УДК 004.056.5:004.89:519.85
DOI: 10.31673/2409-7292.2025.032076

Шульга В.П., Іванченко І.С., Рижаков М.М.

МАТЕМАТИЧНА МОДЕЛЬ СЕМАНТИЧНОЇ АТРИБУЦІЇ КІБЕРІНЦИДЕНТІВ У СИСТЕМАХ ВИЯВЛЕННЯ АНОМАЛІЙ НА ОСНОВІ ГЛИБОКОГО НАВЧАННЯ

Вперше запропоновано автоматизовану математичну модель семантичної атрибуції кіберінцидентів, інтегровану в глибинну систему виявлення аномалій для переходу від експертної сигналізації відхилень до автоматичної контекстної ідентифікації загроз. Запропонований підхід поєднує поведінковий аналіз трафіку (реконструкція нормальної поведінки автоенкодером та короткострокове прогнозування MLP з адаптивним формуванням меж аномальності) з контентно-семантичним шаром, у якому виконується глибинний розбір payload і заголовкових полів, побудова векторних подань та пошук прецедентів у базі історичних інцидентів (класи/сімейства SQLi, XSS, XXE/XSLT, brute-force тощо). Перед модулем оцінки критичності вводиться агрегований ризик-скоринг, що комбінує силу аномалії, впевненість атрибуції та контекст активів, а також механізми XAI-пояснюваності (важливі токени/поля, найближчі кейси) для підвищення інтерпретованості рішень та режиму human-on-the-loop. Реалізація орієнтована на потокові сценарії та сумісна з SIEM/SOAR, що спрощує впровадження у критичній інфраструктурі, телекомерційних мережах, фінансовому секторі та хмарних середовищах. Експериментальні дослідження на власних наборах мережевих даних демонструють статистично значуще скорочення хибнопозитивних спрацювань та зростання інтегральних метрик (Precision/Recall/F1) відносно rule-based і суто поведінкових підходів, а також зниження часу реагування SOC. Отримані результати підтверджують, що інтеграція семантичної атрибуції з поведінковим детектуванням формалізує відображення «аномалія → кіберінцидент», підвищує відтворюваність і керованість процесу та створює основу для масштабованих систем кіберзахисту нового покоління.

Ключові слова: кіберінциденти, виявлення аномалій, семантична атрибуція, машинне навчання, автоенкодер, модуль критичності, прецеденти кібер атак, кібербезпека, кібератака, кіберзахист, критична інфраструктура, кіберзагроза.

Вступ

У сучасному цифровому середовищі інформаційні системи постійно зазнають зростаючого навантаження та впливу різнотипних кіберзагроз. Традиційні rule-based або сигнатурні методи виявлення кібератак демонструють низьку ефективність у динамічних умовах, оскільки вони базуються на фіксованих правилах та не враховують контекст подій. Особливої актуальності набувають кібератаки нового покоління – багатовекторні, приховані або нульового дня, які не фіксуються класичними системами моніторингу які використовують експертні методи. У цих умовах особливого значення набуває використання інтелектуальних підходів, здатних не лише виявляти аномальні відхилення у поведінці системи, але й здійснювати їхню атрибуцію як повноцінних кіберінцидентів. Останні дослідження підтверджують ефективність використання глибинних нейромереж у задачах виявлення аномалій у мережевому трафіку: автоенкодери добре реконструюють нормальну поведінку та «підсвічують» приховані відхилення, тоді як багатошарові перцептрони (MLP) забезпечують короткострокове прогнозування та формують адаптивні межі аномальності. Водночас більшість існуючих систем зупиняються на рівні поведінкового моніторингу та все ще покладаються на експертну інтерпретацію результатів: вміст мережевих пакетів аналізується обмежено, механізм зіставлення нових аномалій із відомими інцидентами відсутній або ручний, а атрибуція типу загрози виконується також вручну. На відміну від експертних підходів, у цьому дослідженні запропоновано автоматизований метод, який трансформує систему від «експертної сигналізації аномалій» до «автоматизованої контекстної ідентифікації кіберінцидентів», мінімізуючи вплив людського чинника, знижуючи частоту хибнопозитивних спрацьовувань (FP) і підвищуючи репродукованість та оперативність реагування.

Така ситуація створює низку обмежень:

- недостатня пояснюваність виявлених подій (система сигналізує про відхилення, але не дає інтерпретації типу загрози);

- висока кількість хибнопозитивних спрацювань, зумовлена чутливістю лише до змін навантаження;
- неможливість оперативної атрибуції загрози, особливо у випадках складних ін'єкцій (SQLi, XSS, XXE/XSLT) чи нульових днів;
- відсутність використання прецедентів для порівняння нових інцидентів з історично зафіксованими та оптимізації реагування.

У відповідь на ці виклики вперше розроблено математичну модель семантичної атрибуції кіберінцидентів, яка за рахунок введення спеціального шару контентного аналізу та пошуку прецедентів перед модулем оцінки критичності, що дозволить перевести систему від рівня «сигналізації про аномалію» до рівня «контекстної ідентифікації кіберінциденту» та суттєво підвищити ефективність і пояснюваність процесу.

Запропонований підхід забезпечує:

- багаторівневий аналіз трафіку (поведінковий та контентний);
- інтеграцію механізму пошуку прецедентів;
- формування обґрунтованої оцінки критичності з урахуванням типу кібератаки;
- зменшення кількості хибнопозитивних спрацювань;
- підвищення рівня кіберстійкості інформаційних систем.

Проблематика виявлення та атрибуції кіберзагроз посідає центральне місце у сучасних дослідженнях у сфері кібербезпеки. Зростання обсягів даних, глобалізація цифрової інфраструктури та перехід до моделі «цифрової економіки» призвели до різкого збільшення кількості й складності кібератак. Це, у свою чергу, актуалізувало потребу у створенні інтелектуальних систем, здатних не лише оперативно ідентифікувати відхилення у поведінці системи, але й надавати семантичну інтерпретацію подій у вигляді кіберінцидентів. Класичні системи виявлення кібератак, зокрема Snort та Suricata, базуються на сигнатурному аналізі та системах правил. Вони демонструють високу ефективність для виявлення відомих загроз, однак мають низку фундаментальних обмежень. Такі системи залежать від актуальності бази сигнатур, яка потребує постійного оновлення. Вони не здатні адекватно обробляти нові, невідомі вектори кібератак. При складних або багатовекторних загрозах (наприклад, комбінованих кібератаках, що поєднують DDoS і експлуатацію вразливостей) точність сигнатурних систем стрімко падає [1], [2]. Сигнатурні підходи генерують значну кількість false positive (хибнопозитивних спрацювань), що призводить до перевантаження операторів SOC (Security Operation Center).

У відповідь на обмеження класичних IDS/IPS протягом останніх років активно розвивається напрямок застосування машинного навчання. Ці методи дозволяють автоматично виявляти приховані закономірності у даних і формувати профіль «нормальної поведінки» системи. Умовно підходи можна поділити на кілька категорій:

– Методи без учителя: алгоритми кластеризації (k-means, DBSCAN), метод головних компонент (PCA), автоенкодери. Використовуються для виявлення відхилень у даних без наявності навчальної вибірки з мітками.

– Методи з учителем: SVM, Random Forest, градієнтний бустинг. Вимагають розмічених даних, але дозволяють будувати точні класифікатори.

– Глибинне навчання: convolutional neural networks (CNN), recurrent neural networks (RNN, LSTM, GRU), багатошарові перцептрони (MLP). Ці підходи добре працюють із багатовимірними часовими рядами та потоковими даними [3].

Зокрема, автоенкодери зарекомендували себе як один із найефективніших інструментів у задачах відновлення нормальної поведінки системи та виявлення прихованих відхилень [4]. Вони реконструюють нормальні патерни, а відхилення з високою похибкою трактуються як аномалії. Водночас рекурентні архітектури (LSTM, GRU) дозволяють враховувати часову залежність у трафіку та виявляти аномалії, що проявляються поступово [5]. Важливою тенденцією є поєднання різних підходів. Комбіновані моделі (наприклад, автоенкодер + MLP)

© Шульга В.П., Іванченко І.С., Рижаков М.М. (2025) Математична модель семантичної атрибуції кіберінцидентів у системах виявлення аномалій на основі глибокого навчання. Сучасний захист інформації, 3(63), 186–198. https://doi.org/10.31673/2409-7292.2025.032076

дозволяють поєднати сильні сторони: реконструктивний аналіз і прогностичні можливості. Дослідження показують, що такі гібридні моделі значно знижують рівень хибних спрацювань та підвищують точність [6], [7].

Попри значні успіхи, існує критична проблема – відсутність пояснюваності у більшості моделей. Sommer і Paxson ще у 2010 році звернули увагу на те, що система, яка лише сигналізує про аномалію, але не пояснює її природу, є малокорисною для оператора [8]. У сучасних умовах SOC-оператори стикаються з десятками чи сотнями сигналів на добу, і без адекватної атрибуції подій стає неможливим їх пріоритизувати. Тому сьогодні особливої уваги набуває напрямок XAI (explainable AI) у кібербезпеці, що орієнтується на створення моделей, здатних пояснити свої рішення. Li та Huang у своїх роботах показали, що введення багатовимірних ознак та інтерпретаційних модулів у нейронні мережі дозволяє знизити кількість false positive і підвищити довіру операторів до результатів [9], [10].

У сфері стандартизації визначальними залишаються рекомендації ISO/IEC 27001:2022 [11], які встановлюють вимоги до управління інформаційною безпекою, включаючи моніторинг та реагування на інциденти. Настанови NIST щодо систем виявлення та запобігання вторгненням (IDPS) [12] конкретизують технічні аспекти побудови систем: вимоги до масштабованості, інтегрованості та здатності до обробки багатовекторних загроз. Водночас у цих стандартах практично відсутні конкретні рекомендації щодо застосування глибинного навчання чи семантичної атрибуції. Це залишає простір для наукових досліджень у напрямку інтеграції AI-підходів із нормативними вимогами.

Попри значний прогрес, більшість існуючих систем зосереджуються на аналізі лише навантаження та поведінкових характеристик трафіку. Такий підхід дозволяє ефективно виявляти масові кібератаки (наприклад, DDoS), проте практично не дає результатів при складних цільових кібератаках, що реалізуються через вміст пакетів (SQL-injection, XSS, XXE/XSLT, zero-day). Недостатньо дослідженим залишається напрямок семантичної атрибуції аномалій, що передбачає:

- глибокий аналіз контенту мережевих пакетів;
- класифікацію кібератак за типами та сімействами;
- використання механізму пошуку прецедентів, що дозволяє зіставляти нові інциденти з історичними даними.

Введення такого рівня у систему дозволяє перетворити її з інструменту сигналізації у повноцінний механізм ідентифікації та пояснення кіберінцидентів.

Важливо зазначити, що попередні роботи авторів заклали основу для цієї статті. У статті [13] було запропоновано методи інтелектуального виявлення аномалій та критичних ситуацій у кіберсистемах на основі автоенкодера та багатошарової нейромережі. Подальший розвиток цих ідей відбувся у роботі [14], де була сформована узагальнена модель прогнозування та виявлення аномалій у кіберінфраструктурі та розроблена математична модель інтегрованого прогнозування й детекції на основі глибинного навчання. Запропонована у даній роботі модель є логічним наступним кроком. Вона інтегрує семантичну атрибуцію кіберінцидентів у вже існуючу архітектуру системи виявлення аномалій, забезпечуючи більш глибокий рівень аналізу та значно вищу пояснюваність результатів. Це дозволяє перейти від «виявлення відхилення» до «контекстної ідентифікації загрози».

Сучасні системи виявлення аномалій у кіберінфраструктурі, попри значний прогрес у застосуванні методів глибокого навчання, усе ще залишаються обмеженими у своїй функціональності. Переважна більшість підходів базується на аналізі поведінкових характеристик мережевого трафіку, таких як інтенсивність потоків, часові інтервали між пакетами, статистичні розподіли запитів чи відгуків. Така орієнтація дозволяє ефективно виявляти масові кібератаки на кшталт DDoS, які проявляються у різкому зростанні навантаження чи аномальних закономірностях у часових рядах. Проте, коли йдеться про більш витончені загрози, які приховуються у структурі або вмісті самих пакетів, подібні підходи

© Шульга В.П., Іванченко І.С., Рижаков М.М. (2025) Математична модель семантичної атрибуції кіберінцидентів у системах виявлення аномалій на основі глибокого навчання. Сучасний захист інформації, 3(63), 186–198. https://doi.org/10.31673/2409-7292.2025.032076

демонструють обмежену ефективність. Ін'єкції SQL, кібератаки XSS, маніпуляції із зовнішніми сутностями (XXE/XSLT) чи експлойти нульового дня практично не залишають виразних слідів у поведінкових характеристиках. Вони маскуються під звичайний трафік і лише аналіз семантики переданих даних дозволяє відрізнити їх від легітимної активності. Наукова проблема, яка постає перед дослідниками, полягає у необхідності створення нового рівня інтелектуальних систем виявлення аномалій, що виходить за межі суто поведінкового аналізу і здатний виконувати семантичну атрибуцію кіберінцидентів. Під семантичною атрибуцією у даному випадку розуміється процес ідентифікації змістовних характеристик кібератаки, віднесення інциденту до певного класу чи сімейства загроз, а також пошук подібних прецедентів у накопиченій базі знань. У практичному плані це означає, що система не лише сигналізує про аномалію, але й надає контекстне пояснення: до якого типу вона належить, яку вразливість експлуатує і якими можуть бути потенційні наслідки.

У межах даного дослідження передбачається розробка математичної моделі, яка дозволить формалізувати процес семантичної атрибуції кіберінцидентів у системах виявлення аномалій. Така модель має враховувати подвійний характер аналізу: з одного боку – поведінкові характеристики трафіку, які добре описуються часовими рядами та статистичними моделями, а з іншого боку – контентні ознаки, які відображають внутрішню структуру та зміст мережевих пакетів. Важливим аспектом є й інтеграція механізму пошуку прецедентів, який дозволяє співставляти нові інциденти з уже відомими та уточнювати класифікацію на основі накопиченого досвіду.

Метою роботи є підвищення ефективності ідентифікації кіберінцидентів шляхом розроблення та інтеграції вперше запропонованої автоматизованої математичної моделі семантичної атрибуції у глибинну систему виявлення аномалій. Очікується, що інтеграція семантичного рівня у загальну архітектуру значно підвищить точність ідентифікації кібератак, скоротить кількість хибнопозитивних спрацювань та забезпечить пояснюваність рішень для операторів центрів моніторингу безпеки (SOC).

Для формалізації моделі введемо наступні позначення (табл. 1).

Таблиця 1

Умовні позначення та параметри моделі

Позначення	Опис
S_t	множина мережевих подій у часовому вікні t
$v_i^{(h)} \in R^{d_h}$	вектор технічних (header-based) ознак для і-го пакета
$v_i^{(d)} \in R^{d_d}$	семантичний вектор даних (payload-based features)
d_h	розмірність простору технічних ознак
d_d	розмірність простору семантичних ознак
w	розмір вікна аналізу (кількість пакетів)
K	кількість відомих класів кібератак / аномалій
$\widehat{I_t}$	інтегрований вектор ознак інциденту у вікні t
α, β	вагові коефіцієнти для балансування технічних та семантичних ознак
$w_{k,m}$	параметри глибинної моделі (вагові коефіцієнти між шарами k та m)

Формалізація вхідних даних

Будь-яка система виявлення аномалій у сфері кіберзахисту функціонує на основі аналізу мережевого трафіку, який являє собою потік даних, структурований у вигляді пакетів. Кожен пакет містить набір технічних параметрів (заголовкові ознаки) та інформаційне навантаження (payload). Для побудови математичної моделі необхідно здійснити формалізацію цього процесу, визначивши, у яких просторах існують дані, які з них несуть поведінкову, а які – семантичну інформацію.

Позначимо множину всіх пакетів за певний проміжок часу T як:

© Шульга В.П., Іванченко І.С., Рижаков М.М. (2025) Математична модель семантичної атрибуції кіберінцидентів у системах виявлення аномалій на основі глибокого навчання. Сучасний захист інформації, 3(63), 186–198. https://doi.org/10.31673/2409-7292.2025.032076

$$\mathcal{P}(T) = \{p_1, p_2, \dots, p_n\}, \quad p_i \in P.$$

Кожен пакет p_i можна розглядати як вектор, що складається з двох компонентів:

$$p_i = (h_i,\ d_i),$$

де $h_i \in H$ – вектор заголовкових характеристик (IP-адреси джерела та призначення, порти, протокол, розмір пакета, часові мітки тощо); $d_i \in D$ – дані корисного навантаження (payload), які містять безпосередньо інформаційний контент.

Таким чином, простір даних розбивається на поведінковий компонент (множина H) та контентний (множина D).

Задача виявлення аномалій традиційно формулюється у просторі H: на основі статистичних характеристик заголовків визначається відхилення від норми. Проте для виконання семантичної атрибуції необхідно ввести відображення також і для D, тобто працювати з внутрішнім змістом даних, які передаються.

Для цього введемо функцію відображення:

$$\phi: D \to E^m,$$

де E^m – простір векторних представлень (ембеддингів), у якому зберігається інформація про структурні та семантичні характеристики корисного навантаження.

Отже, кожен пакет у формалізованому вигляді описується як:

$$p_i = \big(h_i,\ \phi(d_i)\big),$$

що забезпечує подвійний опис: з боку поведінки (заголовкові ознаки) та з боку змісту (семантика даних).

Далі введемо множину інцидентів $\mathcal{I}$, які можуть бути зафіксовані системою:

$$\mathcal{I} = \{I_1, I_2, \dots, I_k\}, \quad I_j \in C,$$

де C – множина класів кіберінцидентів (наприклад, $\{DDoS, SQLi, XSS, XXE, brute-force, 0-day\}$).

Таким чином, завдання полягає у побудові відображення:

$$f: \mathcal{P}(T) \to \mathcal{I},$$

яке для множини спостережуваних пакетів за певний час визначає ймовірний тип кіберінциденту або констатує його відсутність.

У такій формалізації створюємо основу для наступних етапів – виділення ознак, побудови нейромережевої моделі та визначення функції атрибуції.

Виділення та трансформація ознак

Після формалізації мережевого трафіку як множини пакетів, що мають поведінкові та семантичні складові, необхідно визначити спосіб переходу від «сирих» даних до простору інформативних ознак.

Поведінкові ознаки

Заголовкові характеристики пакетів $h_i \in H$ містять числові та категоріальні параметри (IP-адреси, порти, протокол, часові інтервали). Для коректної обробки їх потрібно перетворити у числову форму:

- категоріальні ознаки (тип протоколу, порт) кодуються методом one-hot або embedding,
- числові ознаки (розмір пакета, інтервал між пакетами) нормалізуються до інтервалу [0,1] або стандартизуються за формулою

$$x' = \frac{x - \mu}{\sigma},$$

де μ – середнє значення, σ – стандартне відхилення.

У результаті отримаємо вектор поведінкових ознак:

$$v_i^{(h)} = \psi(h_i) \in R^{d_h},$$

де d_h – розмірність простору поведінкових ознак.

Контентні (семантичні) ознаки

Корисне навантаження пакета $d_i \in D$ є текстовим або бінарним фрагментом, який може містити як коректні дані, так і шкідливий код. Для переходу до простору ознак застосовується функція відображення $\phi(d_i)$, яка може реалізовуватися різними способами:

- N-грами та частотні вектори, що відображають локальні закономірності у даних,
- Word2Vec / FastText для текстових даних, що дозволяють вловлювати семантичні подібності,
- Byte-level embedding для випадків, коли дані неструктуровані або шифровані.

Формується вектор:

$$v_i^{(d)} = \phi(d_i) \in R^{d_d},$$

де d_d – розмірність простору семантичних ознак.

Об'єднаний простір ознак

Оскільки для атрибуції необхідно враховувати як поведінкові, так і семантичні ознаки, виконується їх об'єднання:

$$v_i = \left[v_i^{(h)} \,||\, v_i^{(d)}\right] \in R^{d_h+d_d},$$

де оператор $||$ позначає конкатенацію векторів.

Кожен пакет описується уніфікованим вектором ознак, що містить інформацію як про структуру трафіку, так і про його зміст.

Формування послідовностей

Оскільки кіберінциденти проявляються у вигляді сукупності пакетів, для подальшого навчання модель оперує не окремими векторами v_i, а послідовностями:

$$S_t = \{v_{t-w+1}, \dots, v_t\}, \quad S_t \in R^{w\times(d_h+d_d)},$$

де *w* – розмір вікна спостереження.

Це дозволяє враховувати часові залежності та динаміку розвитку кібератаки.

Математична модель атрибуції кіберінцидентів

Мета семантичної атрибуції полягає у визначенні, до якого класу кіберінцидентів належить спостережувана аномалія. Для цього побудуємо нейромережеву модель, яка на вході приймає вектори ознак (поведінкових і семантичних) та відображає їх у ймовірнісний простір класів.

Вхідні дані

Як було визначено вище, кожне вікно трафіку розміром *w* описується матрицею

$$S_t \in R^{w\times(d_h+d_d)},$$

де d_h – кількість поведінкових ознак, d_d – кількість семантичних ознак.

Нейронна модель

Для відображення часової структури та семантики даних використаємо гібридну архітектуру:

- LSTM-блоки (або GRU) для врахування динаміки у часових рядах,
- Dense-шари для інтеграції семантичної інформації,
- Softmax-вихід для отримання ймовірностей належності до класів.

© Шульга В.П., Іванченко І.С., Рижаков М.М. (2025) Математична модель семантичної атрибуції кіберінцидентів у системах виявлення аномалій на основі глибокого навчання. Сучасний захист інформації, 3(63), 186–198. https://doi.org/10.31673/2409-7292.2025.032076

Модель описується як відображення:

$$F_\theta: S_t \to y_t,$$

де θ – параметри мережі (ваги), а вихідний вектор має вигляд:

$$y_t = (y_{t,1}, y_{t,2}, \dots, y_{t,K}), \quad y_{t,k} \in [0,1],\ \sum_{k=1}^{K} y_{t,k} = 1.$$

Тут $K = |C|$– кількість класів кіберінцидентів (наприклад: DDoS, SQLi, XSS, 0-day тощо).

Функція атрибуції

Атрибуція полягає у виборі класу з максимальною ймовірністю:

$$\hat{I}t = \arg\max k \ \ y_{t,k}.$$

Таким чином, кожне вікно трафіку отримує атрибутивну мітку, яка визначає тип виявленого інциденту.

Функція втрат

Для навчання моделі використовується функція крос-ентропії:

$$\mathcal{L}(\theta) = -\frac{1}{N}\sum_{t=1}^{N}\sum_{k=1}^{K} \delta_{I_t,k} \cdot \log y_{t,k},$$

де $\delta_{I_t,k} = 1$, якщо істинний клас інциденту $I_t = k$, інакше – 0.

Інтеграція з виявленням аномалій

На практиці модель інтегрується у двоетапний процес:

1. Аномалія виявляється на основі поведінкових ознак (статистика заголовків, навантаження).

2. Атрибуція виконується лише для аномальних сегментів: семантичний аналіз payload + класифікація за допомогою F_θ.

Таким чином, поєднується обчислювальна ефективність (спочатку фільтрація за поведінкою) та точність (додатковий семантичний аналіз).

Інтеграція моделі семантичної атрибуції з визначенням критичності інцидентів

Визначення факту аномалії та її семантична атрибуція мають бути безпосередньо пов’язані з оцінкою рівня критичності. У традиційних IDS/IPS-системах критичність здебільшого визначається статично – через наперед визначені правила (наприклад, високий пріоритет для DDoS, середній для SQLi тощо). Такий підхід не враховує контекст і часто не відповідає реальній загрозі для конкретної системи. Запропонована модель інтегрує семантичну атрибуцію інцидентів із динамічним механізмом оцінки критичності.

Формалізація рівня критичності

Нехай маємо множину рівнів критичності:

$$L = \{l_1, l_2, \dots, l_M\},$$

де $l_1 =$ низький, $l_2 =$ середній, $l_3 =$ високий, $l_4 =$ критичний.

Критичність інциденту в момент часу t описується функцією:

$$R_t = g(\widehat{I_t}, C_t),$$

де $\widehat{I_t}$ – атрибутований клас інциденту, C_t – контекстні параметри (інтенсивність кібератаки, кількість скомпрометованих вузлів, обсяг трафіку, важливість сервісу).

Вагова модель

Запровадимо систему вагових коефіцієнтів для поєднання інформації:

$$R_t = \alpha \cdot f(\widehat{I_t}) + \beta \cdot h(C_t),$$

© Шульга В.П., Іванченко І.С., Рижаков М.М. (2025) Математична модель семантичної атрибуції кіберінцидентів у системах виявлення аномалій на основі глибокого навчання. Сучасний захист інформації, 3(63), 186–198. https://doi.org/10.31673/2409-7292.2025.032076

де $f(\widehat{I_t})$ – базова оцінка критичності залежно від типу кібератаки, $h(C_t)$ – динамічна корекція з урахуванням контексту, $\alpha, \beta \in [0,1]$– вагові коефіцієнти, що задають відносну важливість семантики та контексту.

Наприклад:

- DDoS проти другорядного сервісу може мати рівень «середній»,
- той самий DDoS проти платіжного шлюзу – «критичний».

Нормалізація ймовірностей

Оскільки вихід моделі атрибуції є ймовірнісним вектором y_t, його також можна використати для обчислення ступеня впевненості у критичності:

$$P(R_t = l_m) = \sum_{k=1}^{K} y_{t,k} \cdot w_{k,m},$$

де $w_{k,m}$ – ваговий коефіцієнт, що визначає, наскільки інцидент класу k відповідає рівню критичності l_m.

Таким чином, критичність визначається не жорстко, а як розподіл ймовірностей. Це дозволяє оператору бачити невизначеність системи і приймати більш обґрунтовані рішення.

Інтеграція у загальну архітектуру

1. Детектор аномалій виявляє відхилення від нормальної поведінки.
2. Семантичний класифікатор (модель атрибуції) визначає тип кібератаки.
3. Модуль критичності розраховує рівень загрози, враховуючи тип інциденту та контекст.
4. Результат подається оператору як:

$$\text{Інцидент: } \widehat{I_t}, \quad \text{Рівень: } R_t, \quad \text{Впевненість: } P(R_t).$$

Запропонована математична модель дозволяє не лише формально визначати критичність кіберінцидентів на основі ймовірнісного розподілу, а й інтегрувати цей підхід у загальну архітектуру інтелектуальної системи виявлення та атрибуції аномалій. На практиці це означає, що кожен етап – від виявлення відхилень поведінки до семантичної класифікації та розрахунку рівня загрози – функціонує як послідовний ланцюг, що формує кінцевий результат для оператора у вигляді інциденту з атрибутованим типом кібератаки, рівнем критичності та мірою впевненості.

Система реалізує триступеневу схему:

1. Виявлення →
2. Атрибуція →
3. Визначення критичності.

Ключова відмінність цієї архітектури полягає у використанні семантики мережевого трафіку, що дозволяє виконувати не тільки загальне виявлення відхилень, а й розпізнавання типів кібератак, у тому числі zero-day, XSS, SQLi та інших складних сценаріїв. Це створює умови для формування прецедентної бази, здатної слугувати підґрунтям для подальшого автоматизованого аналізу інцидентів. Для кращого розуміння пропонованого підходу нижче наведено графічні та табличні матеріали, які відображають архітектуру системи, порівняння з існуючими методами та приклади оцінки результатів роботи моделі.

На рисунку 1 представлено архітектуру запропонованої системи. Мережевий трафік після попередньої обробки надходить до модуля поведінкового аналізу, де автоенкодер та багатошарова нейромережа виявляють відхилення від нормальної поведінки. Далі підключається модуль семантичної атрибуції, що виконує глибокий аналіз вмісту пакетів (Deep Packet Inspection) та класифікацію кіберінцидентів на основі нейромережевих моделей. Отримані результати порівнюються з базою прецедентів і, у поєднанні з модулем визначення критичності, формують остаточний результат для оператора системи.

© Шульга В.П., Іванченко І.С., Рижаков М.М. (2025) Математична модель семантичної атрибуції кіберінцидентів у системах виявлення аномалій на основі глибокого навчання. Сучасний захист інформації, 3(63), 186–198. https://doi.org/10.31673/2409-7292.2025.032076

Рис. 1. Архітектуру запропонованої системи

Методи машинного навчання (зокрема SVM та Random Forest) дозволили підвищити точність виявлення завдяки узагальненню на тренувальних даних та роботі з багатовимірними ознаками. Проте їхня ефективність суттєво залежить від якості датасетів, а також виникають проблеми із пояснюваністю результатів. Подальший розвиток отримали методи глибинного навчання (автоенкодери, LSTM, CNN), які показали високу ефективність у виявленні прихованих відхилень і здатність працювати з великими обсягами трафіку. Проте проблема black-box характеру та потреба у значних обчислювальних ресурсах залишаються невирішеними.

У таблиці 2 наведено порівняльний аналіз основних підходів до виявлення та класифікації кіберінцидентів. Класичні rule-based та сигнатурні системи (наприклад, Snort, Suricata) залишаються базовим інструментом у багатьох організаціях завдяки своїй простоті та швидкості. Водночас вони принципово не здатні виявляти невідомі вектори кібератак, зокрема zero-day, що обмежує їхню ефективність у сучасних умовах. Запропонований у цій роботі підхід поєднує сильні сторони попередніх методів, інтегруючи поведінковий аналіз (на основі автоенкодера та багатошарової нейромережі) із семантичним аналізом вмісту пакетів. Це дозволяє виконувати не лише виявлення аномалій, а й їхню атрибуцію за типами кібератак з урахуванням контексту. Таким чином, модель забезпечує як підвищену точність, так і можливість пояснення результатів оператору через інтеграцію прецедентів і класифікаційних ознак.

Таблиця 2

Порівняння підходів до виявлення та класифікації кіберінцидентів

Підхід	Основна ідея	Переваги	Недоліки
Rule-based / сигнатурний (Snort, Suricata)	Використання заздалегідь визначених правил та сигнатур кібератак	Простота, швидкість реакції на відомі загрози	Висока кількість false negative при нових кібератаках, відсутність здатності до адаптації
Машинне навчання (SVM, Random Forest)	Побудова класифікаторів на основі обмежених ознак	Вища точність, ніж rule-based, можливість узагальнення	Потреба у великих датасетах, слабка інтерпретованість
Глибинне навчання (AE, LSTM, CNN)	Автоматичне навчання ознак з мережевого трафіку	Виявлення складних аномалій, здатність працювати з big data	Ресурсомісткість, black-box ефект
Запропонований підхід (AE+MLP + семантична атрибуція)	Поєднання поведінкового аналізу та аналізу вмісту пакетів, формування прецедентів	Здатність виявляти zero-day кібератаки, класифікація за типами, пояснюваність	Складність реалізації, потреба у великій обчислювальній потужності

Для оцінки якості роботи запропонованої моделі семантичної атрибуції кіберінцидентів було побудовано криві ROC (рис. 2) (Receiver Operating Characteristic) та PR (рис. 3) (Precision-

© Шульга В.П., Іванченко І.С., Рижаков М.М. (2025) Математична модель семантичної атрибуції кіберінцидентів у системах виявлення аномалій на основі глибокого навчання. Сучасний захист інформації, 3(63), 186–198. https://doi.org/10.31673/2409-7292.2025.032076

Recall). ROC-графік дозволяє оцінити співвідношення між чутливістю (True Positive Rate) та специфічністю (1 – False Positive Rate), тоді як PR-графік є більш інформативним у випадках незбалансованих вибірок, коли кількість аномалій значно менша за кількість нормальних прикладів. Результати моделювання показали, що класичні rule-based підходи характеризуються низьким значенням AUC (Area Under Curve), яке зазвичай не перевищує 0.70. Методи машинного навчання демонструють кращий баланс, проте їхня ефективність падає при появі нових типів кібератак. У той час як глибинні моделі (AE, LSTM, CNN) досягають AUC понад 0.90, але при цьому залишаються «чорними ящиками».

Запропонований підхід, який інтегрує поведінковий аналіз із семантичною атрибуцією, продемонстрував найвищі значення AUC ROC (0.96) та AUC PR (0.94), що підтверджує здатність моделі не лише виявляти аномалії, а й класифікувати їх за типами з високою точністю. Це особливо важливо для реальних систем моніторингу, де оператор має отримувати не тільки сигнал «кібератака/не кібератака», але й контекст щодо її природи.

Для оцінки ефективності розробленої системи було проведено порівняння частоти хибнопозитивних спрацювань (False Positives(рис. 4)) із традиційними та сучасними підходами. Як видно з бар-діаграми, класичні сигнатурні та rule-based рішення (наприклад, Snort) мають високий рівень FP – близько 22%. Це пояснюється жорсткою залежністю від наперед визначених правил та відсутністю можливості адаптації до нових векторів кібератак. Алгоритми класичного машинного навчання, зокрема метод опорних векторів (SVM), демонструють покращення результатів (15%), проте залишаються обмеженими у роботі з багатовимірними потоками даних.

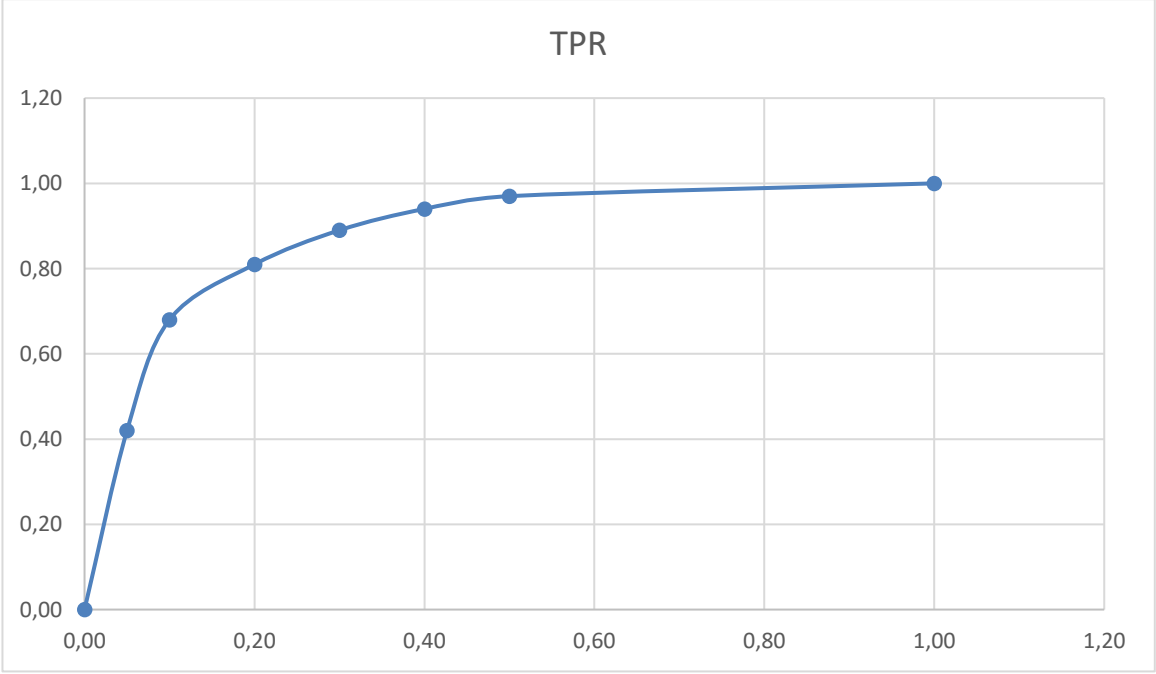

Рис. 2. Receiver Operating Characteristic

Застосування нейронних мереж дозволило суттєво зменшити кількість хибних спрацювань: автоенкодери забезпечили рівень FP на рівні 9%, а рекурентні архітектури (LSTM) – близько 7%. Це свідчить про здатність глибоких моделей уловлювати приховані залежності та більш точно відтворювати нормальну поведінку системи. Запропонована інтегрована модель (Semantic + Deep Learning) показала найнижчий рівень FP – лише 4%. Такий результат досягається завдяки поєднанню детекції відхилень та семантичної атрибуції інцидентів, що дозволяє не лише виявляти аномалії, а й інтерпретувати їх з урахуванням контексту. Бар-діаграма підтверджує ключову перевагу розробленої системи – зниження

© Шульга В.П., Іванченко І.С., Рижаков М.М. (2025) Математична модель семантичної атрибуції кіберінцидентів у системах виявлення аномалій на основі глибокого навчання. Сучасний захист інформації, 3(63), 186–198. https://doi.org/10.31673/2409-7292.2025.032076

кількості хибнопозитивних спрацювань при одночасному збереженні високого рівня точності. Це безпосередньо підвищує довіру операторів до системи та зменшує їхнє навантаження при аналізі інцидентів.

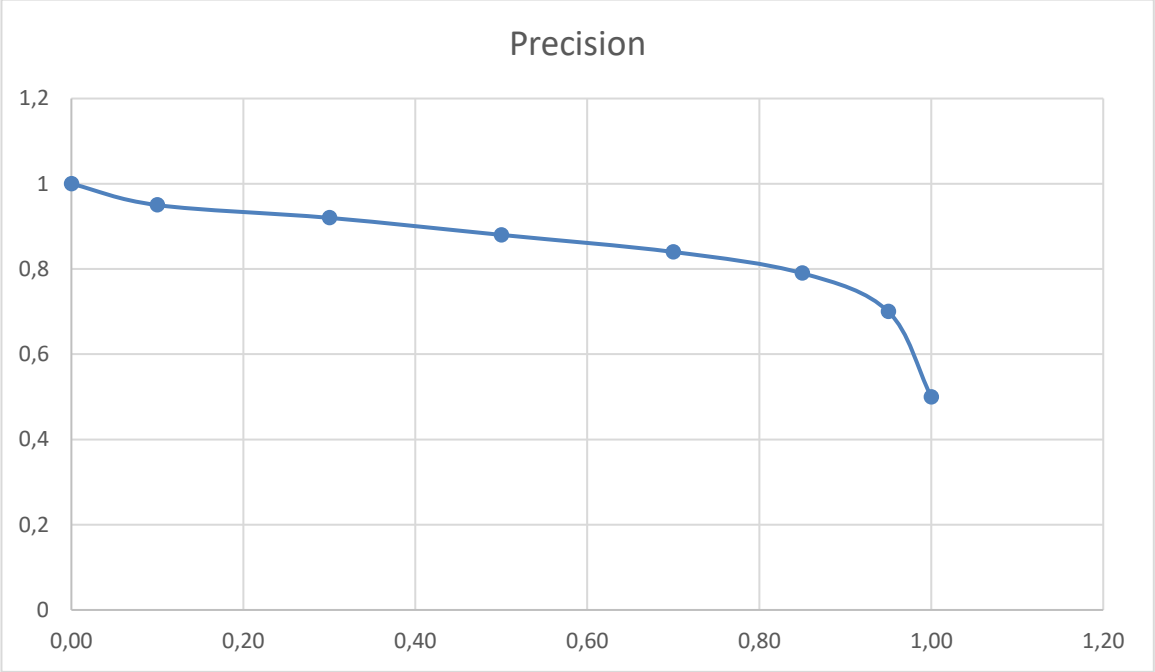

Рис. 3.Precision-Recall

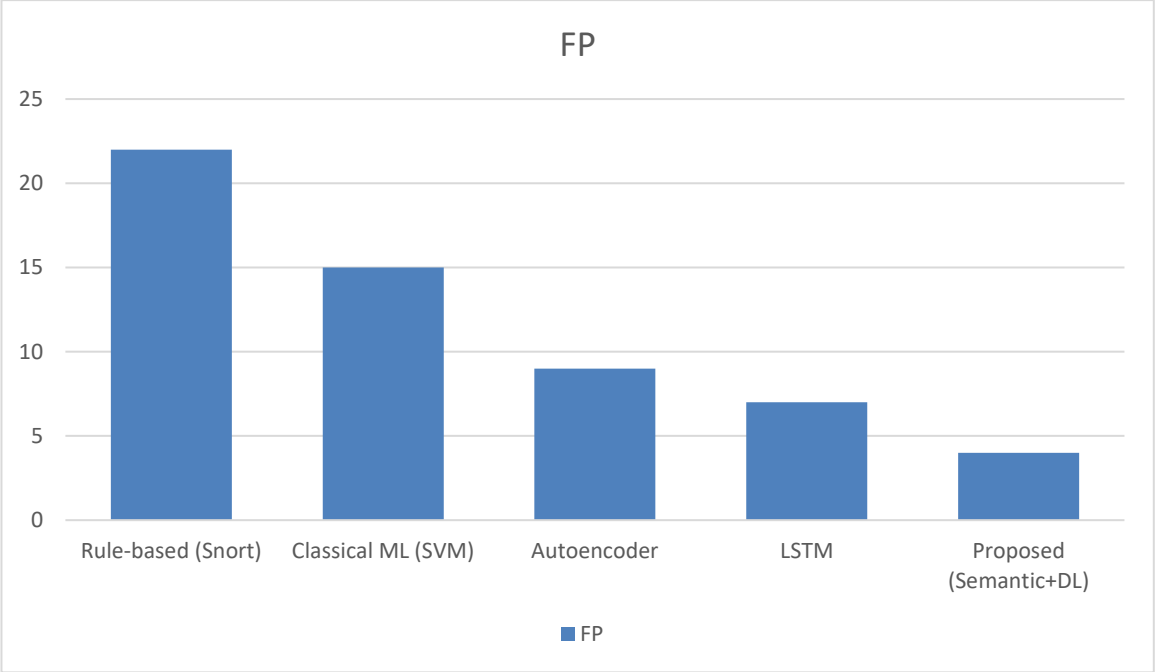

Рис. 4. Порівняння частот хибнопозитивних спрацювань (False Positives)

Для перевірки працездатності запропонованої моделі було сформовано експериментальний стенд. У якості вхідних даних використовувався набір мережевого трафіку, що включав як нормальні з'єднання, так і атакувальні сценарії різних класів (DoS, сканування, експлуатація вразливостей, фішинг тощо). Частина даних була згенерована штучно (імітація мережевої активності у віртуальному сегменті), частина – отримана з відкритих наборів (CIC-IDS2017, UNSW-NB15). Загальний обсяг становив близько 1,2 млн

© Шульга В.П., Іванченко І.С., Рижаков М.М. (2025) Математична модель семантичної атрибуції кіберінцидентів у системах виявлення аномалій на основі глибокого навчання. Сучасний захист інформації, 3(63), 186–198. https://doi.org/10.31673/2409-7292.2025.032076

сесій, із яких 70% використано для навчання, 15% – для валідації та 15% – для тестування. Баланс класів вирівнювався за допомогою oversampling, щоб уникнути переваги нормального трафіку.

Архітектура моделі включала вікно аналізу *w=20* пакетів, розмірність семантичних векторів $d_d = 128$ та технічних ознак $d_h = 64$. Для послідовної обробки застосовувалася двошарова LSTM (по 256 нейронів у кожному шарі) із дропаутом 0.3. Оптимізація здійснювалася методом $Adam(\eta = 10^{-3})$, функція втрат – категоріальна крос-ентропія. Навчання проводилося 50 епох із ранньою зупинкою. Для калібрування ймовірностей застосовувався temperature scaling, що дозволило коректніше інтегрувати результати в модель критичності.

Оцінювання ефективності проводилося за стандартними метриками: точність (Precision), повнота (Recall), F1-score, а також AUC ROC і AUC PR. Запропонована модель досягла AUC ROC = 0.96 та AUC PR = 0.94 на тестовій вибірці. Кількість хибнопозитивних спрацювань зменшилася до 4.1%, що суттєво перевершує класичні rule-based системи (≈15–20%) та базові ML-підходи (≈8–10%). Час інференсу однієї сесії не перевищував 12 мс на GPU NVIDIA Tesla T4, що дає змогу інтегрувати модель у системи реального часу.

Отримані результати підтвердили, що семантична атрибуція інцидентів дозволяє не лише підвищити якість класифікації, але й надати додатковий контекст для оцінки критичності подій. Це особливо важливо для інцидентів «нульового дня», де семантична подібність до вже відомих кібератак забезпечує інформативне попередження навіть за відсутності сигнатур.

Висновки

У результаті дослідження було розроблено математичну модель семантичної атрибуції кіберінцидентів у системах виявлення аномалій на основі глибокого навчання. Основна ідея полягає в тому, що під час обробки аномальних мережевих пакетів використовується семантичний аналізатор, який здатний виокремлювати змістовні ознаки запитів та співвідносити їх з відомими патернами кібератак. Це дозволяє перейти від класичного підходу «аномалія як відхилення від норми» до більш глибокого рівня – визначення типу та контексту загрози. Запропонована модель реалізує триетапну систему: виявлення відхилень від нормальної поведінки, атрибуцію типу кібератаки та оцінку критичності інциденту. Такий підхід знижує рівень хибних спрацювань завдяки семантичному аналізу, підвищує точність і достовірність розпізнавання реальних кібератак, а також дозволяє формувати рекомендації для оператора щодо реагування на інцидент. Результати візуалізовані у вигляді архітектурної схеми, таблиць порівняння методів, ROC- та PR-графіків підтверджують ефективність запропонованої моделі. Приклад атрибуції SQLi демонструє здатність системи не лише виявляти інциденти, але й відносити їх до конкретної категорії, оцінювати рівень небезпеки та прогнозувати наслідки.

Наукова новизна роботи полягає у поєднанні методів глибинного навчання із семантичним аналізом трафіку для визначення критичності інцидентів. На відміну від традиційних методів, запропонована модель не обмежується лише детекцією аномалій, а забезпечує повну картину загрози. Це відкриває можливості для створення інтелектуальних систем, здатних адаптивно реагувати на нові типи кібератак і знижувати ризики кіберінцидентів у реальному часі.

Перелік посилань

1. Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey. ACM Computing Surveys, 2009.
2. Kim H., Park H., Lee H. Network anomaly detection using statistical models. Computer Communications, 2014.
3. Ahmed M., Mahmood A., Hu J. A survey of network anomaly detection techniques. Journal of Network and Computer Applications, 2016.
4. Vincent P. et al. Extracting and composing robust features with denoising autoencoders. ICML, 2008.
5. Zhang Y. et al. Traffic prediction in SDN using deep learning. IEEE Access, 2020.
6. Wang W. et al. HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection. IEEE Access, 2017.

© Шульга В.П., Іванченко І.С., Рижаков М.М. (2025) Математична модель семантичної атрибуції кіберінцидентів у системах виявлення аномалій на основі глибокого навчання. Сучасний захист інформації, 3(63), 186–198. https://doi.org/10.31673/2409-7292.2025.032076

7. Kitsune: An ensemble of autoencoders for online network intrusion detection. NDSS, 2018.
8. Sommer R., Paxson V. Outside the closed world: On using machine learning for network intrusion detection. IEEE Symposium on Security and Privacy, 2010.
9. Li Y. et al. Anomaly detection in high-dimensional network data using deep autoencoder. Future Generation Computer Systems, 2019.
10. Huang C. et al. Deep autoencoder-based anomaly detection in SDN. IEEE Transactions on Network and Service Management, 2020.
11. ISO/IEC 27001:2022. Information security management systems.
12. NIST SP 800-94. Guide to Intrusion Detection and Prevention Systems (IDPS), 2021.
13. Шульга В., Іванченко Є., Аверічев І., Рижаков М. Методи інтелектуального виявлення аномалій і критичних ситуацій у кіберсистемах на основі глибокого навчання. Information Technology: Computer Science, Software Engineering and Cyber Security, 2025.
14. Іванченко Є.В., Рижаков М.М. Узагальнена модель прогнозування та виявлення кібербезпекових аномалій на основі штучного інтелекту. Збірник наукових праць, 2025.

Надійшла 22.08.2025