https://doi.org/10.31891/2219-9365-2026-87-26
УДК 004.75:004.49:004.3

ПАЮК Вадим
Хмельницький національний університет
https://orcid.org/0000-0002-9969-8239
e-mail: vadympaiuk@gmail.com
САВЕНКО Богдан
Хмельницький національний університет
https://orcid.org/0000-0001-5647-9979
e-mail: savenko_bohdan@ukr.net

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

У статті розглянуто задачу підвищення ефективності функціонування систем і засобів протидії комп'ютерним атакам за рахунок удосконалення процесів обробки та оптимізації даних у підсистемах пам'яті. Актуальність дослідження обумовлена постійним зростанням обсягів даних кібермоніторингу, підвищенням складності сучасних комп'ютерних атак та необхідністю забезпечення оперативного доступу до релевантної інформації для своєчасного виявлення загроз і прийняття рішень щодо протидії їм.

Метою роботи є розроблення методу оптимізації даних на основі самонавчання засобів підсистеми пам'яті систем і засобів протидії комп'ютерним атакам, який забезпечує адаптивне управління структурою, розміщенням і використанням інформаційних ресурсів в умовах динамічної зміни параметрів інформаційного середовища.

Запропонований метод, на відміну від існуючих підходів, забезпечує комплексну оптимізацію початкових, накопичених, використаних даних та даних обманних систем із урахуванням критеріїв ресурсної ефективності, структури даних, шарового розміщення, апаратно-мережного розподілу та специфіки функціонування систем кіберзахисту. Це дозволяє зменшити надлишковість інформаційних масивів, підвищити швидкодію доступу до критично важливої інформації та покращити інформативність інформаційного середовища.

Для оцінювання ефективності методу проведено експериментальні дослідження, у межах яких проаналізовано вплив рівня оптимізації даних на час доступу до критичної інформації, швидкість реагування системи на комп'ютерні атаки та точність їх виявлення. Результати експериментів підтвердили наявність стійкої залежності між рівнем оптимізації даних та інтегральними показниками ефективності системи протидії комп'ютерним атакам.

Отримані результати підтверджують, що використання запропонованого методу забезпечує підвищення швидкодії доступу до релевантних даних, скорочення часу реагування на загрози та покращення точності виявлення комп'ютерних атак. Практичне значення роботи полягає у можливості використання запропонованого методу як складової адаптивної архітектури обробки даних у сучасних системах і засобах кіберзахисту.

Перспективними напрямами подальших досліджень визначено розроблення архітектури засобів та методу вибору даних у системах захисту корпоративних мереж.

Ключові слова: корпоративні мережі; системи обману; комп'ютерні атаки; архітектура систем, оптимізація даних.

PAIUK Vadym, SAVENKO Bohdan
Khmelnytskyi National University

OPTIMIZATION OF DATA STORAGE IN COMPUTER ATTACK SYSTEMS AND TOOLS TAKING INTO ACCOUNT THE HISTORICAL ASPECT OF THEIR USE

The article considers the task of increasing the effectiveness of systems and means of countering computer attacks due to the improvement of data processing and optimization processes in memory subsystems. The relevance of the study is due to the constant growth of cyber-monitoring data volumes, the increasing complexity of modern computer attacks, and the need to ensure prompt access to relevant information for timely detection of threats and decision-making to counter them.

The purpose of the work is to develop a data optimization method based on self-learning of the system memory subsystem and countermeasures against computer attacks, which provides adaptive management of the structure, placement and use of information resources in conditions of dynamic changes in the parameters of the information environment.

The proposed method, in contrast to existing approaches, provides comprehensive optimization of initial, accumulated, used data and data of fraudulent systems, taking into account the criteria of resource efficiency, data structure, layered placement, hardware and network distribution, and the specifics of the functioning of cyber protection systems. This makes it possible to reduce the redundancy of information arrays, increase the speed of access to critically important information, and improve the informativeness of the information environment.

To evaluate the effectiveness of the method, experimental studies were conducted, within which the influence of the level of data optimization on the time of access to critical information, the speed of the system's response to computer attacks and the accuracy of their detection were analyzed. The results of the experiments confirmed the presence of a stable dependence between the level of data optimization and the integral indicators of the effectiveness of the system against computer attacks.

The obtained results confirm that the use of the proposed method provides an increase in the speed of access to relevant data, a reduction in the response time to threats, and an improvement in the accuracy of detection of computer attacks. The practical significance of the work lies in the possibility of using the proposed method as a component of the adaptive architecture of data processing in modern systems and means of cyber protection.

The development of the architecture of tools and the method of data selection in corporate network protection systems is defined as promising directions for further research.

Keywords: corporate networks; fraud systems; computer attacks; system architecture, data optimization.

Стаття надійшла до редакції / Received 10.06.2026
Прийнята до друку / Accepted 30.07.2026
Опубліковано / Published 10.09.2026

ПОСТАНОВКА ПРОБЛЕМИ У ЗАГАЛЬНОМУ ВИГЛЯДІ ТА ЇЇ ЗВ'ЯЗОК ІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

Сучасні комп'ютерні атаки (КА) на корпоративні мережі (КМ) характеризуються високою складністю, адаптивністю та здатністю швидко змінювати сценарії реалізації, що суттєво ускладнює процеси їх своєчасного виявлення та ефективної протидії. Під час здійснення КА зловмисники стикаються з багаторівневою системою захисту, яка включає різноманітні апаратно-програмні та програмні засоби забезпечення кібербезпеки [1, 2]. Ефективність функціонування таких засобів значною мірою визначається не лише якістю алгоритмів виявлення загроз, а й ефективністю обробки, зберігання та використання даних, що накопичуються в процесі їх функціонування.

На кожному рівні захисту корпоративних мереж формуються значні обсяги історичних даних, які містять інформацію про попередні стани системи, характеристики трафіку, ознаки атак та результати прийнятих рішень. Такі дані утворюють підсистему пам'яті, яка виступає ключовим джерелом накопиченого досвіду функціонування систем і засобів протидії КА [3]. Проте постійне зростання обсягів даних кібермоніторингу призводить до накопичення інформаційної надлишковості, збільшення часу доступу до критично важливої інформації та зниження оперативності прийняття рішень.

Зловмисники розуміють залежність сучасних систем кіберзахисту від ефективності функціонування підсистем пам'яті та прагнуть використовувати слабкі місця, пов'язані не лише з сенсорами виявлення, а й із процесами організації, обробки та використання історичних даних [4]. У таких умовах особливої актуальності набуває задача оптимізації даних у підсистемах пам'яті систем і засобів протидії КА.

Оптимізація даних повинна забезпечувати адаптивне управління структурою, розміщенням і використанням інформаційних ресурсів з урахуванням особливостей апаратно-програмного середовища, характеру інформаційного навантаження та динаміки змін кіберзагроз. Це дозволяє зменшити надлишковість інформаційних масивів, скоротити час доступу до релевантних даних, підвищити швидкодію обробки інформації та покращити ефективність підтримки процесів виявлення і протидії КА.

Таким чином, актуальним науково-прикладним завданням є розроблення нових методів оптимізації даних у підсистемах пам'яті систем і засобів протидії комп'ютерним атакам, які забезпечують адаптивне керування інформаційними ресурсами та підвищення загальної ефективності функціонування систем кіберзахисту.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Технологія блокчейн характеризується децентралізацією, безпекою та незмінністю даних, що забезпечило її широке застосування в різних сферах [5]. 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 [5].

У роботі [6] 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 [6].

Великі дані мають значний потенціал для підвищення [7] ефективності різних галузей, однак дані, що зберігаються у різних системах, часто є несумісними, що призводить до їх ізольованості 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 [8], 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 атак, 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 [9] 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 [10-13] 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.

ФОРМУЛЮВАННЯ ЦІЛЕЙ СТАТТІ

Метою роботи є підвищення ефективності оптимізації пам’яті з історичними даними щодо функціонування систем та засобів протидії КА в корпоративних мережах.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

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 співвіднесені 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:

$$M_{DI}^{1} = \left\{ m_{DI,1}^{1}, m_{DI,1}^{1}, \dots, m_{DI,N_{M_{DI}^{1}}}^{1} \right\}, \tag{1}$$

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:

$$M_{DI}^{D} = \left\{ V_{DI,1}^{D}, V_{DI,2}^{D}, \dots, V_{DI,N_{M_{DI}^{D}}}^{D} \right\}, \tag{2}$$

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) задамо загальними для всіх джерел інформації поданнями так:

$$V_{DI,i}^{D} = \left(v_{DI,i,1}^{D}, v_{DI,i,2}^{D}, \dots, v_{DI,i,N_{V_{DI,i}^{D}}}^{D} \right), \quad (3)$$

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 (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:

$$M_{DD}^{1} = \left\{ m_{DD,1}^{1}, m_{DD,2}^{1}, \dots, m_{DD,N_{M_{DD}^{1}}}^{1} \right\}, \quad (4)$$

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 (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 (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:

$$M_{DS}^{2} = \left\{ m_{DS,1}^{2}, m_{DS,2}^{2}, \dots, m_{DS,N_{M_{DS}^{2}}}^{2} \right\}, \quad (5)$$

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 (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 (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:

$$M_{DK}^{3} = \left\{ m_{DK,1}^{3}, m_{DK,2}^{3}, \dots, m_{DK,N_{M_{DK}^{3}}}^{3} \right\}, \quad (6)$$

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 (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 (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 непідготовлені некритичні, 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 нукають 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:

$$M_{DA}^4 = \left\{ m_{DA,1}^4, m_{DA,2}^4, \dots, m_{DA,N_{M_{DA}^4}}^4 \right\}, \quad (7)$$

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 (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 (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:

$$M_{DP}^5 = \left\{ m_{DP,1}^5, m_{DP,2}^5, \dots, m_{DP,N_{M_{DP}^5}}^5 \right\}, \quad (8)$$

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

Елементами множини M_{DP}^5 (формула (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) $m^5_{DP,3}$ – відомості про розподіл доступу; 4) $m^5_{DP,4}$ – відомості про сформовані профілі користувачів; 5) $m^5_{DP,5}$ – відомості про топологію корпоративних мереж; 6) $m^5_{DP,6}$ – відомості про розміщення конфіденційної інформації в корпоративних інформаційних системах; 7) $m^5_{DP,7}$ – відомості про довірені ресурси в корпоративних інформаційних системах; 8) $m^5_{DP,8}$ – відомості про трафік комп'ютерних мереж за період функціонування.

Кількість елементів множини M^5_{DP} (формула (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:

$$M^6_{DR} = \left\{ m^6_{DR,1}, m^6_{DR,2}, \dots, m^6_{DR,N_{M^6_{DR}}} \right\}, \tag{9}$$

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

Елементами множини M^6_{DR} (формула (9)) можуть бути такі: 1) $m^6_{DR,1}$ – відомості про КА за їх типами та для різних способів їх здійснення; 2) $m^6_{DR,2}$ – відомості про корпоративні інформаційні організації чи підприємства та їх ресурси; 3) $m^6_{DR,3}$ – відомості про засоби та системи протидії КА перших п'яти рівнів; 4) $m^6_{DR,4}$ – відомості про інциденти під час функціонування від початку встановлення систем та засобів; 5) $m^6_{DR,5}$ – відомості про рівні конфіденційної інформації корпоративних мереж та місця її розміщення; 6) $m^6_{DP,6}$ – відомості про довірені ресурси за периметром корпоративних мереж.

Кількість елементів множини M^6_{DR} (формула (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 сьомому, 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 висуваються гіпотези, перевіряються аномалії, аналізуються патерни поведінки, що в результаті формує нові правила виявлення.

Процес пошуку загроз на сьомому 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 поведкового аналізу, про комунікацію між хостами, топологію мереж тощо. Для етапу виявлення або спростування гіпотези використовують дані індикаторів компроментації, 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. Н**акопичені дані в пам'яті про події та інциденти є одними із ключових елементів в таких системах і засобах, але** важливо *як саме* вони

використовується. 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:

$$M_{DST}^7 = \left\{ m_{DST,1}^7, m_{DST,2}^7, \dots, m_{DST,N_{M_{DST}^7}}^7 \right\}, \quad (10)$$

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 (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 компроментації, 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 (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, зо 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 компроментацію. 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 використані 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 стільки 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.

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 [14] 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:

$$F_{opt}^{D}\left(P_0^{F_{opt}^{D}}, P_1^{F_{opt}^{D}}, \dots, P_{n_{F_{opt}^{D}}}^{F_{opt}^{D}}\right) = \sum_{i=0}^{n_{F_{opt}^{D}}} \alpha_i \cdot F_{opt,i}^{D}\left(p_1^{F_{opt,i}^{D}}, p_2^{F_{opt,i}^{D}}, \dots, p_{n_{F_{opt,i}^{D}}}^{F_{opt,i}^{D}}\right) \to min, \qquad (11)$$

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 (11) 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.**

Функції *Fopt,0Dp1Fopt,0D,p2Fopt,0D,…,pnFopt,0DFopt,0D* визначає базові критерії ресурсної оптимізації. Функції *Fopt,iDp1Fopt,iD,p2Fopt,iD,…,pnFopt,iDFopt,iD* ($i=1,2,\dots,4$) деталізують особливості структури даних, їх розподілу за категоріями, ієрархії шарів пам'яті, апаратного розміщення в корпоративній мережі та параметрів і даних обманних систем. Деталізуємо кожну окрему оптимізаційну функцію для конкретних критеріїв.

В загальному випадку ф**ункції** $F^D_{opt,i}\left(p_1^{F^D_{opt,i}}, p_2^{F^D_{opt,i}}, \dots, p_{n_{F^D_{opt,i}}}^{F^D_{opt,i}}\right)$ $(i = 0,1, \dots, 4)$ задамо так:

$$F^D_{opt,i}\left(p_1^{F^D_{opt,i}}, p_2^{F^D_{opt,i}}, \dots, p_{n_{F^D_{opt,i}}}^{F^D_{opt,i}}\right) = \sum_{j=1}^{n_{F^D_{opt,i}}} \alpha_{i,j} \cdot p_j^{F^D_{opt,i}} \to min, \quad (12)$$

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

Згідно формули 12) задамо ф**ункцію** $F^D_{opt,0}\left(p_1^{F^D_{opt,0}}, p_2^{F^D_{opt,0}}, \dots, p_{n_{F^D_{opt,0}}}^{F^D_{opt,0}}\right)$ поєднанням таких критеріїв оптимізації: 1) критерій використання пам'яті $p_1^{F^D_{opt,0}}$; 2) критерій часу доступу $p_2^{F^D_{opt,0}}$; 3) критерій щодо надлищковості даних $p_3^{F^D_{opt,0}}$; 4) критерій втрати актуальності $p_4^{F^D_{opt,0}}$; 5) критерій інформативності $p_5^{F^D_{opt,0}}$; 6) критерій витрат мережних ресурсів $p_6^{F^D_{opt,0}}$.

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

$$p_1^{F^D_{opt,0}} = \sum_{k=1}^{n_{p_1^{F^D_{opt,0}}}} \frac{v_{z,k}^{p_1^{F^D_{opt,0}}}}{v_{v,k}^{p_1^{F^D_{opt,0}}}}; \ 0 \le \frac{v_{z,k}^{p_1^{F^D_{opt,0}}}}{v_{v,k}^{p_1^{F^D_{opt,0}}}} \le 1, \quad (13)$$

де $v_{v,k}^{p_1^{F^D_{opt,0}}}$ – доступний обсяг пам'яті у k – тому вузлі; $k = 1,2, \dots, n_{p_1^{F^D_{opt,0}}}$; $n_{p_1^{F^D_{opt,0}}}$ – кількість вузлів; $v_{z,k}^{p_1^{F^D_{opt,0}}}$ – зайнятий обсяг пам'яті у k – тому вузлі.

Параметр $p_1^{F^D_{opt,0}}$ знаходиться в межах інтервалу $0 \le \frac{p_1^{F^D_{opt,0}}}{n_{p_1^{F^D_{opt,0}}}} \le 1$ та його оцінку для відображення критерію задамо, наприклад, так:

1) $0 \le \frac{p_1^{F^D_{opt,0}}}{n_{p_1^{F^D_{opt,0}}}} \le 0{,}4$ – низьке оптимальне завантаження;

2) $0{,}4 < \frac{p_1^{F^D_{opt,0}}}{n_{p_1^{F^D_{opt,0}}}} \le 0{,}75$ – середнє завантаження;

3) $0{,}75 < \frac{p_1^{F^D_{opt,0}}}{n_{p_1^{F^D_{opt,0}}}} \le 1$ – високе критичне завантаження.

Критерій часу доступу $p_2^{F^D_{opt,0}}$ задамо так:

$$p_2^{F^D_{opt,0}} = \sum_{k=1}^{n_{p_2^{F^D_{opt,0}}}} \frac{t_{z,k}^{p_2^{F^D_{opt,0}}}}{t_{v,k}^{p_2^{F^D_{opt,0}}}}; \ 0 \le \frac{t_{z,k}^{p_2^{F^D_{opt,0}}}}{t_{v,k}^{p_2^{F^D_{opt,0}}}} \le 1, \quad (14)$$

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

Параметр часу доступу $p_2^{F^D_{opt,0}}$ знаходиться в межах інтервалу $0 \le \frac{p_2^{F^D_{opt,0}}}{n_{p_2^{F^D_{opt,0}}}} \le 1$ та його оцінку для відображення критерію часу доступу $p_2^{F^D_{opt,0}}$ задамо, наприклад, так:

1) $0 \le \frac{p_1^{F^D_{opt,0}}}{n_{p_1^{F^D_{opt,0}}}} \le 0{,}5$ – низький оптимальний час доступу;

2) $0{,}5 < \frac{p_1^{F^D_{opt,0}}}{n_{p_1^{F^D_{opt,0}}}} \le 0{,}75$ – середній час доступу;

3) $0{,}75 < \frac{p_1^{F^D_{opt,0}}}{n_{p_1^{F^D_{opt,0}}}} \le 1$ – високий критичний час доступу.

Крім того, параметр часу доступу $p_2^{F^D_{opt,0}}$ 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 $p_3^{F^D_{opt,0}}$ для визначення надлишковості даних так:

$$p_3^{F^D_{opt,0}} = \frac{n_z^{p_3^{F^D_{opt,0}}}}{n_v^{p_3^{F^D_{opt,0}}}}, \tag{15}$$

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

Тоді, значення $p_3^{F^D_{opt,0}}$ повинно бути близьким до нуля, тобто $0 \le p_3^{F^D_{opt,0}}$, 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:

$$q_k = e^{-\lambda_k \cdot \Delta t_k}, \tag{16}$$

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 $p_4^{F^D_{opt,0}}$ даних з врахуванням формули (3.14) задамо так:

$$p_4^{F^D_{opt,0}} = 1 - \frac{1}{n_D} \cdot \sum_{k=1}^{n_D} q_k, \tag{17}$$

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 $p_4^{F^D_{opt,0}}$ такі: $0 \le p_4^{F^D_{opt,0}} \le 1$, причому актуальність даних можна встановити за такими значеннями:

1) $p_4^{F^D_{opt,0}} = 0$ – усі значення актуальні;

2) $0 < p_4^{F^D_{opt,0}} \le 0{,}5$ – часткова втрата актуальності даних;

3) $0{,}5 < p_4^{F^D_{opt,0}} \le 1$ – 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:

$$g_k = \frac{H_{k,1} - H_{k,2}}{H_{k,1}}, \tag{18}$$

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:

$$H_{k,m} = -\sum_{j=1}^{m_H} P_{k,j}^H \cdot log_2 P_{k,j}^H, \tag{19}$$

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 (19) співвідношення $H_{k,2} \le H_{k,1}$ задає критерій інформативності $p_5^{F_{opt,0}^D}$, згідно якого $0 \le g_k \le 1$. 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 (19) значення критерію інформативності $p_5^{F_{opt,0}^D}$ задамо так:

$$p_5^{F_{opt,0}^D} = \frac{1}{m_{p_5^{F_{opt,0}^D}}} \cdot \sum_{k=1}^{p_5^{F_{opt,0}^D}} g_k, \tag{20}$$

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 пропускної здатності в КМ так:

$$p_6^{F_{opt,0}^D} = \frac{1}{m_{p_6^{F_{opt,0}^D}}} \cdot \sum_{k=1}^{m_{p_6^{F_{opt,0}^D}}} \frac{Q_{k,1}}{Q_{k,1} + Q_{k,2}}, \tag{21}$$

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 (21) 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 (12). 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:

$$p_1^{F_{opt,1}^D} = \frac{Z_{1,2} - Z_{1,1}}{Z_{1,2}}, \tag{22}$$

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:

$$p_2^{F_{opt,1}^D} = \frac{Z_{2,2} - Z_{2,1}}{Z_{2,2}}, \tag{23}$$

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, різнорідні джерела телеметрії та різні типи обманних систем. Критерій визначимо так:

$$p_3^{F_{opt,1}^D} = \frac{Z_{3,2} - Z_{3,1}}{Z_{3,2}}, \tag{24}$$

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 визначаєг наявність в інформаційному масиві незалежних інформаційних сутностей, а не повторення однакових структурних шаблонів.

$$p_4^{F_{opt,1}^D} = \frac{Z_{4,1}}{Z_{4,2}}, \tag{25}$$

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:

$$p_5^{F_{opt,1}^D} = \frac{Z_{5,2} - Z_{5,1}}{Z_{5,2}}, \tag{26}$$

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 $F^D_{opt,2}\left(p_1^{F^D_{opt,2}}, p_2^{F^D_{opt,2}}, \dots, p_{n_{F^D_{opt,2}}}^{F^D_{opt,2}}\right)$ задамо з урахуванням формули (3.10) та виділимо такі критерії шарового розміщення даних $F^D_{opt,2}$: 1) критерій пріоритетності розміщення даних $p_1^{F^D_{opt,2}}$; 2) критерій міграції даних між шарами $p_2^{F^D_{opt,2}}$; 3) критерій балансування шарів $p_3^{F^D_{opt,2}}$; 4) критерій релевантності шару $p_4^{F^D_{opt,2}}$; 5) критерій ефективності використання шарів даних $p_5^{F^D_{opt,2}}$.

Критерій пріоритетності розміщення даних $p_1^{F^D_{opt,2}}$ 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:

$$p_1^{F^D_{opt,2}} = \frac{1}{m_{p_1^{F^D_{opt,2}}}} \cdot \sum_{k=1}^{n_{p_1^{F^D_{opt,2}}}} \frac{|S_{1,k,1} - S_{1,k,2}|}{m_{p_1^{F^D_{opt,2}}}}, \tag{27}$$

де $m_{p_6^{F^D_{opt,0}}}$ – кількість шарів; $n_{p_1^{F^D_{opt,2}}}$ – кількість об'єктів даних; $k = 1,2, \dots, n_{p_1^{F^D_{opt,2}}}$; $S_{1,k,1}$ – фактичний шар розміщення; $S_{1,k,2}$ – рекомендований шар.

Значення $p_1^{F^D_{opt,2}}$ повинно прямувати до нуля в найкращому варіанті.

Критерій міграції даних між шарами $p_2^{F^D_{opt,2}}$ 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:

$$p_2^{F^D_{opt,2}} = \frac{S_{2,1}}{S_{2,2}}, \tag{28}$$

де $S_{2,1}$ – кількість коректних міграцій; $S_{2,2}$ – загальна кількість міграцій.

Критерій балансування шарів $p_3^{F^D_{opt,2}}$ 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 $p_4^{F^D_{opt,2}}$ 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 $p_5^{F^D_{opt,2}}$ 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 обсягу 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 $F^D_{opt,3}\left(p_1^{F^D_{opt,3}}, p_2^{F^D_{opt,3}}, \dots, p_{n_{F^D_{opt,3}}}^{F^D_{opt,3}}\right)$ 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 $F^D_{opt,3}\left(p_1^{F^D_{opt,3}}, p_2^{F^D_{opt,3}}, \dots, p_{n_{F^D_{opt,3}}}^{F^D_{opt,3}}\right)$ задамо з урахуванням формули (12) та виділимо такі критерії апаратно-мережного розподілу $F^D_{opt,3}$: 1) критерій збалансованості обчислювального навантаження $p_1^{F^D_{opt,3}}$; 2) критерій мережної локалізації даних $p_2^{F^D_{opt,3}}$; 3) критерій стійкості мережної інфраструктури $p_3^{F^D_{opt,3}}$; 4) критерій маршрутизації інформаційних потоків $p_4^{F^D_{opt,3}}$; 5) критерій доступності вузлів зберігання $p_5^{F^D_{opt,3}}$; 6) критерій сегментації корпоративної мережі $p_6^{F^D_{opt,3}}$; 7) критерій масштабованості інфраструктури $p_7^{F^D_{opt,3}}$.

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 $F^D_{opt,4}\left(p_1^{F^D_{opt,4}}, p_2^{F^D_{opt,4}}, \dots, p_{n_{F^D_{opt,4}}}^{F^D_{opt,4}}\right)$ 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 обманних 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 обманні 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 обманних систем $F^D_{opt,4}\left(p_1^{F^D_{opt,4}}, p_2^{F^D_{opt,4}}, \dots, p_{n_{F^D_{opt,4}}}^{F^D_{opt,4}}\right)$ задамо з урахуванням формули (3.10) та виділимо такі критерії оптимізації даних обманних систем $F^D_{opt,4}$: 1) критерій правдоподібності інформаційного середовища $p_1^{F^D_{opt,2}}$; 2) критерій адаптивності сценаріїв обману $p_2^{F^D_{opt,2}}$; 3) критерій інформативності даних обманних систем $p_3^{F^D_{opt,2}}$; 4) критерій узгодженості з реальною інфраструктурою $p_4^{F^D_{opt,2}}$; 5) критерій ізоляції інформаційних потоків обманних систем $p_5^{F^D_{opt,2}}$; 6) критерій накопичення знань про сценарії КА $p_3^{F^D_{opt,2}}$; 7) критерій аналітичної цінності даних обманних систем $p_4^{F^D_{opt,2}}$; 8) критерій ресурсної раціональності зберігання даних обманних систем $p_5^{F^D_{opt,2}}$.

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. Порушення 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 (11) 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 (11) та критеріями за формулою (12).

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 (12)). 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 (11).

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 (11). 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 різнорідних 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 мережевими 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.

Експерименти

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:

$$\Delta E_{opt} = E_{opt} - E_{opt,base}, \tag{29}$$

де E_{opt} - ефективність системи після впровадження методу оптимізації даних; $E_{opt,base}$ - ефективність системи без впровадження методу оптимізації даних.

Результати експерименту подамо у вигляді табл. 1.

Таблиця 1

Вплив оптимізації даних на достовірність виявлення КА

Показник	До оптимізації	Після оптимізації	Покращення
Інформаційна надлишковість			
Час доступу до критичних даних	1.8 с	0.9 с	
Час реагування на КА	4.6 с	2.7 с	
Ймовірність виявлення КА			
Хибні спрацювання			

Для аналізу результатів будуємо 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. 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. За апроксимуючу функцію приймемо лінійну модель та використовуючи метод найменших квадратів отримуємо таку функцію:

$$T_{access} = 2{,}1 - 1{,}48\, E_{opt}, \tag{30}$$

де E_{opt} - ефективність системи з використанням методу оптимізації даних; T_{access} – час доступу до даних.

Графік експериментальних і апроксимованих даних зображено на рис. 1. Графік зміни швидкості реагування на КА, зображений на рис. 2, відображає зміну часу реагування системи на КА після впровадження

методу оптимізації даних.

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). Графік залежності точності виявлення КА від показника E_{opt} зображено на рис. 3. 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.

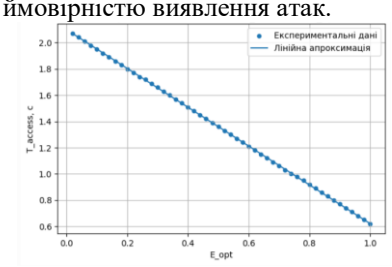

Рис. 1. Графік експериментальних і апроксимованих даних

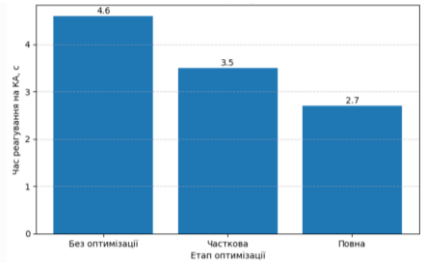

Рис. 2. Графік зміни швидкості реагування на КА

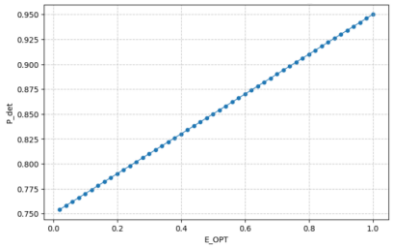

Рис. 3. Графік залежності точності виявлення КА від показника E_{opt}

Скорочення часу реагування підтверджує, що швидкий доступ до оптимізованих даних прискорює процес прийняття рішень щодо нейтралізації КА.

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, експеримент підтверджує, що адаптивна оптимізація структури та розміщення даних є одним із ключових факторів підвищення ефективності сучасних систем кіберзахисту.

ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ

У результаті розроблення методу оптимізації даних на основі самонавчання засобів підсистеми пам'яті систем і засобів протидії комп'ютерним атакам забезпечено адаптивне управління структурою, розміщенням та використанням даних з урахуванням їх типу, ресурсної ефективності та особливостей функціонування інформаційного середовища.

Особливістю запропонованого методу є комплексна оптимізація початкових, накопичених, використаних даних та даних обманних систем із врахуванням критеріїв зменшення інформаційної надлишковості, підвищення швидкодії доступу до даних та покращення ефективності їх подальшого використання у процесах виявлення і протидії комп'ютерним атакам. Результати експериментального дослідження підтвердили ефективність запропонованого методу. Встановлено, що підвищення рівня оптимізації даних забезпечує скорочення часу доступу до критичної інформації, підвищення швидкості реагування системи на комп'ютерні атаки та покращення точності їх виявлення. Аналіз експериментальних даних показав наявність стійкої причинно-наслідкової залежності між рівнем оптимізації даних та інтегральними показниками ефективності системи протидії комп'ютерним атакам.

Таким чином, результати теоретичних і експериментальних досліджень підтвердили, що розроблений метод оптимізації даних є ефективним інструментом підвищення продуктивності систем кіберзахисту та може бути використаний як ключовий компонент адаптивної архітектури обробки даних у системах і засобах протидії комп'ютерним атакам.

Напрямами подальших досліджень є розроблення архітектури засобів вибору даних для систем та засобів протидії КА в КМ, зокрема розроблення методу вибору даних.

Література

1. Kashtalian A., Savenko O., Sachenko A. Agglomerative clustering of data collected by honeypots. 2021 11th IEEE international conference on intelligent data acquisition and advanced computing systems: technology and applications (IDAACS), Cracow, Poland, 22–25 September 2021. 2021. DOI: https://doi.org/10.1109/idaacs53288.2021.9661027

2. 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

3. Kashtalian A., Lysenko S., Sachenko A., Savenko B., Savenko O., Nicheporuk A. Evaluation criteria of centralization options in the architecture of multicomputer systems with traps and baits. Radioelectronic and Computer Systems. 2025. Vol. 1. Pp. 264-297. DOI: https://doi.org/10.32620/reks.2025.1.18

4. Kashtalian A., Lysenko S., Kysil T., Sachenko A., Savenko O., Savenko B. Method and Rules for Determining the Next Centralization Option in Multicomputer System Architecture. International Journal of Computing. 2025. Vol. 24(1), 35-51. DOI: https://doi.org/10.47839/ijc.24.1.3875

5. Koukaras P., Tjortjis C. Data Organisation for Efficient Pattern Retrieval: Indexing, Storage, and Access Structures. Big Data and Cognitive Computing. 2025; 9(10):258. https://doi.org/10.3390/bdcc9100258

5. 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

6. 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

7. Jiang S. Big Data Sharing: A Comprehensive Survey. Data, 2025. 10(11), 182. https://doi.org/10.3390/data10110182

8. 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

9. 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

10. 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. https://doi.org/10.51594/csitrj.v4i3.661

11. 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. https://doi.org/10.1177/14649934221076580

12. 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. https://doi.org/10.1109/PDGC56933.2022.10053276

13. 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. https://doi.org/10.1051/SHSCONF/202110202001

14. PAIUK B., & SAVENKO Б. (2026). ORGANIZATION OF DATA STORAGE IN COMPUTER ATTACK SYSTEMS AND TOOLS TAKING INTO ACCOUNT THE HISTORICAL ASPECT OF THEIR USE. MEASURING AND COMPUTING DEVICES IN TECHNOLOGICAL PROCESSES, (2), 321–336. https://doi.org/10.31891/2219-9365-2026-86-39