархічну модель даних.

Ієрархічна модель даних з урахуванням семи рівнів згідно множини векторів даних (формула (2.2)) M_{DI}^D від джерел інформації різних рівнів та моделі $P_{P,1}$ (формула (2.43)) організації зберігання даних в системах та засобах протидії КА. Кожен рівень містить початкові дані, накопичує поточні дані та генерує для використання нові дані. Це означає, що дані описуються не лише функцією часу, а й функцією рівня обробки. Узагальнимо дані з урахуванням їх поділу так:

$$M_{DI}^{D,3,7} = \bigcup_{i=1}^3 \bigcup_{j=1}^7 M_{DI,i,j}^D, \quad (3.1)$$

де $M_{DI,i,j}^D$ – підмножина даних i -того типу (початкові, накопичувані, використані) для j – того рівня систем та засобів протидії КА.

Дані певного типу та рівня можуть належати різним рівням одночасно.

Структуризація даних за джерелами формування з урахуванням 7-рівневої архітектури систем протидії КА може бути здійснена з використанням певних алгоритмів обробки та розподілу даних, але враховуючи складність завдань, які вирішують системи і засоби, то необхідне їх поєднання. Здійснимо комбіновану структуризацію даних, яка включатиме поетапну багатокритеріальну обробку даних, де кожна подія буде проходити через послідовність взаємопов'язаних механізмів:

- 1) первинне потокове захоплення даних;
- 2) нормалізація та уніфікація структури;
- 3) метадані-орієнтована класифікація;
- 4) контекстне збагачення;
- 5) адаптивна маршрутизація до рівнів 1–7;
- 6) агрегація та збереження історії використання.

Таким чином, дані не обробляються єдиним засобом, а будуть проходити каскадну багаторівневу трансформацію, де кожен етап буде додавати нову інформаційну цінність.

Крок 2. Формування та фіксація початкових даних.

Початкові дані є критично важливою складовою функціонування системи та засобів протидії КА, оскільки саме вони визначають базову конфігурацію системи, політики безпеки, правила виявлення КА, параметри взаємодії між сімома рівнями системи, початковий стан моделей аналізу. Якщо ці дані є нестабільними або змінюються без контролю, то це призводить до некоректного виявлення КА, порушення логіки реагування, розбалансування між рівнями 1–7, втрати відтворюваності результатів аналізу. Тому початкові дані повинні бути однозначно визначені, зафіксовані та незмінні після ініціалізації системи, що забезпечує їх роль як еталонного інформаційного базису. Початкові дані надходять в момент розгортання систем та засобів протидії КА та зберігаються в матрицях M_1^T (формула (2.19)) та $M_{v_{DI,m,n}}^T$ (формула (2.20)). За результатами такого розгортання та надходження початкових даних фіксується базова конфігурація системи.

Крок 3. Накопичення даних.

Перезапис даних призводить до втрати історії, тому необхідно використати подієву модель. Тому, необхідно здійснювати перехід від статичного представлення даних до динамічного, подійно-орієнтованого накопичення інформації, яке відображає реальний стан КМ у часі. У системах та засобах протидії КА кожна подія (логін, мережний запит, аномалія, спрацювання правила) є не тільки значенням, а фактом зміни стану системи, який має бути збережений разом із контекстом виникнення. Таким чином, дані не перезаписуються, а накопичуються як послідовність подій, що утворюють історію поведінки системи. Тоді вирішується проблема того, що нові значення даних замінюють попередні, відомості про попередній стан втрачається, неможливо відтворити послідовність КА, аналіз інцидентів стає неповним. Це є критичним, оскільки КА часто є багатокроковими та розподіленими у часі. Подійна модель передбачає, що кожна дія фіксується як окрема подія, події зберігаються у хронологічному порядку та стан системи відновлюється

через аналіз послідовності подій. Задамо підмножини накопичуваних даних $M_{DI,2,j}^D$ ($j = 1, 2, \dots, 7$) так:

$$M_{DI,2,j}^D(t_n) = M_{DI,2,j}^D(t_{n-1}) \cup \{p_n\}, \quad (3.2)$$

де $M_{DI,2,j}^D(t_n)$ - підмножини накопичуваних даних ($j = 1, 2, \dots, 7$) в момент часу t_n при n -ій ітерації; $M_{DI,2,j}^D(t_{n-1})$ - підмножини накопичуваних даних ($j = 1, 2, \dots, 7$) в момент часу t_n при $n - 1$ -ій ітерації, тобто попереднього кроку; p_n - дані нової події з n -ї ітерації.

Накопичувані дані підмножин $M_{DI,2,j}^D$ ($j = 1, 2, \dots, 7$) зберігаються в матрицях M_1^T (формула (2.19)) та $M_{v_{DI,m,n}^{D,i}}^T$ (формула (2.20)). Впровадження подійної моделі забезпечує збереження повної історії КА, можливість реконструкції сценаріїв компрометації, виявлення багатокрокових атак (АРТ), аналіз поведінки користувачів у часі та основу для машинного навчання та кореляції подій.

Крок 4. Формування використаних даних.

На цьому етапі відбувається виділення змісту, закономірностей і висновків із цих даних. Накопичені дані самі по собі не є достатніми для виявлення складних КА, оцінювання ризиків та прийняття рішень. Необхідно формалізувати процес, який поєднує історичні події, використовує початкові знання та формує нові, більш узагальнені або оцінені дані. Задамо формування використовуваних даних з наявних початкових та накопичених даних так:

$$m_{DI,3,j,k}^D = M_{F,1,2}^D(M_{DI,1,j}^D, M_{DI,2,j}^D), \quad (3.3)$$

де $m_{DI,3,j,k}^D \in M_{DI,3,j}^D$ - елемент підмножини використовуваних даних ($j = 1, 2, \dots, 7$); $M_{DI,1,j}^D$ - підмножини початкових даних ($j = 1, 2, \dots, 7$); $M_{DI,2,j}^D$ - підмножини накопичуваних даних ($j = 1, 2, \dots, 7$); $M_{F,1,2}^D$ - множина функцій для формування використовуваних даних.

Дані можуть бути такими, що були використані, та такими, що підготовлені до використання, тобто використовувані. Множина функцій $M_{F,1,2}^D$ (формула (3.3))

містить функції, які поділяють множину використаних/використовуваних даних на такі підмножини:

1) використані дані, що отримані з початкових та/або накопичених даних з урахуванням глибини горизонту, для вирішення певних завдань в системах та засобах протидії КА, тобто наявні дані без попередньої підготовки та модифікації;

2) використовувані дані, які отримані з початкових та/або накопичених даних з урахуванням глибини горизонту та сформовані на їх основі для вирішення певних завдань в системах та засобах протидії КА, тобто підготовлені дані для їх використання без штучної модифікації;

3) використовувані дані, які отримані з початкових та/або накопичених даних, а також з підготовлених використовуваних даних, з урахуванням глибини горизонту та модифіковані для вирішення певного завдання в системах та засобах протидії КА, тобто штучно створені дані.

Дані, які не змінено, але вони можуть бути використані в іншому контексті, віднесемо до даних третього типу. До даних третього типу віднесено даня, які раніше не існували, а були створені на основі аналізу. Це найбільш цінний тип даних, оскільки він містить нові знання.

Формування таких даних забезпечується використанням різних методів обробки:

1. Кореляційний аналіз, що передбачає об'єднання подій у логічні ланцюги, що дозволяє виявляти багатокрокові атаки та пов'язувати окремі дії в один інцидент.

2. Правильовий (сигнатурний) аналіз передбачає застосування наперед визначених правил (сигнатури КА, шаблони поведінки), що дозволяє швидко виявляти відомі загрози.

3. Поведінковий аналіз і машинне навчання, які використовують для аналізу відхилення від нормальної поведінки користувачів, систем, мережної активності.

4. Контекстний аналіз, при якому до подій додається додаткова інформація про критичність активів, геолокацію, репутацію IP та роль користувача.

5. Агрегація та узагальнення дають змогу здійснити поєднання великої кількості подій в більш компактні представлення.

Оскільки використані дані можуть застосовуватися багаторазово і накопичують результати оцінювання, то наступним етапом є формалізація їх структури та організація збереження історії їх використання, що дозволяє відстежувати ефективність рішень і забезпечує адаптивність системи в часі.

Крок 5. Формалізація структури похідних даних.

Результати аналізу, отримані на попередньому етапі, перестають розглядатися як одноразові висновки і перетворюються на повноцінні інформаційні об'єкти, придатні для довготривалого зберігання, повторного використання та накопичення досвіду. На цьому етапі факт оформлюється таким чином, щоб зберегти не тільки сам результат, а й усі обставини його отримання та подальшого використання. Суть кроку полягає у фіксації чотирьох ключових аспектів кожного результату: набір подій, логів або інших елементів, що стали підставою для висновку; момент часу, коли цей результат було отримано, що дозволяє оцінювати його актуальність; зміст результату (тип інциденту, рівень ризику, класифікація КА, рекомендації); історія використання цього результату, яка відображає, як саме він застосовувався далі, тобто переда

вання його аналітику, підтвердження, використання для реагування, закриття тощо. Таким чином, результат аналізу отримує “життєвий цикл” і перестає бути статичною величиною. У практиці SIEM це відповідає перетворенню окремих подій на інцидент, який має не лише опис, а й зв'язок із джерельними подіями, часову прив'язку та журнал дій. Наприклад, система може на основі кількох підозрілих логінів і доступу до ресурсу сформулювати інцидент компрометації облікового запису. Якщо цей інцидент зберегти лише як текстове повідомлення, то подальший аналіз буде обмеженим. Якщо ж він зберігається разом із переліком використаних подій, часом створення, рівнем ризику та історією обробки (ескалація, підтвердження, блокування, закриття), то він стає повноцінним об'єктом управління безпекою.

На рівні структури даних формується центральна сутність “інцидент”, яка містить основні атрибути (тип, критичність, час створення, результат), пов'язується з набором подій, які стали її причиною, і має окрему історію змін та дій. Окремо можуть зберігатися контекстні характеристики (наприклад, критичність активу, географічне походження, оцінка загрози), які доповнюють інцидент і впливають на прийняття

рішень. У підсумку формується структура, де один інцидент пов'язаний з багатьма подіями (зв'язок “один до багатьох”) і з багатьма записами історії використання (також “один до багатьох”), що повністю відповідає ідеї збереження як джерела, так і наслідків. Зв'язок із 7-рівневою архітектурою систем та засобів протидії КА проявляється в тому, що кожен із цих елементів має чітку прив'язку до рівня.

Таким чином, фактично поєднуються всі рівні систем через єдиний об'єкт, тобто використані дані у вигляді інцидентів із історією. Він забезпечує не лише збереження результатів аналізу, а й їх інтеграцію в повний цикл функціонування системи, тобто від події до стратегічного управління.

Крок 6. Фіксація одноразового використання даних.

Фіксація першого використання сформованих результатів як завершеного та цілісного факту. Його суть у тому, що після того, як певний результат аналізу (наприклад, виявлений інцидент або оцінка ризику) був застосований хоча б один раз, то він повинен бути зафіксований разом із моментом цього використання та отриманим ефектом. Це дозволяє перетворити результат із “потенційного” на “реально використаний”, тобто такий, що вже має підтверджену практичну значущість. Бо є необхідність відокремити просто сформовані результати від тих, що реально були використані в процесі функціонування системи. Саме перше використання є ключовим, оскільки воно визначає початкову оцінку ефективності результату та фіксує його як завершений інформаційний об'єкт. Результатом виконання є створення завершеного, зафіксованого результату, який уже має підтверджений факт використання і може бути надалі оцінений, порівняний або повторно використаний у системах чи засобах протидії КА.

Крок 7. Накопичення історії багаторазового використання даних.

Один і той самий результат (наприклад, інцидент, оцінка або правило) може застосовуватися неодноразово в різних умовах і з різними наслідками, тому кожен факт такого використання повинен фіксуватися. Дані з часом не втрачають цінності одразу, а можуть повторно використовуватись для нових рішень, перевірок або аналізу. При цьому результати їх застосування можуть відрізнятися, що є важливим для оцінювання ефективності та надійності.

На практиці це реалізується шляхом додавання нового запису до історії при кожному використанні даних, тобто із зазначенням часу застосування та отриманого результату. Таким чином формується послідовність використань, яка відображає “життєвий цикл” даного результату. У результаті виконання цього кроку формується повна історія використання даних, що дозволяє аналізувати їх ефективність, повторюваність та зміну результатів у часі. Такі дані зберігаються в матрицях M_1^T (формула (2.19)) та $M_{v_{DI,m,n}^{D,i}}^T$ (формула (2.20)).

Крок 8. Оцінювання часової глибини даних.

Оцінювання часової глибини даних, тобто визначення того, наскільки “старими” є ті вхідні дані, на основі яких сформовано певний результат, необхідно бо актуальність інформації безпосередньо залежить від часу, тобто чим старіші дані, тим меншою може бути їхня цінність для прийняття рішень. Водночас повністю ігнорувати історичні дані не можна, бо вони можуть містити ознаки довготривалих або прихованих КА. Тому виникає потреба не просто зберігати дані, а кількісно оцінювати їхню “віддаленість у часі” відносно моменту формування результату.

Для цього введемо поняття часової глибини використаних даних (глибини горизонту), яку будемо визначати як найбільшу різницю між часом формування результату та часами створення всіх вхідних даних, що були використані. Задамо часову глибину даних T_F^D так:

$$T_F^D(m_{DI,3,j,k}^D) = \max_{m_{DI,3,j,k}^D \in M_{DI,3,j}^D} (t_p - t_m), \quad (3.4)$$

де $m_{DI,3,j,k}^D \in M_{DI,3,j}^D$ – елемент підмножини (формула (3.3)) використовуваних даних ($j = 1, 2, \dots, 7$); $T_F^D(m_{DI,3,j,k}^D)$ – часова глибина k -го використаного (похідного) даного; $M_{DI,3,j}^D$ – множина вхідних даних (подій, записів, логів), на основі яких сформовано результат; t_p – момент часу, коли результат було отримано; t_m – момент часу створення кожного окремого вхідного даного.

Вираз $(t_p - t_m)$ показує “вік” кожного такого даного на момент формування результату, а оператор $\max$ дозволяє вибрати найбільше значення серед них, тобто визначити найстаріший елемент, що вплинув на результат.

Таким чином, можна встановити час, на який система чи засіб повернулись в минуле, щоб отримати даний висновок. Якщо значення часової глибини є малим, то це означає, що результат базується на відносно нових даних і має високу актуальність. Якщо ж це значення є значним, то результат сформовано із залученням старих даних, що може свідчити як про втрату актуальності, так і про наявність довготривалих процесів або КА.

На практиці обчислення цього показника здійснюється автоматично на основі часових міток, які вже присутні у всіх даних. Система визначає час формування результату, порівнює його з часами всіх використаних подій і знаходить максимальну різницю. Отримане значення виступає кількісною оцінкою актуальності результату та може бути використане для подальших рішень щодо обробки даних.

Крок 9. Шарове розділення даних.

Зберігання даних потрібно здійснювати за принципом шарового розділення, тобто розподіляти всі дані на окремі логічні шари залежно від їх типу, призначення та характеру використання. Необхідність такого підходу зумовлена тим, що різні типи даних у системах протидії КА мають суттєво відмінні вимоги до швидкості доступу, частоти використання, обсягу та актуальності. Зокрема, початкові дані є відносно стабільними і рідко змінюються, накопичені дані мають великий обсяг і активно поповнюються, а використані (похідні) дані є найбільш цінними з точки зору аналізу та прийняття рішень і потребують швидкого доступу. Вся множина даних розділяється на три основні шари: шар початкових даних; шар накопичених даних; шар використаних даних. Такий розподіл здійснюється на основі їх функціональної ролі. Тому, кожен із цих шарів має власний життєвий цикл, інтенсивність доступу та вимоги до обробки. Це означає, що дані не зберігаються в єдиному масиві, а розміщуються у різних логічних або фізичних сегментах сховища.

Крок 10. Забезпечення простежуваності даних.

Забезпечення простежуваності даних означає можливості встановлення походження кожного сформованого результату та відновлення процесу його отримання. Необхідність цього зумовлена тим, що в системах протидії КА результати аналізу (інциденти, оцінки, рішення) не можуть розглядатися ізольовано, бо для їх

перевірки, довіри та повторного використання важливо чітко знати, на основі яких саме даних і подій вони були сформовані. Для кожного результату визначається функція походження, яка відображає, з яких саме даних він був отриманий. Таким чином, будь-який результат можна простежити до первинних подій або початкових даних, що забезпечує повну прозорість процесу аналізу.

Реалізуємо простежуваність даних через збереження посилань або ідентифікаторів усіх вхідних даних, які використані при формуванні результату. Результатом буде забезпечення відтворюваності. Система отримає можливість повторити процес формування результату, перевірити його або використати ті самі дані для інших аналітичних задач.

Крок 11. Локальне накопичення даних.

Кожен компонент системи повинен мати можливість самостійно формувати, зберігати та використовувати дані без залежності від зовнішніх сервісів, зберігаючи при цьому повну функціональність. В реальних умовах КМ можливі збої зв'язку, відмови центральних вузлів або навмисні КА на інфраструктуру. У таких ситуаціях система не повинна втрачати здатність виявляти загрози, аналізувати події та реагувати на них. Тому всі ключові процеси, тобто накопичення подій, формування результатів і їх використання, мають підтримуватися на локальному рівні. У кожному вузлі або рівні системи дані накопичуються незалежно, шляхом послідовного додавання нових подій до вже наявного локального набору. Нові дані формуються безпосередньо в процесі роботи компонента (наприклад, сенсора, агента, модуля аналізу) і додаються до локального сховища без перезапису попередніх значень. Це забезпечує збереження історії навіть у відокремленому режимі.

Крок 12. Синхронізація з центральним сховищем.

Всі дані, які були сформовані та збережені в автономному режимі у окремих вузлах або рівнях системи, повинні бути передані до центрального середовища та інтегровані з уже наявною інформацією без втрати цілісності та історії. В умовах тимчасової ізоляції кожен компонент системи накопичує власний фрагмент даних, який відображає лише локальну частину подій. Після відновлення зв'язку ці фрагменти мають бути об'єднані, щоб відновити повну картину функціонування

системи, забезпечити коректний аналіз і уникнути інформаційних розривів. Синхронізація повинна забезпечити узгоджене об'єднання даних із урахуванням часових міток, унікальності подій і вже наявних записів. Нові локальні дані повинні додаватись до центрального сховища таким чином, щоб не виникало дублювання, зберігалася хронологія подій, не втрачалися зв'язки між даними та їх походженням, та враховувалася історія використання. Для цього використаємо механізми синхронізації: буферизація локальних даних до моменту передачі; використання унікальних ідентифікаторів подій; узгодження за часовими мітками; контроль цілісності та повноти переданих даних.

Крок 13. Адаптивний розподіл даних за типами сховищ

Всі дані не зберігаються однаково, а динамічно переміщуються між рівнями зберігання (швидкими або повільнішими) залежно від того, як часто і наскільки недавно вони використовувались. Такий підхід дозволяє досягти балансу між швидкістю доступу та витратами ресурсів. В системах протидії КА обсяг даних постійно зростає, але лише невелика їх частина активно використовується у поточний момент часу. Зберігання всіх даних у високопродуктивному середовищі є неефективним і ресурсоємним, тоді як переміщення рідко використовуваних даних у дешевші або повільніші сховища дозволяє оптимізувати інфраструктуру без втрати інформації. Для реалізації такого розподілу даних введемо показник часу використання даних як різницю між поточним моментом часу та моментом останнього використання конкретного елемента даних так:

$$T_A^D(m_{DI,i,j,k}^D) = \max_{m_{DI,i,j,k}^D \in M_{DI,i,j}^D} (t_{A,p} - t_{A,m}), \quad (3.5)$$

де $m_{DI,i,j,k}^D \in M_{DI,i,j}^D$ – елемент підмножини (формули (3.1)-(3.4)) даних ($j = 1, 2, \dots, 7$; $i = 1, 2, 3$); $T_A^D(m_{DI,i,j,k}^D)$ – час використання k -го даного; $M_{DI,i,j}^D$ – множина вхідних даних (подій, записів, логів), на основі яких сформовано результат; $t_{A,p}$ – момент часу, коли дані було використано; $t_{A,m}$ – час останнього використання конкретного елемента k -го даного.

На основі формули (3.5) здійснюємо розподіл даних між трьома типами сховищ так:

$$F_{A,T}^D(m_{DI,i,j,k}^D) = \begin{cases} m_{DI,i,j,k}^D \in M_{A,T,1}^D, T_A^D(m_{DI,i,j,k}^D) < T_1^A \\ m_{DI,i,j,k}^D \in M_{A,T,2}^D, T_1^A < T_A^D(m_{DI,i,j,k}^D) < T_2^A, \\ m_{DI,i,j,k}^D \in M_{A,T,3}^D, T_2^A < T_A^D(m_{DI,i,j,k}^D) \end{cases} \quad (3.6)$$

де $F_{A,T}^D(m_{DI,i,j,k}^D)$ – функція розподілу даних; $m_{DI,i,j,k}^D \in M_{DI,i,j}^D$ – елемент підмножини (формули (3.1)-(3.4)) даних ($j = 1, 2, \dots, 7$; $i = 1, 2, 3$); $T_A^D(m_{DI,i,j,k}^D)$ – значення часу використання k -го даного; $M_{DI,i,j}^D$ – множина вхідних даних (подій, записів, логів), на основі яких сформовано результат; $t_{A,p}$ – момент часу, коли дані було використано; $t_{A,m}$ – значення часу останнього використання конкретного елемента k -го даного; $M_{A,T,1}^D$ – підмножина даних (сховище), які використовуються часто і потребують швидкого доступу; $M_{A,T,2}^D$ – підмножина даних (сховище), до яких звертаються рідше, але вони все ще мають значення для аналізу; $M_{A,T,3}^D$ – підмножина даних (сховище), які практично не використовуються, але зберігаються для довгострокового аналізу або аудиту; T_1^A – значення нижньої межі часу; T_2^A – значення верхньої межі часу.

Порогові значення T_1^A і T_2^A визначаються політиками системи і можуть адаптуватися залежно від навантаження, типу даних або вимог до безпеки. Розподіл не є статичним, а динамічним. Дані можуть переміщуватися між сховищами в процесі експлуатації системи. Реалізуємо такий підхід через автоматизовані механізми керування життєвим циклом даних, які відстежують частоту та час доступу, переміщують дані між рівнями зберігання і забезпечують прозорий доступ незалежно від фізичного розташування даних. Таким чином, досягається оптимізація використання ресурсів системи. Найбільш важливі та часто використовувані дані доступні з мінімальною затримкою, тоді як рідко використовувані дані не перевантажують швидкі сховища.

Крок 14. Перевірка відповідності метаданих даних характеристикам апаратного середовища зберігання.

Необхідно здійснити перевірку узгодженості між властивостями даних, які зафіксовані у вигляді їх метаданих, та характеристиками апаратних засобів, на яких ці

дані розміщуються. На попередніх етапах для кожного елемента даних вже сформовано набір характеристик. Водночас апаратні засоби також мають формалізовані характеристики, наприклад, продуктивність, надійність, вік, інтенсивність зношення та доступні ресурси. Без зіставлення цих двох груп параметрів неможливо забезпечити ефективне і безпечне розміщення даних. Кожен елемент даних задамо вектором метаданих $V_{DI,1}^D$, а кожен носій даних - вектором $V_{DI,2}^D$ його характеристик. Перевірку відповідності між ними задамо так:

$$V_{DI,1}^D(m_{DI,i,j,k}^D) = (v_{DI,1,i,j,1}^D, v_{DI,1,i,j,2}^D, v_{DI,1,i,j,3}^D, v_{DI,1,i,j,4}^D, v_{DI,1,i,j,5}^D, \dots, v_{DI,1,i,j,n_{V_{DI,1}^D}}^D);$$

$$V_{DI,2}^D(m_{DI,i,j,k}^D) = (v_{DI,2,i,j,1}^D, v_{DI,2,i,j,2}^D, v_{DI,2,i,j,3}^D, v_{DI,2,i,j,4}^D, v_{DI,2,i,j,5}^D, \dots, v_{DI,2,i,j,n_{V_{DI,2}^D}}^D); \quad (3.7)$$

$$q_{DI,i,j,k}^D = F_{DI,Z,V1,V2}^D(V_{DI,1}^D(m_{DI,i,j,k}^D), V_{DI,2}^D(m_{DI,i,j,k}^D)),$$

де $q_{DI,i,j,k}^D$ – значення допустимості поточного розміщення даних в наявних апаратних ресурсах для зберігання даних; $q_{DI,i,j,k}^D \in [0; 1]$; $m_{DI,i,j,k}^D$ – елемент підмножини $M_{DI,i,j}^D$ (формули (3.1)-(3.4)) даних ($j = 1, 2, \dots, 7$; $i = 1, 2, 3$); $V_{DI,1}^D(m_{DI,i,j,k}^D)$ – вектор метаданих даних; $v_{DI,1,i,j,y}^D$ – координата вектору метаданих даних; $y = 1, 2, \dots, n_{V_{DI,1}^D}$; $n_{V_{DI,1}^D}$ – кількість координат вектору метаданих даних; $V_{DI,2}^D(m_{DI,i,j,k}^D)$ – вектор характеристик апаратного ресурсу для зберігання даних; $v_{DI,1,i,j,z}^D$ – координата вектору характеристик апаратного ресурсу для зберігання даних; $z = 1, 2, \dots, n_{V_{DI,2}^D}$; $n_{V_{DI,2}^D}$ – кількість координат вектору характеристик апаратного ресурсу для зберігання даних; $F_{DI,Z,V1,V2}^D$ – функція відповідності, яка визначає допустимість поточного розміщення даних в наявних апаратних ресурсах.

Координати вектору метаданих даних можуть бути такі:

- 1) $v_{DI,1,i,j,1}^D$ – часова глибина;
- 2) $v_{DI,1,i,j,2}^D$ – частота використання;
- 3) $v_{DI,1,i,j,3}^D$ – інтенсивність звернень;
- 4) $v_{DI,1,i,j,4}^D$ – рівень критичності;
- 5) $v_{DI,1,i,j,5}^D$ – рівень/шар даних.

Координати вектору характеристик апаратного ресурсу для зберігання даних можуть бути такі:

- 1) $v_{DI,2,i,j,1}^D$ – продуктивність;
- 2) $v_{DI,2,i,j,2}^D$ – надійність;
- 3) $v_{DI,2,i,j,3}^D$ – вік;
- 4) $v_{DI,2,i,j,4}^D$ – поточне навантаження.

Перевірка виконується шляхом оцінювання цієї функції для кожної пари “дані–носій”. Якщо встановлюється, що характеристики носія не задовольняють вимоги, то сформовані метаданими даних (наприклад, високочастотні або критичні дані розміщені на застарілому або перевантаженому носії), така ситуація фіксується як невідповідність. Це можна реалізувати через механізми моніторингу, які одночасно аналізують метадані даних, сформовані на попередніх кроках, технічні параметри апаратного середовища та поточний стан використання ресурсів. Тоді, система отримує формалізовану оцінку відповідності між даними та середовищем їх зберігання. Це дозволяє не лише виявляти неефективні або ризиковані розміщення, але й забезпечує основу для автоматизованого прийняття рішень щодо їх коригування.

Крок 15. Формування попереднього досвіду використаних даних.

Система повинна не лише зберігати факти використання даних, але й повинна формувати узагальнений досвід, який надалі може бути використаний для прогнозування, адаптації та підтримки прийняття рішень. Необхідність такого підходу обумовлена тим, що результати аналізу, інциденти та сценарії реагування з часом починають повторюватися. Якщо система здатна накопичувати інформацію про попередні використання даних і їх наслідки, то вона отримує можливість швидше оцінювати нові ситуації, прогнозувати розвиток подій, повторно використовувати успішні рішення і зменшувати навантаження на аналітиків. Історія використання даних перетворюється на структурований досвід, який може зберігатися та аналізуватися за допомогою різних математичних підходів.

Для формування попереднього досвіду використаних даних використаємо різні методи та технології, щоб на їх основі забезпечити комбінований підхід до

формування досвіду використання історичних даних. Використаємо методи і технології, які базуються на таких моделях:

- 1) марковська модель;
- 2) баєсівська модель;
- 3) статистичні моделі накопичення досвіду.

У випадку використання марківської моделі кожен стан системи (наприклад, “виявлення події”, “підтвердження інциденту”, “реагування”, “закриття”) розглядаємо як окремий стан марковського процесу, а історія використання даних дозволить оцінити ймовірності переходів між цими станами. Згідно марківської моделі випадковий процес не має пам’яті, тобто збереження тривалої історії для даних не забезпечується. При цьому, наявна можливість збереження ймовірності переходів між станами, яка може бути використана для формування досвіду використання даних на попередньому кроці їх задіявання в системах та засобах протидії КА. В контексті розглядуваної задачі із збереженням даних та досвіду попереднього їх використання наявна їх відповідність такому виду марківського процесу як дискретний процес з дискретним часом. Випадковий процес X_t вважається марковським процесом, якщо для будь-яких моментів часу з відрізка $[0;T]$ функція розподілу останнього значення X_{t_n} при фіксованих значеннях $X_{t_0}, X_{t_1}, X_{t_2}, \dots, X_{t_{n-1}}$ залежить тільки від $X_{t_{n-1}}$. В марківському випадковому процесі майбутній розвиток подій залежить тільки від поточного стану і не залежить від того, коли і яким чином система перейшла в цей стан. Тому зберігання даних про певні процеси будемо базувати на формуванні матриць для кожного блоку даних, якими описуватимемо для кожного з них ймовірності переходу з одного стану в інший за відомий проміжок часу. Множину всіх можливих послідовностей переходів відображає дерево логічних переходів з ймовірностями переходу з одного стану в інший. По дереву логічних переходів можна визначити ймовірність переходу в той чи інший стан після декількох кроків. Нехай маємо множину станів системи $S^{D,P} = \{s_1^{D,P}, s_2^{D,P}, \dots, s_{n^{S^{D,P}}}^{D,P}\}$ ($n^{S^{D,P}}$ – кількість станів системи) та множини ймовірності станів для кожного стану окремо $P_i^{D,P} =$

$\{P_{i,1}^{D,P}, P_{i,2}^{D,P}, \dots, P_{i,n^{S^{D,P}}}^{D,P}\} (j = 1, 2, \dots, n^{S^{D,P}}; i = 1, 2, \dots, n^{S^{D,P}})$ після будь-якого кроку.

Задамо таку інформацію матрицею так:

$$P^{D,P} = \begin{pmatrix} P_{1,1}^{D,P} & P_{1,2}^{D,P} & \dots & P_{1,n^{S^{D,P}}}^{D,P} \\ P_{2,1}^{D,P} & P_{2,2}^{D,P} & \dots & P_{2,n^{S^{D,P}}}^{D,P} \\ \dots & \dots & \dots & \dots \\ P_{n^{S^{D,P}},1}^{D,P} & P_{n^{S^{D,P}},2}^{D,P} & \dots & P_{n^{S^{D,P}},n^{S^{D,P}}}^{D,P} \end{pmatrix}, \quad (3.8)$$

де $P_{i,j}^{D,P}$ – ймовірність передувати i -тому стану до j -того стану системи.

Також, згідно матриці, яку задано за формулою (3.8) можна визначити кількість кроків, за які система повертається до початкового стану. Матрицею $P^{D,P}$ можна задавати дані, які в подальшому використовуються в якості їх попереднього досвіду застосування. При цьому глибина горизонту ретроспективного використання даних для марковської моделі буде найменшою і приймемо її за мінімальну базову глибину горизонту.

Другим варіантом для збереження попереднього досвіду використання даних є використання байєсівських моделей, у яких попередній досвід використовується для уточнення ймовірності появи певних подій або КА. У такому випадку кожне нове використання даних змінює оцінки ймовірності на основі накопиченого досвіду. Це дозволяє системі адаптивно враховувати нові знання та поступово підвищувати точність аналізу. Потужність байєсівської мережі полягає в тому, що апріорні ймовірності можна уточнювати при надходженні додаткової інформації. Застосуємо динамічні байєсівські мережі, оскільки процеси із збереженням даних та їх використання в динамічному режимі протягом тривалого часу функціонування систем та засобів протидії КА в КМ, що повинно враховувати часову зміну станів систем і засобів.

Байєсівська модель для збереження попереднього досвіду використання даних дає змогу не тільки зберігати окремі події або результати аналізу, а формалізувати залежності між ними в часі. Тоді система накопичує не лише дані, а й знання про те, як змінювалися стани, які події передували інцидентам та які рішення були прийняті

і до яких наслідків це призводило. Використання динамічної байєсівської мережі полягає у представленні системи як послідовності часових зрізів. Кожен часовий зріз містить набір вузлів, що описують стан системи в певний момент часу. До таких вузлів введемо такі параметри:

- 1) тип події;
- 2) рівень загрози;
- 3) стан вузла мережі;
- 4) факт виявлення атаки;
- 5) тип реагування;
- 6) результат реагування;
- 7) рівень довіри до даних;
- 8) актуальність даних;
- 9) стан апаратного ресурсу.

Між вузлами сформуємо два типи зв'язків:

- 1) внутрішньопарові, якими задамо залежності між параметрами в межах одного моменту часу;
- 2) міжчасові, якими задамо вплив попереднього стану системи на наступний.

Фактично це дозволяє моделювати еволюцію систем та засобів протидії КА у часі. Наприклад, якщо на попередньому часовому кроці було зафіксовано аномальну активність і високий рівень ризику, то мережа може оцінювати ймовірність виникнення інциденту на наступному кроці. Аналогічно, якщо певний тип даних раніше багаторазово використовувався для успішного виявлення КА, то система підвищує ймовірність його значущості в майбутньому.

Введемо надбудову над історією використання даних на основі таких параметрів для запису і фіксування історії щодо використаних даних:

- 1) час використання;
- 2) результат;
- 3) тип даних;
- 4) рівень системи;
- 5) ефективність рішення на основі набору спостережень для навчання мережі.

Здійснимо побудову динамічної байєсівської мережі у кілька етапів. На першому етапі визначимо часові кроки моделі так: секунди; хвилини; окремі сесії; інциденти; цикли аналізу. Далі формуємо множину вузлів мережі з такими параметрами: тип параметра; можливі стани; джерело даних. Після цього задамо залежності між вузлами, наприклад, так: попередній рівень загрози впливає на наступний; тип події впливає на ймовірність КА; ефективність попереднього реагування впливає на майбутній вибір рішень; часова глибина даних впливає на рівень довіри до них. Наступним етапом є побудова таблиць умовних ймовірностей, які визначають наскільки один стан впливає на інший та найбільш ймовірні наявні переходи. Ці таблиці можна задавати експертно, формувати статистично або отримувати згідно навчання автоматично на історичних даних. Після накопичення історії використання система починає оновлювати параметри мережі відповідно до нового досвіду. Динамічна баєсівська мережа адаптується до нових сценаріїв КА і нових умов функціонування. У контексті зберігання даних вона дозволяє визначати, найбільш цінні наявні дані, прогнозувати ймовірність повторного використання даних, оцінювати актуальність історичних даних, автоматично змінювати пріоритети зберігання та враховувати часову зміну значущості даних.

Третім варіантом для застосування є статистичні моделі накопичення досвіду для оцінювання та узагальнення результатів попереднього використання даних. Система не тільки зберігає окремі факти використання даних, а накопичує статистичні характеристики, які відображають частоту, ефективність, повторюваність та результативність застосування різних типів даних у процесах виявлення і протидії КА. У такому підході кожен випадок використання даних розглядається як статистичне спостереження. При цьому фіксуються тип використаних даних, час використання, рівень системи або засобу протидії КА, тип інциденту, результат використання, ефективність прийнятого рішення, кількість повторних використань, тривалість актуальності даних. На основі накопичення таких спостережень система формує статистичні оцінки, які дозволяють визначати найчастіше використовувані типи даних, найбільш ефективні дані, шаблони повторюваних подій, найчастіші результати успішного реагування та зміну актуальності даних у часі.

Для реалізації третього варіанту можна застосувати декілька різних статистичних методів, формуючи комбінований підхід для досягнення ефективності розроблюваного процесу.

Модель частотного аналізу включає для кожного типу даних накопичення такої статистика: кількості використань; кількості успішних застосувань; середнього часу між використаннями; середньої тривалості актуальності.

Статистична модель розподілів ймовірності дає змогу формувати оцінки таких параметрів: ймовірність повторного використання даних; ймовірність успішного виявлення КА; ймовірність втрати актуальності даних через певний проміжок часу. Для цього використаємо нормальний розподіл, експоненційний розподіл, пуассонівський розподіл.

Моделі часових рядів дає змогу аналізувати зміну активності використання даних у часі, наприклад: сезонність атак; періодичність використання певних даних; тенденції зміни ефективності окремих джерел інформації.

Таким чином, статистичні моделі накопичення досвіду забезпечують перехід від звичайного збереження історії до кількісного оцінювання ефективності даних. Це дозволяє реалізувати адаптивне управління інформацією, підвищити ефективність використання ресурсів і забезпечити більш точне виявлення та прогнозування КА.

Використання лише одного підходу для збереження та оцінювання історичного досвіду функціонування систем і засобів протидії КА є недостатнім через різноманітність даних, різні часові горизонти їх актуальності та різний рівень визначеності процесів у КМ. Тому було доцільно поєднано марковські моделі, динамічні баєсовські мережі та статистичні методи, які можуть застосовуватись як до однакових, так і до різних типів даних та сценаріїв. Марковські моделі доцільні для опису послідовності станів системи та переходів між ними в процесі розвитку КА або реагування на неї. Їх особливістю є можливість накопичення досвіду щодо типових сценаріїв переходів між станами та оцінювання ймовірностей подальшого розвитку подій на основі попередніх випадків використання даних. Динамічні баєсовські мережі доцільно використовувати для врахування причинно-наслідкових зв'язків між подіями, ознаками КА, станами системи та результатами прийняття рішень у часі. Їх

особливістю є можливість працювати з неповними, нечіткими або ймовірнісними даними, а також враховувати зміну взаємозалежностей між параметрами в процесі накопичення нових відомостей. Це дозволяє адаптувати оцінювання ефективності використання даних відповідно до зміни умов функціонування КМ. Статистичні методи забезпечують узагальнення накопичених результатів використання даних, визначення тенденцій, частотних характеристик, аномалій та оцінювання ефективності прийнятих рішень. Їх особливістю є низька обчислювальна складність та можливість швидкого опрацювання великих обсягів історичних даних, що важливо для забезпечення автономності функціонування систем і засобів протидії КА.

Таким чином, спільне використання цих підходів забезпечує комплексне представлення історичного досвіду. Марковські моделі описують динаміку станів, Динамічні баєсівські мережі задають причинно-наслідкові залежності в умовах невизначеності, а статистичні методи - кількісне узагальнення та оцінювання накопичених результатів.

На практиці це означає, що кожен результат аналізу та кожен випадок використання даних не лише зберігається, але й впливає на майбутню поведінку системи. Наприклад, якщо певний тип інциденту раніше неодноразово підтверджувався як критичний, то система може автоматично підвищувати його пріоритет у майбутньому. Аналогічно, якщо деякі дані виявлялися малоефективними або застарілими, то їх значущість може знижуватися.

Таким чином, формується попередній досвід використання даних, який стає основою для адаптивного функціонування систем та засобів протидії КА, що дозволяє перейти від звичайного накопичення інформації до накопичення знань і досвіду та це підвищує ефективність аналізу, прогнозування і реагування в КА.

Таким чином, ефективне функціонування систем та засобів протидії КА потребує організації зберігання даних з урахуванням не лише їх походження та структури, а й історії використання в процесах прийняття рішень. Метод передбачає поділ даних на початкові, накопичувані та використані, що дає змогу враховувати етапи їх формування, повторного застосування та результати оцінювання ефективності такого застосування. Особливістю методу є збереження та опрацювання

історичного досвіду використання даних із застосуванням марковських моделей, динамічних баєсівських мереж і статистичних методів. При цьому марковські моделі забезпечують опис послідовності станів та сценаріїв розвитку подій, динамічні баєсівські мережі - врахування причинно-наслідкових зв'язків та невизначеності даних у часі, а статистичні методи - узагальнення результатів використання даних та оцінювання ефективності прийнятих рішень. Це забезпечує накопичення історичних відомостей щодо використання даних, підвищення ефективності їх повторного застосування, адаптивності систем та автономності їх функціонування навіть в умовах недоступності спеціалізованих засобів зберігання даних.

3.2 Метод оптимізації даних на основі самонавчання засобів підсистеми пам'яті

Тривалий час функціонування систем та засобів протидії КА супроводжується постійним накопиченням даних різного типу. З певним часом їх може збільшитись до такої кількості, що вони будуть суттєво впливати на ефективність засобів та систем, які їх використовують. Частина з цих даних може повторюватись багатократно, бути дуже подібними між собою, рідко використовуватись, втратити актуальність тощо. Адміністратор засобів не може опрацювати великі обсяги таких даних швидко і коректно. Тому системи і засоби протидії КА повинні самі здійснювати їх оптимізацію. Цього можна досягти введенням в їх архітектуру алгоритмів оптимізації даних із можливістю самонавчання. Відсутність самонавчання буде формувати кроки до здійснення оптимізації згідно детермінованих алгоритмів, що відповідатиме за суттю опрацювання даних адміністратором. Самонавчання може базуватись на одиничних даних або на їх поєднаннях з урахуванням організації даних на основі марковської, баєсівської та статистичних моделей. Метою такої оптимізації з самонавчанням є забезпечення раціонального використання ресурсів підсистеми пам'яті шляхом автоматичного виявлення, аналізу, узагальнення та реорганізації накопичених даних для підвищення швидкодії системи, зменшення надлишковості інформації, підтримання її актуальності та адаптації до змін умов функціонування. Суть оптимізації даних полягає у приведенні структури накопиченої інформації до

такого стану, за якого забезпечується максимально ефективне використання пам'яті при збереженні необхідного рівня інформативності. Це передбачає виявлення повторюваних, подібних, застарілих або рідко використовуваних даних, їх класифікацію, об'єднання, архівацію або вилучення з метою скорочення обсягу зберігання та прискорення доступу до релевантної інформації.

У процесі функціонування система може багаторазово фіксувати однакові або майже ідентичні події, наприклад повторні сигнали про одну й ту саму активність КА з незначними відмінностями у часових мітках. У такому випадку оптимізація полягає в об'єднанні подібних записів у єдиний узагальнений інформаційний об'єкт. Наприклад, якщо протягом короткого інтервалу часу отримано серію повідомлень про однаковий тип дії об'єкта, то вони агрегуються в один запис із зазначенням кількості повторень.

Інформація про об'єкти, яка тривалий час не використовувалася або втратила актуальність через зміну умов функціонування, може бути автоматично вилучена або перенесена до шару архівного сховища. Наприклад, якщо дані про певну траєкторію або характеристику КА не використовувалися протягом визначеного проміжку часу, то система оцінює їх коефіцієнт актуальності та, у разі його зниження нижче встановленого порога, виконує архівацію.

Якщо в пам'яті накопичуються великі масиви даних про однотипні події або схожі характеристики КА, то система може групувати їх у кластери. Наприклад, усі записи про об'єкти зі схожими параметрами або режимами функціонування можуть бути об'єднані в окремі групи тоді замість аналізу кожного запису окремо система використовує центри кластерів, що значно прискорює пошук і прийняття рішень.

Система може визначати ступінь значущості інформації залежно від частоти використання, актуальності та важливості для прийняття рішень. Наприклад, дані про об'єкти, що часто аналізуються або пов'язані з підвищеним рівнем загрози, розміщуються у швидкодоступних областях пам'яті. Менш значуща інформація переноситься до повільніших сегментів зберігання.

У випадку накопичення великих обсягів телеметричних або сигнальних даних система може застосовувати алгоритми стиснення. Наприклад, якщо певні

послідовності параметрів повторюються, то вони можуть бути представлені компактними шаблонами або кодами замість повного збереження кожного значення.

На основі механізмів самонавчання система може прогнозувати, які дані з високою ймовірністю втратять актуальність найближчим часом. Наприклад, якщо аналіз статистики показує, що певний тип інформації після конкретного інтервалу більше не використовується, то система автоматично позначає його як кандидата для вилучення або архівації.

Якщо в процесі експлуатації змінюється характер звернень до даних, то система може змінювати власну структуру організації пам'яті. Наприклад, при збільшенні частоти використання певної категорії інформації вона переноситься до кешованої області або індексується додатковими ознаками для прискорення доступу.

Система повинна забезпечувати самостійне уточнення критеріїв відбору. Наприклад, якщо після вилучення певних даних система фіксує погіршення якості прийняття рішень, то вона коригує порогові значення актуальності та частоти використання. У подальшому аналогічні записи вже не будуть видалятися, а зберігатимуться довше.

Таким чином, дані систем та засобів протидії КА є різнотипними та в процесі їх накопичення потребують оптимізації, бо впливатимуть на результати виявлення КА.

Оптимізаційна задача полягає у визначенні такого способу організації та перетворення даних, який забезпечить мінімізацію витрат пам'яті та часу доступу до інформації при одночасному збереженні її цілісності, значущості та можливості подальшого використання для прийняття рішень. Для розв'язання цієї задачі здійснимо послідовне виконання кількох етапів. На початковому етапі здійснюємо аналіз накопичених даних, під час якого можна визначити частоту їх використання, ступінь актуальності, наявність дублювання та подібності між інформаційними об'єктами. Наступним етапом буде побудова моделі організації пам'яті, яка дозволяє встановити взаємозв'язки між даними та оцінити їх значущість для функціонування системи. Після цього формуємо критерії оптимізації, відповідно до яких обираються способи перетворення даних. Далі виконуємо безпосередньо реорганізацію інформаційного масиву шляхом кластеризації, ранжування, компресії або вилучення

надлишкових елементів. Завершальним етапом є оцінювання ефективності проведеної оптимізації та коригування параметрів подальшого функціонування.

Самонавчання в цьому процесі забезпечує собою здатність системи чи засобів протидії КА до самостійної адаптації алгоритмів оптимізації на основі аналізу результатів власної роботи та накопиченого досвіду функціонування без безпосереднього втручання адміністратора. Його сутність полягає в тому, що системи чи засоби не лише виконують визначений набір операцій, а й постійно вдосконалюють правила опрацювання інформації, враховуючи зміни у структурі даних, частоті звернень до них та ефективності попередніх рішень. Процес самонавчання включає накопичення статистики функціонування, виявлення закономірностей у використанні даних, формування внутрішньої моделі прийняття рішень, адаптацію параметрів алгоритмів та оцінювання результатів виконаних дій. На основі отриманого зворотнього зв'язку можна забезпечити уточнення критеріїв оптимізації, що забезпечує поступове підвищення точності прийняття рішень і формування більш ефективних стратегій організації пам'яті. Таким чином, самонавчання забезпечує перехід від статичного, детермінованого способу оптимізації до адаптивного механізму, здатного пристосовуватись до змін середовища та особливостей накопичених даних.

Розглянемо спочатку розв'язування оптимізаційної задачі щодо організації та перетворення даних систем та засобів протидії КА. Для цього приймемо загальний інформаційний масив $M_{DI}^{D,3,7}$ систем та засобів протидії КА з урахуванням поділу даних на початкові, накопичувані, використані згідно з формулою (3.1) та відповідно на підмножини $M_{DI,i,j}^D$ (i - тип, j – рівень систем та засобів протидії КА).

Для розв'язування оптимізаційної задачі щодо організації та перетворення даних систем та засобів протидії КА введемо оптимізаційну функцію. Враховуючи багатокритеріальність цієї оптимізаційної задачі, оптимізаційну функцію задамо системою взаємопов'язаних оптимізаційних функцій конкретних критеріїв. Для цього здійснимо конструювання багатокритеріальної інтегральної оптимізаційної моделі організації та перетворення даних систем та засобів протидії КА так:

$$F_{opt}^D \left(P_0^{F_{opt}^D}, P_1^{F_{opt}^D}, \dots, P_{n_{F_{opt}^D}}^{F_{opt}^D} \right) = \sum_{i=0}^{n_{F_{opt}^D}} \alpha_i \cdot F_{opt,i}^D \left(p_1^{F_{opt,i}^D}, p_2^{F_{opt,i}^D}, \dots, p_{n_{F_{opt,i}^D}}^{F_{opt,i}^D} \right) \rightarrow \min, \quad (3.9)$$

де $F_{opt}^D \left(P_0^{F_{opt}^D}, P_1^{F_{opt}^D}, \dots, P_{n_{F_{opt}^D}}^{F_{opt}^D} \right)$ - багатокритеріальна інтегральна оптимізаційна функція, яка складена з окремих нелінійних функцій, що відповідають різним вхідним факторам; $P_i^{F_{opt}^D} = \left(p_1^{F_{opt,i}^D}, p_2^{F_{opt,i}^D}, \dots, p_{n_{F_{opt,i}^D}}^{F_{opt,i}^D} \right)$; $P_i^{F_{opt}^D}$ - вектор параметрів для i - того критерію; $i = 0, 1, \dots, n_{F_{opt}^D}$; $n_{F_{opt}^D}$ - кількість критеріїв для оптимізації; $j = 1, 2, \dots, n_{F_{opt,i}^D}$; $n_{F_{opt,i}^D}$ - кількість параметрів параметрів для i - того критерію; α_i - ваговий коефіцієнт для i - того критерію; $F_{opt,i}^D \left(p_1^{F_{opt,i}^D}, p_2^{F_{opt,i}^D}, \dots, p_{n_{F_{opt,i}^D}}^{F_{opt,i}^D} \right)$ - оптимізаційна функція для i - того критерію.

Параметри в критеріях можуть бути в різних діапазонах. Найкращим значенням багатокритеріальної інтегральної оптимізаційної функції буде найменше значення від попереднього значення. Головним завданням цієї функції є формування переліку таких параметрів, які призведуть до вибору найкращого рішення щодо організації та перетворення даних систем та засобів протидії КА.

Згідно формули (3.9) введемо окремі оптимізаційні функції для критеріїв, наприклад, так:

- 1) $F_{opt,0}^D \left(p_1^{F_{opt,0}^D}, p_2^{F_{opt,0}^D}, \dots, p_{n_{F_{opt,0}^D}}^{F_{opt,0}^D} \right)$ - базові критерії ресурсної оптимізації;
- 2) $F_{opt,1}^D \left(p_1^{F_{opt,1}^D}, p_2^{F_{opt,1}^D}, \dots, p_{n_{F_{opt,1}^D}}^{F_{opt,1}^D} \right)$ - критерії структури даних;
- 3) $F_{opt,2}^D \left(p_1^{F_{opt,2}^D}, p_2^{F_{opt,2}^D}, \dots, p_{n_{F_{opt,2}^D}}^{F_{opt,2}^D} \right)$ - критерії шарового розміщення;
- 4) $F_{opt,3}^D \left(p_1^{F_{opt,3}^D}, p_2^{F_{opt,3}^D}, \dots, p_{n_{F_{opt,3}^D}}^{F_{opt,3}^D} \right)$ - критерії апаратно-мережного розподілу;
- 5) $F_{opt,4}^D \left(p_1^{F_{opt,4}^D}, p_2^{F_{opt,4}^D}, \dots, p_{n_{F_{opt,4}^D}}^{F_{opt,4}^D} \right)$ - критерії оптимізації даних обманних

систем.

Таким чином, в формулі (3.9) реалізована лінійна інтеграція часткових оптимізаційних функцій. Функція $F_{opt,0}^D \left(p_1^{F_{opt,0}^D}, p_2^{F_{opt,0}^D}, \dots, p_{n_{F_{opt,0}^D}}^{F_{opt,0}^D} \right)$ визначає базові критерії ресурсної оптимізації. Функції $F_{opt,i}^D \left(p_1^{F_{opt,i}^D}, p_2^{F_{opt,i}^D}, \dots, p_{n_{F_{opt,i}^D}}^{F_{opt,i}^D} \right)$ ($i = 1, 2, \dots, 4$) деталізують особливості структури даних, їх розподілу за категоріями, ієрархії шарів пам'яті, апаратного розміщення в корпоративній мережі та параметрів і даних обманих систем. Деталізуємо кожну окрему оптимізаційну функцію для конкретних критеріїв.

В загальному випадку функції $F_{opt,i}^D \left(p_1^{F_{opt,i}^D}, p_2^{F_{opt,i}^D}, \dots, p_{n_{F_{opt,i}^D}}^{F_{opt,i}^D} \right)$ ($i = 0, 1, \dots, 4$) задамо так:

$$F_{opt,i}^D \left(p_1^{F_{opt,i}^D}, p_2^{F_{opt,i}^D}, \dots, p_{n_{F_{opt,i}^D}}^{F_{opt,i}^D} \right) = \sum_{j=1}^{n_{F_{opt,i}^D}} \alpha_{i,j} \cdot p_j^{F_{opt,i}^D} \rightarrow \min, \quad (3.10)$$

де $F_{opt,i}^D \left(p_1^{F_{opt,i}^D}, p_2^{F_{opt,i}^D}, \dots, p_{n_{F_{opt,i}^D}}^{F_{opt,i}^D} \right)$ – оптимізаційна функція для i – того критерію; $i = 0, 1, \dots, n_{F_{opt}}^D$; $n_{F_{opt}}^D$ – кількість критеріїв для оптимізації; $j = 1, 2, \dots, n_{F_{opt,i}^D}$; $n_{F_{opt,i}^D}$ – кількість параметрів для i – того критерію; $\alpha_{i,j}$ – j – тий ваговий коефіцієнт для i – того критерію.

Згідно формули (3.10) задамо функцію $F_{opt,0}^D \left(p_1^{F_{opt,0}^D}, p_2^{F_{opt,0}^D}, \dots, p_{n_{F_{opt,0}^D}}^{F_{opt,0}^D} \right)$

поєднанням таких критеріїв оптимізації:

- 1) критерій використання пам'яті $p_1^{F_{opt,0}^D}$;
- 2) критерій часу доступу $p_2^{F_{opt,0}^D}$;
- 3) критерій щодо надлищковості даних $p_3^{F_{opt,0}^D}$;
- 4) критерій втрати актуальності $p_4^{F_{opt,0}^D}$;
- 5) критерій інформативності $p_5^{F_{opt,0}^D}$;

б) критерій витрат мережних ресурсів $p_6^{F_{opt,0}^D}$.

Критерій використання пам'яті характеризує обсяг зайнятих ресурсів з урахуванням розподілення в корпоративних мережах і задамо його параметром так:

$$p_1^{F_{opt,0}^D} = \sum_{k=1}^{n_{F_{opt,0}^D}} \frac{v_{z,k}^{p_1^{F_{opt,0}^D}}}{v_{v,k}^{p_1^{F_{opt,0}^D}}}; \quad 0 \leq \frac{v_{z,k}^{p_1^{F_{opt,0}^D}}}{v_{v,k}^{p_1^{F_{opt,0}^D}}} \leq 1, \quad (3.11)$$

де $v_{v,k}^{p_1^{F_{opt,0}^D}}$ – доступний обсяг пам'яті у k – тому вузлі; $k = 1, 2, \dots, n_{F_{opt,0}^D}$; $n_{F_{opt,0}^D}$ –

кількість вузлів; $v_{z,k}^{p_1^{F_{opt,0}^D}}$ – зайнятий обсяг пам'яті у k – тому вузлі.

Параметр $p_1^{F_{opt,0}^D}$ знаходиться в межах інтервалу $0 \leq \frac{p_1^{F_{opt,0}^D}}{n_{F_{opt,0}^D}} \leq 1$ та його оцінку

для відображення критерію задамо, наприклад, так:

1) $0 \leq \frac{p_1^{F_{opt,0}^D}}{n_{F_{opt,0}^D}} \leq 0,4$ – низьке оптимальне завантаження;

2) $0,4 < \frac{p_1^{F_{opt,0}^D}}{n_{F_{opt,0}^D}} \leq 0,75$ – середнє завантаження;

3) $0,75 < \frac{p_1^{F_{opt,0}^D}}{n_{F_{opt,0}^D}} \leq 1$ – високе критичне завантаження.

Критерій часу доступу $p_2^{F_{opt,0}^D}$ задамо так:

$$p_2^{F_{opt,0}^D} = \sum_{k=1}^{n_{F_{opt,0}^D}} \frac{t_{z,k}^{p_2^{F_{opt,0}^D}}}{t_{v,k}^{p_2^{F_{opt,0}^D}}}; \quad 0 \leq \frac{t_{z,k}^{p_2^{F_{opt,0}^D}}}{t_{v,k}^{p_2^{F_{opt,0}^D}}} \leq 1, \quad (3.12)$$

де $t_{v,k}^{p_2^{F_{opt,0}^D}}$ – максимальний час доступу до пам'яті у k – тому вузлі за результатами багатократних звернень; $k = 1, 2, \dots, n_{p_2^{F_{opt,0}^D}}$; $n_{p_2^{F_{opt,0}^D}}$ – кількість вузлів; $t_{z,k}^{p_2^{F_{opt,0}^D}}$ – середній час доступу до пам'яті у k – тому вузлі після багатократних звернень.

Параметр часу доступу $p_2^{F_{opt,0}^D}$ знаходиться в межах інтервалу $0 \leq \frac{p_2^{F_{opt,0}^D}}{n_{p_2^{F_{opt,0}^D}}} \leq 1$ та

його оцінку для відображення критерію часу доступу $p_2^{F_{opt,0}^D}$ задамо, наприклад, так:

- 1) $0 \leq \frac{p_1^{F_{opt,0}^D}}{n_{p_1^{F_{opt,0}^D}}} \leq 0,5$ – низький оптимальний час доступу;
- 2) $0,5 < \frac{p_1^{F_{opt,0}^D}}{n_{p_1^{F_{opt,0}^D}}} \leq 0,75$ – середній час доступу;
- 3) $0,75 < \frac{p_1^{F_{opt,0}^D}}{n_{p_1^{F_{opt,0}^D}}} \leq 1$ – високий критичний час доступу.

Крім того, параметр часу доступу $p_2^{F_{opt,0}^D}$ повинен враховувати шарі розміщення даних і він таким чином може бути для кожного шару різним з врахуванням різних вузлів, в яких розглядаються та використовуються різні шари розміщення даних.

В процесі функціонування систем та засобів протидії КА може відбуватись накопичення дубльованих даних, що формує певну надлишковість при їх зберіганні. Її потрібно оцінити та врахувати при здійсненні оптимізації. Введемо параметр $p_3^{F_{opt,0}^D}$ для визначення надлишковості даних так:

$$p_3^{F_{opt,0}^D} = \frac{n_z^{p_3^{F_{opt,0}^D}}}{n_v^{p_3^{F_{opt,0}^D}}}, \quad (3.13)$$

де $n_z^{p_3^{F_{opt,0}^D}}$ – кількість дубльованих записів; $n_v^{p_3^{F_{opt,0}^D}}$ – кількість всіх записів.

Тоді, значення $p_3^{F_{opt,0}^D}$ повинно бути близьким до нуля, тобто $0 \leq p_3^{F_{opt,0}^D}$, інакше організація даних не буде оптимальною.

Частина даних з певним часом втрачає свою актуальність повністю або частково, причому часткова втрата актуальності може призводити до отримання не завжди точних результатів, тобто часткова втрата актуальності впливатиме на результат, але вона може і не впливати. В цьому випадку розглядатимемо дані, при частковій втраті актуальності в яких не відбувається негативного впливу на результат.

Цінність накопичених даних безпосередньо залежить від часу їх останнього використання та швидкості зміни оперативної обстановки. У процесі функціонування системи частина інформації поступово втрачає значущість через зміну характеристик КА, оновлення сценаріїв впливу, зміну параметрів мережного середовища, появу нових шаблонів аномальної поведінки та моральне старіння раніше накопичених сценаріїв реагування. Наприклад, сигнатура втручання, яка була релевантною кілька місяців тому, може втратити практичну цінність після зміни алгоритмів роботи КА. Аналогічно історичні телеметричні шаблони можуть перестати відповідати поточному режиму функціонування системи.

Для кількісного опису цього процесу введемо коефіцієнт q_k актуальності окремого інформаційного об'єкта:

$$q_k = e^{-\lambda_k \cdot \Delta t_k}, \quad (3.14)$$

де λ_k - коефіцієнт старіння інформації для k – того інформаційного об'єкта; t_k - момент останнього використання k – того інформаційного об'єкта; t – поточний час; $\Delta t_k = t - t_k$.

Вибір експоненційної функції зумовлений тим, що втрата актуальності даних у системах та засобах протидії КА має нелінійний характер. На початковому інтервалі після створення або використання запису його релевантність зменшується незначно, однак зі збільшенням часу ймовірність його практичної цінності стрімко зменшується. Експоненційний закон дає змогу врахувати прискорене старіння оперативних даних, плавну деградацію історичних знань та можливість адаптивного налаштування швидкості старіння через параметр λ .

Критерій втрати актуальності $p_4^{F_{opt,0}^D}$ даних з врахуванням формули (3.14) задамо так:

$$p_4^{F_{opt,0}^D} = 1 - \frac{1}{n_D} \cdot \sum_{k=1}^{n_D} q_k, \quad (3.15)$$

де n_D – кількість інформаційних об'єктів; $k = 1, 2, \dots, n_D$; q_k - коефіцієнт актуальності k - того інформаційного об'єкта.

Межі зміни значення $p_4^{F_{opt,0}^D}$ такі: $0 \leq p_4^{F_{opt,0}^D} \leq 1$, причому актуальність даних можна встановити за такими значеннями:

- 1) $p_4^{F_{opt,0}^D} = 0$ – усі значення актуальні;
- 2) $0 < p_4^{F_{opt,0}^D} \leq 0,5$ – часткова втрата актуальності даних;
- 3) $0,5 < p_4^{F_{opt,0}^D} \leq 1$ – інформація суттєво застаріла.

Критерій інформативності $p_5^{F_{opt,0}^D}$ даних характеризує отриманий інформаційний об'єкт щодо зменшення невизначеності поточного або прогнозованого стану КА. Якщо після надходження нових даних невизначеність суттєво зменшується, то такі дані є високоінформативними. Якщо ж вони майже не змінюють уявлення системи про ситуацію, то їх інформативність є низькою. Коефіцієнт інформативності визначимо через зміну ентропії інформаційного стану об'єкта так:

$$g_k = \frac{H_{k,1} - H_{k,2}}{H_{k,1}}, \quad (3.16)$$

де k – номер інформаційного об'єкта; g_k - коефіцієнт інформативності k - го інформаційного об'єкта; $H_{k,2}$ - ентропія стану систем чи засобів протидії КА після їх оброблення для k - го інформаційного об'єкта; $H_{k,1}$ - ентропія стану систем чи засобів протидії КА до оброблення для k - го інформаційного об'єкта.

Визначимо ентропію інформаційного стану об'єкту m - го можливого стану КА так:

$$H_{k,m} = - \sum_{j=1}^{m_H} p_{k,j}^H \cdot \log_2 p_{k,j}^H, \quad (3.17)$$

де $P_{k,j}^H$ - імовірність m - го можливого стану КА; m_H - кількість можливих станів; k - номер інформаційного об'єкта; $j = 1, 2, \dots, m_H$.

Тоді, згідно з формулою (3.17) співвідношення $H_{k,2} \leq H_{k,1}$ задає критерій інформативності $p_5^{F_{opt,0}^D}$, згідно якого $0 \leq g_k \leq 1$. Інтерпретація значення g_k може бути визначена так:

- 1) $g_k \approx 0$, тобто дані практично не містять нової інформації;
- 2) $g_k \approx 0,5$, тобто наявна середня інформативність;
- 3) $g_k \approx 1$, тобто дані майже повністю усувають невизначеність.

В першому випадку система функціонує в штатному режимі, в другому – наявна аномальна активність, в третьому – наявна спроба несанкціонованого впливу.

Згідно з формулою (3.17) значення критерію інформативності $p_5^{F_{opt,0}^D}$ задамо так:

$$p_5^{F_{opt,0}^D} = \frac{1}{m_{p_5^{F_{opt,0}^D}}} \cdot \sum_{k=1}^{m_{p_5^{F_{opt,0}^D}}} g_k, \quad (3.18)$$

де k – номер інформаційного об'єкта; $m_{p_5^{F_{opt,0}^D}}$ – кількість інформаційних об'єктів; $k = 1, 2, \dots, m_{p_5^{F_{opt,0}^D}}$; g_k - коефіцієнт інформативності k - го інформаційного об'єкта.

Таким чином, оцінювання інформативності здійснюється через реальний внесок даних в уточнення стану КА, а не через умовні експертні оцінки.

Критерій витрат мережних ресурсів $p_6^{F_{opt,0}^D}$ задамо з урахуванням пропускну здатності в КМ так:

$$p_6^{F_{opt,0}^D} = \frac{1}{m_{p_6^{F_{opt,0}^D}}} \cdot \sum_{k=1}^{m_{p_6^{F_{opt,0}^D}}} \frac{Q_{k,1}}{Q_{k,1} + Q_{k,2}}, \quad (3.19)$$

де $m_{p_6^{F_{opt,0}^D}}$ – загальна кількість мережних ресурсів; $k = 1, 2, \dots, m_{p_6^{F_{opt,0}^D}}$; $Q_{k,1}$ – витрачені мережні ресурси; $Q_{k,2}$ – вільні мережні ресурси.

Аналогічно з попередніми критеріями можна згідно формули (3.19) встановити значення для трактування результатів критерію витрат мережних ресурсів $p_6^{F_{opt,0}^D}$.

Деталізуємо критерії структур даних $F_{opt,1}^D \left(p_1^{F_{opt,1}^D}, p_2^{F_{opt,1}^D}, \dots, p_{n_{F_{opt,1}^D}}^{F_{opt,1}^D} \right)$. Критерії

структури даних характеризують не ресурсні параметри функціонування системи, а саме якість внутрішньої організації інформаційного середовища систем та засобів протидії КА. Їх призначення полягає у визначенні ефективності побудови взаємозв'язків між інформаційними об'єктами, ступеня впорядкованості даних, їх логічної цілісності та здатності підтримувати процеси аналізу і прийняття рішень. На відміну від критеріїв ресурсної оптимізації, де оцінюються обсяг пам'яті, час доступу або навантаження на обчислювальні ресурси, критерії структури даних описують саме інформаційну організацію масиву даних. Тому до них не включатимемо характеристики, які вже враховані у ресурсних критеріях або критеріях шарового розміщення.

Функції для критеріїв структур даних $F_{opt,1}^D$ задамо з урахуванням формули (3.10). Виділимо такі критерії структур даних $F_{opt,1}^D$:

- 1) критерій зв'язаності структури $p_1^{F_{opt,1}^D}$;
- 2) критерій групування даних $p_2^{F_{opt,1}^D}$;
- 3) критерій однорідності структури $p_3^{F_{opt,1}^D}$;
- 4) критерій унікальності інформаційних об'єктів $p_4^{F_{opt,1}^D}$;
- 5) критерій структурної адаптивності $p_5^{F_{opt,1}^D}$.

Розглянемо кожен з критеріїв.

Критерій характеризує ступінь логічного взаємозв'язку між інформаційними об'єктами. У системах протидії КА дані рідко існують ізольовано. Наприклад, телеметричні записи пов'язані з часовими мітками, журнали аномалій – із конкретними сценаріями реагування, а результати аналізу – з історією прийнятих рішень.

Критерій зв'язаності структури $p_1^{F_{opt,1}^D}$ задамо так:

$$p_1^{F_{opt,1}^D} = \frac{Z_{1,2} - Z_{1,1}}{Z_{1,2}}, \quad (3.20)$$

де $Z_{1,1}$ – кількість встановлених структурних зв'язків; $Z_{1,2}$ – максимально можлива кількість зв'язків.

Малі значення $p_1^{F_{opt,1}^D}$ відображають зв'язаність структур даних.

Критерій групування даних $p_2^{F_{opt,1}^D}$ характеризує ефективність об'єднання подібних інформаційних об'єктів у логічні групи або кластери. Для систем протидії КА це можуть бути групи однотипних телеметричних подій, кластери аномалій, категорії сигнатур та сценарії схожих атак. Критерій визначимо так:

$$p_2^{F_{opt,1}^D} = \frac{Z_{2,2} - Z_{2,1}}{Z_{2,2}}, \quad (3.21)$$

де $Z_{2,1}$ – кількість об'єктів, що входять до структурованих груп; $Z_{2,2}$ – загальна кількість інформаційних об'єктів.

Малі значення $p_2^{F_{opt,1}^D}$ відображають ефективну структуру даних.

Критерій однорідності структури $p_3^{F_{opt,1}^D}$ оцінює ступінь узгодженості форматів, типів та структури інформаційних об'єктів. У системах та засобах протидії КА неоднорідність може виникати через використання різних сенсорів, різні формати журналів, різномірні джерела телеметрії та різні типи обманних систем. Критерій визначимо так:

$$p_3^{F_{opt,1}^D} = \frac{Z_{3,2} - Z_{3,1}}{Z_{3,2}}, \quad (3.22)$$

де $Z_{3,1}$ – кількість структурно несумісних об'єктів; $Z_{3,2}$ – загальна кількість об'єктів.

Малі значення $p_3^{F_{opt,1}^D}$ відображають узгоджену структуру даних.

Критерій унікальності інформаційних об'єктів $p_4^{F_{opt,1}^D}$ на відміну від критерію надлишковості, який оцінює ресурсне дублювання, характеризує структурну унікальність інформаційних об'єктів. Він визначає

наявність в інформаційному масиві незалежних інформаційних сутностей, а не повторення однакових структурних шаблонів.

$$p_4^{F_{opt,1}^D} = \frac{Z_{4,1}}{Z_{4,2}}, \quad (3.23)$$

де $Z_{4,1}$ – кількість унікальних інформаційних структур; $Z_{4,2}$ – загальна кількість структур.

Наприклад, однакові журнали подій мають низьку структурну унікальність, а нові сценарії аномальної поведінки – високу.

Критерій структурної адаптивності $p_5^{F_{opt,1}^D}$ характеризує здатність структури даних перебудовуватись при зміні умов функціонування системи. У системах та засобах протидії КА це проявляється через автоматичне створення нових кластерів, зміну зв'язків між даними, перебудову структури сценаріїв і адаптацію до нових типів загроз. Критерій визначимо так:

$$p_5^{F_{opt,1}^D} = \frac{Z_{5,2} - Z_{5,1}}{Z_{5,2}}, \quad (3.24)$$

де $Z_{5,1}$ – кількість успішних структурних адаптацій; $Z_{5,2}$ – кількість змін умов функціонування.

Таким чином, критерії структури даних дозволяють оцінити ефективність внутрішньої організації інформаційного середовища систем та засобів протидії КА з точки зору зв'язаності, структурованості, однорідності, унікальності та адаптивності інформаційних об'єктів. Їх використання забезпечує формування впорядкованої структури даних, здатної підтримувати процеси аналізу, виявлення загроз і прийняття рішень в умовах постійної зміни складу та характеристик інформаційних потоків без дублювання функцій інших оптимізаційних критеріїв.

Критерії шарового розміщення характеризують ефективність розподілу інформаційних об'єктів між різними рівнями підсистеми пам'яті систем та засобів протидії КА відповідно до їх важливості, частоти використання, актуальності та функціонального призначення. На відміну від базових ресурсних критеріїв, які оцінюють загальні витрати пам'яті та часу доступу, критерії шарового розміщення визначають саме правильність ієрархічного розподілу даних між рівнями зберігання. Також вони не дублюють критерії структури даних, оскільки не описують внутрішню

організацію інформаційних зв'язків, а оцінюють відповідність даних конкретним шарам пам'яті. Функції для критеріїв шарового розміщення

$F_{opt,2}^D \left(p_1^{F_{opt,2}^D}, p_2^{F_{opt,2}^D}, \dots, p_{n_{F_{opt,2}^D}}^{F_{opt,2}^D} \right)$ задамо з урахуванням формули (3.10) та виділимо

такі критерії шарового розміщення даних $F_{opt,2}^D$:

- 1) критерій пріоритетності розміщення даних $p_1^{F_{opt,2}^D}$;
- 2) критерій міграції даних між шарами $p_2^{F_{opt,2}^D}$;
- 3) критерій балансування шарів $p_3^{F_{opt,2}^D}$;
- 4) критерій релевантності шару $p_4^{F_{opt,2}^D}$;
- 5) критерій ефективності використання шарів даних $p_5^{F_{opt,2}^D}$.

Критерій пріоритетності розміщення даних $p_1^{F_{opt,2}^D}$ характеризує відповідність між важливістю інформації та рівнем шару, у якому вона розміщена. У системах та засобах протидії КА найбільш критичні дані повинні знаходитись у шарах швидкого доступу, тоді як менш актуальні – у проміжних або архівних. Задамо критерій так:

$$p_1^{F_{opt,2}^D} = \frac{1}{m_{p_1^{F_{opt,2}^D}}} \cdot \sum_{k=1}^{n_{p_1^{F_{opt,2}^D}}} \frac{|S_{1,k,1} - S_{1,k,2}|}{m_{p_1^{F_{opt,2}^D}}}, \quad (3.25)$$

де $m_{p_1^{F_{opt,2}^D}}$ – кількість шарів; $n_{p_1^{F_{opt,2}^D}}$ – кількість об'єктів даних; $k = 1, 2, \dots, n_{p_1^{F_{opt,2}^D}}$;

$S_{1,k,1}$ – фактичний шар розміщення; $S_{1,k,2}$ – рекомендований шар.

Значення $p_1^{F_{opt,2}^D}$ повинно прямувати до нуля в найкращому варіанті.

Критерій міграції даних між шарами $p_2^{F_{opt,2}^D}$ характеризує ефективність переміщення інформації між шарами пам'яті при зміні її актуальності або інтенсивності використання. У системах та засобах протидії КА це може включати перенесення застарілої телеметрії до архіву, переміщення важливих сценаріїв до оперативного шару та автоматичне кешування активних даних. Задамо його так:

$$p_2^{F_{opt,2}^D} = \frac{S_{2,1}}{S_{2,2}}, \quad (3.26)$$

де $S_{2,1}$ – кількість коректних міграцій; $S_{2,2}$ – загальна кількість міграцій.

Критерій балансування шарів $p_3^{F_{opt,2}^D}$ оцінює рівномірність завантаження шарів пам'яті. Надмірне перевантаження окремого шару призводить до деградації функціонування системи навіть за наявності вільних ресурсів інших рівнів. Критерій можна визначити з урахуванням середньоквадратичного відхилення завантаження шарів та середнє завантаження.

Критерій релевантності шару $p_4^{F_{opt,2}^D}$ характеризуватиме відповідність типу даних функціональному призначенню шару. Наприклад, оперативні телеметричні потоки повинні зберігатись у швидкодоступних шарах, історичні журнали – в архівних, а моделі самонавчання – у проміжних шарах. Критерій визначатимемо через кількість об'єктів, що розміщені у функціонально відповідних шарах та через загальну кількість об'єктів.

Критерій ефективності використання шарів даних $p_5^{F_{opt,2}^D}$ оцінює доцільність використання кожного рівня пам'яті відповідно до його функціонального призначення. Він дозволяє виявляти недовантажені шари, невикористовувані області пам'яті, надлишкове дублювання між шарами, нераціональний розподіл інформаційних потоків. Критерій визначатимемо як ефективно використаний обсяг певного шару з врахуванням загального обсягу шару.

Такий підхід дозволяє оцінити ефективність ієрархічного розподілу інформації між шарами пам'яті систем та засобів протидії КА з урахуванням пріоритетності даних, коректності їх міграції, балансування ресурсів та функціональної відповідності рівнів зберігання без дублювання базових ресурсних або структурних критеріїв.

Критерії апаратно-мережного розподілу $F_{opt,3}^D \left(p_1^{F_{opt,3}^D}, p_2^{F_{opt,3}^D}, \dots, p_{n_{F_{opt,3}^D}}^{F_{opt,3}^D} \right)$ характеризують ефективність просторового та функціонального розміщення компонентів підсистеми пам'яті систем та засобів протидії КА в межах корпоративної мережної інфраструктури. Їх призначення полягає у забезпеченні узгодженого

функціонування серверів, вузлів оброблення даних, мережних сегментів, сховищ інформації та периферійних засобів моніторингу в умовах розподіленого функціонування системи. На відміну від критеріїв ресурсної оптимізації, які оцінюють загальні витрати пам'яті та часу доступу до інформації, а також критеріїв шарового розміщення, що визначають логіку розподілу даних між рівнями пам'яті, критерії апаратно-мережного розподілу орієнтовані саме на оцінювання фізичного та мережного розташування апаратних ресурсів, каналів взаємодії та інформаційних потоків.

У системах та засобах протидії КА інформаційні ресурси можуть бути розподілені між центральними серверами, вузлами оперативного аналізу, периферійними сенсорами, мережними шлюзами, спеціалізованими вузлами обманних систем, резервними дата-центрами та віддаленими сегментами корпоративної мережі. Ефективність такого розподілу значною мірою визначає здатність системи забезпечувати оперативне оброблення даних, підтримувати стійкість функціонування та здійснювати своєчасне реагування на загрози.

Функції для критеріїв апаратно-мережного розподілу $F_{opt,3}^D \left(p_1^{F_{opt,3}^D}, p_2^{F_{opt,3}^D}, \dots, p_{n_{F_{opt,3}^D}}^{F_{opt,3}^D} \right)$ задамо з урахуванням формули (3.10) та виділимо

такі критерії апаратно-мережного розподілу $F_{opt,3}^D$:

- 1) критерій збалансованості обчислювального навантаження $p_1^{F_{opt,3}^D}$;
- 2) критерій мережної локалізації даних $p_2^{F_{opt,3}^D}$;
- 3) критерій стійкості мережної інфраструктури $p_3^{F_{opt,3}^D}$;
- 4) критерій маршрутизації інформаційних потоків $p_4^{F_{opt,3}^D}$;
- 5) критерій доступності вузлів зберігання $p_5^{F_{opt,3}^D}$;
- 6) критерій сегментації корпоративної мережі $p_6^{F_{opt,3}^D}$;
- 7) критерій масштабованості інфраструктури $p_7^{F_{opt,3}^D}$.

Одним із ключових аспектів апаратно-мережного розподілу є забезпечення збалансованості обчислювального навантаження між вузлами системи. Надмірне перевантаження окремих серверів або вузлів аналізу може призвести до збільшення часу оброблення телеметрії, накопичення черг даних, втрати частини інформаційних потоків та зниження ефективності реагування системи на зміни оперативної обстановки. Тому розподіл обчислювальних задач повинен забезпечувати рівномірне використання процесорних ресурсів, оперативної пам'яті та систем зберігання даних у межах КМ.

Важливе значення має також мережне розміщення інформаційних ресурсів відносно вузлів, які їх найчастіше використовують. У системах протидії КА це стосується передусім потоків оперативної телеметрії, журналів аномальної активності, спектральних вимірювань та результатів функціонування засобів моніторингу. Розташування таких даних у віддалених сегментах мережі може спричинити збільшення затримок передачі, перевантаження магістральних каналів зв'язку та дублювання мережного трафіку. Тому інформаційні ресурси повинні розміщуватись максимально наближено до вузлів їх інтенсивного використання.

Критерії апаратно-мережного розподілу також враховують стійкість мережної інфраструктури до відмов, перевантажень або навмисного впливу. Для систем та засобів протидії КА цей аспект є особливо важливим, оскільки втрата окремих вузлів або каналів зв'язку може суттєво обмежити можливість аналізу поточної ситуації. Забезпечення стійкості досягається шляхом резервування серверів, дублювання каналів зв'язку, використання резервних сховищ даних, географічного рознесення вузлів та реалізації механізмів динамічного перенаправлення інформаційних потоків.

Окрему роль відіграє ефективність маршрутизації інформаційних потоків між компонентами системи. У межах КМ одночасно можуть функціонувати телеметричні потоки, журнали подій, результати спектрального аналізу, службові потоки самонавчання та дані обманних систем. Нераціональна маршрутизація таких потоків здатна спричинити перевантаження окремих сегментів мережі, накопичення затримок, втрату пакетів і погіршення загальної продуктивності системи. Тому

організація мережних маршрутів повинна забезпечувати мінімізацію часу передачі даних та рівномірний розподіл навантаження між мережними вузлами.

Важливим аспектом є забезпечення доступності вузлів зберігання даних. У системах та засобах протидії КА це визначає можливість безперервного доступу до критично важливих інформаційних ресурсів, історичних сценаріїв реагування, моделей самонавчання та журналів аномальної активності. Доступність залежить від рівня резервування вузлів, стабільності мережних з'єднань, ефективності синхронізації між сховищами та здатності системи підтримувати функціонування при часткових відмовах інфраструктури.

Критерії апаратно-мережного розподілу також враховують рівень сегментації КМ. Сегментація дозволяє ізолювати окремі компоненти системи, локалізувати потенційні атаки, відокремити вузли обманних систем від критичних ресурсів та обмежити поширення загроз між мережними сегментами. Недостатня сегментація може призвести до неконтрольованого поширення атак, порушення стабільності функціонування та збільшення навантаження на інформаційну інфраструктуру.

Крім цього, важливою характеристикою є масштабованість апаратно-мережної інфраструктури. У процесі функціонування систем та засобів протидії КА постійно збільшуються обсяги телеметрії, кількість сенсорів, журнальних записів, сценаріїв реагування та даних обманних систем. Тому інфраструктура повинна забезпечувати можливість додавання нових серверів, вузлів оброблення даних, мережних сегментів та сховищ інформації без суттєвої перебудови архітектури системи.

Таким чином, критерії апаратно-мережного розподілу дозволяють оцінити ефективність фізичної та мережної організації підсистеми пам'яті систем та засобів протидії КА з точки зору збалансованості навантаження, локалізації інформаційних ресурсів, стійкості інфраструктури, ефективності маршрутизації, доступності вузлів зберігання, сегментації корпоративної мережі та можливості подальшого масштабування системи.

Критерії оптимізації даних обманних систем $F_{opt,4}^D \left(p_1^{F_{opt,4}^D}, p_2^{F_{opt,4}^D}, \dots, p_{n_{F_{opt,4}^D}}^{F_{opt,4}^D} \right)$

характеризують ефективність організації, зберігання, використання та адаптації

інформаційних ресурсів, що формуються або використовуються засобами обману в системах та засобах протидії КА. На відміну від загальних критеріїв структури даних чи апаратно-мережного розподілу, ці критерії орієнтовані саме на специфіку функціонування обманных компонентів інформаційного середовища, для яких характерними є динамічна зміна сценаріїв взаємодії, імітація реальних інформаційних ресурсів, накопичення даних про поведінку потенційного зловмисника та постійна адаптація до змін зовнішнього середовища.

У системах та засобах протидії КА обманные системи можуть формувати значні обсяги інформації, до яких належать журнали взаємодії з пастками, дані про мережну активність зловмисника, хибні сигнатури, сценарії імітації функціонування КА, моделі поведінки атакуючої сторони, результати аналізу дій зловмисника та історія попередніх сценаріїв обману. Частина цих даних має короткочасний характер і використовується лише в оперативному режимі, тоді як інша інформація накопичується для подальшого навчання системи та вдосконалення механізмів протидії.

Функції для критеріїв оптимізації даних обманных систем $F_{opt,4}^D \left(p_1^{F_{opt,4}^D}, p_2^{F_{opt,4}^D}, \dots, p_{n_{F_{opt,4}^D}}^{F_{opt,4}^D} \right)$ задамо з урахуванням формули (3.10) та виділимо

такі критерії оптимізації даних обманных систем $F_{opt,4}^D$:

- 1) критерій правдоподібності інформаційного середовища $p_1^{F_{opt,2}^D}$;
- 2) критерій адаптивності сценаріїв обману $p_2^{F_{opt,2}^D}$;
- 3) критерій інформативності даних обманных систем $p_3^{F_{opt,2}^D}$;
- 4) критерій узгодженості з реальною інфраструктурою $p_4^{F_{opt,2}^D}$;
- 5) критерій ізоляції інформаційних потоків обманных систем $p_5^{F_{opt,2}^D}$;
- 6) критерій накопичення знань про сценарії КА $p_3^{F_{opt,2}^D}$;
- 7) критерій аналітичної цінності даних обманных систем $p_4^{F_{opt,2}^D}$;
- 8) критерій ресурсної раціональності зберігання даних обманных систем $p_5^{F_{opt,2}^D}$.

Одним із ключових аспектів оптимізації даних обманних систем є забезпечення правдоподібності інформаційного середовища, яке імітується для потенційного зловмисника. Дані обманних систем повинні бути достатньо наближеними до реальної структури інформаційних ресурсів, інакше зловмисник може виявити факт використання засобів обману. Тому оптимізація спрямована не лише на зменшення обсягів даних або прискорення доступу до них, а й на підтримання відповідного рівня реалістичності інформаційного середовища.

Важливим критерієм є також адаптивність даних обманних систем до зміни сценаріїв КА і поведінки зловмисника. У процесі функціонування системи накопичуються нові дані про способи взаємодії зловмисника з пастками, особливості його поведінки, використовувані мережні маршрути та характер інформаційних запитів. На основі цих даних система повинна автоматично перебудовувати сценарії обману, змінювати структуру імітаційних ресурсів та формувати нові шаблони взаємодії.

Критерії оптимізації також враховують ступінь корисності даних обманних систем для виявлення загроз і накопичення знань про поведінку потенційного зловмисника. Значна частина інформації, що формується обманними компонентами, може бути дубльованою, випадковою або не містити нових ознак активності. Тому оптимізація повинна забезпечувати відокремлення інформаційно цінних даних від другорядних записів, які не впливають на ефективність аналізу.

Окрему роль відіграє узгодженість даних обманних систем із реальною інформаційною інфраструктурою. Структура журналів, часові характеристики подій, параметри мережної активності та сценарії функціонування імітаційних вузлів повинні відповідати характеристикам реальних компонентів КМ. Порухення цієї узгодженості може призвести до виявлення обманної системи та втрати її ефективності.

Важливим аспектом оптимізації є контроль ізоляції даних обманних систем від критичних інформаційних ресурсів. Незважаючи на необхідність правдоподібної інтеграції в корпоративне середовище, інформаційні потоки обманних компонентів повинні бути логічно та мережно відокремлені від основних функціональних

сегментів системи. Це дозволяє мінімізувати ризики поширення загроз або несанкціонованого доступу до критичних ресурсів через компоненти обманної інфраструктури.

Критерії оптимізації даних обманних систем також враховують ефективність накопичення знань про сценарії атак та можливість їх подальшого використання в процесах самонавчання. Дані про поведінку зловмисника, послідовність його дій, часові закономірності активності та реакції на різні типи пасток формують основу для побудови моделей прогнозування та адаптивного вдосконалення механізмів протидії КА.

Крім цього, оптимізація повинна враховувати співвідношення між обсягом сформованих даних та їх реальною аналітичною цінністю. Надмірне накопичення журналів взаємодії або дубльованих сценаріїв обману може призвести до збільшення навантаження на підсистему пам'яті та зниження швидкості аналізу. Тому важливим завданням є автоматичне виділення найбільш інформативних фрагментів даних, узагальнення подібних сценаріїв та перенесення застарілої інформації до архівних шарів пам'яті.

Таким чином, критерії оптимізації даних обманних систем дозволяють оцінити ефективність функціонування обманних компонентів систем та засобів протидії КА з точки зору правдоподібності інформаційного середовища, адаптивності сценаріїв обману, інформативності накопичених даних, узгодженості з реальною інфраструктурою, ізоляції інформаційних потоків, ефективності накопичення знань про дії зловмисника та раціональності використання ресурсів підсистеми пам'яті.

Згідно введеної функції оптимізації за формулою (3.9) задамо основні кроки методу оптимізації даних на основі самонавчання засобів підсистеми пам'яті

Крок 1. Збір та первинний аналіз накопичених даних.

На цьому етапі здійснюється оцінювання структури інформаційного масиву, визначення частоти звернення до окремих даних, встановлення рівня їх актуальності, а також виявлення дубльованих, подібних або застарілих інформаційних об'єктів. Результатом цього кроку є формування початкового уявлення про стан підсистеми пам'яті та виявлення потенційних напрямів її оптимізації.

Крок 2. Побудова моделі організації даних у підсистемі пам'яті.

На цьому етапі встановлюються взаємозв'язки між окремими інформаційними об'єктами, визначаються закономірності їх використання та оцінюється їх значущість для забезпечення функціонування системи.

Крок 3. Формування критеріїв оптимізації.

На основі результатів попереднього аналізу визначаються показники, відповідно до яких буде оцінюватися ефективність подальших перетворень даних. Цільова функція оптимізації задана за формулою (3.9) та критеріями за формулою (3.10).

Крок 4. Вибір способів перетворення даних.

На цьому етапі система визначає операції, які доцільно застосувати до інформаційного масиву. Для кластеризації використовується критерій мінімізації внутрішньокластерної відстані.

Крок 5. Безпосередня реорганізація підсистеми пам'яті.

На цьому етапі здійснюється реалізація обраних процедур оптимізації згідно функцій, якими визначені критерії оптимізації (задані формулою (3.10)). Ефективність ущільнення даних можна оцінити коефіцієнтом компресії.

Крок 6. Оцінювання результатів проведеної оптимізації.

На цьому етапі виконується порівняння параметрів функціонування підсистеми до і після перетворень за формулою (3.9).

Крок 7. Адаптація алгоритмів на основі самонавчання.

Підсистема накопичує статистику щодо результатів виконаних дій, аналізує ефективність застосованих рішень та формує оновлені правила подальшого опрацювання інформації. Оновлення параметрів виконується за формулою (3.9). Цей крок забезпечує перехід від фіксованого алгоритмічного підходу до адаптивного механізму, здатного самостійно пристосовуватися до змін у структурі даних і умов функціонування систем та засобів протидії КА на основі самонавчання.

Таким чином, метод оптимізації даних на основі самонавчання засобів підсистеми пам'яті забезпечує комплексне розв'язання задачі організації, перетворення та адаптивного управління інформаційними ресурсами систем та

засобів протидії КА в умовах постійного накопичення різномірних даних. На відміну від традиційних детермінованих підходів, запропонований метод враховує не лише ресурсні характеристики підсистеми пам'яті, а й структурні особливості даних, принципи їх шарового розміщення, апаратно-мережний розподіл та специфіку функціонування обманних систем. Реалізація методу базується на послідовному виконанні етапів збору та аналізу накопичених даних, побудови моделі організації підсистеми пам'яті, формування критеріїв оптимізації, вибору способів перетворення інформаційних масивів, безпосередньої реорганізації структури пам'яті, оцінювання результатів оптимізації та подальшої адаптації алгоритмів на основі самонавчання. Такий підхід дозволяє забезпечити зменшення надлишковості даних, підвищення швидкодії доступу до інформації, раціональний розподіл інформаційних ресурсів між шарами пам'яті та мережними вузлами, а також підтримання високого рівня інформативності даних для задач протидії КА.

Особливістю методу є використання адаптивного механізму самонавчання, який забезпечує накопичення статистики функціонування системи, аналіз ефективності прийнятих рішень та автоматичне коригування параметрів оптимізації відповідно до змін структури інформаційних потоків і умов функціонування систем та засобів протидії КА. Це дозволяє забезпечити перехід від статичного способу організації пам'яті до інтелектуального механізму адаптивного управління даними.

Метод створює основу для побудови самонавчальних підсистем пам'яті систем та засобів протидії КА, здатних функціонувати в умовах динамічного зростання обсягів інформації, зміни характеру загроз та ускладнення структури інформаційного середовища.

3.3 Метод вибору даних з урахуванням глибини горизонту, самонавчання та попереднього досвіду використання

Використання глибини горизонту, механізмів самонавчання та попереднього досвіду при здійсненні вибору даних є необхідним через високу динамічність процесів в середовищі КМ, в якому ефективність функціонування систем значною

мірою залежить не лише від обсягу доступної інформації, а й від здатності правильно визначати її актуальність, практичну цінність та перспективність використання. Орієнтація лише на поточний стан даних не забезпечує достатньої стійкості прийняття рішень, оскільки інформація швидко втрачає актуальність, а умови середовища можуть змінюватися швидше, ніж система встигає адаптуватися до них. Саме тому виникає необхідність врахування глибини горизонту, яка дозволяє оцінювати не тільки поточну корисність інформації, а й її потенційну значущість у майбутніх сценаріях функціонування системи. Такий підхід забезпечує можливість прогнозування ефективності використання даних у короткостроковій, середньостроковій та довгостроковій перспективах, що особливо важливо для адаптивних та інтелектуальних систем.

Водночас орієнтація виключно на прогнозований горизонт не є достатньою, оскільки система повинна враховувати накопичений досвід попереднього використання даних. Попередній досвід дозволяє оцінювати реальну ефективність інформації на основі історичних результатів її застосування. Аналіз успішних та неуспішних сценаріїв дає змогу виявляти закономірності, визначати стабільно корисні джерела інформації та мінімізувати повторення помилкових рішень. Використання попереднього досвіду підвищує надійність відбору даних, оскільки система отримує можливість опиратися не лише на теоретичну оцінку релевантності, а й на фактичні результати функціонування в аналогічних умовах.

Однак навіть поєднання горизонту аналізу та попереднього досвіду не забезпечує достатньої гнучкості без механізмів самонавчання. Самонавчання необхідне для постійного коригування критеріїв оцінювання, адаптації вагових коефіцієнтів та оновлення моделей вибору даних відповідно до змін середовища. У реальних умовах характеристики інформації, поведінка об'єктів та параметри інформаційних потоків можуть змінюватися настільки швидко, що накопичений досвід частково втрачає актуальність. Самонавчання дозволяє системі виявляти нові закономірності, автоматично враховувати результати власної роботи та підвищувати точність подальших рішень без необхідності повного ручного переналаштування системи.

Поєднання глибини горизонту, попереднього досвіду та самонавчання є не лише можливим, а й доцільним, оскільки ці компоненти взаємно доповнюють один одного. Глибина горизонту забезпечує стратегічне бачення та орієнтацію на майбутні умови, попередній досвід формує основу для оцінювання практичної ефективності, а самонавчання забезпечує адаптивність та здатність до безперервного вдосконалення. Разом вони формують багаторівневий механізм вибору даних, який поєднує прогнозування, накопичення знань та динамічну адаптацію.

Разом з тим між цими компонентами можуть виникати певні протиріччя. Одним із основних є конфлікт між історичним досвідом та новими умовами функціонування. Дані, які були ефективними в минулому, можуть втратити актуальність у зміненому середовищі КМ, тоді як система, надмірно орієнтована на попередній досвід, може продовжувати надавати їм високий пріоритет. Іншим протиріччям є суперечність між довгостроковим горизонтом прогнозування та необхідністю швидкої адаптації до поточних змін. Орієнтація на довгострокову перспективу може знижувати чутливість системи до оперативних змін середовища, тоді як надмірна адаптивність самонавчання здатна призвести до нестабільності моделей та втрати стратегічної узгодженості. Також можливе протиріччя між стабільністю накопичених знань і постійним оновленням параметрів навчання, оскільки надто агресивне самонавчання може призводити до деградації корисного історичного досвіду.

Незважаючи на зазначені суперечності, їх можна мінімізувати за рахунок збалансованого механізму інтеграції компонентів. Використання адаптивних вагових коефіцієнтів, багаторівневого оцінювання даних та механізмів контролю стабільності навчання дозволяє забезпечити компроміс між стратегічною стійкістю та оперативною адаптивністю. У результаті поєднання глибини горизонту, самонавчання та попереднього досвіду дозволяє досягти підвищення точності вибору даних, зменшення інформаційної надлишковості, покращення адаптації до динамічного середовища, прискорення прийняття рішень, зниження кількості помилкових рішень та накопичення ефективної бази знань для подальшого функціонування системи. Таке поєднання характерних властивостей в методі забезпечуватиме формування інтелектуального механізму відбору даних, здатного ефективно функціонувати в

умовах невизначеності, зміни інформаційних потоків та постійної еволюції середовища.

Основні кроки методу вибору даних з урахуванням глибини горизонту, самонавчання та попереднього досвіду використання.

Крок 1. Формування множини вхідних даних.

На цьому етапі система збирає всі доступні дані з різних джерел і формує єдиний інформаційний простір згідно даних з формули (2.2).

Крок 2. Визначення стратегії вибору даних та формування оптимізаційної моделі.

На цьому етапі система формує єдиний механізм прийняття рішень, який перетворює множину різнорідних даних у впорядковану систему оцінювання та вибору. Стратегія визначає не лише критерії важливості даних, але й спосіб їхнього порівняння, нормалізації та оптимізації в умовах обмежених ресурсів.

2.1. Формування формального опису вхідного інформаційного простору даних

Задамо кожен елемент даних з формули (2.2) вектором ознак так:

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

$$V_{DI,i}^{P,D} = (v_{DI,i,1}^{P,D}, v_{DI,i,2}^{P,D}, v_{DI,i,3}^{P,D}, v_{DI,i,4}^{P,D}, v_{DI,i,5}^{P,D}, v_{DI,i,6}^{P,D}, v_{DI,i,7}^{P,D}), \quad (3.27)$$

де $F_{DI}^{P,D}$ – функція формування векторів вхідного інформаційного простору даних;
 $V_{DI,i}^{P,D}$ – вектор вхідного інформаційного простору даних; $v_{DI,i,1}^{P,D}$ – актуальність (горизонт); $v_{DI,i,2}^{P,D}$ – ефективність попереднього використання; $v_{DI,i,3}^{P,D}$ – інформативність; $v_{DI,i,4}^{P,D}$ – надлишковість; $v_{DI,i,5}^{P,D}$ – вартість обробки; $v_{DI,i,6}^{P,D}$ – часові характеристики; $v_{DI,i,7}^{P,D}$ – контекстна відповідність; $M_{DI}^D = \{V_{DI,1}^D, V_{DI,2}^D, \dots, V_{DI,N_{M_{DI}^D}}^D\}$;
 $N_{M_{DI}^D}$ – кількість векторів для забезпечення зберігання даних від джерел інформації різних рівнів; $V_{DI,i}^D$ – i -тий вектор; $i = 1, 2, \dots, N_{M_{DI}^D}$.

2.2. Нормалізація та підготовка критеріїв.

Оскільки всі характеристики мають різну природу, то їх потрібно привести до єдиної шкали: уніфікація діапазону значень значень; усунення розмірностей; згладжування шуму; фільтрація аномалій.

2.3. Формування стратегії оптимізації на основі оптимізаційної функції.

Суть цього етапу полягає в тому, що система формує формалізовану стратегію поведінки, яка одночасно враховує корисність, ризики, ресурси та досвід попереднього використання. Базова стратегія визначає баланс між корисністю та витратами.

$$F_{DI}^{O,D}(V_{DI,i}^{P,D}) = \sum_{i_1=1}^{n_{i_1}} a_{DI,i,i_1}^{O,D} \cdot A_{DI,i,i_1}^{O,D} - \sum_{i_2=1}^{n_{i_2}} b_{DI,i,i_2}^{O,D} \cdot B_{DI,i,i_2}^{O,D}, \quad (3.28)$$

де $F_{DI}^{O,D}$ – оптимізаційна функція для формування стратегій вибору; $a_{DI,i,i_1}^{O,D}$ – коефіцієнти для значень локальних функцій, які накопичують результат; $b_{DI,i,i_2}^{O,D}$ – коефіцієнти для значень локальних функцій, які зменшують результат; $i_1 = 1, 2, \dots, n_{i_1}$; $i_2 = 1, 2, \dots, n_{i_2}$; n_{i_1} – кількість локальних функцій, які накопичують результат; n_{i_2} – кількість локальних функцій, які зменшують результат; $A_{DI,i,i_1}^{O,D}$ – значення локальних функцій, які накопичують результат; $B_{DI,i,i_2}^{O,D}$ – значення локальних функцій, які зменшують результат; $V_{DI,i}^{P,D}$ – вектор вхідного інформаційного простору даних (формула (3.27)).

Корисність в формулі (3.28) задана додатніми значеннями суми, а витрати – від’ємними. Деталізуємо функції в формулі (3.28).

Модель актуальності (горизонтальна динаміка) залежить від часу та сценарію використання і задамо її так:

$$A_{DI,i,1}^{O,D} = e^{-\mu \cdot t_i} \cdot H_{DI,i,1}^{O,D}(t_i), \quad (3.29)$$

де t_i – тривалість перебування в системі сформованих даних; μ – швидкість старіння інформації; $H_{DI,i,1}^{O,D}$ – функція горизонту (оперативний / тактичний / стратегічний режим).

Функція $H_{DI,i,1}^{O,D}$ приймає, наприклад, три значення: {1}; {2}; {3}.

Дані швидко втрачають вагу, якщо вони короткострокові, але можуть залишатися важливими у стратегічному горизонті.

Модель досвіду (підкріплювальна оцінка) базується на ймовірності успіху в аналогічних умовах. Задамо її функцією так:

$$A_{DI,i,2}^{O,D} = P_{DI,2}^{O,D}(V_{DI,i}^{P,D}, S_{DI,i}^{P,D}), \quad (3.30)$$

де $P_{DI,2}^{O,D}$ - ймовірність успіху в аналогічних умовах; $V_{DI,i}^{P,D}$ – вектор вхідного інформаційного простору даних (формула (3.27)); $S_{DI,i}^{P,D}$ – стан середовища КМ.

Досвід передбачає, що дані зберігають інформацію про те, як вони працювали в подібних ситуаціях, і, тоді, система адаптується до контексту, а не тільки до історії.

Модель інформативності забезпечує зменшення невизначеності ентропії системи і задамо її функцією $A_{DI,i,3}^{O,D}$ так:

$$A_{DI,i,3}^{O,D} = P_{DI,3}^{O,D}(V_{DI,i}^{P,D}, S_{DI,i}^{P,D}) - P_{DI,3}^{O,D}(V_{DI,i}^{P,D}, S_{DI,i,p}^{P,D}), \quad (3.31)$$

де $P_{DI,3}^{O,D}$ - ймовірність визначення стану системи; $V_{DI,i}^{P,D}$ – вектор вхідного інформаційного простору даних (формула (3.27)); $S_{DI,i}^{P,D}$ – стан середовища КМ; $S_{DI,i,p}^{P,D}$ – стан середовища КМ після події, тобто після стану $S_{DI,i}^{P,D}$.

Ця модель дає змогу встановити те, що чим більше дані прояснюють ситуацію в системі, то тим їх цінність краща.

2.4. Модель надлишковості (структурна залежність)

Надлишковість визначається як залежність від уже наявних даних:

$$B_{DI,i,1}^{O,D} = \sum_{r=1, r < i}^{N_{M_{DI}^D}} P_{DI,4}^{O,D}(V_{DI,r}^{P,D}, V_{DI,i}^{P,D}), \quad (3.32)$$

де $V_{DI,i}^{P,D}$ – вектор вхідного інформаційного простору даних (формула (3.27)); $M_{DI}^D = \{V_{DI,1}^D, V_{DI,2}^D, \dots, V_{DI,N_{M_{DI}^D}}^D\}$; $N_{M_{DI}^D}$ – кількість векторів для забезпечення зберігання даних від джерел інформації різних рівнів; $V_{DI,i}^D$ - i -тий вектор; $i = 1, 2, \dots, N_{M_{DI}^D}$; $P_{DI,4}^{O,D}$ – кореляція або схожість між даними двох різних вхідних векторів $V_{DI,r}^{P,D}, V_{DI,i}^{P,D}$.

Таким чином, якщо дані повторюють інші, то їх цінність зменшується, бо система прагне різноманітності інформації.

Модель вартості, тобто ресурсну функцію $B_{DI,i,2}^{O,D}$ задамо так:

$$B_{DI,i,2}^{O,D} = \sum_{r=1}^3 b_{DI,5,r}^{O,D} \cdot B_{DI,5,r}^{O,D}(V_{DI,i}^{P,D}), \quad (3.33)$$

де $V_{DI,i}^{P,D}$ – вектор вхідного інформаційного простору даних (формула (3.27)); $B_{DI,5,1}^{O,D}$ – час обробки; $B_{DI,5,2}^{O,D}$ – обчислювальна складність; $B_{DI,5,3}^{O,D}$ – пам'ять/ресурси; $r = 1,2,3$; $b_{DI,5,r}^{O,D}$ – ваги ресурсів.

Цінні дані знижують пріоритет, бо система здійснює більші витрати за їх використання.

Таким чином, визначення стратегії вибору даних буде здійснено за рахунок формування оптимізаційної моделі для того, щоб система мала числові значення для порівняння даних і могла визначати їх пріоритет.

Крок 3. Аналіз глибини горизонту.

На цьому етапі система визначає корисність даних з урахуванням часу та сценарію використання. Однакові дані можуть мати різну цінність залежно від горизонту застосування. Інформація, яка критично важлива для оперативного реагування, може бути малокорисною у стратегічному аналізі, і навпаки. Тому система повинна не просто оцінити дані, а визначити час корисності, значущість та можливість використання їх в поточному сценарії.

Аналіз глибини горизонту використання даних система оцінює в контексті корисності в майбутньому, зокрема з використанням формул (3.4), (3.5) і (3.29). Для цього здійснюємо аналіз часових характеристик даних, час виникнення, швидкість старіння, частоту оновлення та тривалість життєвого циклу. Додатково оцінюємо динаміку зміни актуальності, тобто темп втрати цінності інформації залежно від умов середовища та часу. Система також визначає відповідність даних різним горизонтам використання, тобто оперативному, тактичному або стратегічному, що дозволяє адаптувати вибір інформації до типу задачі. Окремо аналізуємо сценарій функціонування середовища, зокрема стабільність умов, наявність критичних подій та ймовірність зміни контексту. Для підвищення точності враховується попередній досвід використання аналогічних даних, їх ефективність у схожих ситуаціях та прогнозована майбутня корисність.

Крок 4. Аналіз попереднього досвіду використання

Система звертається до історії застосування даних і аналізує корисність даних при попередніх використаннях. Використаємо результати кроку 15 «Формування попереднього досвіду використаних даних» з методу організації зберігання даних в системах та засобах протидії комп'ютерним атакам з урахуванням історичного аспекту їх використання. В ньому використано методи і технології, які базуються на моделях марковській (формула (3.8)), баєсівській та статистичних моделях накопичення досвіду. Для цього аналізуємо історію використання інформації, частоту її залучення до процесу прийняття рішень, рівень успішності отриманих результатів та вплив на точність функціонування системи. Оцінюємо досягнення цільових показників систем та засобів протидії КА, зменшення невизначеності та підвищенню якості прогнозування з певними даними. Додатково враховуватимемо помилки, що виникали під час використання даних, рівень хибних спрацьовувань, стабільність результатів у різних сценаріях та здатність даних адаптуватися до змін середовища. Система також аналізує повторюваність успішних випадків застосування інформації, тривалість її практичної корисності та залежність ефективності від контексту використання. На основі накопиченої статистики формуємо оцінку надійності та перспективності даних, що дозволяє визначати ступінь довіри до них у майбутніх циклах використання.

Крок 5. Оцінювання інформативності даних

Система визначає змістовну цінність даних з урахуванням їх здатності впливати на якість прийняття рішень, формування нових залежностей та уточнення поточного стану середовища. На відміну від попередніх етапів, де оцінювались часові характеристики та досвід використання, тут основна увага приділяється структурним і семантичним властивостям інформації. Система аналізує рівень унікальності даних, ступінь їх зв'язку з іншими елементами інформаційного простору, наявність прихованих закономірностей, а також внесок у покращення результатів прогнозування, класифікації або ідентифікації станів системи.

Задамо функцію оцінювання інформативності даних так:

$$F_{DI,O}^{O,D}(V_{DI,i}^{P,D}) = \sum_{r=1}^3 u_{DI,r}^{O,D} \cdot C_{DI,r}^{O,D}(V_{DI,i}^{P,D}), \quad (3.34)$$

де $F_{DI,0}^{O,D}$ – функція оцінювання інформативності даних; $u_{DI,r}^{O,D}$ – коефіцієнти для значень локальних функцій $C_{DI,r}^{O,D}$; $C_{DI,r}^{O,D}$ – значення локальних функцій; $V_{DI,i}^{P,D}$ – вектор вхідного інформаційного простору даних (формула (3.27)).

Локальні функції задамо з урахуванням таких властивостей:

- 1) $C_{DI,1}^{O,D}$ – функція для визначення семантичної значущості;
- 2) $C_{DI,2}^{O,D}$ – функція для визначення структурного внеску;
- 3) $C_{DI,3}^{O,D}$ – функція для визначення приросту ефективності.

У результаті система визначає рівень змістовної корисності даних, тобто здатність формувати нові знання та покращувати функціонування системи, що дозволить підвищити обґрунтованість подальшого вибору інформації.

Деталізуємо кожну з локальних функцій.

Локальну функцію задамо на основі використання косинусної подібності. Косинусна подібність у векторному просторі дає змогу виміряти схожість двох об'єктів, але не через порівняння самих слів або символів, а через порівняння їхніх числових семантичних представлень. Кожен текст, фрагмент даних або контекст перетворюється спеціальною моделлю у багатовимірний вектор. Такий вектор несе приховану семантичну інформацію. Близькі за змістом об'єкти отримують близькі напрямки у просторі, навіть якщо вони формально різні за словами. Після цього порівнюється не довжина векторів і не їхні абсолютні значення, а кут між ними. Саме цей кут і відображає семантичну схожість. Якщо два вектори спрямовані майже в один бік, то це означає, що відповідні дані описують схожі поняття або ситуації, навіть якщо формулювання різні. Якщо ж вектори перпендикулярні, то це означає відсутність змістовного зв'язку. А якщо вони спрямовані у протилежні боки, це може означати концептуальну протилежність або несумісність змісту. Таким чином, косинусна подібність вимірює не синтаксичну схожість, а семантичну близькість у прихованому змістовому просторі, де кожна координата відображає певну абстрактну ознаку значення, вивчену моделлю автоматично з великих обсягів даних. Такий підхід дає змогу розпізнавати подібність між даними навіть тоді, коли вони не мають спільних слів, але описують одну й ту саму ситуацію в КМ. Тоді, задамо функцію так:

$$C_{DI,1}^{O,D}(V_{DI,i}^{P,D}) = \frac{\overrightarrow{V_{DI,i}^{P,D}} \cdot \overrightarrow{V_{DI,j}^{P,D}}}{|\overrightarrow{K_{DI,i}^{P,D}}| \cdot |\overrightarrow{K_{DI,j}^{P,D}}|}, \quad (3.35)$$

де $V_{DI,i}^{P,D}$ – вектор вхідного інформаційного простору даних (формула (3.27)); $K_{DI,i}^{P,D}$ – векторне подання поточного контексту; $|\overrightarrow{V_{DI,i}^{P,D}}|$ – довжина (норма) вектора даних $V_{DI,i}^{P,D}$; $|\overrightarrow{K_{DI,i}^{P,D}}|$ – довжина (норма) вектора контексту $K_{DI,i}^{P,D}$; $\overrightarrow{V_{DI,i}^{P,D}} \cdot \overrightarrow{V_{DI,j}^{P,D}}$ – скалярний добуток векторів; $\overrightarrow{V_{DI,j}^{P,D}}$ – вектор вхідного інформаційного простору даних (формула (3.27)), з яким порівнюється i – тий вектор.

Згідно отриманих значень з формули (2.35) можна встановити наступне:

- 1) якщо значення функції $C_{DI,1}^{O,D}(V_{DI,i}^{P,D})$ близьке до нуля, то семантичний зв'язок майже відсутній у даних $V_{DI,i}^{P,D}$ та $V_{DI,j}^{P,D}$;
- 2) якщо значення функції $C_{DI,1}^{O,D}(V_{DI,i}^{P,D})$ менше нуля, то наявний протилежний зміст у даних $V_{DI,i}^{P,D}$ та $V_{DI,j}^{P,D}$;
- 3) якщо значення функції $C_{DI,1}^{O,D}(V_{DI,i}^{P,D})$ близьке до одиниці, то дані $V_{DI,i}^{P,D}$ та $V_{DI,j}^{P,D}$ максимально близькі до контексту.

Таким чином, косинусна подібність дозволяє оцінювати не буквальну схожість даних, а ступінь їх змістовної відповідності поточному контексту системи.

Функцію $C_{DI,2}^{O,D}$ для визначення структурного внеску даних $V_{DI,i}^{P,D}$ у побудову інформаційних залежностей задамо так:

$$C_{DI,2}^{O,D}(V_{DI,i}^{P,D}) = \sum_{k=1}^{n_{C_{DI,2}^{O,D}}} u_{DI,2,k}^{O,D} \cdot C_{DI,2,k}^{O,D}(V_{DI,i}^{P,D}), \quad (3.36)$$

де $V_{DI,i}^{P,D}$ – вектор вхідного інформаційного простору даних (формула (3.27)); $n_{C_{DI,2}^{O,D}}$ – кількість ознак; $u_{DI,2,k}^{O,D}$ – ваги ознак; $C_{DI,2,k}^{O,D}(V_{DI,i}^{P,D})$ – значення ознак; $k = 1, 2, \dots, n_{C_{DI,2}^{O,D}}$.

Функцію $C_{DI,3}^{O,D}$ для визначення приросту ефективності задамо так:

$$C_{DI,3}^{O,D}(V_{DI,i}^{P,D}) = C_{DI,3,n}^{O,D}(V_{DI,i}^{P,D}) - C_{DI,3,p}^{O,D}(V_{DI,i}^{P,D}), \quad (3.37)$$

де $V_{DI,i}^{P,D}$ – вектор вхідного інформаційного простору даних (формула (3.27)); $C_{DI,3,p}^{O,D}(V_{DI,i}^{P,D})$ – значення якості системи без використання даних; $C_{DI,3,n}^{O,D}(V_{DI,i}^{P,D})$ – значення якості системи після інтеграції даних.

Крок 6. Зменшення показників вибору (редукція простору даних).

На цьому етапі система виконує скорочення надлишкової інформації та зменшення розмірності простору даних з метою підвищення ефективності подальшого вибору, зменшення обчислювальної складності та усунення інформаційного шуму. Не всі ознаки, параметри або елементи даних однаково впливають на результат прийняття рішень. Частина інформації може дублюватися, мати слабкий вплив або створювати надлишкові залежності, що ускладнює аналіз і погіршує якість оптимізації. Тому система аналізує структуру даних, рівень кореляції між ознаками, ступінь повторюваності інформації та внесок кожного параметра у формування результату. На основі цього виконується відбір найбільш інформативних компонентів і усунення другорядних або залежних ознак.

Для зменшення показників вибору застосуємо схему пошуку аномалій даних в класах. При цьому обов'язковим є групування близьких значень для зменшення в подальшому на наступних етапах кількості вхідних даних. Отримана схема з профілями відповідає випадку багатомірного аналізу, де розглядаються декілька об'єктів за багатьма ознакам. При опрацюванні статичних даних спостережень в багатомірному аналізі застосуємо таксономічні методи, які не потребують експертних оцінок, а базуються тільки на результатах спостережень. Спочатку для дослідження будуюмо матрицю спостережень так:

$$X = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{1k} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2k} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ik} & \dots & x_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{s1} & x_{s2} & \dots & x_{sk} & \dots & x_{sn} \end{bmatrix}, \quad (3.38)$$

де n – кількість ознак, що спостерігаються на об'єктах; s – кількість об'єктів; x_{ik} – кількість проявів k – того признаку у i – того об'єкту за період спостережень.

Для нормалізації ознак здійснимо перетворення так:

$$V_{ik} = \frac{x_{ik}}{\sum_{i=1}^s x_{ik}}, \quad (3.39)$$

і здійснимо перехід до матриці V так:

$$V = \begin{bmatrix} V_{11} & V_{21} & \dots & V_{1k} & \dots & V_{1n} \\ V_{21} & V_{22} & \dots & V_{2k} & \dots & V_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ V_{i1} & V_{i2} & \dots & V_{ik} & \dots & V_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ V_{s1} & V_{s2} & \dots & V_{sk} & \dots & V_{sn} \end{bmatrix}, \quad (3.40)$$

Тоді можна визначити ізотонічний показник W , який буде відображати позицію S – го об'єкту за сукупністю усіх ознак:

$$W = \begin{bmatrix} W_1 \\ W_2 \\ \dots \\ W_i \\ \dots \\ W_s \end{bmatrix}, \quad (3.41)$$

де $W_i = \sum_{i=1}^n V_{ik}$.

Згідно з матрицею (3.40) упорядкуємо об'єкти за ізотонічним показником W_i , тобто за рангом, який характеризує об'єкт за сукупністю ознак. Визначаємо групи об'єктів за ізотонічним показником. Далі проводимо структурне (ізоморфне) упорядкування об'єктів. Для цього з матриці X формуємо матрицю Z за формулами (3.39) та (3.40) так:

$$Z_{ik} = \frac{x_{ik}}{\sum_{i=1}^s x_{ik}} / \sum_{i=1}^n \frac{x_{ik}}{\sum_{i=1}^s x_{ik}}, \quad (3.42)$$

$$Z = \begin{bmatrix} Z_{11} & Z_{12} & \dots & Z_{1k} & \dots & Z_{1n} \\ Z_{21} & Z_{22} & \dots & Z_{2k} & \dots & Z_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Z_{i1} & Z_{i2} & \dots & Z_{ik} & \dots & Z_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ Z_{s1} & Z_{s2} & \dots & Z_{sk} & \dots & Z_{sn} \end{bmatrix}, \quad (3.43)$$

Далі будуємо матрицю відстані в D з використанням формул (3.42) та (3.43) так:

$$d_{im} = \left\{ \frac{1}{N} \sum_{k=1}^n (Z_{ik} - Z_{mk})^2 \right\}^{\frac{1}{2}}, \quad (3.44)$$

Можна також використати середню абсолютну різницю значень ознак, яку визначимо так:

$$d_{im} = \frac{1}{N} \sum_{k=1}^n |Z_{ik} - Z_{mk}|, \quad (3.45)$$

$$D = \begin{bmatrix} 0 & d_{12} & \dots & d_{1k} & \dots & d_{1s} \\ d_{21} & 0 & \dots & d_{2k} & \dots & d_{2s} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ d_{i1} & d_{i2} & \dots & 0 & \dots & d_{is} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ d_{s1} & d_{s2} & \dots & d_{sk} & \dots & 0 \end{bmatrix}. \quad (3.46)$$

Тобто отримана квадратна матриця $S \times S$, де кількість рядків та кількість стовпців дорівнює кількості об'єктів. В цій матриці можна буде виділити підмножини на однорідність. Для цього застосуємо, наприклад, метод Чекановського.

Числові значення в матриці відстанів заміняємо позначками, наприклад, так:

- × - відстань $B < 0,4$;
- - відстань $0,4 < B < 0,8$;
- - відстань $B > 0,8$.

В результаті буде отримана невпорядкована діаграма Чекановського. Для отримання впорядкованої діаграми рядки та стовпці переставляються таким чином, щоб ці позначки були якомога ближче до головної діагоналі. Наприклад, нехай для матриці 8×8 отримана така діаграма (табл. 3.1).

Таблиця 3.1 - Упорядкована діаграма Чекановського

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

З діаграми встановлено об'єкти 1, 2, 6, 7, які згруповані і мають найменше віддалення (групою) від головної діагоналі. Найбільш віддалені об'єкти відносяться до номерів 4, 8.

Таким чином, можна здійснити поділ на класи а, також, в подальшому вже в класах здійснювати аналіз відхилень для пошуку аномалій.

Крок 7. Інтегральне ранжування даних.

Система виконує узагальнену оцінку кожного елемента даних шляхом інтеграції результатів попередніх етапів аналізу. Здійснюється формування єдиного інтегрального показника, який відображає загальну корисність даних з урахуванням часової актуальності, попереднього досвіду використання, семантичної значущості, ресурсних витрат та результатів редукції простору даних. Таким чином, система переходить від аналізу окремих характеристик до комплексного порівняння всіх елементів інформаційного простору.

Результатом повинна бути пріоритетна множина даних, яка найбільш повно відповідає поточному контексту, цілям системи та умовам функціонування.

Крок 8. Оптимальний вибір даних

Система виконує остаточний вибір підмножини даних, які мають найбільшу інтегральну цінність для поточного процесу прийняття рішень. Основна мета полягає у формуванні компактного набору інформації, що забезпечує максимальну корисність за мінімальної надлишковості та допустимих ресурсних витрат. Після інтегрального ранжування система переходить від оцінювання окремих елементів до задачі оптимального відбору множини даних.

Крок 9. Використання вибраних даних

Відібрана множина даних використовується для виконання цільових задач системи, зокрема прогнозування, аналізу ситуації, класифікації станів, виявлення закономірностей або підтримки прийняття рішень. Основна мета полягає у практичному застосуванні оптимально сформованого набору інформації для отримання результату з максимальною точністю, стійкістю та адаптивністю до умов середовища КМ.

Крок 10. Самонавчання системи

Самонавчання системи здійснюватимемо щодо вибору та оптимізації інтегрованих сукупностей даних. Після використання сформованої множини інформації система аналізує ефективність отриманого результату, оцінює внесок

окремих даних і їх комбінацій у якість прогнозування, аналізу або прийняття рішень та накопичує досвід попереднього функціонування. На основі цього коригуються ваги критеріїв, параметри оптимізаційної функції та структура вибору даних. Система поступово визначає, які поєднання інформації забезпечують найкращі результати, які дані є надлишковими, а які є найбільш стійкими до змін середовища. Це дозволяє адаптувати процес вибору до нових сценаріїв, підвищувати точність рішень, зменшувати вплив шуму та формувати оптимальні інтегровані набори даних для наступних циклів роботи системи.

Крок 11. Оновлення бази знань

Оновлення бази знань системи здійснюватимемо шляхом накопичення інформації про результати вибору даних, ефективність інтегрованих сукупностей та особливості функціонування в різних сценаріях середовища. Після завершення циклу роботи система зберігає оцінки якості рішень, параметри оптимізаційної функції, успішні комбінації даних, рівень їх інформативності, актуальності та семантичної значущості. Додатково фіксуємо помилки, надлишкові структури та зміни поведінки середовища. На основі накопиченого досвіду формуємо базу знань, яка буде використовуватись для подальшого самонавчання, адаптації критеріїв вибору та прогнозування ефективності майбутніх рішень. Це дозволяє системі поступово накопичувати структуровані знання про ефективні стратегії роботи з даними, скорочувати час адаптації до нових умов та підвищувати точність оптимального вибору інформації в наступних циклах функціонування.

Розроблений метод вибору даних з урахуванням глибини горизонту, самонавчання та попереднього досвіду використання реалізує комплексний підхід до формування оптимальної інформаційної сукупності для виконання цільових задач системи. Суть методу полягає у поєднанні часової актуальності даних, їх семантичної значущості, накопиченого досвіду використання та адаптивної оптимізації процесу вибору. Метод враховує, що ефективність функціонування системи залежить не лише від окремих елементів інформації, а й від структури інтегрованих сукупностей даних, їх взаємозв'язків та здатності забезпечувати найбільш корисний результат у конкретному сценарії середовища.

3.4 Висновки до третього розділу

Розроблено комплекс взаємопов'язаних методів організації, оптимізації та вибору даних у системах і засобах протидії КА, які забезпечують перехід від статичного керування інформацією до адаптивного, історично керованого та самонавчального підходу. Метод організації зберігання даних базується на їх поділі на початкові, накопичувані та використані з одночасним урахуванням історії застосування через збереження оцінок ефективності на основі марковських моделей, динамічних баєсівських мереж і статистичних методів, що дозволило формалізувати динаміку станів системи, врахувати причинно-наслідкові зв'язки та забезпечити накопичення досвіду. Метод оптимізації даних на основі самонавчання підсистеми пам'яті забезпечує комплексну адаптивну оптимізацію різних типів даних з урахуванням ресурсної ефективності, структурної організації, шарового розміщення та мережно-апаратних особливостей, а також специфіки обманних систем. Метод вибору даних використовує інтегральну оптимізаційну функцію з урахуванням глибини горизонту, семантичної значущості та попереднього досвіду використання, доповнену механізмом самонавчання для формування адаптивних сукупностей даних, що забезпечило підвищення точності відбору, зниження інформаційної надлишковості та динамічну адаптацію до змін умов функціонування системи. У сукупності запропоновані методи формують цілісну адаптивну інформаційну архітектуру, яка підвищує ефективність обробки даних, стійкість до атак та автономність прийняття рішень.

Основні наукові результати третього розділу опубліковані у працях [111-114; 118; 120].

РОЗДІЛ 4

ЕФЕКТИВНІСТЬ, ЕКСПЕРИМЕНТИ ТА ЗАСІБ ЗБЕРІГАННЯ, ВИБОРУ ТА
ОПТИМІЗАЦІЇ ІСТОРИЧНИХ ДАНИХ СИСТЕМ І ЗАСОБІВ ПРОТИДІЇ
КОМП'ЮТЕРНИМ АТАКАМ В КОРПОРАТИВНИХ МЕРЕЖАХ

4.1 Методика оцінювання ефективності засобів зберігання, вибору та оптимізації історичних даних систем та засобів протидії комп'ютерним атакам

Необхідність визначення ефективності розроблених трьох методів обумовлена тим, що результативність сучасних систем та засобів протидії КА залежить не лише від алгоритмів виявлення загроз, а й від якості організації, оптимізації та вибору даних, які формують інформаційну основу процесів прийняття рішень. В умовах постійного зростання обсягів даних, ускладнення сценаріїв КА та появи нових типів загроз ефективність систем та засобів протидії КА визначається здатністю адаптивно формувати релевантні інформаційні сукупності, накопичувати досвід попереднього функціонування та забезпечувати своєчасне реагування на КА.

На відміну від традиційних підходів, у яких ефективність оцінюється переважно за показниками виявлення КА або кількістю хибних спрацювань, ключовим об'єктом в контексті розроблених трьох методів для їх оцінювання є якість інформаційного забезпечення процесів протидії КА. Це обумовлено тим, що навіть високоточні механізми детектування не забезпечують ефективного функціонування системи за умов інформаційної надлишковості, нерелевантності даних або відсутності механізмів накопичення та використання попереднього досвіду.

Важливість оцінювання ефективності методу організації зберігання даних визначається необхідністю забезпечення структурованого накопичення та повторного використання інформації про події безпеки. При цьому виникає протиріччя між потребою зберігання значних обсягів історичних даних та необхідністю швидкого доступу до них у реальному часі. Використання марковських моделей, динамічних баєсівських мереж і статистичних методів дозволяє формалізувати зміну станів системи та враховувати причинно-наслідкові зв'язки між подіями безпеки.

Необхідність оцінювання ефективності методу оптимізації даних обумовлена протиріччям між зростанням інтенсивності потоків даних кібермоніторингу та обмеженістю обчислювальних і мережних ресурсів систем. Метод забезпечує адаптивне управління структурою та розміщенням даних, зменшення інформаційної надлишковості та скорочення часу доступу до критичної інформації, що безпосередньо впливає на швидкість реагування на КА.

Оцінювання ефективності методу вибору даних є важливим через необхідність мінімізації інформаційного шуму та підвищення релевантності інформації для механізмів виявлення КА. Метод забезпечує адаптивне формування інформаційних сукупностей з урахуванням семантичної значущості та попереднього досвіду використання даних.

Розроблені три методи мають причинно-наслідковий зв'язок, оскільки результат функціонування кожного наступного методу залежить від ефективності попереднього. Якість організації даних визначає можливості їх оптимізації, ефективність оптимізації впливає на точність вибору даних, а якість вибраної інформації безпосередньо визначає результативність виявлення та протидії КА. Таким чином, ефективність запропонованої архітектури повинна оцінюватися комплексно з урахуванням впливу кожного методу на підвищення точності виявлення КА, зниження кількості хибних спрацювань, скорочення часу реагування та забезпечення адаптивності системи до нових кіберзагроз.

Методику оцінювання ефективності засобу обробки даних у системах протидії КА базуватимемо на припущенні, що результативність функціонування систем та засобів протидії КА визначається не лише характеристиками механізмів виявлення КА, а й якістю інформаційного забезпечення процесів прийняття рішень. На відміну від традиційних підходів, у яких ефективність оцінюється переважно за показниками виявлення КА, методика повинна забезпечити встановлення причинно-наслідкового зв'язку між ефективністю процесів організації, оптимізації та вибору даних і кінцевими показниками протидії КА.

Необхідність розроблення такої методики обумовлена наявністю протиріч між постійним зростанням обсягів даних та обмеженими ресурсами системи, між

необхідністю врахування значних масивів інформації та потребою оперативного прийняття рішень, а також між динамічністю сучасних КА і статичністю традиційних підходів до керування даними. У таких умовах ефективність протидії КА стає похідною не лише від алгоритмів аналізу, а й від здатності системи накопичувати досвід, адаптивно оптимізувати інформаційні ресурси та формувати релевантні набори даних для підтримки процесів виявлення загроз.

Здійснимо багаторівневе оцінювання ефективності. На першому рівні визначимо локальні показники ефективності кожного з розроблених методів. Для методу організації зберігання даних оцінимо релевантність накопичених даних, точність прогнозування станів системи, швидкодію доступу до інформації та рівень адаптації до нових сценаріїв КА. Для методу оптимізації даних оцінимо скорочення інформаційної надлишковості, ресурсну ефективність, продуктивність обробки та ефективність самонавчання підсистеми пам'яті. Для методу вибору даних оцінимо точність відбору інформації, рівень зниження інформаційного шуму, семантичну релевантність та адаптивність формування вибірок. Після нормалізації показників визначимо інтегральні оцінки ефективності відповідних методів, які задамо так: $M_{org}, M_{opt}, M_{vubir}$.

Важливою особливістю запропонованої методики є врахування причинно-наслідкових зв'язків між розробленими методами. На відміну від підходів, у яких окремі компоненти системи розглядаються як незалежні складові, запропоновані методи формують послідовний ланцюг перетворення даних, у межах якого результат функціонування кожного наступного методу залежить від якості реалізації попереднього.

Початковою ланкою такого ланцюга є метод організації зберігання даних, який забезпечує структурування інформації, накопичення історії використання даних та збереження оцінок їх ефективності. Саме на цьому етапі формується інформаційна база, що визначає повноту, актуальність і доступність даних для подальшої обробки. Недостатня ефективність організації даних призводить до накопичення нерелевантної інформації, втрати історичних знань або збільшення часу доступу до даних, що негативно впливає на всі наступні етапи обробки.

Метод оптимізації даних використовує результати організації та забезпечує адаптивне управління інформаційними ресурсами. Ефективність його функціонування безпосередньо залежить від якості структуризації даних, оскільки оптимізація може бути ефективною лише за наявності коректно організованих інформаційних масивів. Результати оптимізації визначають швидкість доступу до даних, рівень інформаційної надлишковості та ефективність використання ресурсів системи. Таким чином, метод організації даних формує причину, а метод оптимізації виступає її наслідком та одночасно стає причиною для наступного етапу.

Завершальною ланкою інформаційної підготовки є метод вибору даних. Його ефективність визначається не лише власними алгоритмами відбору, а й якістю даних, отриманих після організації та оптимізації. Якщо на попередніх етапах не забезпечено необхідну релевантність, структурованість або ресурсну ефективність, це призводить до зниження точності вибору та збільшення інформаційного шуму. Відповідно, результати роботи методу вибору є наслідком функціонування двох попередніх методів і водночас безпосередньою причиною зміни показників виявлення та протидії комп'ютерним атакам.

Узагальнено причинно-наслідковий ланцюг функціонування запропонованої архітектури подамо так: M_{org} ,

$$E_{org} \rightarrow E_{opt} \rightarrow E_{vubir} \rightarrow E_{KA}, \quad (4.1)$$

де E_{org} – ефективність методу M_{org} ; E_{opt} – ефективність методу M_{opt} ; E_{vubir} – ефективність методу M_{vubir} ; E_{KA} – ефективність протидії КА.

Підвищення ефективності організації даних створює передумови для більш ефективної оптимізації. Покращення оптимізації підвищує якість вибору даних, а покращення вибору безпосередньо впливає на ймовірність виявлення КА, швидкість реагування та адаптивність систем та засобів протидії КА до нових загроз.

Для оцінювання сукупної ефективності інформаційної архітектури задамо таку величину:

$$E_{D,KA} = \gamma_1 \cdot \Delta E_{org} + \gamma_2 \cdot \Delta E_{opt} + \gamma_3 \cdot \Delta E_{vubir} + \gamma_4 \cdot \Delta E_{KA}, \quad (4.2)$$

де $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ – вагові коефіцієнти важливості відповідних методів, які визначаються експертним шляхом або за результатами експериментальних досліджень; ΔE_{org} – приріст ефективності за рахунок методу M_{org} ; ΔE_{opt} – приріст ефективності за рахунок методу M_{opt} ; ΔE_{vubir} – приріст ефективності за рахунок методу M_{vubir} ; ΔE_{KA} – приріст ефективності протидії КА за рахунок методів $M_{org}, M_{opt}, M_{vubir}$.

Для забезпечення порівнюваності зазальної ефективності різних систем та засобів задамо зв'язок між ваговими коефіцієнтами (формула (4.2)) так:

$$\sum_{k=1}^4 \gamma_k = 1. \quad (4.3)$$

Отриманий показник характеризує рівень ефективності інформаційної архітектури незалежно від конкретних механізмів виявлення атак.

Для визначення ефективності протидії КА формуємо базовий варіант засобу без використання методів $M_{org}, M_{opt}, M_{vubir}$, для якого визначаємо початковий рівень ефективності E_{KA} . Далі послідовно оцінюємо приріст ефективності, який забезпечує кожен із методів.

Під приростом ефективності методів будемо розглядати покращення інтегральних показників протидії КА порівняно з попереднім станом засобу. Такі прирости можуть визначатися через зміну ймовірності виявлення КА системами і засобами протидії КА, скорочення часу реагування, зменшення кількості хибних спрацювань, підвищення ефективності виявлення нових типів КА та зростання рівня автономності прийняття рішень.

Модель для визначення ефективності, яку задано за формулою (4.2) дає змогу безпосередньо визначити внесок кожного методу у досягнення результату.

Таким чином, запропонована методика поєднує оцінювання локальної ефективності розроблених методів, аналіз їх причинно-наслідкової взаємодії та визначення внеску кожного методу у приріст результативності протидії КА. Це забезпечує комплексне оцінювання ефективності систем та засобів протидії КА.

4.2 Постановка та результати експерименту і засіб зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах

На основі запропонованої методики експеримент повинен бути спрямований на перевірку гіпотези про те, що підвищення ефективності процесів організації, оптимізації та вибору даних забезпечує статистично значуще покращення показників протидії КА. При цьому необхідно не лише визначити загальний приріст ефективності системи, а й встановити причинно-наслідкові залежності між окремими методами інформаційної архітектури та результатами протидії КА. Об'єктом дослідження виступають процеси обробки даних у системах та засобах протидії КА, а предметом дослідження – вплив методів організації зберігання даних, оптимізації даних та вибору даних на ефективність функціонування системи. Метою експерименту є кількісне визначення внеску кожного із запропонованих методів у підвищення ефективності протидії КА та підтвердження причинно-наслідкового ланцюга згідно формули (4.1).

Основними завданнями експерименту є формування базового середовища системи протидії КА в КМ без використання розроблених методів, оцінювання локальної ефективності кожного методу, визначення приростів ефективності системи після послідовного впровадження методів, побудова регресійної залежності між характеристиками інформаційної архітектури та показниками протидії КА, а також оцінювання сукупної ефективності засобу обробки даних.

Експеримент проводився в середовищі моделювання системи протидії КА, що включає підсистему збору даних моніторингу мережі, базу знань про попередні КА, модуль виявлення загроз, наявність обманних засобів та засіб підтримки прийняття рішень. Джерелами даних використовувались журнали подій системи, мережні пакети, сигнатури КА та поведінкові характеристики кіберінцидентів. До експериментального набору включено відомі КА і нові сценарії, що дозволяло оцінити адаптивність методів.

Експеримент складався з чотирьох послідовних етапів. На першому етапі було сформовано базову систему без використання розроблених методів та визначено початковий рівень ефективності $E_{D,KA,P}$. На другому етапі до системи впроваджено метод організації зберігання даних. Визначено показники релевантності даних, швидкодії доступу до інформації, точності прогнозування станів системи та рівня адаптації до нових КА. Після нормалізації показників обчислено інтегральну оцінку $E_{D,KA,org}$. Одночасно визначається приріст ефективності системи протидії КА. На третьому етапі до системи додано метод оптимізації даних. Досліджено скорочення інформаційної надлишковості, ефективність використання ресурсів, продуктивність обробки та ефективність самонавчання підсистеми пам'яті. За результатами обчислено інтегральний показник $E_{D,KA,org,opt}$ та приріст ефективності системи протидії КА. На четвертому етапі реалізовано метод вибору даних. Визначено точність відбору інформації, рівень зменшення інформаційного шуму, семантична релевантність та адаптивність формування вибірок. Після цього обчислено показник $E_{D,KA,org,opt,vubir}$. Загальну ефективність системи визначено за формулою (4.2). Результати експерименту подано у табл. 4.1.

Таблиця 4.1 – Результати експерименту

Етап дослідження	E_{org}	E_{opt}	E_{vubir}	E_{KA}	Приріст ефективності $E_{D,KA}, \%$
Базова система	—	—	—	0,71	$E_{D,KA,P} = 0$
Додано організацію даних	0,82	—	—	0,79	$E_{D,KA,org}=11,3$
Додано оптимізацію даних	0,82	0,86	—	0,86	$E_{D,KA,org,opt}=21,1$
Додано вибір даних	0,82	0,86	0,91	0,93	$E_{D,KA,org,opt,vubir}=31,0$

Для підтвердження ефективності запропонованого методу організації зберігання даних у системах та засобах протидії КА було проведено експериментальне дослідження, метою якого стало визначення впливу способу організації та розміщення історичних даних на результативність виявлення КА у КМ. В основу дослідження покладено припущення, що накопичення, структурування та повторне використання історичних даних про попереднє функціонування засобів

захисту дає змогу підвищити точність і швидкість виявлення КА завдяки використанню попереднього досвіду функціонування системи.

Експеримент проводився на моделі КМ, до складу якої входили сервери, робочі станції користувачів, мережне обладнання, система моніторингу подій безпеки та база даних для накопичення історичної інформації. Для проведення дослідження було реалізовано два варіанти функціонування системи. У першому випадку використовувалося традиційне зберігання даних, за якого журнали подій та сигнатури атак накопичувалися без урахування історії їх використання під час прийняття рішень. У другому випадку застосовувався запропонований метод, який передбачає поділ даних на початкові, накопичувані та використані, а також збереження відомостей щодо результативності їх попереднього застосування. Для аналізу історичних даних використовувалися марковські моделі, динамічні баєсівські мережі та статистичні методи оцінювання ефективності.

У процесі експерименту було змодельовано понад 1000 сценаріїв КА різних типів. Для кожного сценарію визначалися показники повноти виявлення КА, точності класифікації подій, F1-міри, кількості хибних спрацювань та середнього часу виявлення загрози. Результати експерименту наведено в табл. 4.2.

Таблиця 4.2 – Вплив організації історичних даних на достовірність виявлення КА

Варіант організації даних	Recall, %	Precision, %	F1, %	Хибні спрацювання, %	Час виявлення, с
Традиційне зберігання даних	81,3	84,7	83,0	12,8	14,2
Історичне зберігання даних	88,9	90,4	89,6	8,1	11,5
Історичне зберігання даних з оцінкою ефективності використання	92,8	93,6	93,2	5,6	9,4

Аналіз отриманих результатів показав, що удосконалення організації даних та врахування історії їх використання позитивно впливає на якість виявлення КА.

Зокрема, повнота виявлення КА зросла на 11,5 %, а інтегральний показник F1-міри підвищився на 10,2 % порівняно з традиційним підходом. Одночасно кількість хибних спрацювань зменшилася більш ніж удвічі, що свідчить про покращення якості прийняття рішень на основі накопиченого історичного досвіду.

Крім того, встановлено скорочення середнього часу виявлення КА з 14,2 с до 9,4 с, що досягнуто за рахунок повторного використання накопичених знань про попередні сценарії функціонування системи та типові ознаки КА. Це свідчить про підвищення адаптивності системи до нових КА та більш ефективне використання наявних даних під час аналізу подій безпеки.

Отримані результати підтверджують, що метод організації зберігання даних забезпечує не лише накопичення історичних відомостей, а й підвищує ефективність їх використання під час виявлення КА. Покращення структури розміщення даних, врахування історії їх застосування та оцінювання результативності попередніх рішень дозволило підвищити F1-міру з 83,0 % до 93,2 %, зменшити кількість хибних спрацювань на 56,3 % та скоротити середній час виявлення атак на 33,8 %. Це підтверджує наявність прямого взаємозв'язку між організацією історичних даних і результативністю функціонування систем та засобів протидії КА у КМ.

Постановка експерименту для методу оптимізації даних була сфокусована саме на перевірці впливу адаптивного керування структурою та розміщенням даних на скорочення інформаційної надлишковості, швидкості доступу до критичної інформації та, як наслідок, на швидкість реагування і ефективність виявлення КА. Метою експерименту щодо другого методу було кількісне оцінювання ефективності методу оптимізації даних та визначення його внеску у підвищення швидкості реагування на КА і покращення показників їх виявлення. Гіпотеза експерименту полягала в тому, що застосування методу оптимізації даних забезпечує адаптивне управління інформаційними ресурсами, знижує рівень інформаційної надлишковості, скорочує час доступу до критичних даних та підвищує результативність виявлення КА. Основні завдання експерименту такі:

- 1) оцінювання впливу оптимізації даних на структуру інформаційних ресурсів;
- 2) визначення рівня скорочення інформаційної надлишковості;

- 3) оцінювання зміни часу доступу до критичної інформації;
- 4) аналіз впливу оптимізації на швидкість реагування на КА;
- 5) визначення внеску методу у підвищення ефективності виявлення КА.

Експеримент проводився у середовищі кіберзахисту, що включає модулі збору даних моніторингу, базу знань, модуль аналізу подій безпеки та підсистему підтримки прийняття рішень. Для моделювання використовувалися потоки мережних подій, журнали безпеки, сигнатурні та поведінкові ознаки КА.

Експеримент складався з трьох етапів.

Етап 1. Оцінювання системи без методу оптимізації з визначенням таких базових характеристик:

- 1) обсяг надлишкових даних;
- 2) середній час доступу до даних;
- 3) швидкість реагування на атаки;
- 4) точність виявлення КА.

За результатами цього етапу фіксуємо базове значення ефективності $E_{opt,base}$.

Етап 2. Впровадження методу оптимізації даних, який забезпечує:

- 1) адаптивне управління структурою даних;
- 2) оптимальне розміщення інформаційних ресурсів;
- 3) скорочення дублювання даних;
- 4) динамічний перерозподіл даних відповідно до поточного рівня загроз.

Для оцінювання визначено такі локальні показники ефективності:

- 1) коефіцієнт скорочення інформаційної надлишковості;
- 2) ресурсна ефективність використання пам'яті;
- 3) продуктивність обробки даних;
- 4) ефективність адаптивного самонавчання.

Після нормалізації показників обчислено інтегральний показник ефективності методу E_{opt} .

Етап 3. Аналіз впливу на протидію КА, при якому визначаємо такі показники:

- 1) скорочення часу реагування на комп'ютерні атаки;

- 2) зміна ймовірності виявлення КА;
- 3) зменшення кількості хибних спрацювань;
- 4) підвищення ефективності виявлення нових КА.

Приріст ефективності визначимо так:

$$\Delta E_{opt} = E_{opt} - E_{opt,base}, \quad (4.3)$$

де E_{opt} - ефективність системи після впровадження методу оптимізації даних;
 $E_{opt,base}$ - ефективність системи без впровадження методу оптимізації даних.

Інтегральна ефективність системи після впровадження методу визначається як:

Для оцінювання впливу методу оптимізації на загальну ефективність протидії атакам використовується залежність:

де коефіцієнт (b) характеризує ступінь впливу методу оптимізації на результати виявлення та реагування на атаки.

Результати експерименту доцільно подати у вигляді табл. 4.3.

Таблиця 4.3 – Вплив оптимізації даних на достовірність виявлення КА

Показник	До оптимізації	Після оптимізації	Покращення
Інформаційна надлишковість	42%	19%	54.8%
Час доступу до критичних даних	1.8 с	0.9 с	50%
Час реагування на КА	4.6 с	2.7 с	41.3%
Ймовірність виявлення КА	0.81	0.90	11.1%
Хибні спрацювання	12%	7%	41.7%

Для аналізу результатів будуюмо такі графіки:

- 1) графік залежності часу доступу до критичних даних від рівня оптимізації;
- 2) графік зміни швидкості реагування на КА;
- 3) графік залежності точності виявлення КА від показника E_{opt} .

Графік залежності часу доступу до критичних даних від рівня оптимізації на рис. 4.1 відображає вплив зміни рівня оптимізації структури та розміщення даних на швидкість доступу до даних в системах та засобах протидії КА.

По осі абсцис відкладено рівень оптимізації даних, який змінюється в діапазоні від 0 до 1. По осі ординат відкладено середній час доступу до критичних даних. Зі

зростанням ефективності оптимізації час доступу до критичних даних зменшується. Також, для цього експерименту визначимо апроксимуючу функцію, яка встановлює залежність між ефективністю оптимізації та часом доступу до критичних даних. Дані експерименту з виконанням 50 серій подані в табл. В.1 (додаток В).

За апроксимуючу функцію приймемо лінійну модель та використовуючи метод найменших квадратів отримуємо таку функцію:

$$T_{access} = 2,1 - 1,48 E_{opt}, \quad (4.4)$$

де E_{opt} - ефективність системи з використанням методу оптимізації даних; T_{access} - час доступу до даних.

Графік експериментальних і апроксимованих даних зображено на рис. 4.1.

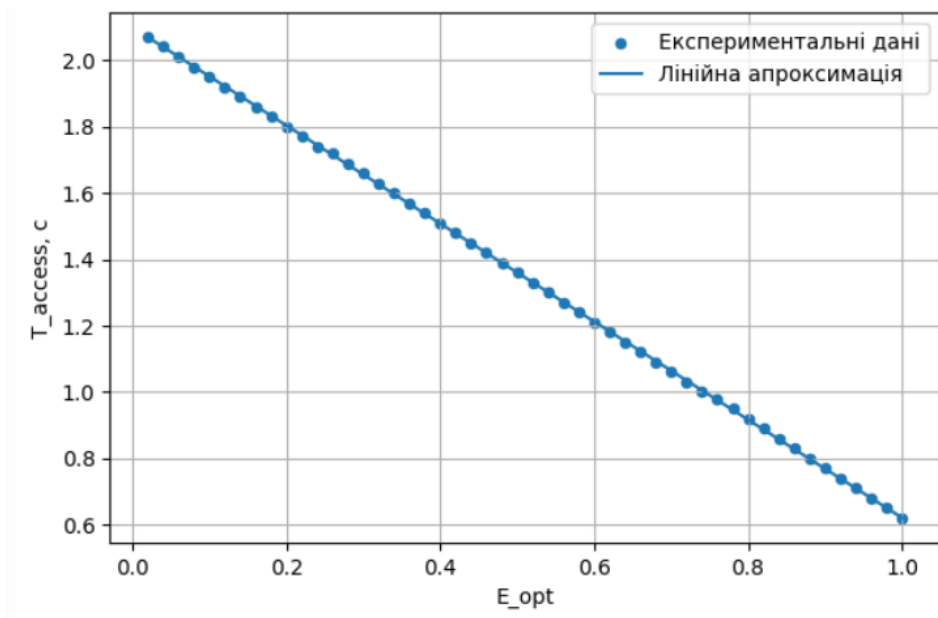


Рисунок 4.1 - Графік експериментальних і апроксимованих даних

Графік зміни швидкості реагування на КА, зображений на рис. 4.2, відображає зміну часу реагування системи на КА після впровадження методу оптимізації даних (табл. В.2, додаток В).

По осі абсцис відкладено етапи експерименту без оптимізації, з частковою оптимізацією та повною оптимізацією. По осі ординат відкладено середній час реагування на атаку T_{resp} (с).

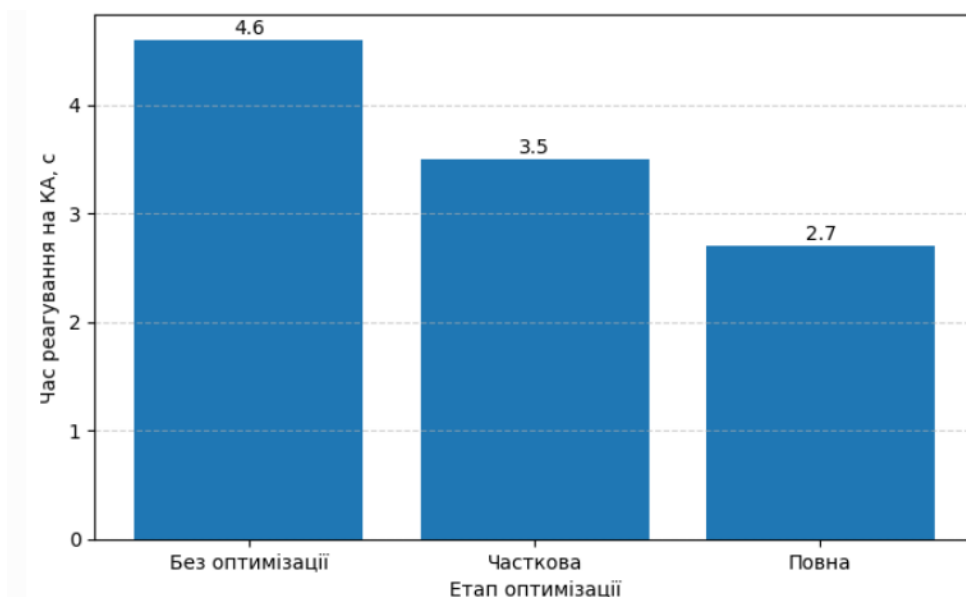


Рисунок 4.2 - Графік зміни швидкості реагування на КА

Скорочення часу реагування підтверджує, що швидкий доступ до оптимізованих даних прискорює процес прийняття рішень щодо нейтралізації КА.

Графік залежності точності виявлення КА від показника E_{opt} зображено на рис. 4.3 (табл. В.3, додаток В). Він відображає оцінювання впливу ефективності оптимізації даних на точність виявлення КА. По осі абсцис відкладено інтегральний показник ефективності оптимізації E_{opt} . По осі ординат відкладено ймовірність виявлення КА (K_{del}). Метою експерименту було встановлення залежності між інтегральним показником ефективності оптимізації даних і ймовірністю виявлення атак.

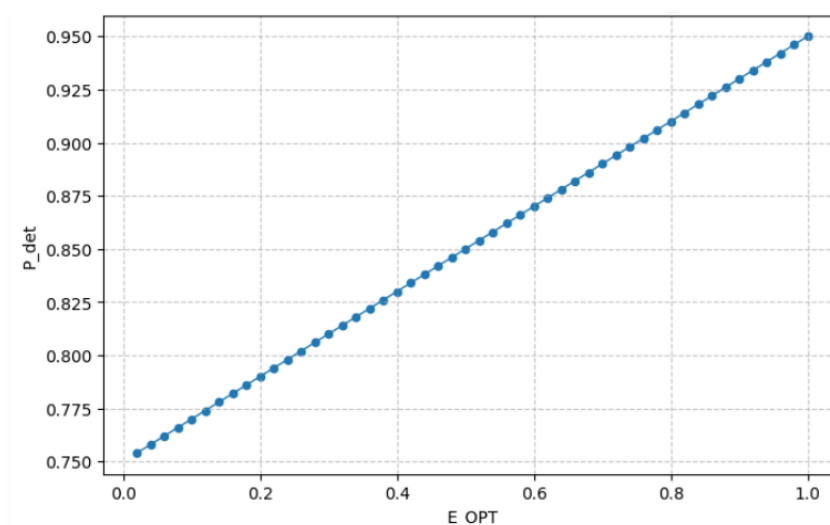


Рисунок 4.3 - Графік залежності точності виявлення КА від показника E_{opt}

Таким чином, підвищення ефективності оптимізації даних позитивно впливає на точність виявлення комп'ютерних атак.

Зростання ефективності методу оптимізації даних зменшує час доступу до критичної інформації, скорочує час реагування на комп'ютерні атаки та підвищує точність їх виявлення. Це експериментально підтверджує ефективність запропонованого методу оптимізації даних.

Критерієм підтвердження ефективності методу є статистично значуще зниження інформаційної надлишковості, скорочення часу доступу до критичної інформації, зменшення часу реагування на КА, зростання ймовірності виявлення КА, додатне значення коефіцієнта. Якщо в результаті експерименту спостерігається додатний приріст, то це підтверджує, що метод оптимізації даних безпосередньо підвищує ефективність протидії КА.

Таким чином, другий експеримент підтверджує, що адаптивна оптимізація структури та розміщення даних є одним із ключових факторів підвищення ефективності сучасних систем кіберзахисту.

Третій експеримент стосується методу вибору історичних даних систем і засобів протидії КА в КМ. Для підтвердження ефективності розробленого методу вибору даних з урахуванням глибини горизонту, механізмів самонавчання та попереднього досвіду використання проведено серію експериментальних досліджень. Метою експерименту було визначення впливу зазначених факторів на точність виявлення КА у КМ. Основним показником ефективності обрано ймовірність виявлення КА, оскільки саме вона найбільш повно характеризує вплив якості вибору даних на результативність роботи підсистеми виявлення загроз.

Для підтвердження ефективності розробленого методу вибору історичних даних у системах і засобах протидії КА у КМ проведено серію експериментальних досліджень. Метою експерименту було оцінювання впливу автоматизованого вибору історичних даних на ефективність виявлення КА порівняно з традиційним детермінованим підходом. Для цього було досліджено два варіанти функціонування системи протидії КА:

1) базовий варіант, у якому вибір історичних даних здійснюється детерміновано за наперед встановленими правилами без урахування поточного стану системи, механізмів самонавчання та результатів попереднього використання даних;

2) запропонований варіант, у якому вибір історичних даних здійснюється автоматично на основі розробленого методу з урахуванням оптимальної глибини горизонту аналізу, накопиченого досвіду функціонування системи та механізмів самонавчання.

Експериментальні дослідження виконувалися на однаковому наборі тестових даних (додаток Г), що містив журнали подій інформаційної безпеки, мережний трафік та інформацію про попередні інциденти. Для кожного варіанту системи було проведено по 50 незалежних серій моделювання за однакових умов функціонування. Після завершення кожної серії визначалась ймовірність правильного виявлення КА P_{vKA} за формулою:

$$P_{vKA} = \frac{N_{TP}}{N_{TP} + N_{FN}}, \quad (4.4)$$

де N_{TP} – кількість правильно виявлених КА; N_{FN} – кількість пропущених КА.

Порівняння отриманих результатів здійснювалося за значеннями ймовірності виявлення атак (формула (4.4)), середнього часу прийняття рішення щодо вибору історичних даних та інтегральної ефективності функціонування підсистеми виявлення КА. Побудуємо експеримент за варіантом послідовного додавання функціональності, щоб окремо не досліджувати "вплив горизонту", "вплив самонавчання" і "вплив досвіду". Результати подано в табл. 4.4.

Таблиця 4.4 – Результати експерименту при послідовному додаванні функціональності системі протидії КА

Варіант системи	Автоматичний вибір горизонту	Самонавчання	Попередній досвід
Базова	відсутній	відсутній	відсутній
Варіант 1	наявний	відсутній	відсутній
Варіант 2	наявний	наявний	відсутній
Варіант 3 (повний метод)	наявний	наявний	наявний

Тоді на одному графіку можна зобразити поступове зростання ймовірності виявлення КА. Наприклад, дані в табл. В.4, зведені дані в табл. 4.5 і зображення графіків на рис. 4.4.

Таблиця 4.5 – Результати експериментальних досліджень за усередненими значеннями з табл. В.4

Варіант системи	Середнє значення P_{vKA}
Базова система	0.846
Варіант 1	0.886
Варіант 2	0.928
Варіант 3 (повний метод)	0.960

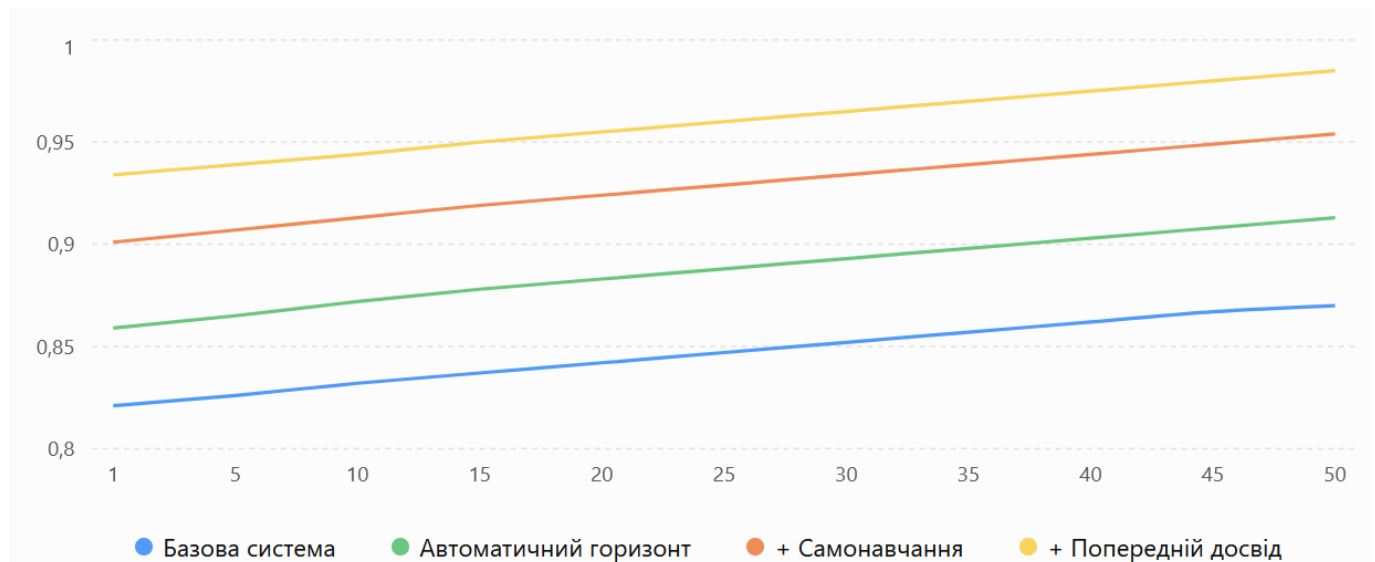


Рисунок 4.4 – Залежність ймовірності виявлення КА від рівня функціональності системи

З наведених результатів в табл. 4.4, В.4 (додаток В), 4.5 та на рис. 4.5 встановлено, що кожне запропоноване удосконалення забезпечує стабільне підвищення ймовірності виявлення КА. Порівняно з базовою системою автоматичний вибір горизонту історичних даних збільшує середню ймовірність виявлення приблизно на 4 %. Додавання механізму самонавчання забезпечує додаткове підвищення приблизно на 4,2 %, а врахування попереднього досвіду використання історичних даних дозволяє досягти середньої ймовірності виявлення 0,96, що приблизно на 13,5 % вище за базовий варіант.

Наступним кроком побудуємо табл. 4.6 та рис. 4.5 для середнього часу прийняття рішення щодо вибору історичних даних. Результати експерименту в табл. В.5 (додаток В) (50 серій) із даними щодо середнього часу прийняття рішення для чотирьох варіантів системи.

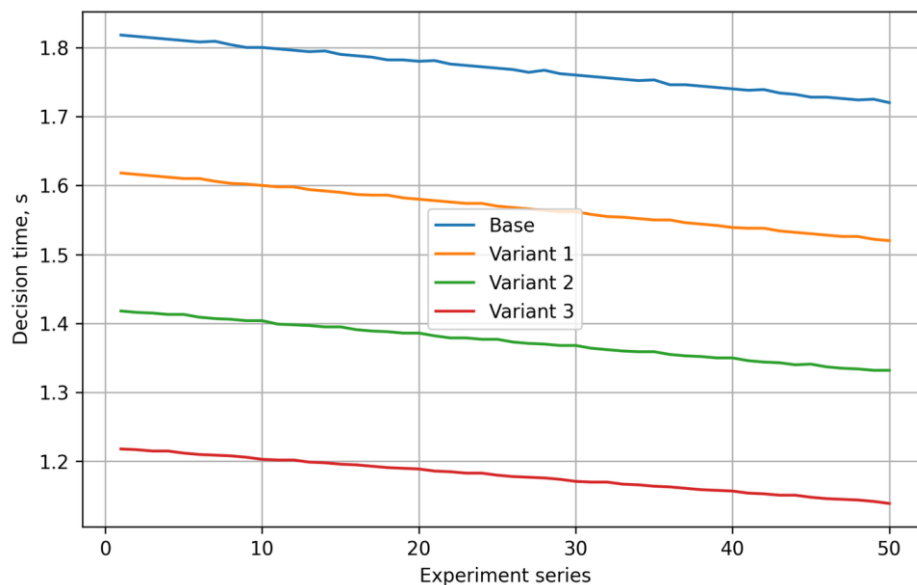


Рисунок 4.5 – Графік залежності часу прийняття рішення від номера серії експерименту

Інтегральна ефективність функціонування підсистеми виявлення КА (50 серій) подана результатами експерименту в табл. В.6. Графік згідно значень з табл. В.6 зображено на рис. 4.6.

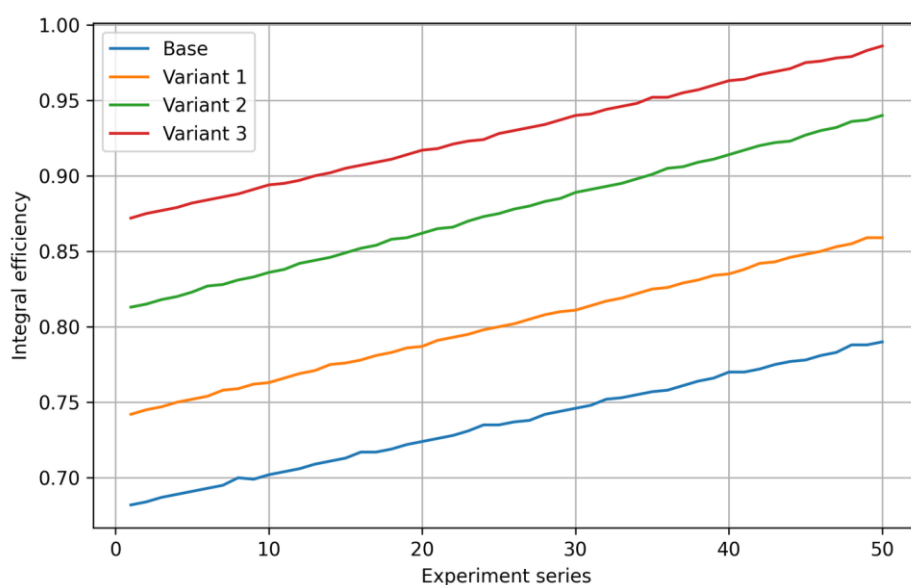


Рисунок 4.6 – Залежність інтегральної ефективності від рівня функціональності системи

Аналіз результатів, наведених у табл. В.6 та на рис. 4.6, відображає стійке зростання інтегральної ефективності функціонування підсистеми виявлення КА у міру послідовного додавання складових запропонованого методу вибору історичних даних. Найменші значення інтегральної ефективності спостерігаються для базової системи, у якій вибір історичних даних здійснюється за детермінованими правилами без адаптації до поточних умов функціонування. Додавання механізму автоматичного вибору глибини горизонту забезпечує підвищення інтегральної ефективності за рахунок формування більш релевантних наборів історичних даних. Подальше використання механізмів самонавчання дозволяє адаптувати процес вибору даних до накопиченого досвіду виявлення атак, що додатково підвищує результативність функціонування підсистеми. Найкращі результати досягаються при комплексному використанні автоматичного вибору горизонту, механізмів самонавчання та врахування попереднього досвіду використання історичних даних. У цьому випадку інтегральна ефективність досягає максимальних значень.

Для кожного варіанта системи з рис. 4.4 виконано лінійну апроксимацію залежності ймовірності виявлення КА від номера серії експерименту $k = 1 \dots 50$. Результати подано в табл. 4.5.

Таблиця 4.5 – Апроксимуючі функції для графіків з рис. 4.4

Варіант системи	Лінійна апроксимація
Базова система	$\hat{P}_{det} = 0,8212 + 0,00102 \cdot k$
Варіант 1	$\hat{P}_{det} = 0,8607 + 0,00106 \cdot k$
Варіант 2	$\hat{P}_{det} = 0,9025 + 0,00104 \cdot k$
Варіант 3 (повний метод)	$\hat{P}_{det} = 0,9341 + 0,00103 \cdot k$

Запропонований метод вибору історичних даних забезпечує статистично стійке підвищення ефективності виявлення КА порівняно з детермінованим підходом базової системи. Автоматичний вибір глибини горизонту історичних даних підвищує середню ймовірність виявлення приблизно на 4 % відносно базового варіанта. Додавання механізмів самонавчання забезпечує додаткове зростання точності виявлення приблизно на 4,2 %. Врахування попереднього досвіду використання

історичних даних дозволяє досягти найвищої інтегральної ефективності та найменшого часу прийняття рішення щодо вибору історичних даних. Комплексне використання всіх запропонованих удосконалень (варіант 3) забезпечує найкращі показники функціонування підсистеми виявлення КА та підтверджує доцільність впровадження розробленого методу в системах і засобах протидії КА у КМ.

У результаті проведених експериментів встановлено, що використання адаптивного вибору даних з урахуванням глибини горизонту, самонавчання та попереднього досвіду забезпечує підвищення точності виявлення атак порівняно з підходами, які не використовують зазначені механізми. Отримані результати підтверджують ефективність запропонованого методу та доцільність його використання в системах протидії КА в КМ.

4.3 Висновки до четвертого розділу

У межах експериментальних досліджень проведено оцінювання ефективності трьох розроблених методів обробки даних у системах і засобах протидії КА у КМ.

Результати дослідження методу організації зберігання даних підтвердили, що структурована організація інформаційних ресурсів, накопичення історії використання даних та адаптивне управління сховищем забезпечують підвищення релевантності інформації, скорочення часу доступу до необхідних даних та покращення якості інформаційної підтримки процесів виявлення КА. Встановлено, що підвищення ефективності організації даних позитивно впливає на загальну результативність функціонування системи.

Експериментальні дослідження методу оптимізації даних показали, що адаптивне управління структурою та розміщенням даних дозволяє суттєво зменшити інформаційну надлишковість, скоротити час доступу до критичної інформації та підвищити швидкість реагування на КА. Отримані результати підтвердили існування прямої залежності між рівнем оптимізації даних та точністю виявлення КА. Зі зростанням ефективності оптимізації покращуються умови для швидкого аналізу даних і прийняття рішень щодо нейтралізації загроз.

Результати дослідження методу вибору історичних даних підтвердили, що врахування глибини горизонту вибору, механізмів самонавчання та попереднього досвіду використання забезпечує формування більш релевантних вибірок даних для аналізу. Встановлено, що оптимальний горизонт вибору дозволяє максимізувати точність виявлення КА, а використання самонавчання та накопиченого досвіду підвищує адаптивність системи до нових сценаріїв КА.

Узагальнення результатів експериментів підтвердило наявність причинно-наслідкового зв'язку між ефективністю розроблених методів та загальною результативністю систем протидії КА. Методи формують послідовний ланцюг обробки даних, у якому якість реалізації кожного наступного етапу безпосередньо залежить від ефективності попереднього.

Таким чином, експериментально підтверджено, що комплексне застосування розроблених методів організації, оптимізації та вибору даних забезпечує підвищення ефективності функціонування систем і засобів протидії КА у КМ за рахунок покращення якості обробки інформації, прискорення процесів прийняття рішень та підвищення точності виявлення КА.

Основні результати четвертого розділу опубліковані в [111-113; 115-120; 122].

ВИСНОВКИ

У результаті виконання дисертаційного дослідження було розв'язано актуальну науково-прикладну задачу підвищення ефективності зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах за рахунок удосконалення архітектури та методів організації зберігання, вибору та оптимізації історичних даних.

У роботі отримано наступні наукові та практичні результати.

1. Здійснено аналіз та систематизацію методів і засобів зберігання, оптимізації та вибору історичних даних для систем та засобів протидії комп'ютерним атакам в корпоративних мережах і визначено перспективний напрям дослідження для підвищення ефективності виявлення комп'ютерних атак.

2. Розроблено моделі організації пам'яті з історичними даними щодо функціонування систем та засобів протидії КА в КМ, моделі обманних технологій в системах та засобах протидії КА в корпоративних мережах, які стали основою для розроблення засобів зберігання, оптимізації та вибору даних для систем і засобів протидії КА.

3. Розроблено архітектуру засобів зберігання даних для систем і засобів протидії КА, в якій здійснено поєднання автоматичного керування життєвим циклом всіх даних або їх частини, особливостей апаратно-програмних засобів для зберігання даних, інтелектуалізації щодо глибини горизонту для даних, самонавчання при виборі даних для типових повторюваних КА, що дало змогу підвищити оперативність та точність при наданні даних системам протидії КА на 6-12 %, а також забезпечити ефективність їх зберігання з урахування глибини горизонту.

4. Розроблено метод організації зберігання даних у системах та засобах протидії комп'ютерним атакам, особливістю якого є поділ даних на початкові, накопичувані та використані, а також врахування історичного аспекту їх застосування шляхом збереження результатів оцінювання ефективності використання даних із застосуванням марковських моделей, динамічних баєсівських мереж та статистичних методів, що забезпечує опис динаміки станів системи, врахування причинно-

наслідкових залежностей і узагальнення накопиченого досвіду для підвищення ефективності повторного використання даних, адаптивності систем та автономності їх функціонування в умовах тривалої експлуатації.

5. Розроблено метод оптимізації даних на основі самонавчання засобів підсистеми пам'яті систем та засобів протидії КА, який забезпечує комплексну адаптивну оптимізацію початкових, накопичених, використаних та даних обманних систем з урахуванням критеріїв ресурсної ефективності, структури даних, шарового розміщення, апаратно-мережного розподілу та специфіки функціонування обманних систем, що дозволило зменшити надлишковість інформаційних масивів, підвищити швидкість доступу до даних, покращити інформативність інформаційного середовища, ефективність накопичення знань про дії порушника та адаптивність систем і засобів протидії КА до змін умов функціонування.

6. Розроблено метод вибору даних, який базується на статичному оцінюванні ознак або фіксованих моделях ранжування, використовує інтегральну оптимізаційну функцію з урахуванням глибини горизонту, семантичної значущості та попереднього досвіду використання, доповнену механізмом самонавчання для адаптивного формування та корекції інтегрованих сукупностей даних, що дозволило підвищити точність відбору, зменшити інформаційну надлишковість і забезпечити динамічну адаптацію до змін середовища.

7. Розроблено засоби для зберігання, оптимізації та вибору даних для систем та засобів протидії КА і КМ і для експериментального дослідження таких засобів з метою покращення їх характеристик, здійснено постановку експерименту, проведено експериментальні дослідження, які дозволили встановити підвищення ефективності виявлення КА.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Biswas J. Analysis of client honeypots. (*IJCSIT*) *international journal of computer science and information technologies*. 2014. Vol. 5, no. 4. P. 5776–5780. URL: <https://ijcsit.com/docs/Volume%205/vol5issue04/ijcsit20140504209.pdf>
2. Viola V. From honeypots to distributed deception platforms: theory and testing of emerging technologies for IT security. 2019. URL: <https://api.semanticscholar.org/CorpusID:226826721>
3. Svanadze V., Gnatyuk S. Challenges and solutions for cybersecurity and information security management in organizations. *CEUR-WS*. 2024. Vol. 3654. P. 497–504. URL: <https://ceur-ws.org/Vol-3654/short20.pdf>
4. An adaptive honeypot deployment algorithm based on learning automata / Y. Zhang et al. 2017 *IEEE second international conference on data science in cyberspace (DSC)*, Shenzhen, China, 26–29 June 2017. 2017. URL: <https://doi.org/10.1109/dsc.2017.52>
5. Karpinski M., Kuznetsov O., Oliynykov R. Security, privacy, confidentiality, and trust in the blockchain: from theory to applications. *Electronics*. 2025. Vol. 14, no. 3. P. 581. URL: <https://doi.org/10.3390/electronics14030581>
6. Ткачов В.М., Коваленко А.А., Фесенко Т.Г. Оптимізація мережного алгоритму функціонування комп'ютерних мереж підвищеної живучості на мобільній платформі на етапі їх проектування. *Системи управління, навігації та зв'язку. Збірник наукових праць*. Полтава: ПНТУ, 2021. Т. 3 (65). С. 143-147. DOI: <https://doi.org/10.26906/SUNZ.2021.3.143>
7. Yevseiev S., Melenti Y., Voitko O., Hrebenuk V., Korchenko A., Mykus S., Milov O., Prokopenko O., Sievierinov O., Chopenko D. Development of a concept for building a critical infrastructure facilities security system, *Eastern-European Journal of Enterprise Technologies*, 2021, vol. 3, pp. 63–83. DOI: <https://doi.org/10.15587/1729-4061.2021.233533>
8. Lobanchykova N., Pilkevych I., Korchenko O. Analysis of attacks on components of IoT systems and cybersecurity technologies. *QuaInT 2021: Workshop on the Quantum Information Technologies, April 11, 2021, Zhytomyr, Ukraine*. *CEUR-WS*. 2021. Vol. 2850. P. 83–96. URL: <https://ceur-ws.org/Vol-2850/paper6.pdf>

9. Летичевський О. О. Сучасні наукові проблеми кібербезпеки. *Visnik Nacionalnoi akademii nauk Ukraini*. 2023. № 2. С. 12–20. DOI: <https://doi.org/10.15407/visn2023.02.012>
10. Mukhin V., Kornaga Y., Zavgorodnii V., Bazaka Y., Zavgorodnya A., Mukhin O., Method of Data Processing System Synthesis for Heterogeneous Distributed Databases Based on Network-Centric Control. *2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Dortmund, Germany, 2023, pp. 607-612, DOI: <https://doi.org/10.1109/IDAACS58523.2023.10348667>
11. Rowe N. C. Honey-pot deception tactics. *Autonomous cyber deception*. Cham, 2019. P. 35–45. DOI: https://doi.org/10.1007/978-3-030-02110-8_3
12. 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>
13. Kashtalian A., Sochor T. K-Means clustering of honeynet data with unsupervised representation learning. International workshop on intelligent information technologies & systems of information security. *CEUR workshop proceedings*. Vol. 2853. 2021. P. 539–549. URL: <https://ceur-ws.org/Vol-2853/paper48.pdf>
14. Security and availability models for smart building automation systems / V. Kharchenko et al. *International Journal of Computing*. 2017. Vol. 16(4). P. 194–202. DOI: <https://doi.org/10.47839/ijc.16.4.907>
15. Model-driven deception for control system environments / W. Hofer et al. 2019 *IEEE international symposium on technologies for homeland security (HST)*, Woburn, MA, USA, 5–6 November 2019. 2019. DOI: [10.1109/HST47167.2019.9032927](https://doi.org/10.1109/HST47167.2019.9032927)
16. Khoroshko V., Khokhlachova Y., Vyshnevskaya N. Decomposition of computer network technology in their design. *Ukrainian scientific journal of information security*. 2023. Vol. 29, no. 3. P. 130–137. DOI: <https://doi.org/10.18372/2225-5036.29.18072>
17. Walter E. C., Ferguson-Walter K. J., Ridley A. D. Incorporating deception into cyberbattlesim for autonomous defense. 2021. DOI: <https://doi.org/10.48550/arXiv.2108.13980>

18. Gnatyuk S., Sydorenko V., Yudin O., Zhyharevych O., Polozhentsev A. Method for calculating the criticality level of sectoral information and telecommunication systems. *CEUR Workshop Proc.* 2022. Vol. 3347 Pp. 234–245. URL: https://ceur-ws.org/Vol-3347/Paper_20.pdf

19. Про основні засади забезпечення кібербезпеки України : Закон України від 05.10.2017 № 2163-VIII. URL: <https://zakon.rada.gov.ua/laws/show/2163-19#Text>

20. Методичні рекомендації щодо забезпечення кіберзахисту автоматизованих систем управління технологічними процесами : Наказ від 29.05.2023 № 463. URL: <https://zakon.rada.gov.ua/rada/show/v0463519-23#top>

21. Про захист інформації в інформаційно-комунікаційних системах : Закон України від 05.07.1994 № 80/94-ВР : станом на 16 груд. 2020 р. URL: <https://zakon.rada.gov.ua/laws/show/80/94-%D0%B2%D1%80#Text>

22. ДСТУ ISO/IEC 27001:2023. Інформаційна безпека, кібербезпека та захист конфіденційності. Системи керування інформаційною безпекою. Вимоги (ISO/IEC 27001:2022, IDT). Вид. офіц.

23. ДСТУ ISO/IEC 27034-1:2017. Інформаційні технології. Методи захисту. Безпека прикладних програм. Частина 1. Огляд і загальні поняття (ISO/IEC 27034-1:2011; Cor 1:2014, IDT). Вид. офіц.

24. ДСТУ ISO/IEC 30111:2016. Інформаційні технології. Методи захисту. Процеси оброблення вразливостей (ISO/IEC 30111:2013, IDT). Вид. офіц.

25. ДСТУ ISO/IEC TR 24774:2016. Інженерія систем і програмних засобів. Керування життєвим циклом. Настанови щодо опису процесу (ISO/IEC TR 24774:2010, IDT). Вид. офіц.

26. ДСТУ ISO/IEC 25001:2016. Інженерія систем і програмних засобів. Вимоги до якості систем і програмних засобів та її оцінювання (SQuaRE). Планування та керування (ISO/IEC 25001:2014, IDT). Вид. офіц.

27. ДСТУ ISO/IEC 15026-3:2018. Інженерія систем і програмних засобів. Гарантії стосовно систем і програмних засобів. Частина 3. Рівні цілісності системи (ISO/IEC 15026-3:2015, IDT). Вид. офіц.

28. ДСТУ ISO/IEC 15026-4:2018. Інженерія систем і програмних засобів. Гарантії стосовно систем і програмних засобів. Частина 4. Гарантування в життєвому циклі (ISO/IEC 15026-4:2012, IDT). Вид. офіц.

29. ДСТУ ISO/IEC 15026-2:2018. Інженерія систем і програмних засобів. Гарантії стосовно систем і програмних засобів. Частина 2. Сценарій гарантування (ISO/IEC 15026-2:2011, IDT). Вид. офіц.

30. Snort. Network Intrusion Detection & Prevention System [Електронний ресурс]. – Режим доступу: <https://www.snort.org/>. – Дата звернення: 22.07.2026.

31. Suricata. Open Source IDS/IPS/NSM Engine [Електронний ресурс]. – Режим доступу: <https://suricata.io/>. – Дата звернення: 22.07.2026.

32. Zeek. The Network Security Monitor [Електронний ресурс]. – Режим доступу: <https://zeek.org/>. – Дата звернення: 22.07.2026.

33. Cisco Secure Network Analytics [Електронний ресурс]. – Режим доступу: <https://www.cisco.com/site/us/en/products/security/security-analytics/secure-network-analytics/index.html>. – Дата звернення: 22.07.2026.

34. Darktrace. AI Cyber Security Platform [Електронний ресурс]. – Режим доступу: <https://darktrace.com/>. – Дата звернення: 22.07.2026.

35. Vectra AI. AI-Driven Network Detection and Response Platform [Електронний ресурс]. – Режим доступу: <https://www.vectra.ai/>. – Дата звернення: 22.07.2026.

36. Corelight. Open Network Detection and Response Platform [Електронний ресурс]. – Режим доступу: <https://corelight.com/>. – Дата звернення: 22.07.2026.

37. IBM QRadar Network Detection and Response [Електронний ресурс]. – Режим доступу: <https://www.ibm.com/products/qradar-siem/ndr>. – Дата звернення: 22.07.2026.

38. Microsoft Defender XDR [Електронний ресурс]. – Режим доступу: <https://www.microsoft.com/security/business/siem-and-xdr/microsoft-defender-xdr>. – Дата звернення: 22.07.2026.

39. Palo Alto Networks. Cortex XDR [Електронний ресурс]. – Режим доступу: <https://www.paloaltonetworks.com/cortex/cortex-xdr>. – Дата звернення: 22.07.2026.

40. Trend Micro Vision One [Электронный ресурс]. – Режим доступа: https://www.trendmicro.com/en_us/business/products/detection-response.html. – Дата звернення: 22.07.2026.
41. Logpoint. Network Detection and Response [Электронный ресурс]. – Режим доступа: <https://www.logpoint.com/>. – Дата звернення: 22.07.2026.
42. OPSWAT MetaDefender Network Detection and Response [Электронный ресурс]. – Режим доступа: <https://www.opswat.com/products/network-detection-and-response>. – Дата звернення: 22.07.2026.
43. Secureworks Taegis XDR [Электронный ресурс]. – Режим доступа: <https://www.secureworks.com/products/taegis-xdr>. – Дата звернення: 22.07.2026.
44. NETSCOUT Omnis Cyber Intelligence [Электронный ресурс]. – Режим доступа: <https://www.netscout.com/product/cyber-intelligence>. – Дата звернення: 22.07.2026.
45. Savenko O., Sachenko A., Lysenko S., Markowsky G., Vasylykiv N. BOTNET DETECTION APPROACH BASED ON THE DISTRIBUTED SYSTEMS. *International Journal of Computing*. 2020. № 19(2), 190-198. DOI: <https://doi.org/10.47839/ijc.19.2.1761>
46. Bobrovnikova K., Lysenko S., Savenko B., Gaj P., Savenko O. Technique for IoT malware detection based on control flow graph analysis. *Radioelectronic and Computer Systems*, 2022(1), pp. 141–153. DOI: <https://doi.org/10.32620/reks.2022.1.11>
47. Savenko O., Nicheporuk A., Hurman I., Lysenko S. Dynamic signature-based malware detection technique based on API call tracing. *CEUR-WS*. 2019. Vol. 2393. P.633-643, ISSN: 1613-0073. URL: https://ceur-ws.org/Vol-2393/paper_278.pdf
48. Lysenko S., Bobrovnikova K., Matiukh S., Hurman I., Savenko O. Detection of the botnets' low-rate DDoS attacks based on self-similarity. *International Journal of Electrical and Computer Engineering*, Vol. 10, Issue 4, 2020. Pp. 3651-3659. DOI: <http://doi.org/10.11591/ijece.v10i4>.
49. Lysenko S., Savenko O., Bobrovnikova K., Kryshchuk A., Savenko B. Information technology for botnets detection based on their behaviour in the corporate area

network / Communications in Computer and Information Science, ISSN: 1865–0929. – 2017. – Vol. 718. – Pp. 166–181. DOI: http://doi.org/10.1007/978-3-319-59767-6_14

50. Savenko B., Lysenko S., Bobrovnikova K., Savenko O., Markowsky G. Detection DNS Tunneling Botnets. *Proceedings of the 2021 IEEE 11th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, IDAACS'2021, Cracow, Poland, September 22-25, 2021. DOI: <https://doi.org/10.1109/IDAACS58523.2023.10348773>

51. Koukaras P., Tjortjis C. Data Organisation for Efficient Pattern Retrieval: Indexing, Storage, and Access Structures. *Big Data and Cognitive Computing*. 2025; 9(10):258. DOI: <https://doi.org/10.3390/bdcc9100258>

52. Koukaras P., Tjortjis C. Data Preprocessing and Feature Engineering for Data Mining: Techniques, Tools, and Best Practices. *AI*. 2025; 6(10):257. DOI: <https://doi.org/10.3390/ai6100257>

53. Koukaras P. Data Integration and Storage Strategies in Heterogeneous Analytical Systems: Architectures, Methods, and Interoperability Challenges. *Information*. 2025; 16(11):932. DOI: <https://doi.org/10.3390/info16110932>

54. Karapiperis D., Tjortjis C., Verykios V. (2026). LSBlock: A Hybrid Blocking System Combining Lexical and Semantic Similarity Search for Record Linkage. In: Chrysanthos P.K., Nørvåg K., Stefanidis K., Zhang Z. (eds) *Advances in Databases and Information Systems. ADBIS 2025. Lecture Notes in Computer Science*, vol 16043. Springer, Cham. DOI: https://doi.org/10.1007/978-3-032-05281-0_9

55. Kandpal M., Goswami V., Priyadarshini R., Barik R.K. Towards Data Storage, Scalability, and Availability in Blockchain Systems: A Bibliometric Analysis. *Data*. 2023; 8(10):148. DOI: <https://doi.org/10.3390/data8100148>

56. Kandpal M., Goswami V., Pritwani Y., Barik R.K., Saikia M.J. *BS-GeoEduNet 1.0*: Blockchain-Assisted Serverless Framework for Geospatial Educational Information Networks. *ISPRS International Journal of Geo-Information*. 2024; 13(8):274. DOI: <https://doi.org/10.3390/ijgi13080274>

57. Oliveira N.R., Santos Y.R., Mendes A.C.R., Barbosa G.N.N., Oliveira M.T., Valle R., Medeiros D.S.V., Mattos D.M.F. Storage Standards and Solutions, Data Storage,

Sharing, and Structuring in Digital Health: A Brazilian Case Study. *Information*. 2024; 15(1):20. DOI: <https://doi.org/10.3390/info15010020>

58. Silva J.V.V., de Oliveira N.R., Medeiros D.S.V. *et al.* A statistical analysis of intrinsic bias of network security datasets for training machine learning mechanisms. *Ann. Telecommun.*. 2022. 77, 555–571. DOI: <https://doi.org/10.1007/s12243-021-00904-5>

59. Saif M. B., Migliorini S., Spoto F. Efficient and Secure Distributed Data Storage and Retrieval Using Interplanetary File System and Blockchain. *Future Internet*. 2024; 16(3):98. DOI: <https://doi.org/10.3390/fi16030098>

60. Saif M.B., Migliorini S., Spoto F. A Survey on Data Availability in Layer 2 Blockchain Rollups: Open Challenges and Future Improvements. *Future Internet*. 2024; 16(9):315. DOI: <https://doi.org/10.3390/fi16090315>

61. Anwar R.W., Qureshi K.N., Nagmeldin W., Abdelmaboud A., Ghafoor K.Z., Javed I.T., Crespi N. Data Analytics, Self-Organization, and Security Provisioning for Smart Monitoring Systems. *Sensors*. 2022; 22(19):7201. DOI: <https://doi.org/10.3390/s22197201>

62. Qureshi, K.N., Ahmad, A., Piccialli, F. *et al.* Nature-inspired algorithm-based secure data dissemination framework for smart city networks. *Neural Comput & Applic* 33, 10637–10656. 2021. DOI: <https://doi.org/10.1007/s00521-020-04900-z>

63. Lu J., Bai J., Shen S. Design and Implementation of a High-Speed Storage System Based on SATA Interface. *Electronics*. 2026; 15(2):452. DOI: <https://doi.org/10.3390/electronics15020452>

64. Li Q., Shan H., Feng Y., Teng L. The Design and Implementation of a High-Speed and Large-Capacity NAND Flash Storage System. *Applied Mechanics and Materials*. 2014. 543-547. 568-571. DOI: 10.4028/www.scientific.net/AMM.543-547.568

65. Wang Y., Fan R., Liang X., Li P., Hei X. Trusted Data Storage Architecture for National Infrastructure. *Sensors*. 2022; 22(6):2318. DOI: <https://doi.org/10.3390/s22062318>

66. Gao W., Hei X., Wang Y. The Data Privacy Protection Method for Hyperledger Fabric Based on Trustzone. *Mathematics*. 2023; 11(6):1357. DOI: <https://doi.org/10.3390/math11061357>

67. Talluri L.S.R.K., Kommana K., Tella S.S., Nandyala P., Mandhala V.N. Efficient Cryptographic Technique to Secure Data in Cloud Environment. In: Bhattacharyya, D., Thirupathi Rao, N. (eds) Machine Intelligence and Soft Computing. *Advances in Intelligent Systems and Computing*. 2021. Vol 1280. Springer, Singapore. DOI: https://doi.org/10.1007/978-981-15-9516-5_31
68. Talluri L.S.R.K., Thirumalaisamy R., Kota R., Sadi R.P.R., K.C.U., Naha R.K., Mahanti A. Providing Consistent State to Distributed Storage System. *Computers*. 2021; 10(2):23. DOI: <https://doi.org/10.3390/computers10020023>
69. Khan A.Q., Nikolov N., Matskin M., Prodan R., Roman D., Sahin B., Bussler C., Soylu A. Smart Data Placement Using Storage-as-a-Service Model for Big Data Pipelines. *Sensors*. 2023; 23(2):564. DOI: <https://doi.org/10.3390/s23020564>
70. Corodescu A.-A., Nikolov N., Khan A.Q., Soylu A., Matskin M., Payberah A.H., Roman D. Big Data Workflows: Locality-Aware Orchestration Using Software Containers. *Sensors*. 2021; 21(24):8212. DOI: <https://doi.org/10.3390/s21248212>
71. Alsulbi KA., Khemakhem MA., Basuhail AA., Eassa FE., Jambi KM., Almarhabi KA. A Proposed Framework for Secure Data Storage in a Big Data Environment Based on Blockchain and Mobile Agent. *Symmetry*. 2021; 13(11):1990. DOI: <https://doi.org/10.3390/sym13111990>
72. Almasabi A.M., Alkhodre A.B., Khemakhem M., Eassa F., Abi Sen A.A., Harbaoui A. Internet of Things-Based Anomaly Detection Hybrid Framework Simulation Integration of Deep Learning and Blockchain. *Information*. 2025; 16(5):406. DOI: <https://doi.org/10.3390/info16050406>
73. Kashtalian A., Lysenko S., Savenko O., Nicheporuk A., Sochor T., Avsiyevych V. Multi-computer malware detection systems with metamorphic functionality. *Radioelectronic and Computer Systems*. 2024. 2024(1), Pp. 152-175. DOI: <https://doi.org/10.32620/reks.2024.1.13>
74. 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. 24(1), 35-51. DOI: <https://doi.org/10.47839/ijc.24.1.3875>

75. Герасіна О. В. Моделювання трафіку інформаційно-комунікаційних мереж із використанням блочно-орієнтованих структур для систем виявлення атак. *Вісник Херсонського національного технічного університету*. 2025. Т. 2, № 3 (94). С. 131–138. DOI: <https://doi.org/10.35546/kntu2078-4481.2025.3.2.16>

76. Герасіна О. В. Фрактальний аналіз трафіку інформаційно-комунікаційних мереж для систем виявлення атак. *Information Technology: Computer Science, Software Engineering and Cyber Security*. 2025. № 3. С. 62–70. DOI: <https://doi.org/10.32782/IT/2025-3-7>

77. Celebi M. Efficiency optimization of a conventional boost DC/DC converter. *Electrical Engineering*, Vol. 100, Iss. 2, pp. 803-809. DOI: 10.1007/s00202-017-0552-0

78. Балацька В., Ткачук, Р., Маслова, Н. Еволюція КСЗІ та інтеграція блок-чейн-технологій у кіберзахисті державних інформаційних систем України. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*. 2025. № 2(30), С. 316–332. <https://doi.org/10.28925/2663-4023.2025.30.975>

79. Яцків В. В., Івасьєв С. В., Давлетова А. Я., Тимошенко Л. М. Методологія впровадження системи управління інформаційною безпекою на основі багаторівневої моделі кіберзахисту згідно з вимогами законодавства України. *Інформатика та математичні методи в моделюванні*. 2025. Т. 15, № 1. С. 137–149. DOI: <https://doi.org/10.15276/imms.v15.no1.137>

80. Тітова В. Ю., Кльоц Ю. П., Лакоцін З., Шлапак О., Троц В. Порівняльний аналіз моделей атак на інформаційну безпеку. *Measuring and Computing Devices in Technological Processes*. 2024. № 4. С. 174–178. DOI: <https://doi.org/10.31891/2219-9365-2024-80-21>

81. Yatskiv V., Kulyna S., Ivasiev S. Data Encryption Method Based on the McEliece Cryptosystem and the Redundant Residue Number System. *Advances in Cyber-Physical Systems*. 2025. Vol. 10, No. 2. P. 214–222. DOI: <https://doi.org/10.23939/acps2025.02.214>

82. Кльоц Ю. П., Мостовий С. В., Сікорський П. І., Остапчук І. О. Система виявлення аномалій у DNS-запитах. *Herald of Khmelnytskyi National University. Technical Sciences*. 2024. Т. 345, № 6(2). С. 141–148. DOI: <https://doi.org/10.31891/2307-5732-2024-345-6-22>

83. Kashtalian, A., Lysenko, S., Sachenko, A., Savenko, B., Savenko, O., & Nicheporuk, A. (2025). Evaluation criteria of centralization options in the architecture of multicomputer systems with traps and baits. *Radioelectronic and Computer Systems*, 2025(1), 264-297. DOI: <https://doi.org/10.32620/reks.2025.1.18>
84. Wang Y., Wang H., Cao Y. Comprehensive Review of Storage Optimization Techniques in Blockchain Systems. *Applied Sciences*, 2025. 15(1), 243. <https://doi.org/10.3390/app15010243>
85. Wang Q., Yu D., Zhou J., Jin C. Data Storage Optimization Model Based on Improved Simulated Annealing Algorithm. *Sustainability*, 2023. 15(9), 7388. <https://doi.org/10.3390/su15097388>
86. Jiang S. Big Data Sharing: A Comprehensive Survey. *Data*, 2025. 10(11), 182. <https://doi.org/10.3390/data10110182>
87. Abdullah M., Nawaz M. M., Saleem B., Zahra M., Ashfaq E. b., Muhammad Z. Evolution Cybercrime—Key Trends, Cybersecurity Threats, and Mitigation Strategies from Historical Data. *Analytics*, 2025. 4(3), 25. <https://doi.org/10.3390/analytics4030025>
88. Bennouk K., Ait Aali N., El Bouzekri El Idrissi Y., Sebai B., Faroukhi A. Z., Mahouachi D. A Comprehensive Review and Assessment of Cybersecurity Vulnerability Detection Methodologies. *Journal of Cybersecurity and Privacy*, 2024. 4(4), 853-908. <https://doi.org/10.3390/jcp4040040>
89. Akindote Odunayo. Adegbite Abimbola, Dawodu Samuel, Omotosho Adedolapo, Anyanwu Anthony. INNOVATION IN DATA STORAGE TECHNOLOGIES: FROM CLOUD COMPUTING TO EDGE COMPUTING. *Computer Science & IT Research Journal*. 2023. 4. 273-299. 10.51594/csitrj.v4i3.661.
90. Cieslik K., Margócsy D. Datafication, power and control in development: A historical perspective on the perils and longevity of data. *Progress in Development Studies*, 2022. 22(4), 352-373. DOI: 10.1177/14649934221076580
91. Gupta V., Verma V., Kaur S., Kaur I. Amalgamation of Cloud computing with big data. In 2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC). 2022. Pp. 67-70) IEEE. DOI: 10.1109/PDGC56933.2022.10053276

92. Wilhelm A., Ziegler W. Extending semantic context analysis using machine learning services to process unstructured data. In SHS web of conferences. 2021. Vol. 102, p. 02001. EDP Sciences. DOI: [10.1051/SHSCONF/202110202001](https://doi.org/10.1051/SHSCONF/202110202001)

93. Герасіна О. В. Нейро-нечітке прогнозування самоподібного трафіку інформаційно-комунікаційних мереж для систем виявлення атак. *Системні технології*. 2025. № 6 (161). С. 3–17. DOI: <https://doi.org/10.34185/1562-9945-5-161-2025-01>

94. Klots Y. P., Petliak N. S., Titova V. Y. A Hybrid Method for Detecting Anomalous Traffic in Computer Networks. *Applied Aspects of Information Technology*. 2025. Vol. 8, No. 2. P. 191–201. DOI: <https://doi.org/10.15276/aait.08.2025.13>

95. Savenko O., Rusyn B., Lysenko S., Ciszewski T., Savenko B., Drozd A., Nicheporuk A., Sachenko A. Synthesis of a Moth and Flame Algorithm for Incorporation into the Architecture of Deceptive Systems with Baits and Traps. *Applied Sciences*. 2026. № 16(5), 2415. DOI: <https://doi.org/10.3390/app16052415>

96. 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>

97. 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>

98. 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>

100. Szczypiorski K. Cybersecurity and Data Science. *Electronics*. 2022; 11(15):2309. DOI: <https://doi.org/10.3390/electronics11152309>

101. 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>

102. 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>.

103. 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>

104. 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>

105. 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

106. 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>

107. 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>.

108. 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>

109. Івасьєв С. В., Бараннік Б. О., Цаволик Т. Г. Побудова віртуальної машини для динамічного аналізу шкідливого програмного забезпечення. *Herald of Khmelnytskyi National University. Technical Sciences*. 2026. Т. 363, № 2. С. 379–385. DOI: <https://doi.org/10.31891/2307-5732-2026-363-5>

110. Isaiev T., Kysil T. Method of Creating Custom Dataset to Train Convolutional Neural Network. *Computer Systems and Information Technologies*. 2024. № 4. С. 37–44. DOI: <https://doi.org/10.31891/csit-2024-4-5>

111. Паюк В.П., Савенко Б. О. Організація зберігання даних в системах та засобах протидії комп'ютерним атакам з урахуванням історичного досвіду їх

використання. *Вимірювальна та обчислювальна техніка в технологічних процесах*. 2026. № 2. С. 321–336. DOI: <https://doi.org/10.31891/2219-9365-2026-86-39>

112. PAIUK V., KASHTALIAN A. SELECTION OF DATA IN SYSTEMS AND TOOLS FOR COMPUTER ATTACK DEFENSE TAKING INTO ACCOUNT THE HISTORICAL ASPECT OF THEIR USE. *Computer Systems and Information Technologies*. 2026. № 3. Pp. 186–204. DOI: <https://doi.org/10.31891/csit-2026-3-19>

113. Паюк В. П., Савенко Б. О. Оптимізація зберігання даних в системах та засобах протидії комп'ютерним атакам з урахуванням історичного аспекту їх використання. *Вимірювальна та обчислювальна техніка в технологічних процесах*. 2026. № 3. С. 217–236. DOI: <https://doi.org/10.31891/2219-9365-2026-87-26>

114. Паюк В. П., Гейдарова О. В., Косенков В. Д., Савенко О. С. Застосування багатофакторного аналізу з метою виявлення незадокументованих закладок програмного забезпечення в локальних комп'ютерних мережах. *Вісник Хмельницького національного університету. Технічні науки*. 2020. № 5. С. 280-285. DOI 10.31891/2307-5732-2020-289-5-280-285

115. Sierhieiev Y., Paiuk V., Sachenko A., Nicheporuk A., Kwiecien A. A graph-based vulnerability detection method. (2024) *CEUR Workshop Proceedings*, 3675, pp. 388-401. *The 5th International Workshop on Intelligent Information Technologies & Systems of Information Security (IntelITSIS-2024)* : CEUR-Workshop Proceedings. Vol. 3675. (Khmelnyskyi, March 2024). Khmelnytskyi, 2024. Pp. 343-355. URL: <https://ceur-ws.org/Vol-3675/>

116. Sierhieiev Y., Paiuk V., Nicheporuk A., Kwiecien A., Huralnyk O. Detection and prediction of the vulnerabilities in software systems based on behavioral analysis with machine learning. *The 1th Proceedings of the 1st International Workshop on Intelligent & CyberPhysical Systems (ICyberPhyS 2024) Khmelnytskyi, Ukraine, June 28, 2024* : CEUR-Workshop Proceedings. Vol. 3736. 2024. Pp. 239-254. URL: <https://ceur-ws.org/Vol-3736/>

117. Sierhieiev Y., Savenko O., Paiuk V., Drozd A. Effectiveness and improvement of Static Application Security Testing (SAST) in the context of SQL Injection vulnerabilities // *Proceedings of 2024 IEEE 14th International Conference on Dependable Systems, Services and Technologies (DeSSerT-2024, Athens, Greece, October 11-13, 2024)*.

DOI: [10.1109/DESSERT65323.2024.11122171](https://doi.org/10.1109/DESSERT65323.2024.11122171)

118. Huralnyk O., Kwiecien A., Paiuk V., Klein O., Lyhun O. Method and Means of Critical Information Processing in Corporate Networks. *Intelitsis '25: The 6th International Workshop on Intelligent Information Technologies & Systems of Information Security*, Aprile 04, 2025, Khmelnytskyi, Ukraine. Pp. 314-328. URL: <https://ceur-ws.org/Vol-3963/paper25.pdf>

119. Stetsyuk Y., Stetsyuk M., Voznyi K., Paiuk V., Kvassay M. A model of a centralized security system, as an information technology for the synthesis of an OS architecture protected against the leakage of confidential information. *1st International Workshop on Advanced Applied Information Technologies: (AdvAIT-2024)*, Khmelnytskyi, Ukraine and Zilina, Slovakia, December 5, 2024. Vol. 3899. P. 224-233. URL: <https://ceur-ws.org/Vol-3899/paper20.pdf>

120. Paiuk V., Kosenkov V., Savenko O., Nicheporuk A., Geidarova O. Application of multifactor analysis for the purpose of detecting malicious software implants of the software in local computer networks. *2nd International Workshop on Intelligent Information Technologies and Systems of Information Security*, March 24–26, 2021. *Ceur Workshop Proceedings*. Vol. 2853. Pp. 517–526. URL: <https://ceur-ws.org/Vol-2853/paper46.pdf>

121. Nicheporuk A., Paiuk V., Savenko B., Savenko O., Geidarova O. Detecting Software Malicious Implant Based on Anomalies Research on Local Area Networks. *Ceur Workshop Proceedings, 1nd International Workshop on Intelligent Information Technologies and Systems of Information Security*, June 10-12, 2020. Khmelnytskyi, Ukraine 2020, Vol. 2623. Pp. 194–207. URL: <https://ceur-ws.org/Vol-2623/paper18.pdf>

122. Свідоцтво про реєстрацію авторського права на твір (програму) № 150528 Україна. Комп'ютерна програма «Програмне забезпечення зберігання, оптимізації і вибору даних в системах та засобах протидії комп'ютерним атакам з урахуванням історичного аспекту їх використання та глибини горизонту» («SECMON») / Нічепорук А. О., Паюк В.П., Савенко Б.О., Каштальян А.С. Дата реєстрації 03.09.2026.

ДОДАТКИ

ДОДАТОК А

Список публікацій здобувача

Наукові праці, в яких опубліковані основні наукові результати дисертації

1. Паюк В.П., Савенко Б.О. Організація зберігання даних в системах та засобах протидії комп'ютерним атакам з урахуванням історичного досвіду їх використання. *Вимірювальна та обчислювальна техніка в технологічних процесах*. 2026. № 2. С. 321–336. DOI: <https://doi.org/10.31891/2219-9365-2026-86-39>

2. PAIUK V., KASHTALIAN A. SELECTION OF DATA IN SYSTEMS AND TOOLS FOR COMPUTER ATTACK DEFENSE TAKING INTO ACCOUNT THE HISTORICAL ASPECT OF THEIR USE. *Computer Systems and Information Technologies*. 2026. № 3. Pp. 186–204. DOI: <https://doi.org/10.31891/csit-2026-3-19>

3. Паюк В. П., Савенко Б.О. Оптимізація зберігання даних в системах та засобах протидії комп'ютерним атакам з урахуванням історичного аспекту їх використання. *Вимірювальна та обчислювальна техніка в технологічних процесах*. 2026. № 3. С. 217–236. DOI: <https://doi.org/10.31891/2219-9365-2026-87-26>

4. Паюк В. П., Гейдарова О. В., Косенков В. Д., Савенко О. С. Застосування багатofакторного аналізу з метою виявлення незадокументованих закладок програмного забезпечення в локальних комп'ютерних мережах. *Вісник Хмельницького національного університету. Технічні науки*. 2020. № 5. С. 280-285. DOI 10.31891/2307-5732-2020-289-5-280-285

Праці, які засвідчують апробацію матеріалів дисертації

5. Sierhieiev Y., Paiuk V., Sachenko A., Nicheporuk A., Kwiecien A. A graph-based vulnerability detection method. (2024) *CEUR Workshop Proceedings*, 3675, pp. 388-401. *The 5th International Workshop on Intelligent Information Technologies & Systems of Information Security (IntelITSIS-2024)* : CEUR-Workshop Proceedings. Vol. 3675. (Khmelnyskyi, March 2024). Khmelnyskyi, 2024. Pp. 343-355. URL: <https://ceur-ws.org/Vol-3675/>

6. Sierhieiev Y., Paiuk V., Nicheporuk A., Kwiecien A., Huralnyk O. Detection and prediction of the vulnerabilities in software systems based on behavioral analysis with machine learning. *The 1th Proceedings of the 1st International Workshop on Intelligent & CyberPhysical Systems (ICyberPhyS 2024) Khmelnytskyi, Ukraine, June 28, 2024 : CEUR-Workshop Proceedings*. Vol. 3736. 2024. Pp. 239-254. URL: <https://ceur-ws.org/Vol-3736/>
7. Sierhieiev Y., Savenko O., Paiuk V., Drozd A. Effectiveness and improvement of Static Application Security Testing (SAST) in the context of SQL Injection vulnerabilities // *Proceedings of 2024 IEEE 14th International Conference on Dependable Systems, Services and Technologies (DeSSerT-2024, Athens, Greece, October 11-13, 2024)*. DOI: [10.1109/DESSERT65323.2024.11122171](https://doi.org/10.1109/DESSERT65323.2024.11122171)
8. Huralnyk O., Kwiecien A., Paiuk V., Klein O., Lyhun O. Method and Means of Critical Information Processing in Corporate Networks. *Intelitsis '25: The 6th International Workshop on Intelligent Information Technologies & Systems of Information Security*, Aprile 04, 2025, Khmelnytskyi, Ukraine. Pp. 314-328. URL: <https://ceur-ws.org/Vol-3963/paper25.pdf>
9. Stetsyuk Y., Stetsyuk M., Voznyi K., Paiuk V., Kvassay M. A model of a centralized security system, as an information technology for the synthesis of an OS architecture protected against the leakage of confidential information. *1st International Workshop on Advanced Applied Information Technologies: (AdvAIT-2024)*, Khmelnytskyi, Ukraine and Zilina, Slovakia, December 5, 2024. Vol. 3899. P. 224-233. URL: <https://ceur-ws.org/Vol-3899/paper20.pdf>
10. Paiuk V., Kosenkov V., Savenko O., Nicheporuk A., Geidarova O. Application of multifactor analysis for the purpose of detecting malicious software implants of the software in local computer networks. *2nd International Workshop on Intelligent Information Technologies and Systems of Information Security*, March 24–26, 2021. *Ceur Workshop Proceedings*. Vol. 2853. Pp. 517–526. URL: <https://ceur-ws.org/Vol-2853/paper46.pdf>
11. Nicheporuk A., Paiuk V., Savenko B., Savenko O., Geidarova O. Detecting Software Malicious Implant Based on Anomalies Research on Local Area Networks. *Ceur Workshop Proceedings, 1nd International Workshop on Intelligent Information*

Technologies and Systems of Information Security, June 10-12, 2020. Khmelnytskyi, Ukraine 2020, Vol. 2623. Pp. 194–207. URL: <https://ceur-ws.org/Vol-2623/paper18.pdf>

Публікації, які додатково відображають наукові результати дисертації

12. Свідоцтво про реєстрацію авторського права на твір (програму) № 150528 Україна. Комп'ютерна програма «Програмне забезпечення зберігання, оптимізації і вибору даних в системах та засобах протидії комп'ютерним атакам з урахуванням історичного аспекту їх використання та глибини горизонту» («SECMON») / Нічепорук А. О., Паюк В.П., Савенко Б.О., Каштальян А.С. Дата реєстрації 03.09.2026.

ДОДАТОК Б
АКТИ ВПРОВАДЖЕННЯ



АКТ

про впровадження результатів дисертаційної роботи
аспіранта кафедри комп'ютерної інженерії та інформаційних систем Хмельницького
національного університету Паюка Вадима Павловича
«Методи та засоби зберігання, вибору та оптимізації історичних даних систем і засобів
протидії комп'ютерним атакам в корпоративних мережах»

Результати дисертаційної роботи Паюка В. аспіранта кафедри комп'ютерної інженерії та інформаційних систем Хмельницького національного університету впроваджені на ТОВ «ІТТ».

При розробленні засобів зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах були отримані на ТОВ «ІТТ» такі результати, які одержані Паюком В. особисто:

1) архітектура засобів зберігання даних для систем і засобів протидії КА, в якій здійснено поєднання автоматичного керування життєвим циклом всіх даних або їх частини, особливостей апаратно-програмних засобів для зберігання даних, інтелектуалізації щодо глибини горизонту для даних, самонавчання при виборі даних для типових повторюваних КА, що дає змогу підвищити оперативність та точність при наданні даних системам протидії КА, а також забезпечити ефективність їх зберігання з урахування глибини горизонту;

2) метод організації зберігання даних у системах та засобах протидії комп'ютерним атакам, особливістю якого є поділ даних на початкові, накопичувані та використані, а також врахування історичного аспекту їх застосування шляхом збереження результатів оцінювання ефективності використання даних із застосуванням марковських моделей, динамічних баєсівських мереж та статистичних методів, що забезпечує опис динаміки станів системи, врахування причинно-наслідкових залежностей і узагальнення накопиченого досвіду для підвищення ефективності повторного використання даних, адаптивності систем та автономності їх функціонування в умовах тривалої експлуатації.

Отриманні результати дозволили покращити протидію зловмисним діям у корпоративній мережі ТОВ «ІТТ» шляхом удосконалення архітектури і методів зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах приблизно на 3-9 %.

Цей акт не є підставою для фінансових розрахунків.

Заступник директора

Іванов О.В.

Системний адміністратор

Веремєєнко В.А.

ДОВІДКА

про впровадження результатів дисертаційної роботи
аспіранта кафедри комп'ютерної інженерії та інформаційних систем
Хмельницького національного університету Паюка Вадима Павловича
«Методи та засоби зберігання, вибору та оптимізації історичних даних систем і
засобів протидії комп'ютерним атакам в корпоративних мережах»

Результати дисертаційної роботи аспіранта кафедри комп'ютерної інженерії та інформаційних систем Хмельницького національного університету Паюка В. впроваджені в ПП «Авіві».

При розробленні засобів зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах, були використані в ПП «Авіві» такі результати, які одержані Паюком В. особисто:

1) архітектура засобів зберігання даних для систем і засобів протидії КА, в якій на відміну відомих, здійснено поєднання автоматичного керування життєвим циклом всіх даних або їх частини, особливостей апаратно-програмних засобів для зберігання даних, інтелектуалізації щодо глибини горизонту для даних, самонавчання при виборі даних для типових повторюваних КА, що дає змогу підвищити оперативність та точність при наданні даних системам протидії КА, а також забезпечити ефективність їх зберігання з урахування глибини горизонту;

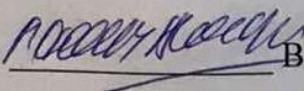
2) метод оптимізації даних на основі самонавчання засобів підсистеми пам'яті систем та засобів протидії КА, який, на відміну від існуючих, забезпечує комплексну адаптивну оптимізацію початкових, накопичених, використаних та даних обманних систем з урахуванням критеріїв ресурсної ефективності, структури даних, шарового розміщення, апаратно-мережевого розподілу та специфіки функціонування обманних систем, що дозволяє зменшити надлишковість інформаційних масивів, підвищити швидкодію доступу до даних, покращити інформативність інформаційного середовища, ефективність накопичення знань про дії зловмисника та адаптивність систем і засобів протидії КА до змін умов функціонування.

Отриманні результати дозволили покращити протидію комп'ютерним атакам у корпоративній мережі ПП «Авіві» шляхом удосконалення засобів зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах та підвищити точність виявлення приблизно на 4-9,5 %.

Ця довідка не є підставою для фінансових розрахунків.

29.06.2026 р.

Директор ПП «Авіві»



В. В. Аскеров





«Затверджую»
В.о. ректора Хмельницького
національного університету
Віктор ЛОПАТОВСЬКИЙ
«26» червня 2026 р.

АКТ

про впровадження в освітній процес Хмельницького національного університету результатів дисертаційної роботи аспіранта кафедри комп'ютерної інженерії та інформаційних систем Паюка Вадима Павловича
«Методи та засоби зберігання, вибору та оптимізації історичних даних систем і засобів протидії комп'ютерним атакам в корпоративних мережах»

Ми, комісія в складі: декана факультету інформаційних технологій, професора кафедри комп'ютерної інженерії та інформаційних систем, д.т.н., професора Говорущенко Т. О. (голова комісії), завідувача кафедри комп'ютерної інженерії та інформаційних систем, д.ф., доцента Павлової О. О., доцента кафедри комп'ютерної інженерії та інформаційних систем, к.т.н., доцента Капустян М. В. склала акт про те, що результати дисертаційної роботи Паюка В. впроваджені та використовуються в освітньому процесі Хмельницького національного університету при викладанні дисциплін на кафедрі комп'ютерної інженерії та інформаційних систем для здобувачів спеціальності F7 Комп'ютерна інженерія, зокрема в курсах «Безпека та захист комп'ютерних систем», «Моделювання та методи оптимізації в наукових та експериментальних дослідженнях», «Методології забезпечення якості, надійності, гарантоздатності та безпеки комп'ютерних систем та мереж».

При викладанні цих дисциплін викладачами кафедр використовувалися наступні матеріали досліджень, отримані Паюком В. особисто:

1) архітектура засобів зберігання даних для систем і засобів протидії КА, в якій на відміну відомих, здійснено поєднання автоматичного керування життєвим циклом всіх даних або їх частини, особливостей апаратно-програмних засобів для зберігання даних, інтелектуалізації щодо глибини горизонту для даних, самонавчання при виборі даних для типових повторюваних КА, що дає змогу підвищити оперативність та точність при надані даних системам протидії КА, а також забезпечити ефективність їх зберігання з урахування глибини горизонту;

2) метод організації зберігання даних у системах та засобах протидії комп'ютерним атакам, особливістю якого є поділ даних на початкові, накопичувані та використані, а також врахування історичного аспекту їх застосування шляхом збереження результатів оцінювання ефективності використання даних із застосуванням марковських моделей, динамічних баєсівських мереж та статистичних методів, що забезпечує опис динаміки станів системи, врахування причинно-наслідкових залежностей і узагальнення накопиченого досвіду для підвищення ефективності повторного використання даних, адаптивності систем та автономності їх функціонування в умовах тривалої експлуатації;

3) метод оптимізації даних на основі самонавчання засобів підсистеми пам'яті систем та засобів протидії КА, який, на відміну від існуючих, забезпечує комплексну адаптивну оптимізацію початкових, накопичених, використаних та даних обманих систем з урахуванням критеріїв ресурсної ефективності, структури даних, шарового розміщення, апаратно-мережевого розподілу та специфіки функціонування обманих систем, що дозволяє зменшити надлишковість інформаційних масивів, підвищити швидкодію доступу до даних, покращити інформативність інформаційного середовища, ефективність накопичення знань про дії злоумисника та адаптивність систем і засобів протидії КА до змін умов функціонування;

4) метод вибору даних, який, на відміну від відомих, що базуються на статичному оцінюванні ознак або фіксованих моделях ранжування, використовує інтегральну оптимізаційну функцію з урахуванням глибини горизонту, семантичної значущості та попереднього досвіду використання, доповнену механізмом самонавчання для адаптивного формування та корекції інтегрованих сукупностей даних, що дозволило підвищити точність відбору, зменшити інформаційну надлишковість і забезпечити динамічну адаптацію до змін середовища.

Отримані матеріали досліджень дозволили розробити лабораторні практикуми з використанням архітектури засобів зберігання даних для систем і засобів протидії КА, методу організації, методу оптимізації та методу вибору даних.



Говорущенко Т. О.

Павлова О. О.

Капустян М. В.

ДОДАТОК В

Дані експериментів

Таблиця В.1 – Дані для графіку залежності часу доступу до критичних даних від рівня оптимізації

Номер серії, k	E_{opt}	$T_{access,k}$, с	Номер серії, k	E_{opt}	$T_{access,k}$, с
1	0.02	2.07	26	0.52	1.33
2	0.04	2.04	27	0.54	1.30
3	0.06	2.01	28	0.56	1.27
4	0.08	1.98	29	0.58	1.24
5	0.10	1.95	30	0.60	1.21
6	0.12	1.92	31	0.62	1.18
7	0.14	1.89	32	0.64	1.15
8	0.16	1.86	33	0.66	1.12
9	0.18	1.83	34	0.68	1.09
10	0.20	1.80	35	0.70	1.06
11	0.22	1.77	36	0.72	1.03
12	0.24	1.74	37	0.74	1.00
13	0.26	1.72	38	0.76	0.98
14	0.28	1.69	39	0.78	0.95
15	0.30	1.66	40	0.80	0.92
16	0.32	1.63	41	0.82	0.89
17	0.34	1.60	42	0.84	0.86
18	0.36	1.57	43	0.86	0.83
19	0.38	1.54	44	0.88	0.80
20	0.40	1.51	45	0.90	0.77
21	0.42	1.48	46	0.92	0.74
22	0.44	1.45	47	0.94	0.71
23	0.46	1.42	48	0.96	0.68
24	0.48	1.39	49	0.98	0.65
25	0.50	1.36	50	1.00	0.62

Таблиця В.2 – Дані для графіку зміни швидкості реагування на КА

Номер серії, k	E_{opt}	T_{resp}, c	Номер серії, k	E_{opt}	T_{resp}, c
1	0.02	4.56	26	0.52	3.61
2	0.04	4.52	27	0.54	3.57
3	0.06	4.49	28	0.56	3.54
4	0.08	4.45	29	0.58	3.50
5	0.10	4.41	30	0.60	3.46
6	0.12	4.37	31	0.62	3.42
7	0.14	4.33	32	0.64	3.38
8	0.16	4.30	33	0.66	3.35
9	0.18	4.26	34	0.68	3.31
10	0.20	4.22	35	0.70	3.27
11	0.22	4.18	36	0.72	3.23
12	0.24	4.14	37	0.74	3.19
13	0.26	4.11	38	0.76	3.16
14	0.28	4.07	39	0.78	3.12
15	0.30	4.03	40	0.80	3.08
16	0.32	3.99	41	0.82	3.04
17	0.34	3.95	42	0.84	3.00
18	0.36	3.92	43	0.86	2.97
19	0.38	3.88	44	0.88	2.93
20	0.40	3.84	45	0.90	2.89
21	0.42	3.80	46	0.92	2.85
22	0.44	3.76	47	0.94	2.81
23	0.46	3.73	48	0.96	2.78
24	0.48	3.69	49	0.98	2.74
25	0.50	3.65	50	1.00	2.70

Таблиця В.3 – Дані для графіку залежності точності виявлення КА від показника E_{opt}

Номер серії, k	E_{opt}	K_{del}	Номер серії, k	E_{opt}	K_{del}
1	0.02	0.754	26	0.52	0.854
2	0.04	0.758	27	0.54	0.858
3	0.06	0.762	28	0.56	0.862
4	0.08	0.766	29	0.58	0.866
5	0.10	0.770	30	0.60	0.870
6	0.12	0.774	31	0.62	0.874
7	0.14	0.778	32	0.64	0.878
8	0.16	0.782	33	0.66	0.882
9	0.18	0.786	34	0.68	0.886
10	0.20	0.790	35	0.70	0.890
11	0.22	0.794	36	0.72	0.894
12	0.24	0.798	37	0.74	0.898
13	0.26	0.802	38	0.76	0.902
14	0.28	0.806	39	0.78	0.906
15	0.30	0.810	40	0.80	0.910
16	0.32	0.814	41	0.82	0.914
17	0.34	0.818	42	0.84	0.918
18	0.36	0.822	43	0.86	0.922
19	0.38	0.826	44	0.88	0.926
20	0.40	0.830	45	0.90	0.930
21	0.42	0.834	46	0.92	0.934
22	0.44	0.838	47	0.94	0.938
23	0.46	0.842	48	0.96	0.942
24	0.48	0.846	49	0.98	0.946
25	0.50	0.850	50	1.00	0.950

Таблиця В.4 – Результати експериментальних досліджень ефективності вибору історичних даних (50 серій)

№	Базова система	Варіант 1 (автоматичний горизонт)	Варіант 2 (+ самонавчання)	Варіант 3 (+ попередній досвід)
1	0.821	0.859	0.901	0.934
2	0.822	0.861	0.903	0.935
3	0.824	0.862	0.904	0.937
4	0.823	0.864	0.906	0.938
5	0.826	0.865	0.907	0.939
6	0.827	0.866	0.908	0.940
7	0.828	0.868	0.910	0.941
8	0.829	0.869	0.911	0.942
9	0.830	0.870	0.912	0.943
10	0.832	0.872	0.913	0.944
11	0.833	0.873	0.914	0.945
12	0.834	0.874	0.916	0.947
13	0.835	0.875	0.917	0.948
14	0.836	0.876	0.918	0.949
15	0.837	0.878	0.919	0.950
16	0.838	0.879	0.920	0.951
17	0.839	0.880	0.921	0.952
18	0.840	0.881	0.922	0.953
19	0.841	0.882	0.923	0.954
20	0.842	0.883	0.924	0.955
21	0.843	0.884	0.925	0.956
22	0.844	0.885	0.926	0.957
23	0.845	0.886	0.927	0.958
24	0.846	0.887	0.928	0.959
25	0.847	0.888	0.929	0.960
26	0.848	0.889	0.930	0.961
27	0.849	0.890	0.931	0.962
28	0.850	0.891	0.932	0.963
29	0.851	0.892	0.933	0.964
30	0.852	0.893	0.934	0.965
31	0.853	0.894	0.935	0.966
32	0.854	0.895	0.936	0.967
33	0.855	0.896	0.937	0.968
34	0.856	0.897	0.938	0.969
35	0.857	0.898	0.939	0.970
36	0.858	0.899	0.940	0.971
37	0.859	0.900	0.941	0.972
38	0.860	0.901	0.942	0.973
39	0.861	0.902	0.943	0.974

Кінець таблиці В.4

40	0.862	0.903	0.944	0.975
41	0.863	0.904	0.945	0.976
42	0.864	0.905	0.946	0.977
43	0.865	0.906	0.947	0.978
44	0.866	0.907	0.948	0.979
45	0.867	0.908	0.949	0.980
46	0.868	0.909	0.950	0.981
47	0.869	0.910	0.951	0.982
48	0.870	0.911	0.952	0.983
49	0.869	0.912	0.953	0.984
50	0.870	0.913	0.954	0.985

Таблиця В.5 – Результати експериментальних досліджень із даними щодо середнього часу прийняття рішення для чотирьох варіантів системи (50 серій)

№	Базова система	Варіант 1 (автоматичний горизонт)	Варіант 2 (+ самонавчання)	Варіант 3 (+ попередній досвід)
1	1,818	1,618	1,418	1,218
2	1,816	1,616	1,416	1,217
3	1,814	1,614	1,415	1,215
4	1,812	1,612	1,413	1,215
5	1,81	1,61	1,413	1,212
6	1,808	1,61	1,409	1,21
7	1,809	1,606	1,407	1,209
8	1,804	1,603	1,406	1,208
9	1,8	1,602	1,404	1,206
10	1,8	1,6	1,404	1,203
11	1,798	1,598	1,399	1,202
12	1,796	1,598	1,398	1,202
13	1,794	1,594	1,397	1,199
14	1,795	1,592	1,395	1,198
15	1,79	1,59	1,395	1,196
16	1,788	1,587	1,391	1,195
17	1,786	1,586	1,389	1,193
18	1,782	1,586	1,388	1,191
19	1,782	1,582	1,386	1,19
20	1,78	1,58	1,386	1,189
21	1,781	1,578	1,382	1,186
22	1,776	1,576	1,379	1,185
23	1,774	1,574	1,379	1,183
24	1,772	1,574	1,377	1,183
25	1,77	1,57	1,377	1,18
26	1,768	1,568	1,373	1,178
27	1,764	1,566	1,371	1,177
28	1,767	1,564	1,37	1,176
29	1,762	1,562	1,368	1,174
30	1,76	1,562	1,368	1,171
31	1,758	1,558	1,364	1,17
32	1,756	1,555	1,362	1,17
33	1,754	1,554	1,36	1,167
34	1,752	1,552	1,359	1,166
35	1,753	1,55	1,359	1,164
36	1,746	1,55	1,355	1,163
37	1,746	1,546	1,353	1,161
38	1,744	1,544	1,352	1,159
39	1,742	1,542	1,35	1,158
40	1,74	1,539	1,35	1,157
41	1,738	1,538	1,346	1,154
42	1,739	1,538	1,344	1,153

Кінець таблиці В.45

43	1,734	1,534	1,343	1,151
44	1,732	1,532	1,34	1,151
45	1,728	1,53	1,341	1,148
46	1,728	1,528	1,337	1,146
47	1,726	1,526	1,335	1,145
48	1,724	1,526	1,334	1,144
49	1,725	1,522	1,332	1,142
50	1,72	1,52	1,332	1,139

Таблиця В.6 – Інтегральна ефективність функціонування підсистеми виявлення КА (50 серій)

№	Базова система	Варіант 1 (автоматичний горизонт)	Варіант 2 (+ самонавчання)	Варіант 3 (+ попередній досвід)
1	0,682	0,742	0,813	0,872
2	0,684	0,745	0,815	0,875
3	0,687	0,747	0,818	0,877
4	0,689	0,75	0,82	0,879
5	0,691	0,752	0,823	0,882
6	0,693	0,754	0,827	0,884
7	0,695	0,758	0,828	0,886
8	0,7	0,759	0,831	0,888
9	0,699	0,762	0,833	0,891
10	0,702	0,763	0,836	0,894
11	0,704	0,766	0,838	0,895
12	0,706	0,769	0,842	0,897
13	0,709	0,771	0,844	0,9
14	0,711	0,775	0,846	0,902
15	0,713	0,776	0,849	0,905
16	0,717	0,778	0,852	0,907
17	0,717	0,781	0,854	0,909
18	0,719	0,783	0,858	0,911
19	0,722	0,786	0,859	0,914
20	0,724	0,787	0,862	0,917
21	0,726	0,791	0,865	0,918
22	0,728	0,793	0,866	0,921
23	0,731	0,795	0,87	0,923
24	0,735	0,798	0,873	0,924
25	0,735	0,8	0,875	0,928
26	0,737	0,802	0,878	0,93
27	0,738	0,805	0,88	0,932
28	0,742	0,808	0,883	0,934
29	0,744	0,81	0,885	0,937
30	0,746	0,811	0,889	0,94
31	0,748	0,814	0,891	0,941
32	0,752	0,817	0,893	0,944
33	0,753	0,819	0,895	0,946
34	0,755	0,822	0,898	0,948
35	0,757	0,825	0,901	0,952
36	0,758	0,826	0,905	0,952
37	0,761	0,829	0,906	0,955
38	0,764	0,831	0,909	0,957
39	0,766	0,834	0,911	0,96
40	0,77	0,835	0,914	0,963
41	0,77	0,838	0,917	0,964
42	0,772	0,842	0,92	0,967
43	0,775	0,843	0,922	0,969
44	0,777	0,846	0,923	0,971

Кінець таблиці В.6

45	0,778	0,848	0,927	0,975
46	0,781	0,85	0,93	0,976
47	0,783	0,853	0,932	0,978
48	0,788	0,855	0,936	0,979
49	0,788	0,859	0,937	0,983
50	0,79	0,859	0,94	0,986

ДОДАТОК Г

Дослідження вибірки даних

1. Характеристика історичних даних, що використовуються у процесі вибору.

Таблиця Г.1 – Історичні дані, що використовуються системою протидії КА

№	Тип історичних даних	Джерело даних	Формат	Використання у запропонованому методі
1	Журнали автентифікації користувачів	Active Directory, LDAP, VPN	структуровані	виявлення повторюваних спроб несанкціонованого доступу
2	Журнали подій операційних систем	Windows Event Log, Linux Syslog	структуровані	аналіз аномальної активності вузлів
3	Журнали мережного трафіку	NetFlow, IPFIX, PCAP	напівструктуровані	формування історії мережних взаємодій
4	Журнали міжмережних екранів	Firewall Logs	структуровані	визначення характерних шаблонів мережних атак
5	Журнали IDS/IPS	Snort, Suricata, Zeek	структуровані	виявлення повторюваних сигнатур атак
6	Події SIEM	SIEM-система	структуровані	кореляція історичних інцидентів
7	Історія виявлених КА	База інцидентів	структуровані	врахування попереднього досвіду
8	Журнали антивірусного захисту	EDR/XDR	структуровані	аналіз шкідливої активності
9	Індикатори компрометації (IoC)	Threat Intelligence	структуровані	оновлення знань про загрози
10	Результати попередніх рішень системи	База знань	структуровані	самонавчання алгоритму
11	Характеристики вузлів мережі	CMDB, інвентаризація	структуровані	визначення критичності ресурсів
12	Статистика використання сервісів	Системний моніторинг	структуровані	аналіз нормальної поведінки
13	Поведінкові профілі користувачів (UEBA)	UEBA	структуровані	виявлення поведінкових аномалій
14	Журнали виконання процесів	EDR	структуровані	аналіз запуску підозрілих процесів
15	Історія реагування на інциденти	SOC	структуровані	оцінювання ефективності попередніх заходів реагування

Класифікація історичних даних на три групи.

1. Дані про події інформаційної безпеки, які характеризують факти виникнення подій:

1.1) журнали автентифікації;

1.2) журнали ОС;

1.3) міжмережних екранів;

1.4) /IPS;

1.5) SIEM.

2. Дані про мережну активність, які характеризують поведінку мережі:

1) NetFlow;

2) PCAP;

3) DNS;

4) HTTP;

5) VPN;

6) маршрутизація;

7) використання сервісів.

3. Дані про попередній досвід. Ця група автоматично вибирається розробленим методом і містить такі дані:

1) історія атак;

2) історія реагування;

3) результати роботи IDS;

4) база знань;

5) історія успішних рішень;

6) індикатори компрометації;

7) результати самонавчання.

Для статистичного експерименту використано шість агрегованих характеристик.

Таблиця Г.2 – Історичні дані, які сформовано з основних джерел інформації

Позначення	Показник
N_{log}	кількість історичних журналів подій
N_{net}	кількість історичних мережних записів
N_{IoC}	кількість історичних IoC
N_{inc}	кількість історичних інцидентів
N_{resp}	кількість успішних реагувань
N_{kb}	кількість правил бази знань

Дані з табл. Г.2 охоплюють усі основні джерела інформації, що використовуються сучасними системами SIEM, SOAR, EDR/XDR, IDS/IPS та іншими

засобами протидії КА. Крім того, вони безпосередньо пов'язані з трьома компонентами розробленого методу:

- 1) автоматичним вибором глибини горизонту (кількість доступних історичних записів),
- 2) самонавчанням (накопичені правила та результати попередніх рішень),
- 3) використанням попереднього досвіду (історія інцидентів, реагувань та індикаторів компрометації).

2. Діапазони значень

Таблиця Г.3 – Діапазони значень історичних даних, які сформовано з основних джерел інформації

Показник	Мін	Макс
N_{log}	1200	2200
N_{net}	900	1800
N_{IoC}	180	420
N_{inc}	60	170
N_{resp}	50	160
N_{kb}	250	700

3. Вектори історичних даних

Таблиця Г.4 – Вектори історичних даних

№	N_{log}	N_{net}	N_{IoC}	N_{inc}	N_{resp}	N_{kb}
1	1856	1506	331	118	85	399
2	1845	1503	309	145	87	501
3	1344	1273	259	138	103	473
4	1674	1316	213	131	92	505
5	1802	1346	326	120	101	475
6	1777	1363	366	96	108	434
7	1515	1428	286	140	74	386
8	1564	1573	287	136	112	530
9	1747	1421	302	118	73	445
10	1478	1403	296	108	104	488
11	1662	1304	341	129	137	376
12	1395	1228	348	111	116	415

Кінець таблиці Г.4

13	1702	1233	333	123	85	390
14	1461	1243	268	137	78	524
15	1716	1362	320	108	114	442
16	1948	1509	301	119	113	487
17	1955	1212	300	105	82	546
18	1913	1187	229	96	114	612
19	1386	1395	261	113	97	451
20	1803	1750	354	102	111	416
21	1592	1160	331	136	116	342
22	1451	1166	301	112	83	404
23	1734	1296	349	109	92	326
24	1833	1288	301	105	118	340
25	2111	1237	212	122	115	579
26	1648	1405	276	120	82	399
27	1944	1182	257	112	93	560
28	1629	1041	317	82	103	456
29	1694	1391	325	105	112	476
30	1576	1236	248	86	98	450
31	1894	1401	270	124	119	416
32	1648	1345	339	126	109	591
33	1892	1389	287	141	107	437
34	1900	1237	297	122	109	556
35	1503	1182	281	108	97	553
36	1664	1174	308	84	90	407
37	1681	1397	346	109	77	607
38	1658	1012	369	70	98	408
39	1534	1372	281	110	91	453
40	1540	1339	285	134	88	366
41	1785	1523	300	144	87	428
42	1518	1070	260	100	95	544
43	1592	1330	229	127	128	533
44	1859	1418	316	109	97	497
45	1843	1171	224	119	113	495
46	1700	1586	288	109	100	408
47	1456	1194	334	131	105	550
48	1299	1214	315	113	73	461
49	1723	1656	260	136	117	456
50	1690	1537	302	119	103	412

4. Дослідження розподілу вхідних даних на основі опрацювання статистичного аналізу

Для кожного типу історичних даних визначено основні описові статистики та перевірено відповідність нормальному закону розподілу за критерієм Шапіро–Уїлка при рівні значущості $\alpha=0,05$. Перевірка нормальності є необхідною умовою подальшого застосування параметричних статистичних критеріїв для аналізу результатів експерименту.

Таблиця Г.5 – Результати статистичної обробки вхідних даних експерименту

Показник	Середнє значення	Медіанне значення	Стандартне відхилення	Мінімальне значення	Максимальне значення	Коефіцієнт варіації, %	Статистика критерію Шапіро–Уїлка	p-value	Висновок
N_{log}	1682,68	1685,5	179,95	1299	2111	10,69	0,9888	0,9144	Normal
N_{net}	1330,08	1334,5	154,57	1012	1750	11,62	0,9802	0,5589	Normal
N_{loc}	296,76	300,5	39,02	212	369	13,15	0,9766	0,4193	Normal
N_{inc}	116,34	118	16,7	70	145	14,36	0,9696	0,2236	Normal
N_{resp}	100,02	100,5	14,83	73	137	14,83	0,9783	0,4831	Normal
N_{kb}	464,1	454,5	71,57	326	612	15,42	0,9768	0,4247	Normal

ДОДАТОК Д

Програмний код

З метою верифікації пропонованих методів було реалізовано систему організації зберігання даних в системах та засобах протидії КА з урахуванням історичного аспекту їх використання та урахуванням глибини горизонту. Розроблена система централізованою платформою моніторингу мережної безпеки із ієрархічним методом організації зберігання даних та крос-джерельною кореляцією подій. Система побудована на клієнт-серверній архітектурі з розподіленими агентами збору даних і централізованим вузлом обробки та зберігання. Загальна архітектура реалізованої системи складається із таких компонентів: сервера, агентів, бази даних та інформаційної панелі.

Система реалізує метод організації зберігання даних з урахуванням історичного аспекту їх використання та метод вибору даних з урахуванням глибини горизонту, самонавчання та попереднього досвіду використання.

Відповідно до першого методу вхідний потік подій не зберігається монолітно, а розподіляється між трьома ієрархічними шарами відповідно до підтвердженої цінності даних. Оперативний шар містить події останніх 72 годин і забезпечує швидкий доступ для поточного аналізу. Аналітичний шар формується динамічно, тобто події переводяться до нього автоматично у момент, коли крос-джерельна кореляція підтверджує їх причетність до інциденту, підвищуючи значення поля *effectiveness*. Поле *effectiveness* є кількісним показником підтвердженої аналітичної цінності події безпеки, що зберігається як число з плаваючою точкою подвійної точності. За замовчуванням воно має нульове значення, яке означає відсутність підтвердженого зв'язку події з реальним інцидентом. Значення поля підвищується автоматично у двох випадках: по-перше, коли механізм крос-джерельної кореляції встановлює причетність події до підтвердженого інциденту, і по-друге, через механізм ретроактивного оновлення, коли нова подія тригерує кореляцію і система автоматично підвищує *effectiveness* усіх раніше зафіксованих подій з того самого IP-адресу в межах заданого часового вікна. Таким чином, *effectiveness* виконує роль

індикатора накопиченого досвіду системи щодо конкретної події. Ненульове значення сигналізує про те, що подія виявилася аналітично значущою в ретроспективі, і саме цей показник є ключовим критерієм при прийнятті рішення про переміщення даних між ієрархічними шарами зберігання. Щоденний скрипт обслуговування реалізує автоматичне переміщення даних між шарами згідно з визначеними горизонтами.

Відповідно до методу вибору даних з урахуванням глибини горизонту, самонавчання та попереднього досвіду використання, система автоматично визначає застарілість даних і переміщує їх в архів або видаляє відповідно до налаштованих часових порогів (72 години для оперативного шару, 90 днів для аналітичного, 365 днів для архіву). Елемент попереднього досвіду реалізований через поле *effectiveness* та механізм оновлення, тобто коли нова подія тригерує кореляцію, система ретроактивно підвищує *effectiveness* усіх пов'язаних подій з того самого IP-адресу в межах часового вікна, тим самим враховуючи накопичений досвід виявлення атак при прийнятті рішень про зберігання.

На рівні клієнтських вузлів функціонують програмні агенти, реалізовані мовою Python, які здійснюють паралельне збирання подій. Для даної системи реалізовано два незалежні джерела, що включають журнал безпеки операційної системи Windows (канал Security, події з ідентифікаторами 5152, 5154, 5156, 5157, що відповідають операціям блокування та дозволу мережних з'єднань Windows Firewall) та системи виявлення вторгнень Snort 2.9, що працює в режимі пасивного прослуховування мережного інтерфейсу через бібліотеку Npcap. За потреби є можливість реалізувати інші типи джерел. Агент реалізує трипотоківу модель виконання, де окремі потоки відповідають за збір подій, їх батчеву відправку на сервер та періодичну реєстрацію вузла через механізм серцебиття. Транспортний рівень побудований на протоколі HTTP REST без використання брокерів повідомлень.

Серверний компонент реалізований на основі фреймворку FastAPI з асинхронною обробкою запитів і використовує PostgreSQL як основну систему керування базами даних. Взаємодія з базою даних здійснюється через ORM-бібліотеку SQLAlchemy із пулом з'єднань, налаштованим на 20 активних і 40

додаткових з'єднань для забезпечення стабільності при пікових навантаженнях. Нормалізація та валідація вхідних даних виконується засобами бібліотеки Pydantic.

ERD діаграму для спроектованої БД наведено на рис. Д.1.

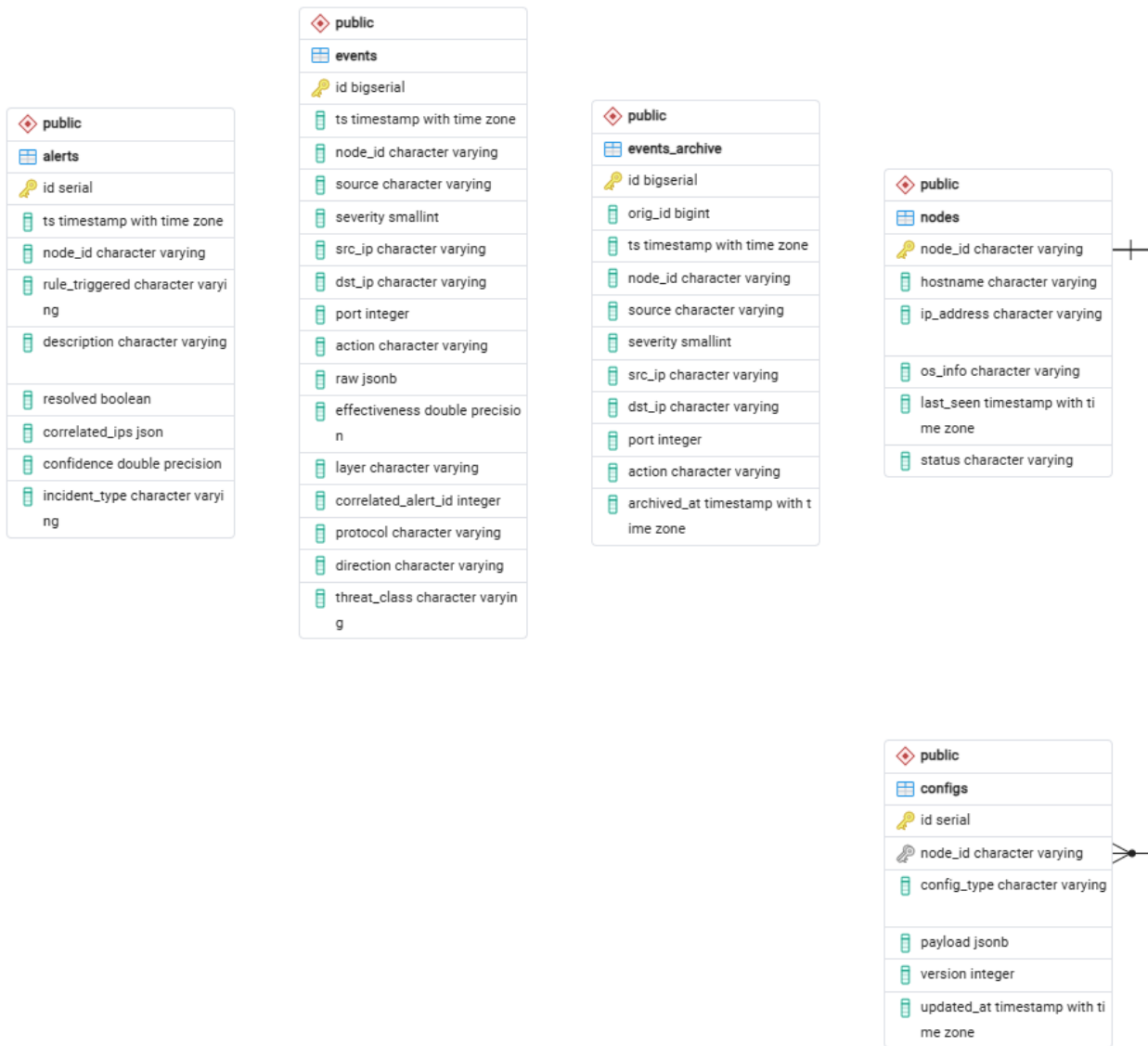


Рисунок Д.1 – ERD діаграма для спроектованої БД

Розроблена база даних складається з п'яти таблиць. Таблиця **nodes** виконує роль центрального реєстру мережних вузлів, на яких розгорнуто агенти збору даних, та зберігає ідентифікаційні й статусні характеристики кожного вузла, зокрема мережне ім'я, IP-адресу, інформацію про операційну систему, час останнього з'єднання та поточний стан. Таблиця **events** є основним сховищем поточних подій безпеки і фіксує набір атрибутів мережного інциденту, такі як адреси та порти джерела й одержувача,

протокол, напрям трафіку, клас загрози, рівень серйозності, вжиту дію та необроблені дані у форматі JSONB, що забезпечує підтримку різнорідних джерел без зміни схеми. З метою реалізації історичного аспекту зберігання та керування глибиною горизонту даних передбачено таблицю `events_archive`, до якої переносяться події після завершення їх активного циклу обробки; при цьому зберігається посилання на оригінальний запис через поле `orig_id` та фіксується мітка часу архівування. Таблиця `alerts` акумулює агреговані інциденти безпеки, що формуються внаслідок кореляції подій, і містить атрибути спрацьованого правила, типу інциденту, рівня достовірності, переліку скорельованих IP-адрес у форматі JSON, а також прапор вирішеності інциденту. Таблиця `configs` забезпечує централізоване керування конфігураціями агентів із підтримкою версіонування, де кожен запис прив'язаний до конкретного вузла через поле `node_id` і містить тип конфігурації та її вміст у форматі JSONB.

Використання типу `bigserial` як первинного ключа в таблицях `events` та `events_archive` обумовлено очікуваними великими обсягами даних в умовах інтенсивного мережного моніторингу. Застосування формату JSONB у полях збереження сирих даних подій і конфігураційних параметрів забезпечує структурну гнучкість при роботі з різнотипними джерелами, такими як журнал безпеки Windows та система виявлення вторгнень Snort. Дворівнева організація зберігання подій через пару таблиць `events` і `events_archive` безпосередньо реалізує запропонований метод керування глибиною горизонту, дозволяючи розмежувати оперативні та історичні дані без втрати цілісності інформації.

Також реалізований механізм зворотного зв'язку між центральним сервером та агентами на основі періодичного опитування. Сервер надає ендпоінт `GET /api/configs/{node_id}`, який повертає актуальну політику для конкретного вузла і яка складається із набору керованих параметрів, що включають мінімальний поріг критичності подій, список портів що вважаються шумом, інтервали відправки та контрольних повідомлень, а також прапорець увімкнення Snort. Кожна конфігурація має версію, яка інкрементується при кожному оновленні через `POST /api/configs/{node_id}`. Агент під час кожного циклу інтервалу контрольних

повідомлень паралельно запитує сервер і порівнює отриману версію з поточною. Якщо версія нова, він застосовує нові параметри в реальному часі без перезапуску процесу. Це дозволяє адміністратору централізовано керувати поведінкою всіх агентів у мережі: наприклад, підвищити поріг фільтрації на вузлі що генерує забагато шуму, або вимкнути Snort на конкретній машині і зміна набуде чинності протягом одного інтервалу контрольних повідомлень (до 30 секунд) без жодного ручного втручання на клієнтській стороні.

```
[agent] heartbeat OK (DESKTOP-G1L0I70)
[agent] конфіг оновлено: версія 0 → 1, flush=20с, min_severity=3, snort=увімк
```

```
curl -X POST "http://localhost:8000/api/configs/DESKTOP-G1L0I70" -H "Content-Type: application/json" -d "{\"min_severity\": 3, \"snort_enabled\": true, \"flush_interval\": 20}"
```

Для забезпечення ефективного аналізу даних, накопичених у системі організації зберігання інформації про події інформаційної безпеки, розроблено веборієнтовану інформаційну панель, яка забезпечує централізоване відображення поточного стану системи, історичних даних та результатів їх обробки (рис. Д.2).

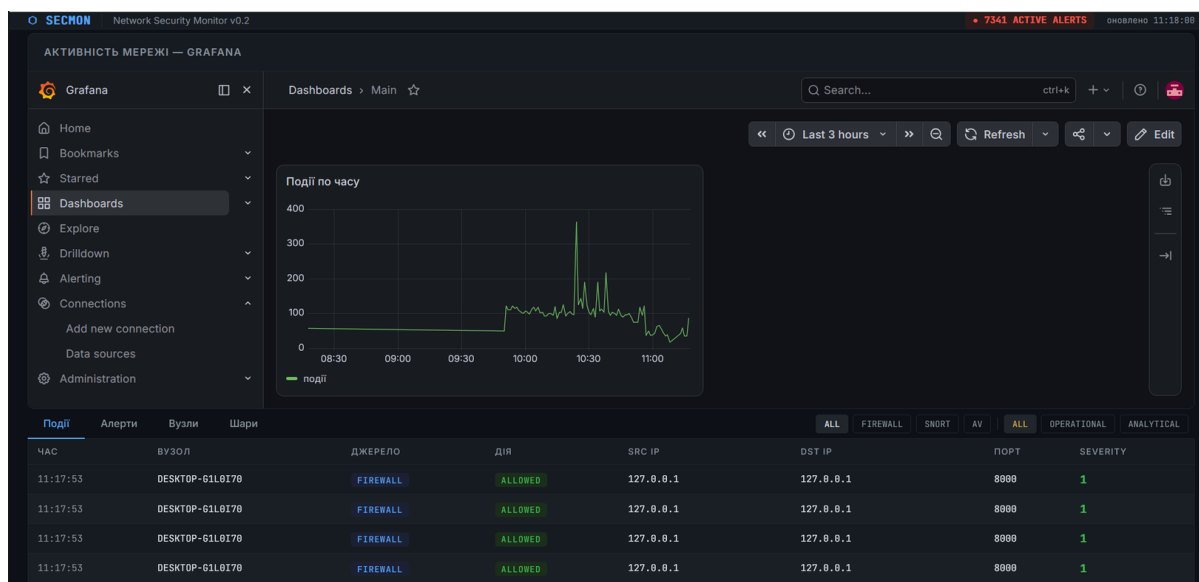


Рисунок Д.2 – Інформаційна панель реалізованої системи організації зберігання даних в системах та засобах протидії комп'ютерним атакам з урахуванням історичного аспекту їх використання та урахуванням глибини горизонту

Інтерфейс реалізовано як односторінковий вебзастосунок на основі React, який взаємодіє з серверною частиною через REST API. Візуалізація часових рядів здійснюється шляхом інтеграції із платформою Grafana, що отримує дані з бази даних PostgreSQL та дозволяє аналізувати накопичену інформацію за довільний часовий горизонт.

Головний екран системи містить інтегровану область візуалізації часової активності мережі, де відображається інтенсивність реєстрації подій безпеки протягом обраного інтервалу часу. Графік формується на основі історичних записів, що дозволяє аналізувати зміни інтенсивності подій, виявляти пікові навантаження, визначати аномальні періоди та оцінювати тенденції розвитку мережної активності.

Панель підтримує зміну глибини аналізу шляхом вибору часового горизонту, що дає можливість виконувати дослідження як короткострокових подій (останні години або доба), так і довгострокових тенденцій, накопичених у базі даних. Такий підхід забезпечує використання історичного аспекту зберігання інформації та дозволяє проводити ретроспективний аналіз інцидентів.

Поточний стан системи характеризується агрегованими статистичними показниками, зокрема загальною кількістю активних сповіщень та часом останнього оновлення даних. Значення оновлюються автоматично відповідно до встановленого інтервалу синхронізації із сервером.

Основним елементом інформаційної панелі є журнал подій безпеки, який формується у вигляді таблиці та містить структуровані записи про всі зареєстровані події. Для кожного запису відображаються:

- час реєстрації події;
- вузол, на якому зафіксовано подію;
- джерело формування повідомлення (Firewall, Snort, антивірус тощо);
- результат обробки або виконана дія;
- IP-адреса джерела;
- IP-адреса призначення;
- мережний порт;
- рівень критичності.

Для спрощення аналізу реалізовано механізми фільтрації подій за джерелом інформації та функціональним призначенням. Зокрема, підтримується вибір повідомлень, сформованих міжмережним екраном, системою виявлення вторгнень Snort або антивірусним програмним забезпеченням. Додатково передбачено групування записів за категоріями Operational та Analytical, що дозволяє розділяти оперативний моніторинг і поглиблений аналіз історичних даних.

Структура інформаційної панелі також передбачає окремі функціональні вкладки для відображення подій, сповіщень, вузлів мережі та рівнів архітектури системи (рис. Д.3 – Д.6).

ЧАС	ВУЗОЛ	ДЖЕРЕЛО	ДІЯ	SRC IP	DST IP	ПОРТ	SEVERITY
11:55:50	DESKTOP-61L0I70	FIREWALL	ALLOWED	192.168.0.103	224.0.0.252	-	1
11:55:50	DESKTOP-61L0I70	FIREWALL	ALLOWED	127.0.0.1	127.0.0.1	8080	1
11:55:50	DESKTOP-61L0I70	FIREWALL	ALLOWED	127.0.0.1	127.0.0.1	8080	1
11:55:50	DESKTOP-61L0I70	FIREWALL	ALLOWED	127.0.0.1	127.0.0.1	8080	1
11:55:50	DESKTOP-61L0I70	FIREWALL	ALLOWED	127.0.0.1	127.0.0.1	8080	1

Рисунок Д.3 – Вкладка Події

ЧАС	ВУЗОЛ	ПРАВИЛО	ОПИС	STATUS
29.06, 11:53:56	DESKTOP-61L0I70	high_severity_block	Заблоковано high-severity з 192.168.0.103	ACTIVE
29.06, 11:53:56	DESKTOP-61L0I70	high_severity_block	Заблоковано high-severity з 192.168.0.103	ACTIVE
29.06, 11:53:56	DESKTOP-61L0I70	high_severity_block	Заблоковано high-severity з 192.168.0.1	ACTIVE
29.06, 11:53:56	DESKTOP-61L0I70	high_severity_block	Заблоковано high-severity з 192.168.0.1	ACTIVE
29.06, 11:53:46	DESKTOP-61L0I70	high_severity_block	Заблоковано high-severity з 192.168.0.103	ACTIVE

Рисунок Д.4 – Вкладка Алерти

Ім'я вузла	IP	Статус
test-laptop	192.168.0.99	OFFLINE
DESKTOP-61L0I70	192.168.0.103	ONLINE

Рисунок Д.5 – Вкладка Вузли

Категорія	Кількість	Відсоток
Operational	357 505	(99%)
Analytical	4 236	(1%)
Archive	319	(0%)

Клас загрози	Кількість	Відсоток
noise	210 658	(58%)
reconnaissance	150 972	(42%)
lateral_movement	111	(0%)

Рисунок Д.6 – Вкладка Шари

Для верифікації розроблених методів було сформовано змодельований набір даних мережних подій безпеки, що охоплює 230 записів і відтворює типові сценарії функціонування системи моніторингу в умовах як штатної роботи мережі, так і цілеспрямованих атак. Набір даних структурований за чотирма послідовними фазами, що чергуються відповідно до реалістичного часового профілю. Перша і третя фази відображають нормальну мережну активність вузла DESKTOP-G1L0I70 з IP-адресою 192.168.0.2 і включають loopback-з'єднання на порт 8000, ICMP-запити від шлюзу 192.168.0.1 та мультикаст трафік на адреси 224.0.0.x, зафіксовані брандмауером Windows з рівнем severity 1. Друга фаза моделює атаку типу DoS Slowloris з вузла 185.220.101.47, що характеризується масовим встановленням повільних HTTP/HTTPS-з'єднань на порти 80 і 443 цільового вузла; при цьому рівень severity поступово зростає від 2 до 5, а після автоматичного виявлення аномалії система переходить від дії allowed до blocked, причому частина подій дублюється джерелом Snort, що імітує крос-джерельну кореляцію. Четверта фаза відтворює горизонтальне сканування портів з вузла 91.189.88.142, під час якого фіксується послідовний перебір портів починаючи з 20-го, що реєструється системою виявлення вторгнень Snort із рівнем severity 4 та дією blocked.

Сформований набір даних дозволяє перевірити коректність розподілу подій між ієрархічними шарами зберігання. Події нормальної активності з нульовим значенням effectiveness після завершення 72-годинного горизонту мають переміщуватись до архівного шару зі стисненням поля raw, тоді як події атаківих фаз, скорельовані між собою та пов'язані зі сформованими алертами, отримують ненульове значення effectiveness і зберігаються в аналітичному шарі протягом 90 днів. Таким чином, змодельований журнал слугує контрольним сценарієм для оцінювання точності механізму ретроактивного оновлення та правильності застосування часових горизонтів зберігання відповідно до підтвердженої аналітичної цінності кожної події. Фрагмент журнал мережних подій безпеки наведено у табл. Д.1.

Таблиця Д.1 – Журнал мережних подій безпеки

Час	Вузол	Джерело	Дія	Src IP	Dst IP	Порт	Severity
12:02:18	DESKTOP-G1L0I70	firewall	allowed	192.168.0.2	192.168.0.2	8000	1
12:02:21	DESKTOP-G1L0I70	firewall	allowed	192.168.0.2	192.168.0.2	8000	1
12:02:24	DESKTOP-G1L0I70	firewall	allowed	192.168.0.1	224.0.0.1	—	1
12:02:27	DESKTOP-G1L0I70	firewall	allowed	192.168.0.2	192.168.0.2	8000	1
12:03:00	DESKTOP-G1L0I70	snort	allowed	185.220.101.47	192.168.0.2	443	2
12:03:03	DESKTOP-G1L0I70	firewall	allowed	185.220.101.47	192.168.0.2	81	2
12:03:06	DESKTOP-G1L0I70	firewall	allowed	185.220.101.47	192.168.0.2	82	2
12:03:09	DESKTOP-G1L0I70	snort	allowed	185.220.101.47	192.168.0.2	83	2

Вихідний програмний код, що був використаний для проведення тестування та експериментальних досліджень, розміщено у відкритому репозиторії на GitHub.

Vadym-Paiuk Secmon: Програмне забезпечення організації, оптимізації та вибору зберігання даних в системах та засобах протидії КА з урахуванням історичного аспекту їх використання та глибини горизонту. GitHub. URL: <https://github.com/Vadym-Paiuk/secmon> (дата звернення: 14.07.2026).

На рис. Д.7 подано структуру репозиторію із вихідним програмним кодом та конфігураційними файлами, необхідними для відтворення експериментальних досліджень.

Розроблене програмне забезпечення реалізує процеси організації, оптимізації та вибору зберігання даних в системах та засобах протидії КА з урахуванням історичного аспекту їх використання та глибини горизонту

Процес організації даних реалізований через трирівневу ієрархічну структуру зберігання. Кожна подія при надходженні отримує атрибути рівень зберігання, ефективність, клас загрози та напрямок. Таблиця events у БД є основним оперативним

шаром, `events_archive` – архівним. Переміщення між шарами виконує `maintenance.py` за часовими горизонтами (72 год – в архів, 90 днів – стиснення).

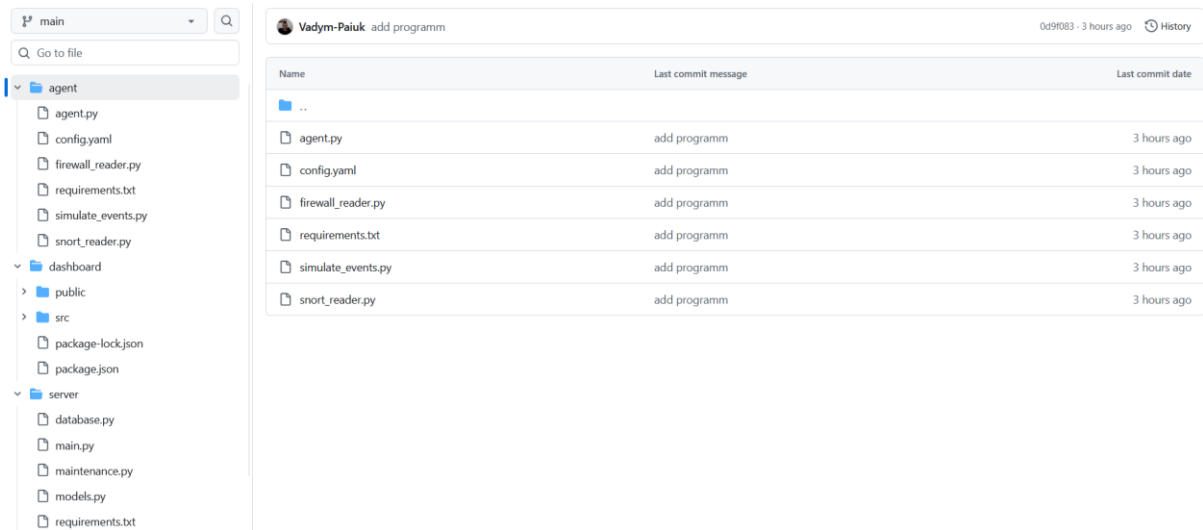


Рисунок Д.7 – Структура репозиторію GitHub із вихідним програмним кодом та конфігураційними файлами

Процес оптимізації даних реалізований на двох рівнях. На рівні агента реалізована фільтрація шуму ще до відправки даних. Зокрема відкидаються дозволені події на типових портах (80, 443, 53...) з пороговим значенням. Параметри оптимізації (інтервали оновлення, рівень прийняття події, рівень шуму) керуються централізовано через `POST /api/configs/{node_id}` і доставляються агентам без перезапуску.

Процес вибору даних реалізований через механізм оцінювання ефективності та горизонт зберігання. При крос-кореляції (коли одна й та сама IP-адреса джерела з'являється у міжмережному екрані та системі виявлення вторгнень Snort протягом 10 хвилин) система підвищує значення ефективності пов'язаних подій до 0,4–1,0 і переводить їх до аналітичного рівня зберігання, де вони не потрапляють під архівування незалежно від віку. Події з нульовою ефективністю після 72 годин автоматично переміщуються до архіву як неактуальні. Таким чином, горизонт зберігання є динамічним: підтверджені інциденти зберігаються роками, тоді як шумові події видаляються.

Реалізований механізм зворотного зв'язку між центральним сервером та агентами на основі підходу періодичного опитування. Сервер надає програмний

інтерфейс GET /api/configs/{node_id}, який повертає актуальну політику для конкретного вузла – набір керованих параметрів, що включає мінімальний поріг критичності подій, список портів, які вважаються шумовими, інтервали передавання даних та контрольних повідомлень, а також прапорець увімкнення системи Snort. Кожна конфігурація має номер версії, який автоматично збільшується при кожному оновленні через програмний інтерфейс POST /api/configs/{node_id}. Агент під час кожного циклу надсилання контрольного повідомлення паралельно звертається до сервера та порівнює отриманий номер версії з поточним. Якщо виявлено новішу версію, нові параметри застосовуються в реальному часі без перезапуску процесу.

Структура репозиторію зображена на рис. Д.8.

```

secmon/
├── server/
│   ├── main.py          # FastAPI сервер, кореляція, API
│   └── models.py        # SQLAlchemy ORM (Node, Event, EventArchive, Alert,
Config)
│   ├── schemas.py       # Pydantic схеми
│   ├── database.py      # підключення до БД PostgreSQL
│   ├── maintenance.py   # щоденне обслуговування шарів зберігання
│   ├── requirements.txt
│   └── .env.example      # шаблон змінних середовища
├── agent/
│   └── agent.py          # головний агент (3 потоки: збір / flush /
heartbeat+config)
│   ├── firewall_reader.py # читання Windows Firewall Event Log
│   ├── snort_reader.py   # читач Snort логу
│   ├── simulate_events.py # генератор тестових подій
│   ├── config.yaml       # локальний конфігураційний файл агента
│   └── requirements.txt
└── dashboard/           # React додаток
    └── src/
        ├── App.jsx
        └── components/
            ├── StatCard.jsx
            ├── Badge.jsx
            ├── EventsTable.jsx
            ├── AlertsTable.jsx
            ├── NodesPanel.jsx
            └── LayersPanel.jsx

```

Рисунок Д.8 – Структура репозиторію

Процес розгортання включає виконання наступних кроків.

Крок 1: Встановити залежності

- [Python 3.13+](https://python.org)
- [Node.js 22+](https://nodejs.org)
- [PostgreSQL 16](https://postgresql.org/download/windows)
- [Snort 2.9-WIN64](https://snort.org/downloads) + [Npcap](https://npcap.com)
- [Grafana OSS](https://grafana.com/grafana/download)

Крок 2: База даних

В pgAdmin або psql створити БД:

```
CREATE DATABASE secmon;
```

Крок 3: Сервер

```
cd server
```

```
pip install -r requirements.txt
```

```
copy .env.example .env
```

Відкрити файл `.env` і вписати свій пароль PostgreSQL:

```
DATABASE_URL=postgresql://postgres: ПАРОЛЬ@localhost:5432/secmon
```

```
SERVER_HOST=0.0.0.0
```

```
SERVER_PORT=8000
```

Запусти сервер – він автоматично створить таблиці:

```
python main.py
```

Крок 4: Додати колонки в БД

Після першого запуску сервера в pgAdmin (Query Tool) виконати:

```
ALTER TABLE events ADD COLUMN IF NOT EXISTS effectiveness FLOAT  
DEFAULT 0.0;
```

```
ALTER TABLE events ADD COLUMN IF NOT EXISTS layer VARCHAR DEFAULT  
'operational';
```

```
ALTER TABLE events ADD COLUMN IF NOT EXISTS correlated_alert_id  
INTEGER;
```

```

ALTER TABLE events ADD COLUMN IF NOT EXISTS protocol VARCHAR;
ALTER TABLE events ADD COLUMN IF NOT EXISTS direction VARCHAR;
ALTER TABLE events ADD COLUMN IF NOT EXISTS threat_class VARCHAR;
ALTER TABLE alerts ADD COLUMN IF NOT EXISTS correlated_ips JSON;
ALTER TABLE alerts ADD COLUMN IF NOT EXISTS confidence FLOAT
DEFAULT 0.5;
ALTER TABLE alerts ADD COLUMN IF NOT EXISTS incident_type VARCHAR;

```

Крок 5: Індокси в БД

```

CREATE INDEX IF NOT EXISTS idx_events_src_ip ON events(src_ip);
CREATE INDEX IF NOT EXISTS idx_events_ts ON events(ts DESC);
CREATE INDEX IF NOT EXISTS idx_events_source ON events(source);
CREATE INDEX IF NOT EXISTS idx_events_layer ON events(layer);

```

Крок 6: Увімкнути аудит Windows Firewall (один раз, від адміністратора), powershell:

```

auditpol /set /subcategory:"Filtering Platform Connection" /success:enable
/failure:enable
auditpol /set /subcategory:"Filtering Platform Packet Drop" /success:enable
/failure:enable

```

Крок 7: Агент (від імені адміністратора), powershell:

```

cd agent
pip install -r requirements.txt
Відкрий `config.yaml` і вказати IP сервера:
server_url: "http://192.168.0.X:8000"
Виконати запуск агенту із powershell:
python agent.py

```

Крок 8: Snort (від адміністратора)

Дізнатися номер мережного інтерфейсу, powershell:

```
C:\Snort\bin\snort.exe -W
```

Запусти, powershell:

```
C:\Snort\bin\snort.exe -i <НОМЕР ІНТЕРФЕЙСУ> -c C:\Snort\etc\snort.conf -l  
C:\Snort\log -A fast -K ascii
```

Крок 9: Дашборд, powershell

```
cd dashboard
```

```
npm install
```

```
npm start
```

Відкриється на `http://localhost:3001`.

Крок 10: Перевірка, powershell:

```
cd agent
```

```
python simulate_events.py
```

Відкрити `http://localhost:8000/api/stats/summary` – буде повернуто статистику з подіями.

Правила snort не включені в репозитарій.

Для тестування системи було використано наступні правила (файл local.rules)

```
# Будь-який ICMP трафік (ping в будь-який бік)
```

```
alert icmp any any -> any any (msg:"ICMP ping detected"; sid:1000001; rev:1;)
```

```
# TCP з'єднання на RDP (3389) - в будь-який бік
```

```
alert tcp any any -> any 3389 (msg:"RDP connection attempt"; sid:1000002; rev:1;)
```

```
# TCP з'єднання на SSH (22) - в будь-який бік
```

```
alert tcp any any -> any 22 (msg:"SSH connection attempt"; sid:1000003; rev:1;)
```

Детекція HTTP трафік (80)

alert tcp any any -> any 80 (msg:"HTTP traffic detected"; sid:1000004; rev:1;)

Детекція HTTPS (443)

alert tcp any any -> any 443 (msg:"HTTPS traffic detected"; sid:1000005; rev:1;)

alert tcp any any -> 192.168.0.103 8000 (msg:"SecMon API access detected";
sid:1000006; rev:1;)