ISSN 2219-9365

https://doi.org/10.31891/2219-9365-2026-86-39
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

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

У статті розглянуто проблему організації зберігання даних у системах та засобах протидії комп'ютерним атакам на корпоративні мережі з урахуванням історичного аспекту їх використання. Показано, що під час проведення комп'ютерних атак зловмисники стикаються з багаторівневою системою захисту, яка відповідає рівням піраміди болю зловмисника та включає різноманітні апаратно-програмні й програмні засоби. Ефективність функціонування таких засобів значною мірою залежить від наявності та організації підсистем пам'яті, що накопичують історичні дані про попереднє функціонування систем захисту. З огляду на це зловмисники дедалі частіше намагаються обходити не лише сенсори та механізми виявлення, а й елементи пам'яті, які містять відомості про попередні інциденти, сценарії реагування та результати використання даних.

Для підвищення рівня кібербезпеки запропоновано моделі зберігання даних та метод організації їх зберігання в системах і засобах протидії комп'ютерним атакам. Метод передбачає поділ даних на початкові, накопичувані та використані, що забезпечує врахування повного життєвого циклу даних, тобто від моменту їх формування до повторного застосування та оцінювання результативності такого використання. Особливістю запропонованого підходу є збереження та аналіз історичного досвіду використання даних із застосуванням марковських моделей, динамічних баєсівських мереж і статистичних методів. Марковські моделі використовуються для опису послідовностей станів і сценаріїв розвитку подій, динамічні баєсівські мережі забезпечують врахування причинно-наслідкових зв'язків та невизначеності у часі, а статистичні методи дають змогу узагальнювати результати та оцінювати ефективність прийнятих рішень.

Запропонований метод сприяє накопиченню історичних відомостей щодо використання даних, підвищенню ефективності їх повторного застосування, адаптивності та автономності функціонування систем протидії комп'ютерним атакам, зокрема в умовах обмеженої доступності спеціалізованих засобів зберігання даних.

Перспективними напрямами подальших досліджень визначено розроблення архітектури засобів зберігання, методів оптимізації та вибору даних, а також інтеграцію обманних технологій у процеси організації та використання історичних 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

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

The article examines the problem of organizing data storage in systems and means of countering computer attacks on corporate networks, taking into account the historical aspect of their use. It is shown that when conducting computer attacks, attackers encounter a multi-level defense system that corresponds to the levels of the attacker's pain pyramid and includes various hardware and software tools. The effectiveness of the functioning of such means largely depends on the presence and organization of memory subsystems that accumulate historical data on the previous functioning of protection systems. With this in mind, attackers are increasingly trying to bypass not only sensors and detection mechanisms, but also memory elements that contain information about previous incidents, response scenarios, and data usage results.

To increase the level of cyber security, data storage models and a method of organizing their storage in systems and means of countering computer attacks are proposed. The method involves the division of data into initial, accumulated and used data, which ensures consideration of the full life cycle of data, i.e. from the moment of their formation to repeated application and evaluation of the effectiveness of such use. A feature of the proposed approach is the preservation and analysis of historical data usage experience using Markov models, dynamic Bayesian networks, and statistical methods. Markov models are used to describe sequences of states and scenarios of the development of events, dynamic Bayesian networks ensure consideration of cause-and-effect relationships and uncertainty in time, and statistical methods make it possible to generalize results and evaluate the effectiveness of decisions made.

The proposed method contributes to the accumulation of historical information on the use of data, increasing the efficiency of their reuse, adaptability and autonomy of the functioning of systems for countering computer attacks, in particular, in conditions of limited availability of specialized data storage facilities.

The development of the architecture of storage facilities, methods of optimization and data selection, as well as the integration of deception technologies into the processes of organization and the use of historical data in the protection systems of corporate networks are defined as promising directions for further research.

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

ISSN 2219-9365

Стаття надійшла до редакції / Received 30.03.2026
Прийнята до друку / Accepted 28.04.2026
Опубліковано / Published 31.05.2026

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

Зловмисники при здійсненні комп'ютерних атак (КА) на корпоративні мережі (КМ) стикаються з протидією їм різних апаратно-програмних та програмних систем [1, 2], що відображено в класифікації індикаторів компрометації піраміди болю зловмисника. Долаючи перешкоди при проникненні в корпоративні мережі зловмисники витрачають багато ресурсів та часу і в цьому процесі для них зростають складності, які пов'язані з проведенням КА. На кожному рівні піраміди болю зловмисника для забезпечення безпеки та захисту корпоративних мереж використовуються різні апаратно-програмні та програмні засоби і системи. Всі вони залежать від даних щодо свого попереднього функціонування, які формують підсистему пам'яті з історичними даними [3]. В залежності від етапу, на якому розміщуються засоби та системи протидії КА організація їх пам'яті з історичними даними є різною. Зловмисники розуміють ієрархію засобів та систем у співвіднесенні до рівнів піраміди болю зловмисника, а також особливо щодо організації в них підсистем пам'яті з історичними даними. Вважаючи елементи чи підсистеми пам'яті одними з основних джерел попереднього досвіду функціонування для протидії КА [4], зловмисники прагнуть обходити не стільки сенсори засобів та систем протидії КА, а саме пам'яті з історичними даними щодо функціонування таких засобів та систем. Тому, організація елементів чи підсистем пам'яті з історичними даними щодо функціонування потребує постійного удосконалення за різними рівнями та формами, зокрема і з використанням обманних технологій, які активно інтегруються в апаратно-програмні та програмні засоби і системи протидії КА на корпоративні мережі, для заплутування зловмисників під час проведення ними КА. Це дало б змогу підвищити рівень кібербезпеки в корпоративних мережах за рахунок змінюваної організації елементів чи підсистем пам'яті з історичними даними щодо попереднього функціонування систем і засобів та з використанням відповідних обманних технологій.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Зростаючий масштаб і складність результатів інтелектуального аналізу даних, таких як часті набори елементів, правила асоціації, послідовності та підграфи, зробили ефективне пошук шаблонів критичним, але недостатньо вивченим завданням. У [5] розглядаються стратегії організації, індексування та доступу, які забезпечують масштабоване та швидке пошук структурованих шаблонів. Було досліджено базові типи даних і вихідних шаблонів, загальні операції пошуку та різноманітність типів запитів, які зустрічаються на практиці. Розглянуто ключові структури індексування, включаючи дерева префіксів, інвертовані індекси, підходи на основі хешу та методи на основі растрових зображень, кожен з яких підходить для різних представлень шаблонів і робочого навантаження. Розглянуто стратегії оптимізації запитів, наголошуючи на обході з урахуванням індексу, кешуванні та механізмах ранжирування. Огляд в [5] забезпечує комплексну основу для розробки систем пошуку шаблонів нового покоління, які є масштабованими, гнучкими та тісно інтегрованими в аналітичні робочі процеси. Пропонований аналіз для всіх сфер застосування, включаючи кібербезпеку, де важливе значення має надійний пошук із можливістю інтерпретації.

У поточному сценарії універсальної доступності даних стикаються з надзвичайно складними проблемами, які пов'язані з інтеграцією та обробкою різноманітних наборів даних для задоволення своїх аналітичних потреб [6]. У роботі аналізуються традиційні та інноваційні методи, що використовуються для зберігання та інтеграції даних, з особливим акцентом на їх вплив на масштабованість, узгодженість і взаємодію в аналітичній екосистемі. Зокрема, він вносить міжрівневу таксономію, що пов'язує механізми інтеграції (відповідність схем, роздільна здатність об'єктів і семантичне збагачення) до субстратів зберігання/запитів (сховища рядків/стовпців, NoSQL, lakehouse та об'єднання), разом із порівняльними таблицями та рисунками, які синтезують компроміси та важелі ефективності/керування. Завдяки рішенням відображення схем, які вирішують проблеми, спричинені структурною неоднорідністю, архітектурою зберігання даних, що варіюється від традиційних рішень до хмарних рішень, а також конвеєрної інтеграції ETL за допомогою об'єднаних процесорів запитів, дослідження приділяє особливу увагу застосуванню керування метаданими, зосереджуючись на семантичному збагаченні за допомогою онтологій і керування лініями для забезпечення наскрізної відстежуваності та управління. Він також охоплює гарячі точки продуктивності та методи кешування, а також компроміси узгодженості, що виникають у розподілених системах.

В останні роки дослідження блокчейну привернули увагу з усього світу [7]. Це децентралізована компетенція, яка розповсюджена та невизначена. Кілька країн і вчених вже успішно застосували блокчейн у багатьох сферах. Блокчейн необхідний у делікатних ситуаціях, оскільки він захищає дані та запобігає їх зміні чи підробці. Крім того, підвищений попит ринку на дані стимулює попит на масштабування даних у всіх галузях. Дослідники з багатьох країн з часом використовували блокчейн у різних секторах, таким чином

приділяючи особливу увагу цьому новому ескалаційному домену блокчейну. Кожен дослідницький проект починається з глибоких знань про робочу область, а нова цікава інформація про блокчейн досить розрізнена. У цьому дослідженні аналізується академічна література про технологію блокчейн, наголошуючи на трьох ключових аспектах: сховище блокчейну, масштабованість і доступність. Це критичні області в ширшому полі технології блокчейн. У цьому дослідженні використовуються CiteSpace і VOSviewer, щоб всебічно зрозуміти поточний стан досліджень у цих областях. Це інструменти бібліометричного аналізу, які зазвичай використовуються в академічних дослідженнях для вивчення моделей і зв'язків у науковій літературі. Таким чином, щоб візуалізувати спосіб зберігання даних із масштабованістю та доступністю, зберігаючи синхронізацію безпеки блокчейну, було проведено необхідні дослідження щодо зберігання, масштабованості та доступності даних у середовищі блокчейну. Кінцева мета полягала в сприянні розробці безпечних та ефективних рішень для зберігання даних у рамках технології блокчейн.

Пандемія COVID-19 підкреслила необхідність гнучких служб охорони здоров'я, які забезпечують надійний і безпечний обмін інформацією, але досягнення належного, приватного та безпечного обміну EMR залишається проблемою через різноманітні формати даних і фрагментовані записи в багатьох накопичувачах даних, що призводить до ускладненої координації між командами охорони здоров'я, можливих медичних помилок і затримок у догляді за пацієнтами. У той час як централізовані системи EMR створюють ризики конфіденційності, а різноманітність форматів даних ускладнює взаємодію, технологія блокчейн пропонує багатообіцяюче рішення, забезпечуючи децентралізоване зберігання, гарантуючи цілісність даних, покращуючи контроль доступу, усуваючи посередників і підвищуючи ефективність охорони здоров'я. У роботі [8] досліджено значення стандартів EMR, проблеми безпеки та підходи на основі блокчейну для сприяння взаємодії та безпечного обміну даними.

Технологія блокчейн була успішно застосована в останні роки для сприяння незмінності, відстежуваності та автентичності раніше зібраних і збережених даних [9]. Однак обсяг даних, що зберігаються в блокчейні, зазвичай обмежений з економічних і технологічних причин. Зокрема, блокчейн зазвичай зберігає лише відбиток даних, наприклад хеш даних, тоді як повна необроблена інформація зберігається поза ланцюгом. Зазвичай цього достатньо, щоб гарантувати незмінність і відстежуваність, але не підтримує іншу важливу властивість, тобто доступність даних. Це особливо потрібно, коли традиційна централізована база даних вибирається для зберігання поза мережею. З цієї причини багато пропозицій намагаються правильно поєднати блокчейн із децентралізованим сховищем IPFS. Однак зберігання даних на IPFS може спричинити деякі проблеми конфіденційності. У роботі запропоновано рішення, яке належним чином поєднує блокчейн, IPFS і методи шифрування, щоб гарантувати незмінність, відстежуваність, доступність і конфіденційність даних.

Доступність Інтернету та його інтеграція з інтелектуальними технологіями віддали перевагу повсякденним предметам і речам і запропонували нові сфери, такі як Інтернет речей (IoT). IoT відноситься до концепції, коли розумні пристрої або речі підключаються та створюють мережу. Ця нова область страждає від обробки великих даних і проблем із безпекою [10]. Існує потреба в розробці моделі аналізу даних з використанням нових технологій, архітектури та моделі безпеки 5G. Надійний обмін даними за наявності законних вузлів завжди є однією з проблем у цих мережах. Інфіковані вузли генерують неточну інформацію та порушують безпеку користувача. У роботі запропоновано модель аналізу даних і архітектура самоорганізації для мереж IoT, щоб зрозуміти різні рівні технології і процеси. У роботі також запропоновано модель безпеки, яка заснована на механізмі автентифікації, виявлення та прогнозування для мереж IoT. Запропонована модель підвищує безпеку та захищає мережу від DoS та DDoS атак.

Під час випробувань [11], щоб відповідати вимогам послідовного отримання та зберігання кількох цілей, кількох систем і кількох типів даних, різні типи даних обробляються в потоки даних імпульсно-кодової модуляції (PCM) із використанням кодування PCM для зберігання. Відповідно до вимог зберігання в режимі реального часу потоків даних PCM з високою бітовою швидкістю розроблено систему зберігання великої ємності на основі Serial Advanced Technology Attachment 3.0 (SATA3.0). Система використовує програмовану вентильну матрицю серії Kintex 7 (FPGA) як ядро керування, отримує потоки даних PCM через низьковольтний диференціальний інтерфейс низьковольтної диференціальної сигналізації (LVDS), зберігає отримані потоки даних PCM на диск mSATA через шину передачі SATA3.0 і передає збережені дані назад на головний комп'ютер через інтерфейс USB3.0 для аналізу. Тим часом, щоб вирішити проблему експорту складних даних, система зберігання створює файлову систему FAT32 через програмне ядро MicroBlaze для оптимізації керування та роботи системи зберігання великої ємності.

Через проблеми із складним збором даних [12], тривалим періодом реалізації проекту, складними даними, поганою безпекою, складною можливістю відстеження та взаємозв'язку даних, управління архівами більшості національної інфраструктури все ще перебуває в доінформаційній епосі. Для вирішення цих проблем у роботі запропоновано архітектуру зберігання даних для національної інфраструктури на основі блокчейну. Вибір даних, що зберігаються в ланцюжку, і дані кількох регіонів або полів моделюються спільно за допомогою федеративного навчання. Параметри та результати зберігаються в ланцюжку, а інформація кожного вузла спільно використовується для вирішення проблеми обміну даними.

У хмарних системах зберігання користувачі повинні мати можливість вимикати програму, коли вона не використовується, і перезапускати її з останнього узгодженого стану, коли потрібно [13]. BlobSeer - це програма для зберігання даних, спеціально розроблена для розподілених систем, створена як альтернативне рішення для існуючої популярної системи зберігання з відкритим кодом розподіленої файлової системи Hadoop (HDFS). У хмарній моделі всі компоненти повинні зупинятися та перезапускатися з узгодженого стану, коли цього вимагає користувач. Одним із обмежень BlobSeer DFS є можливість втрати даних під час перезавантаження системи. Таким чином, важливо забезпечити послідовний стан запуску та зупинки компонентів BlobSeer під час використання в хмарному середовищі, щоб запобігти втраті будь-яких даних. У роботі досліджено можливість BlobSeer забезпечити систему розподіленого зберігання даних узгодженого стану з інтеграцією функції перезапуску контрольних точок. Щоб продемонструвати доступність узгодженого стану, було створено кластер із кількома машинами та розгорнуто сутності BlobSeer із функцією контрольних точок на різних машинах.

Конвеєри великих даних розроблені для обробки даних, що характеризуються однією або декількома з трьох особливостей великих даних, широко відомих як три V (об'єм, швидкість і різноманітність), за допомогою серії кроків (наприклад, вилучення, трансформація та переміщення), створюючи основу для використання розширеної аналітики та методів ML/AI [14]. Обчислювальний континуум (тобто хмара/туман/край) дозволяє отримати доступ до практично нескінченної кількості ресурсів, де конвеєри даних можуть виконуватися в масштабі. Однак реалізація конвеєрів даних у континуумі є складним завданням, яке потребує врахування обчислювальних ресурсів, каналів передачі даних, тригерів, методів передачі даних, інтеграції черг повідомлень тощо. Завдання стає ще більш складним, якщо зберігання даних розглядається як частина конвеєрів даних. Локальне сховище дороге, складне в обслуговуванні та пов'язане з кількома проблемами (наприклад, доступність даних, безпека даних і резервне копіювання). Використання хмарного сховища, тобто сховища як послуги (StaaS), замість локального сховища має потенціал для забезпечення більшої гнучкості з точки зору масштабованості, відмовостійкості та доступності. Запропоновано загальний підхід до інтеграції StaaS з конвеєрами даних, тобто обчислення на локальному сервері або в певній хмарі, але інтеграцію зі StaaS, і розроблено метод ранжирування доступних варіантів зберігання на основі п'яти ключових параметрів: вартість, близькість, продуктивність мережі, шифрування на стороні сервера та ваги/уподобання користувачів. Проведена оцінка демонструє ефективність запропонованого підходу з точки зору продуктивності передачі даних, корисності окремих параметрів і можливості динамічного вибору варіанта зберігання на основі чотирьох основних сценаріїв користувача.

Сума великих даних [15], створених із різних джерел, значно збільшується з кожним днем до такої міри, що для традиційних методів зберігання стає складно зберігати цю велику кількість даних. З цієї причини більшість організацій вирішили використовувати сторонні хмарні сховища для зберігання даних. Останнім часом хмарне сховище прогресує, але все ще стикається з численними проблемами щодо безпеки та конфіденційності. У цьому документі розглядаються проблеми безпеки та конфіденційності великих даних, а також мінімальні вимоги, які мають забезпечувати майбутні рішення. Основна мета роботи полягає в розробленні нової технічної структури для контролю та управління ризиками безпеки та конфіденційності великих даних. Запропонований фреймворк використовує переваги технології Blockchain для забезпечення безпечного зберігання великих даних шляхом керування їх метаданими та політиками та усунення зовнішніх сторін для підтримки безпеки та конфіденційності даних. Крім того, він використовує технологію мобільного агента, щоб скористатися перевагами, пов'язаними з продуктивністю системи в цілому.

ФОРМУЛЮВАННЯ ЦІЛЕЙ СТАТТІ

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

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ

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

Елементи чи підсистеми пам'яті в засобах чи системах протидії КА відповідно до поділу на рівні із зростанням рівнів, тобто для вищих рівнів порівняно з нижчими, містять більше даних про попередні події, зберігають дані триваліший час, а також дані для прийняття рішень, тобто загалом в них дані мають довшу історію, складніший математичний апарат їх опрацювання, ширший контекст прийняття рішень на їх основі та менша залежність від окремих сигнатур. Ефективність функціонування систем та засобів протидії КА в корпоративних мережах суттєво залежить від організації збереження даних в пам'яті та доступу до них, бо процеси в мережах відбуваються швидко і актуальність опрацьовуваних даних повинна бути належною. Тобто використання наявних в поточний момент часу даних, які зберігались в системах та засобах на початку їх встановлення, накопичувались і зберігаються протягом певного часу, включно і про події в контексті КА впливають на прийняття рішень системами і засобами протидії КА в корпоративних мережах.

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

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

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

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

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

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

1) $m_{vp,1}^{p}$ – обсяг дискових ресурсів для зберігання даних;
2) $m_{vp,2}^{p}$ – класифікація даних за приналежністю до завдань систем та засобів;
3) $m_{vp,3}^{p}$ – структурування та представлення класифікованих даних;
4) $m_{vp,4}^{p}$ – індексування даних в класах;
5) $m_{vp,5}^{p}$ – наявність різних методів пошуку даних за вимогами;
6) $m_{vp,6}^{p}$ – розподілення даних в мережі;
7) $m_{vp,7}^{p}$ – розміщення даних;
8) $m_{vp,8}^{p}$ – наявність різних методів опрацювання даних;
9) $m_{vp,9}^{p}$ – топології мережі;
10) $m_{vp,10}^{p}$ – розподіл ролей користувачів корпоративних мереж
11) $m_{vp,11}^{p}$ – технічна організація політик доступу в корпоративних мережах;
12) $m_{vp,12}^{p}$ – оновлення програмного забезпечення в комп'ютерних станціях;
13) $m_{vp,13}^{p}$ – доповнення / вилучення комп'ютерних станцій в корпоративних мережах;
14) $m_{vp,14}^{p}$ – заміна операційних систем в комп'ютерних станціях;
15) $m_{vp,15}^{p}$ – оновлення систем та засобів забезпечення безпеки і захисту корпоративних мереж та їх вузлів повністю або частково;
16) $m_{vp,16}^{p}$ – часткова заміна мережного обладнання корпоративних мереж;
17) $m_{vp,17}^{p}$ – оновлення баз сигнатур, баз правил;
18) $m_{vp,18}^{p}$ – відсутність оновлень баз сигнатур, баз правил протягом тривалого часу;
19) $m_{vp,19}^{p}$ – відсутність змін в топології та апаратно-програмних і програмних засобах корпоративних мереж, включно із засобами та системами забезпечення безпеки і захисту, протягом тривалого часу експлуатації корпоративних мереж;
20) $m_{vp,20}^{p}$ – автоматичне формування нових сигнатур для баз сигнатур засобами та системами забезпечення безпеки і захисту корпоративних мереж;
21) $m_{vp,21}^{p}$ – застосування методів протидії КА, які оперативно виконують завдання;
22) $m_{vp,22}^{p}$ – застосування методів протидії КА, які здійснюють аналіз ситуацій в корпоративних мережах і, відповідно, можуть повільно виконати завдання.

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

Обсяг дискових ресурсів для зберігання даних суттєво впливає на збереження даних, оскільки вони постійно накопичуються та потребують значного простору для розміщення. Дані, які зберігаються, мають різні типи, структуру, отримані з різних джерел та позиціонуються за різним призначенням для їх подальшого

використання, тому для їх зберігання потрібно здійснювати класифікацію нових даних за приналежністю до завдань систем та засобів з подальшим їх розміщенням відповідно до приналежності визначеним класам. Для того, щоб класифіковані дані було зручно і швидко отримувати з місць їх розміщення потрібно забезпечити належне структурування та представлення. Нові дані повинні доповнювати наявні дані і, тому, їх узгодження щодо порядку повинно реалізовано через індексування даних у відповідних класах. Також, можуть варіанти з даними, які віднесені до певних класів, а в них надмірна або недостатня кількість параметрів порівняно з наявними даними в класах, що потребує додаткового їх приведення у відповідність, бо впливатиме на їх використання. В системах та засобах протидії КА наявні різні методи пошуку даних за вимогами, які впливають на організацію індексування та формати розміщення даних. При розміщенні даних в корпоративних мережах в зв'язку з потребою ефективного використання обсягів пам'яті комп'ютерних станцій та серверів вони можуть бути розміщені за класами в різних вузлах мережі, а може бути частина класів подана в різних вузлах в межах даних кожного класу, що впливатиме на збереження та актуальність даних в різних аспектах їх використання та забезпечення доступу до них. Тобто дані певного класу можуть бути фізично розміщені в різних місцях. Обсяги даних визначають також швидкість опрацювання, а також ця швидкість залежить від наявності різних методів опрацювання даних, які реалізовані в системах і засобах протидії КА. Наявність різних методів опрацювання даних має таку особливість, що продуктивність частини з них знижується із збільшенням обсягів даних.

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

Введемо множину функцій $M_{F_{vp,i}^p}$ для оцінювання впливів елементів множини впливів M_{vp}^p на збереження та актуальність даних в різних аспектах їх використання та забезпечення доступу і актуальності (формула (1)) так:

$$v_{vp,i}^p = F_{vp,i}^p\left(m_{vp,i}^p\right), \tag{2}$$

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

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

$$V_{vp,t}^p = \left(t, v_{vp,1}^p, v_{vp,2}^p, \dots, v_{vp,n_{M_{vp}^p}}^p\right), \tag{3}$$

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

Також, задамо множину функцій для оцінювання впливів елементів множини впливів M_{vp}^p вектором її елементів, тобто функцій впливів $F_{vp,i}^p$ ($i = 1,2,\dots,n_{M_{vp}^p}$; $n_{M_{vp}^p}$ – кількість елементів множини впливів M_{vp}^p) аналогічно до вектору значень оцінювання впливів (формула (3)) так:

$$F_{vp,t}^p = \left(t, F_{vp,1}^p, F_{vp,2}^p, \dots, F_{vp,n_{M_{vp}^p}}^p\right), \tag{4}$$

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

Оцінювання за формулами (2) – (4) може бути здійснено за різними варіантами. Наприклад, можуть бути оцінені крайні останні значення елементів множини впливів M_{vp}^p (формула (1)) або середні значення з

усіх накопичених чи з певного проміжку часу тощо.

Таким чином, визначено множину впливів на збереження та актуальність даних в різних аспектах їх використання та забезпечення доступу і актуальності (формула (1)) та функції і вектор значень оцінювання таких впливів (формули (2) – (4)). Вони необхідні для забезпечення організації пам'яті з історичними даними при функціонуванні систем протидії КА в корпоративних мережах. Значення оцінок використовуватимуться для прийняття рішень щодо їх оптимізації, доповнення, вилучення, переміщення та опрацювання. Також, вони необхідні для безпосереднього врахування при здійсненні протидії КА наявними методами.

Деталізуємо функцію $F^p_{vp,1}$ оцінювання елементу $m^p_{vp,i}$ так:

$$v^p_{vp,1} = F^p_{vp,1}\left(m^p_{vp,1}\right) = \frac{\sum_{l=1}^{n_{vp,1}} d_{vp,1,l}}{\sum_{l=1}^{n_{z,vp1}} d_{z,vp,1,l}}, \tag{5}$$

де $F^p_{vp,1}$ – функція оцінювання елементу $m^p_{vp,1}$; $d_{z,vp,1,l}$ – обсяг l – того дискового ресурсу для зберігання даних із загальної кількості таких ресурсів; $l = 1,2,\dots,n_{z,vp,1}$; $n_{z,vp,1}$ – загальна кількість наявних окремих дискових ресурсів для зберігання даних;

$d_{vp,1,l}$ – обсяг l – того дискового ресурсу для зберігання даних із визначених таких ресурсів та задіяних для зберігання; $l = 1,2,\dots,n_{vp,1}$; $n_{vp,1}$ – кількість наявних задіяних окремих дискових ресурсів для зберігання даних.

Також, наприклад, деталізуємо функцію $F^p_{vp,6}$ оцінювання елементу $m^p_{vp,6}$, який відображає розподілення даних в мережі, так:

$$v^p_{vp,6} = F^p_{vp,6}\left(m^p_{vp,6}\right) = 1 - D^{n_{vp,6}}_{l=1}\frac{k_{vp,6,l}}{k_{z,vp.6,l}}, \tag{6}$$

де D – добуток; $l = 1,2,\dots,n_{vp,6}$; $n_{vp,6}$ – загальна кількість класів даних, які потребують збереження і відрізняються між собою представленням даних, їх структуруванням та індексуванням; $F^p_{vp,6}$ – функція оцінювання елементу $m^p_{vp,6}$; $k_{z,vp.6,l}$ – загальна кількість окремих дискових ресурсів, в яких може бути розміщений l – тий клас даних; $k_{vp,6,l}$ - фактична кількість окремих дискових ресурсів, в яких розміщено l – тий клас даних в поточний момент часу t (формула (3)); $m^p_{vp,6}$ – розподілення даних в мережі.

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

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

$$v^p_{vp,7} = F^p_{vp,7}\left(m^p_{vp,7}\right) = \frac{\sum_{l=1}^{n_{vp,6}} \frac{1}{n_{vp,7,l}}}{n_{vp,6}}, \tag{7}$$

де $n_{vp,6}$ – загальна кількість класів даних, які потребують збереження і відрізняються між собою представленням даних, їх структуруванням та індексуванням; $l = 1,2,\dots,n_{vp,6}$; $F^p_{vp,7}$ – функція оцінювання елементу $m^p_{vp,7}$; $n_{vp,7,l}$ – загальна кількість окремих дискових ресурсів, в яких розміщено l - клас даних в поточний момент часу t (формула (3)); $m^p_{vp,7}$ – розміщення даних.

Решту функцій оцінювання $F^p_{vp,i}$ для елементу $m^p_{vp,i}$ ($i = 1,2,\dots,n_{M^p_{vp}}$; $n_{M^p_{vp}}$ – кількість елементів множини впливів M^p_{vp}) можна деталізувати аналогічно.

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

$$M^D_K = \bigcup_{i=1}^{n_{M^D_K}} K^D_i, \tag{8}$$

де K^D_i – клас даних від джерел інформації різних рівнів (формули (2) та (3)); $i = 1,2,\dots,n_{M^D_K}$; $n_{M^D_K}$ – кількість класів; M^D_K – множина класів даних.

Кожен клас даних містить унікальну для нього інформацію і має відповідне для неї структурування та індексування. А також, може бути використаний винятково на певному одному чи декількох рівнях або може бути використаний для всіх рівнів, включно з тим, що частково на певних рівня або повністю на усіх

рівнях. Така вимога щодо класів даних актуалізує необхідність ефективної їх організації, збереження та актуалізації в пам'яті для засобів чи систем протидії КА. Класи даних узгоджуються з векторами для забезпечення зберігання даних від джерел інформації різних рівнів. Формування класу даних від джерел інформації різних рівнів, які задано за формулою (8), задамо введенням координати $v^D_{DI,i,18}$ у вектор $V^D_{DI,i}$ ($i = 1,2,\dots,N_{M^D_{DI}}$; $N_{M^D_{DI}}$ – кількість векторів для забезпечення зберігання даних від джерел інформації різних рівнів), в які буде відображено структуру даних, що потребують збереження. Тоді, вектор $V^D_{DI,i}$ будемо включати до класу K^D_i ($i = 1,2,\dots,n_{M^D_K}$; $n_{M^D_K}$ – кількість класів) з множини M^D_K (формула (8)), якщо сформована структура даних задана координатою $v^D_{DI,i,18}$ збіжна зі структурою даних класу. Якщо для сформованого вектору $V^D_{DI,i}$ координата $v^D_{DI,i,18}$ не буде відповідати жодному класу K^D_i, тоді такий вектор надсилаються адміністратору систему з відповідного змісту поясненням для прийняття ним рішення.

Множина функцій для оцінювання впливів елементів множини впливів M^p_{vp} на збереження та актуальність даних в різних аспектах їх використання та забезпечення доступу і актуальності (формула (1)), а також вектори, які задані за формулами (3) і (4), стосуються оцінювання в поточний момент часу t, тому введемо матрицю оцінювання з урахуванням того, що доповнювані дані до класів будуть фіксуватись разом з часом їх отримання і формування. Це дасть змогу при виборі даних для використання в системах і засобах протидії КА здійснювати вибір за певні часові проміжки або враховувати інтервали доповнення до класів. Розглянемо особливості даних і впливів в контексті різних часових термінів їх зберігання та використання. Перші вісім елементів ($m^p_{vp,1}$ - $m^p_{vp,8}$) множини впливів M^p_{vp} відносяться до ресурсних обмежень, наступні три ($m^p_{vp,9}$ - $m^p_{vp,11}$) – характеризують семантичну актуальність даних, одинадцять наступних елементів ($m^p_{vp,11}$ - $m^p_{vp,22}$) – відносяться до зберігання та оновлення сигнатур різної природи та призначення.

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

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

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

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

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

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

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

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

Системи та засоби протидії КА всіх семи рівнів піраміди болю зловмисника по різному використовують дані, які зберігаються для їх функціонування, включно з даними, що накопичуються в процесі їх тривалої експлуатації. Отримані дані від джерел інформації різних рівнів формують базу множини векторів. Також для даних різних рівнів тривалість збереження даних є різною, але в контексті застосування засобів всіх семи рівнів є потреба в збереженні всіх наявних та отримуваних даних. Крім того, частина даних могла бути використана при прийнятті рішень, що дає змогу оцінити ефективність такого застосування і, відповідно, при подальшому збереженні таких даних є потреба в доповненні їх результатами оцінювання. Такі результати можуть теж накопичуватись з часом.

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

1) початкові дані;

2) накопичувані дані;

3) використані дані.

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

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

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

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

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

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

$$M_{DI}^{D,3,7} = \bigcup_{i=1}^{3} \bigcup_{j=1}^{7} M_{DI,i,j}^{D}, \tag{9}$$

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

Дані певного типу та рівня можуть належати різним рівням одночасно.

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

1) первинне потокове захоплення даних;

2) нормалізація та уніфікація структури;

3) метадані-орієнтована класифікація;

4) контекстне збагачення;

5) адаптивна маршрутизація до рівнів 1–7;

6) агрегація та збереження історії використання.

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

Крок 2. Формування та фіксація початкових даних.

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

Крок 3. Накопичення даних.

Перезапис даних призводить до втрати історії, тому необхідно використати подійну модель. Тому, необхідно здійснювати перехід від статичного представлення даних до динамічного, подійно-орієнтованого накопичення інформації, яке відображає реальний стан КМ у часі. У системах та засобах протидії КА кожна подія (логін, мережний запит, аномалія, спрацювання правила) є не тільки значенням, а фактом зміни стану системи, який має бути збережений разом із контекстом виникнення. Таким чином, дані не перезаписуються, а накопичуються як послідовність подій, що утворюють історію поведінки системи. Тоді вирішується проблема того, що нові значення даних замінюють попередні, відомості про попередній стан втрачається, неможливо відтворити послідовність КА, аналіз інцидентів стає неповним. Це є критичним, оскільки КА часто є багатокроковими та розподіленими у часі. Подійна модель передбачає, що кожна дія фіксується як окрема подія, події зберігаються у хронологічному порядку та стан системи відновлюється через аналіз послідовності подій. Задамо підмножини накопичуваних даних $M_{DI,2,j}^D$ $(j = 1,2,\dots,7)$ так:

$$M_{DI,2,j}^D(t_n) = M_{DI,2,j}^D(t_{n-1}) \cup \{p_n\}, \tag{10}$$

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

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

Крок 4. Формування використаних даних.

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

$$m_{DI,3,j,k}^D = M_{F,1,2}^D\left(M_{DI,1,j}^D, M_{DI,2,j}^D\right), \tag{11}$$

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

Дані можуть бути такими, що були використані, та такими, що підготовлені до використання, тобто використовувані. Множина функцій $M_{F,1,2}^D$ містить функції, які поділяють множину використаних/використовуваних даних на такі підмножини:

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

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

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

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

Формування таких даних забезпечується використанням різних методів обробки:

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

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

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

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

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

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

Крок 5. Формалізація структури похідних даних.

Результати аналізу, отримані на попередньому етапі, перестають розглядатися як одноразові висновки і перетворюються на повноцінні інформаційні об'єкти, придатні для довготривалого зберігання, повторного використання та накопичення досвіду. На цьому етапі факт оформлюється таким чином, щоб зберегти не тільки сам результат, а й усі обставини його отримання та подальшого використання. Суть кроку полягає у фіксації чотирьох ключових аспектів кожного результату: набір подій, логів або інших елементів, що стали підставою для висновку; момент часу, коли цей результат було отримано, що дозволяє оцінювати його актуальність; зміст результату (тип інциденту, рівень ризику, класифікація КА, рекомендації); історія використання цього результату, яка відображає, як саме він застосовувався далі, тобто переданвання його аналітику, підтвердження, використання для реагування, закриття тощо. Таким чином, результат аналізу отримує "життєвий цикл" і перестає бути статичною величиною. У практиці SIEM це відповідає перетворенню окремих подій на інцидент, який має не лише опис, а й зв'язок із джерельними подіями, часову прив'язку та журнал дій. Наприклад, система може на основі кількох підозрілих логінів і доступу до ресурсу сформувати інцидент компрометації облікового запису. Якщо цей інцидент зберегти лише як текстове повідомлення, то подальший аналіз буде обмеженим. Якщо ж він зберігається разом із переліком використаних подій, часом створення, рівнем ризику та історією обробки (ескалація, підтвердження, блокування, закриття), то він стає повноцінним об'єктом управління безпекою.

Крок 6. Фіксація одноразового використання даних.

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

Крок 7. Накопичення історії багаторазового використання даних.

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

Крок 8. Оцінювання часової глибини даних.

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

Крок 9. Шарове розділення даних.

Зберігання даних потрібно здійснювати за принципом шарового розділення, тобто розподіляти всі

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

Крок 10. Забезпечення простежуваності даних.

Забезпечення простежуваності даних означає можливості встановлення походження кожного сформованого результату та відновлення процесу його отримання. Необхідність цього зумовлена тим, що в системах протидії КА результати аналізу (інциденти, оцінки, рішення) не можуть розглядатися ізольовано, бо для їх перевірки, довіри та повторного використання важливо чітко знати, на основі яких саме даних і подій вони були сформовані. Для кожного результату визначається функція походження, яка відображає, з яких саме даних він був отриманий. Таким чином, будь-який результат можна простежити до первинних подій або початкових даних, що забезпечує повну прозорість процесу аналізу.

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

Крок 11. Локальне накопичення даних.

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

Крок 12. Синхронізація з центральним сховищем.

Всі дані, які були сформовані та збережені в автономному режимі у окремих вузлах або рівнях системи, повинні бути передані до центрального середовища та інтегровані з уже наявною інформацією без втрати цілісності та історії. В умовах тимчасової ізоляції кожен компонент системи накопичує власний фрагмент даних, який відображає лише локальну частину подій. Після відновлення зв'язку ці фрагменти мають бути об'єднані, щоб відновити повну картину функціонування системи, забезпечити коректний аналіз і уникнути інформаційних розривів. Синхронізація повинна забезпечити узгоджене об'єднання даних із урахуванням часових міток, унікальності подій і вже наявних записів. Нові локальні дані повинні додаватись до центрального сховища таким чином, щоб не виникало дублювання, зберігалася хронологія подій, не втрачалися зв'язки між даними та їх походженням, та враховувалася історія використання. Для цього використаємо механізми синхронізації: буферизація локальних даних до моменту передачі; використання унікальних ідентифікаторів подій; узгодження за часовими мітками; контроль цілісності та повноти переданих даних.

Крок 13. Адаптивний розподіл даних за типами сховищ

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

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

Необхідно здійснити перевірку узгодженості між властивостями даних, які зафіксовані у вигляді їх метаданих, та характеристиками апаратних засобів, на яких ці дані розміщуються. На попередніх етапах для кожного елемента даних вже сформовано набір характеристик. Водночас апаратні засоби також мають формалізовані характеристики, наприклад, продуктивність, надійність, вік, інтенсивність зношення та доступні ресурси. Без зіставлення цих двох груп параметрів неможливо забезпечити ефективне і безпечне

розміщення даних. Кожен елемент даних задамо вектором метаданих $V_{DI,1}^{D}$, а кожен носій даних - вектором $V_{DI,2}^{D}$ його характеристик.

Крок 15. Формування попереднього досвіду використаних даних.

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

Для формування попереднього досвіду використаних даних використаємо різні методи та технології, щоб на їх основі забезпечити комбінований підхід до формування досвіду використання історичних даних. Використаємо методи і технології, які базуються на таких моделях:

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

У випадку використання марківської моделі кожен стан системи (наприклад, “виявлення події”, “підтвердження інциденту”, “реагування”, “закриття”) розглядаємо як окремий стан марковського процесу, а історія використання даних дозволить оцінити ймовірності переходів між цими станами. Згідно марківської моделі випадковий процес не має пам’яті, тобто збереження тривалої історії для даних не забезпечується. При цьому, наявна можливість збереження ймовірності переходів між станами, яка може бути використана для формування досвіду використання даних на попередньому кроці їх задіювання в системах та засобах протидії КА. В контексті розглядуваної задачі із збереженням даних та досвіду попереднього їх використання наявна їх відповідність такому виду марківського процесу як дискретний процес з дискретним часом. Випадковий процес X_t вважається марковським процесом, якщо для будь-яких моментів часу з відрізку [0;T] функція розподілу останнього значення X_{t_n} при фіксованих значеннях $X_{t_0}, X_{t_1}, X_{t_2}, \dots, X_{t_{n-1}}$ залежить тільки від $X_{t_{n-1}}$. В марківському випадковому процесі майбутній розвиток подій залежить тільки від поточного стану і не залежить від того, коли і яким чином система перейшла в цей стан. Тому зберігання даних про певні процеси будемо базувати на формуванні матриць для кожного блоку даних, якими описуватимемо для кожного з них ймовірності переходу з одного стану в інший за відомий проміжок часу. Множину всіх можливих послідовностей переходів відображає дерево логічних переходів з ймовірностями переходу з одного стану в інший. По дереву логічних переходів можна визначити ймовірність переходу в той чи інший стан після декількох кроків. Нехай маємо множину станів системи $S^{D,P} = \left\{S_1^{D,P}, S_2^{D,P}, \dots, S_{n^{S^{D,P}}}^{D,P}\right\}$ ($n^{S^{D,P}}$ – кількість станів системи) та множини ймовірності станів для кожного стану окремо $P_i^{D,P} = \left\{P_{i,1}^{D,P}, P_{i,2}^{D,P}, \dots, P_{i,n^{S^{D,P}}}^{D,P}\right\}$ ($j = 1,2,\dots,n^{S^{D,P}}$; $i = 1,2,\dots,n^{S^{D,P}}$) після будь-якого кроку.

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

Баєсівська модель для збереження попереднього досвіду використання даних дає змогу не тільки зберігати окремі події або результати аналізу, а формалізувати залежності між ними в часі. Тоді система накопичує не лише дані, а й знання про те, як змінювалися стани, які події передували інцидентам та які рішення були прийняті і до яких наслідків це призводило. Використання динамічної байєсівської мережі полягає у представленні системи як послідовності часових зрізів. Кожен часовий зріз містить набір вузлів, що описують стан системи в певний момент часу. До таких вузлів введемо такі параметри:

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

9) стан апаратного ресурсу.

Між вузлами сформуємо два типи зв'язків:

1) внутрішньошарові, якими задамо залежності між параметрами в межах одного моменту часу;

2) міжчасові, якими задамо вплив попереднього стану системи на наступний.

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

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

1) час використання;

2) результат;

3) тип даних;

4) рівень системи;

5) ефективність рішення на основі набору спостережень для навчання мережі.

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

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

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

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

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

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

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

Використання лише одного підходу для збереження та оцінювання історичного досвіду функціонування систем і засобів протидії КА є недостатнім через різнорідність даних, різні часові горизонти їх актуальності та різний рівень визначеності процесів у КМ. Тому було доцільно поєднано марковські моделі, динамічні баєсівські мережі та статистичні методи, які можуть застосовуватись як до однакових, так і до різних типів даних та сценаріїв. Марковські моделі доцільні для опису послідовності станів системи та

переходів між ними в процесі розвитку КА або реагування на неї. Їх особливістю є можливість накопичення досвіду щодо типових сценаріїв переходів між станами та оцінювання ймовірностей подальшого розвитку подій на основі попередніх випадків використання даних. Динамічні баєсівські мережі доцільно використовувати для врахування причинно-наслідкових зв'язків між подіями, ознаками КА, станами системи та результатами прийняття рішень у часі. Їх особливістю є можливість працювати з неповними, нечіткими або ймовірнісними даними, а також враховувати зміну взаємозалежностей між параметрами в процесі накопичення нових відомостей. Це дозволяє адаптувати оцінювання ефективності використання даних відповідно до зміни умов функціонування КМ. Статистичні методи забезпечують узагальнення накопичених результатів використання даних, визначення тенденцій, частотних характеристик, аномалій та оцінювання ефективності прийнятих рішень. Їх особливістю є низька обчислювальна складність та можливість швидкого опрацювання великих обсягів історичних даних, що важливо для забезпечення автономності функціонування систем і засобів протидії КА.

Отже, спільне використання цих підходів забезпечує комплексне представлення історичного досвіду. Марковські моделі описують динаміку станів, Динамічні баєсівські мережі задають причинно-наслідкові залежності в умовах невизначеності, а статистичні методи - кількісне узагальнення та оцінювання накопичених результатів. При цьому формується попередній досвід використання даних, який стає основою для адаптивного функціонування систем та засобів протидії КА, що дозволяє перейти від звичайного накопичення інформації до накопичення знань і досвіду та це підвищує ефективність аналізу, прогнозування і реагування в КА.

Експерименти

Для підтвердження ефективності запропонованого методу організації зберігання даних у системах та засобах протидії КА було проведено експериментальне дослідження, метою якого стало визначення впливу способу організації та розміщення історичних даних на результативність виявлення КА у КМ. В основу дослідження покладено припущення, що накопичення, структурування та повторне використання історичних даних про попереднє функціонування засобів захисту дає змогу підвищити точність і швидкість виявлення КА завдяки використанню попереднього досвіду функціонування системи.

Експеримент проводився на моделі КМ, до складу якої входили сервери, робочі станції користувачів, мережеве обладнання, система моніторингу подій безпеки та база даних для накопичення історичної інформації. Для проведення дослідження було реалізовано два варіанти функціонування системи. У першому випадку використовувалося традиційне зберігання даних, за якого журнали подій та сигнатури атак накопичувалися без урахування історії їх використання під час прийняття рішень. У другому випадку застосовувався запропонований метод, який передбачає поділ даних на початкові, накопичувані та використані, а також збереження відомостей щодо результативності їх попереднього застосування. Для аналізу історичних даних використовувалися марковські моделі, динамічні баєсівські мережі та статистичні методи оцінювання ефективності.

У процесі експерименту було змодельовано понад 1000 сценаріїв КА різних типів. Для кожного сценарію визначалися показники повноти виявлення КА, точності класифікації подій, F1-міри, кількості хибних спрацювань та середнього часу виявлення загрози. Результати експерименту наведено в таблиці 1.

Таблиця 1

Вплив організації історичних даних на ефективність виявлення КА

Варіант організації даних				**Хибні спрацювання, %**	**Час виявлення, с**
Традиційне зберігання даних					
Історичне зберігання даних					
Історичне зберігання даних з оцінкою ефективності використання					

Аналіз отриманих результатів показав, що удосконалення організації даних та врахування історії їх використання позитивно впливає на якість виявлення КА. Зокрема, повнота виявлення атак зросла на 11,5 %, а інтегральний показник F1-міри підвищився на 10,2 % порівняно з традиційним підходом. Одночасно кількість хибних спрацювань зменшилася більш ніж удвічі, що свідчить про покращення якості прийняття рішень на основі накопиченого історичного досвіду.

Крім того, спостерігається скорочення середнього часу виявлення атак з 14,2 до 9,4 с, що пояснюється можливістю повторного використання накопичених знань про попередні сценарії функціонування системи та типові ознаки атак. Це свідчить про підвищення адаптивності системи до нових кіберзагроз та більш ефективне використання наявних даних під час аналізу подій безпеки.

Отримані результати підтверджують, що запропонований метод організації зберігання даних забезпечує не лише накопичення історичних відомостей, а й підвищує ефективність їх використання під час виявлення комп'ютерних атак. Покращення структури розміщення даних, врахування історії їх застосування

та оцінювання результативності попередніх рішень дозволило підвищити F1-міру з 83,0 % до 93,2 %, зменшити кількість хибних спрацювань на 56,3 % та скоротити середній час виявлення атак на 33,8 %. Це підтверджує наявність прямого взаємозв'язку між організацією історичних даних і результативністю функціонування засобів протидії КА у КМ.

ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ

Розроблено моделі зберігання даних та метод організації зберігання даних в системах та засобах протидії комп'ютерним атакам з урахуванням історичного аспекту їх використання. Ефективне функціонування систем та засобів протидії КА потребує організації зберігання даних з урахуванням не лише їх походження та структури, а й історії використання в процесах прийняття рішень. Метод передбачає поділ даних на початкові, накопичувані та використані, що дає змогу враховувати етапи їх формування, повторного застосування та результати оцінювання ефективності такого застосування. Особливістю методу є збереження та опрацювання історичного досвіду використання даних із застосуванням марковських моделей, динамічних баєсівських мереж і статистичних методів. При цьому марковські моделі забезпечують опис послідовності станів та сценаріїв розвитку подій, динамічні баєсівські мережі - врахування причинно-наслідкових зв'язків та невизначеності даних у часі, а статистичні методи - узагальнення результатів використання даних та оцінювання ефективності прийнятих рішень. Це забезпечує накопичення історичних відомостей щодо використання даних, підвищення ефективності їх повторного застосування, адаптивності систем та автономності їх функціонування навіть в умовах недоступності спеціалізованих засобів зберігання даних.

Напрямами подальших досліджень є розроблення архітектури засобів зберігання, оптимізації та вибору даних для систем та засобів протидії КА в КМ, зокрема розроблення методів оптимізації та вибору даних з урахуванням також і засобів, що використовують обманні технології.

References

1. 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

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., 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

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

6. Koukaras P. Data Integration and Storage Strategies in Heterogeneous Analytical Systems: Architectures, Methods, and Interoperability Challenges. *Information*. 2025; 16(11):932. https://doi.org/10.3390/info16110932

7. Kandpal M., Goswami V., Priyadarshini R., Barik RK. Towards Data Storage, Scalability, and Availability in Blockchain Systems: A Bibliometric Analysis. *Data*. 2023; 8(10):148. https://doi.org/10.3390/data8100148

8. Oliveira NRd., Santos YdRd., Mendes ACR., Barbosa GNN., Oliveira MTd., Valle R., Medeiros DSV., Mattos DMF. Storage Standards and Solutions, Data Storage, Sharing, and Structuring in Digital Health: A Brazilian Case Study. *Information*. 2024; 15(1):20. https://doi.org/10.3390/info15010020

9. Bin Saif M., Migliorini S., Spoto F. Efficient and Secure Distributed Data Storage and Retrieval Using Interplanetary File System and Blockchain. *Future Internet*. 2024; 16(3):98. https://doi.org/10.3390/fi16030098

10. Anwar RW, Qureshi KN., Nagmeldin W., Abdelmaboud A., Ghafoor KZ., Javed IT., Crespi N. Data Analytics, Self-Organization, and Security Provisioning for Smart Monitoring Systems. *Sensors*. 2022; 22(19):7201. https://doi.org/10.3390/s22197201

11. Lu J., Bai J., Shen S. Design and Implementation of a High-Speed Storage System Based on SATA Interface. *Electronics*. 2026; 15(2):452. https://doi.org/10.3390/electronics15020452

12. Wang Y., Fan R., Liang X., Li P., Hei X. Trusted Data Storage Architecture for National Infrastructure. *Sensors*. 2022; 22(6):2318. https://doi.org/10.3390/s22062318

13. Talluri LSRK., Thirumalaisamy R., Kota R., Sadi RPR., KCU., Naha RK., Mahanti A. Providing Consistent State to Distributed Storage System. *Computers*. 2021; 10(2):23. https://doi.org/10.3390/computers10020023

14. Khan AQ., Nikolov N., Matskin M., Prodan R., Roman D., Sahin B., Bussler C., Soylu A. Smart Data Placement Using Storage-as-a-Service Model for Big Data Pipelines. *Sensors*. 2023; 23(2):564. https://doi.org/10.3390/s23020564

15. Alsulbi KA., Khemakhem MA., Basuhail AA., Eassa FE., Jambi KM., Almarhabi KA. A Proposed Framework for Secure Data Storage in a Big Data Environment Based on Blockchain and Mobile Agent. *Symmetry*. 2021; 13(11):1990. https://doi.org/10.3390/sym13111990