УДК 004.056.5:004.891.2:004.75:004.942
DOI: 10.31673/2409-7292.2025.032156

Хохлачова Ю.Є., Флоров С.В., Черкаський О.В., Черкаський Д.О., Переметчик Д.О., Білан М.В.

МОДЕЛЮВАННЯ НЕЙРОВІРУСНИХ КАМПАНІЙ У МЕРЕЖЕВОМУ ТРАФІКУ НА ОСНОВІ ПОЄДНАННЯ ПОТОКОВИХ, ЖУРНАЛЬНИХ І БАЙТОВИХ ОЗНАК

У роботі досліджено задачу виявлення «нейровірусів» - шкідливих програмних структур , які поєднують класичні мережеві методи з технологіями машинного навчання для обфускації та адаптивності. Основна мета полягає у підвищенні точності детекції нових і прихованих атак у мережевому трафіку шляхом комплексного аналізу даних різної природи: агрегованих потоків (NetFlow), детальних записів мережевого аналізатора Zeek, журналів систем управління інформацією та подіями безпеки (Security Information and Event Management, далі SIEM) і байтових артефактів прошивок. Запропоновано архітектури на основі згорткових і рекурентних нейронних мереж (Convolutional Neural Network + Long Short-Term Memory, далі CNN+LSTM) для моделювання часових послідовностей та автоенкодерів із довготривалою короткочасною пам'яттю (Autoencoder + LSTM, далі AE+LSTM) для безнаглядного виявлення відхилень. Байтові послідовності перетворюються методом Byte2Image у зображення сталої розмірності у відтінках сірого, що уніфікує обробку артефактів протоколів Transport Layer Security / Secure Sockets Layer (TLS/SSL) та Simple Network Management Protocol (SNMP), у тому числі з уразливих прошивок мережевого обладнання. Методика передбачає синхронізацію часових вікон, балансування класів і навчання моделі на зваженій функції втрат з урахуванням вартості різних типів помилок. Експерименти проведено на підмножинах NetFlow, Zeek і SIEM з відтворенням сценаріїв гібридних атак: скритого сканування, каналів керування через TLS зі зниженням версії, експлуатації SNMPv2c з типовими community-рядками та ін'єкцій у прошивки. Порівняння з базовими методами (Random Forest, Isolation Forest, рекурентні нейронні мережі) показало підвищення інтегральної оцінки F1 до 0.94 для невідомих сімейств атак і зменшення середньої затримки спрацювання на 27% у режимі реального часу. Запропонована архітектура узгоджується з принципами нульової довіри (Zero Trust), підтримує кореляцію з матрицею MITRE ATT&CK та забезпечує відтворюваність. Практичний внесок полягає у підвищенні стійкості операторських і корпоративних мереж до хвильових гібридних атак та формуванні регламентованого пакета з методики, специфікації моделі, карти даних, протоколу випробувань і креслення топології. Результати можуть стати основою для автоматизації політик доступу, адаптивного відбору телеметрії й інтеграції з платформами розвідданих про загрози.

Ключові слова: нейровірус, мультимодальні нейронні мережі, згорткова і рекурентна модель, автоенкодер, Byte2Image, NetFlow, Zeek, система управління інформацією та подіями безпеки, система виявлення вторгнень, Zero Trust, SNMP, firmware-атаки.

Вступ

Гібридні кібератаки, що поєднують мережеві прийоми, вплив на прошивки (firmware) активного обладнання і використання розвідувальних даних про загрози, зумовлюють потребу у мультимодальних підходах до виявлення порушень. Модель «нульової довіри» (Zero Trust) передбачає відмову від припущення існування «довіреної» зони та вимагає постійного підтвердження атрибутів доступу, контексту та поведінки користувачів і пристроїв [1]. Європейське агентство з кібербезпеки (European Union Agency for Cybersecurity, ENISA) фіксує зростання частки складних, багатоетапних кампаній у критичних секторах, де шифрування трафіку та інжиніринг протоколів істотно ускладнюють інспекцію [2]. Для прив'язки ознак до відомих технік зловмисників використовується матриця MITRE ATT&CK [3]. У роботі під «нейровірусами» розуміються кампанії, у яких застосовуються модулі машинного навчання для адаптивного обходу механізмів виявлення або приховування каналів керування та ексфільтрації даних. Надійне виявлення таких загроз у потоках електронних комунікацій потребує поєднання кількох джерел: агрегованих мережевих потоків (NetFlow), детальних подій рівня сесій і протоколів (система Zeek, відома також як Bro), журналів систем управління інформацією та подіями безпеки (Security Information and Event Management, SIEM) та байтових артефактів з прошивок і сесій (наприклад, фрагментів рукостискань протоколів Transport Layer Security / Secure Sockets Layer, TLS/SSL, і витягів з Simple Network Management Protocol, SNMP). У стратегії раннього виявлення важливо враховувати вразливості SSL/TLS (пониження версії, застарілі набори шифрів) та SNMP (версії 1 і 2c без

шифрування, використання типових community-рядків), особливо коли вони закріплені у прошивках комутаторів, маршрутизаторів, контролерів або модулів управління базовою системою (Baseboard Management Controller, BMC) [4–7].

Постановка проблеми

Сучасні мережі електронних комунікацій перебувають під постійною загрозою складних багатоступеневих атак, що поєднують класичні методи компрометації протоколів (TLS/SSL, SNMP), модифікації прошивок мережевого обладнання та застосування алгоритмів машинного навчання для обходу існуючих засобів захисту. Такі атаки, які в роботі визначено як «нейровірусні кампанії», характеризуються високою адаптивністю, здатністю змінювати поведінкові патерни у часі та маскувати канали керування й ексфільтрації даних у зашифрованому трафіку [7]. Проблема ускладнюється переходом операторських і корпоративних систем до віртуалізованих функцій та розподілених архітектур, що істотно збільшує обсяг та різнорідність телеметрії. Використання лише одного джерела даних — наприклад, NetFlow чи логів Zeek — виявляється недостатнім для виявлення прихованих загроз. Аналіз шифрованого трафіку додатково обмежує можливості класичних методів інспекції, змушуючи дослідників звертатися до непрямих ознак, статистичних характеристик та контексту із систем управління інформацією і подіями безпеки (SIEM) [7].

Традиційні підходи до виявлення аномалій (наприклад, Random Forest, ізоляційні ліси чи класичні LSTM-моделі) демонструють прийнятний рівень ефективності на відкритих наборах даних, проте у реальних умовах характеризуються високими показниками хибнопозитивних спрацювань та недостатньою здатністю до генералізації на нові або обфусковані сімейства атак [7]. Зростаючі вимоги до обробки потоків у режимі реального часу вимагають побудови архітектур, здатних синхронізувати кілька модальностей даних, адаптивно оновлювати пороги прийняття рішень та інтегруватися у концепцію нульової довіри (Zero Trust).

Таким чином, основна проблема полягає у розробці мультимодальних методів виявлення загроз, які б одночасно враховували статистику мережевих потоків, журнальні події та байтові артефакти, забезпечуючи при цьому прийнятний баланс між точністю, швидкодією та масштабованістю системи.

Аналіз останніх досліджень і публікацій

Останні роки характеризуються активним розвитком мультимодальних підходів до виявлення вторгнень: поєднання потокових характеристик, протокольних журналів і байтових артефактів демонструє переваги над одномодальними рішеннями. Роботи з мультимодального глибинного навчання для NIDS підтверджують ефективність ансамблів CNN/LSTM та автоенкодерів для виявлення складних патернів і аномалій у динаміці трафіку [4], [5]. В окремій гілці досліджень розвиваються автоенкодери для зашифрованого трафіку, які спираються на часові та статистичні ознаки замість глибинної інспекції пакетів [2]. Разом із тим, типові обмеження — чутливість до зсувів розподілів, висока кількість FP у реальному середовищі, а також обмежена узагальнюваність на нові сімейства атак — залишаються актуальними й сьогодні [7].

Широко застосовувані еталонні датасети (UNSW-NB15, CIC-IDS2017, Kitsune) забезпечують порівнюваність результатів і швидке прототипування, але не повною мірою репрезентують промислові навантаження, варіативність протоколів і сучасні техніки обфускації [14]–[16]. Для підсилення семантики подій дослідники залучають телеметрію рівня сесій/протоколів через Zeek і метадані потоків NetFlow, а також інтегрують SIEM для кореляції інцидентів і відповідності нормативним вимогам [6], [7], [10]. На рівні стандартів суттєве значення мають RFC 5424 (системні журнали), RFC 3411–3418 (SNMPv3) та RFC 8446 (TLS 1.3), які визначають формат і видимість ознак у зашифрованих каналах [10-12].

Окремий вектор стосується подання байтових послідовностей як зображень (Byte2Image) для застосування згорткових мереж до двійкових артефактів; такий підхід розширює клас ознак для TLS/SSL-рукостискань, SNMP-фрагментів і firmware-сегментів [13],

[18], [28]. Паралельно триває інтенсифікація класичних і глибинних методів: Random Forest, LSTM, CNN, варіаційні автоенкодери, а також узагальнювальні огляди щодо ризик-орієнтованої оцінки та індустріальних систем (ICS) підкреслюють компроміс точність/швидкодія та потребу в адаптивній калібровці порогів [19-23], [24], [31]. Вразливості SSL (наприклад, DROWN) ілюструють важливість протокол-специфічних індикаторів і контролю даунгрейду навіть за наявності сучасних версій TLS [29].

На рівні політик і архітектур помітним стає перехід до Zero Trust, рекомендацій ISO/IEC 27002, напрацювань ENISA/ETSI щодо багатошарових захисних моделей і кореляції з матрицею MITRE ATT&CK; це спрямовує дослідження до інтегрованих рішень, де ML-детекція безпосередньо впливає на керування доступом і телеметрією [1-3], [9], [25], [30]. Водночас прогалини залишаються: рідкісна інтеграція firmware-артефактів у поточні NIDS, недостатня вартісна (cost-sensitive) оптимізація рішень у реальному часі та обмежена узгодженість з контурами Zero Trust. Саме ці розриви адресує дане дослідження, поєднуючи NetFlow/Zeek/SIEM із Byte2Image та двома взаємодоповнювальними архітектурами (CNN+LSTM, AE+LSTM) у єдиній рамці з ризик-скорингом і пороговою оптимізацією [7].

Метою дослідження є підвищення точності та швидкодії виявлення нових і обфускованих кібератак шляхом інтегрованого використання кількох джерел телеметрії: агрегованих мережевих потоків (NetFlow), протокольних подій Zeek, журналів SIEM та байтових артефактів із сесій і прошивок. Запропонований підхід базується на поєднанні згорткових і рекурентних моделей (CNN+LSTM) для аналізу часових послідовностей і автоенкодерів із рекурентними шарами (AE+LSTM) для безнаглядного виявлення відхилень.

Завдання дослідження

Завдання дослідження полягають у формуванні уніфікованого простору ознак і часових вікон для синхронізації даних різних модальностей; у конструюванні байтових зображень (Byte2Image) для уніфікованої обробки TLS/SSL-рукостискань, SNMP-фрагментів та firmware-сегментів; у розробці та калібруванні ймовірнісного оцінювача ризику для врахування вартості пропусків та хибнопозитивних спрацювань; в оптимізації порогів прийняття рішень за допомогою вартісних функцій для досягнення збалансованості між точністю та пропускною здатністю; а також в інтеграції результатів моделювання з принципами Zero Trust і узгодженні їх із матрицею MITRE ATT&CK для підвищення керованості ризиками та відповідності сучасним стандартам безпеки [1-3], [25], [30], [7].

Виклад основного матеріалу дослідження

Методологія дослідження побудована на інтеграції даних різної природи — мережевих потоків, протокольних подій та байтових артефактів. Такий мультимодальний підхід дозволяє виявляти як класичні атаки, так і адаптивні кампанії, що змінюють свою поведінку у часі [4], [5]. Основними етапами методики є нормалізація потоків і журналів, побудова часових вікон, формування уніфікованого простору ознак [6], [7], а також застосування глибинних нейронних архітектур для класифікації та пошуку аномалій [20-23]. Для цього використано дві взаємодоповнювальні стратегії: (1) згорткові та рекурентні моделі для аналізу послідовностей трафіку (CNN+LSTM) [21]; (2) автоенкодери з рекурентними шарами для безнаглядного виявлення відхилень (AE+LSTM) [22]. Додатково застосовано метод візуалізації байтових послідовностей у форматі зображень, що уніфікує обробку артефактів різних протоколів і прошивок (Byte2Image) [13], [18].

Використовуються три модальності: (i) агреговані мережеві потоки (NetFlow) з полями src/dst IP/port, proto, bytes, packets, duration, flags; (ii) події Zeek (conn, ssl, dns, http, snmp) зі штампами часу й атрибутами сесій; (iii) журнали SIEM (узагальнені поля host, facility, severity, event_id, tag) згідно з RFC 5424 [10]. Для прошивок та фрагментів сесій із TLS/SSL і SNMP використано перетворення Byte2Image: байтова послідовність довжини N переводиться у матрицю $H \times W$ у відтінках сірого.

$$\Phi:\ b_{1:N} \mapsto I \in \mathbb{R}^{H\times W},\quad I_{i,j} = \frac{b_{(i-1)W+j}}{255},\quad N = HW,\ b_k \in \{0,\dots,255\}. \tag{1}$$

Часові вікна синхронізуються до кроку Δt (напр., 10 с) із подальшою побудовою послідовностей $\{x_t^F\}_{t=1}^T$ (ознаки потоків), $\{x_t^L\}_{t=1}^T$ (логи SIEM/Zeek) та $\{x_t^B\}_{t=1}^T$ (вектори ознак від Ф).

Оцінимо апостеріорну вірогідність атаки $p(y=1|x^F, x^L, x^B)$ через латентну змінну z (стан кампанії/етапу):

$$p(y=1|x^F, x^L, x^B) = \sigma(\eta + \log \int p(x^F|z)\, p(x^L|z)\, p(x^B|z)\, p(z;\theta)\, dz), \tag{2}$$

де $\sigma(\cdot)$ — сигмоїда, η — зсув, $p(\cdot)$ — щільності, параметризовані θ.

Кінцеве рішення приймається за правилом $s_\theta(x) \geq \tau$, де s_θ — скоринговий вихід моделі, а τ — поріг. Оптимізацію виконуємо за зваженою вартісною функцією:

$$J(\theta, \tau) = \alpha\, \mathbb{E}[\mathbf{1}\{y=1,\ s_\theta(x) < \tau\}] + \beta\, \mathbb{E}[\mathbf{1}\{y=0,\ s_\theta(x) \geq \tau\}] + \lambda \parallel \theta \parallel_2^2, \tag{3}$$

де α і β відображають відносну вартість пропуску та хибного спрацювання, λ — регуляризація. Для практичної калібровки використовуємо валідаційні криві ROC/PR і мінімізуємо (3) grid-пошуком по τ.

Приклад розрахунку. Для поширеності $p = 0.01$, коефіцієнтів $\alpha = 20$, $\beta = 1$, та пари $(\mathrm{TPR}, \mathrm{FPR}) = (0.94, 0.06)$ маємо ризик $R \approx \alpha p(1-\mathrm{TPR}) + \beta(1-p)\mathrm{FPR} = 0.012 + 0.0594 = 0.0714$. Зсув порога дає (0.91,0.03) і $R \approx 0.018 + 0.0297 = 0.0477$, що вигідніше.

По кожному часовому кроці t обчислюємо згорткові ознаки f_{CNN} від x_t^F та $I_t = \Phi(x_t^B)$, потім пропускаємо через LSTM:

$$h_t = \mathrm{LSTM}(h_{t-1}, f_{\mathrm{CNN}}(x_t^F, I_t)), \qquad s_\theta(x) = \sigma(w^T h_T + b), \tag{4}$$

із бінарною крос-ентропією:

$$\mathcal{L}_{\mathrm{clf}} = -y \log s_\theta(x) - (1-y)\log(1 - s_\theta(x)) + \lambda \parallel \theta \parallel_2^2. \tag{5}$$

Ініціалізацію мережі здійснюємо з урахуванням збалансованих міні-пакетів і ваг класів для дисбалансу (class weights).

AE+LSTM для безнаглядного виявлення відхилень

Автоенкодер із LSTM навчається реконструювати нормальну поведінку багатомодальних векторів $\mathrm{x_t} = [\mathrm{x_t^F}; \mathrm{x_t^L}; \psi(\mathrm{I_t})]$, де ψ — плаский вектор ознак із зображення $\mathrm{I_t}$ (наприклад, середні та дисперсії блоків). Аномалії визначаються за реконструкційною похибкою $\mathrm{r_t} = \parallel \mathrm{x_t} - \hat{\mathrm{x}}_\mathrm{t} \parallel_2$:

$$\mathcal{L}_{\mathrm{AE}} = \frac{1}{\mathrm{T}} \sum_{\mathrm{t=1}}^{\mathrm{T}} \parallel \mathrm{x_t} - \hat{\mathrm{x}}_\mathrm{t} \parallel_2^2 + \beta_1 \sum_\ell \parallel \mathrm{W}_\ell \parallel_1, \qquad \mathrm{S(x)} = \mu(\mathrm{r}) + \kappa\, \mathrm{MAD(r)}, \tag{6}$$

де $\mathrm{S(x)}$ — сумарний скоринг, μ — середнє, MAD — медіанне абсолютне відхилення. Поріг для $\mathrm{S(x)}$ добирається за (3).

Кореляція з SIEM і Zero Trust

Події Zeek (ssl.log, snmp.log) та syslog/SIEM уніфікуються до спільної часової шкали. Для TLS/SSL контролюємо ознаки: версія (TLS 1.0/1.2/1.3), набір шифрів, розмір сертифіката, SNI; для SNMP — версія (v1/v2c/v3), тип події (get, set, trap), ознаки community та авторизації [11, 12, 17]. Політики Zero Trust на рівні контролера (Policy Decision Point) використовують $\mathrm{s_\theta(x)}$ та $\mathrm{S(x)}$ для адаптивного доступу [1, 25].

Запропонований набір методів забезпечує формування цілісної архітектури для детекції «нейровірусів», у якій поєднано ймовірнісні моделі ризику [24], оптимізацію порогів за вартісними функціями [19], а також мультимодальний аналіз телеметрії [14-16]. Це дозволяє врахувати як часову динаміку атак, так і специфічні ознаки протоколів Transport Layer Security (TLS) [12] чи Simple Network Management Protocol (SNMP) [11]. Подальша частина роботи

присвячена оцінюванню ефективності обраних моделей на контрольних наборах даних [14-16], порівнянню з базовими підходами [19], [20] та аналізу отриманих показників продуктивності [29-31].

Результати. Було сформовано синтетично-реалістичні підвибірки на основі UNSW-NB15, CIC-IDS2017 і Kitsune [14, 15, 16], доповнені штучними сценаріями: (i) схований скан; (ii) C2 через TLS зі зниженням версії (downgrade) [29, 12]; (iii) SNMPv2c з типовими community-рядками «public/private» [11]; (iv) модифікації прошивок, виділені за допомогою Binwalk [18, 28]. Потоки зведено в вікна по $\Delta t = 10$ с; для Byte2Image використано 64×64 та 32×32.

Для кількісної оцінки ефективності застосованих методів було проведено порівняння класичних та сучасних моделей машинного навчання і глибинного навчання. Основну увагу зосереджено на показниках точності (Precision), повноти (Recall), узагальненої метрики F1 та затримки обчислення, що є критично важливими параметрами для систем виявлення аномалій у режимі реального часу.

Таблиця 1

Порівняння точності для нових/обфускованих атак (усереднення по сценаріях)

Модель	**Точність (Precision)**	**Повнота (Recall)**	**F1**	**Затримка, мс**
Random Forest[4]	0.88	0.79	0.83	12
LSTM[7]	0.90	0.85	0.87	18
AE+LSTM (ця робота)	0.92	0.88	0.90	9
CNN+LSTM (ця робота)	0.95	0.93	0.94	11

У таблиці 1 наведено результати порівняння чотирьох моделей: базового алгоритму Random Forest, рекурентної нейронної мережі LSTM, а також двох мультимодальних архітектур, запропонованих у цій роботі (AE+LSTM та CNN+LSTM). Як видно з результатів, запропоновані архітектури демонструють вищі значення метрик точності та повноти при меншій або співмірній затримці. Зокрема, модель CNN+LSTM досягає найкращого значення F1-міри (0.94), тоді як AE+LSTM показує найнижчу затримку (9 мс), що робить її доцільною для сценаріїв з підвищеними вимогами до швидкодії. Таким чином, використання мультимодальних моделей підтверджує їхню перевагу у виявленні складних кібератак порівняно з традиційними підходами.

Абляція Byte2Image і ознак Zeek

Для оцінювання впливу різних конфігурацій перетворення Byte2Image та використаних наборів ознак було проведено експериментальне порівняння. Метою цього аналізу стало визначення компромісу між якістю виявлення аномалій та продуктивністю системи, оскільки обидва аспекти є критично важливими у задачах режиму реального часу.

Таблиця 2

Вплив конфігурацій Byte2Image і наборів ознак на якість та продуктивність

Конфігурація	**F1**	**TPR**	**FPR**	**Пропускна здатність, под./с**
+ базові NetFlow	0.91	0.90	0.07	18,500
+ NetFlow+Zeek (ssl, snmp)	0.94	0.93	0.05	12,900
+ Zeek повний (conn, dns, http, ssl, snmp)	0.93	0.92	0.06	11,700
+ лише NetFlow (без Zeek)	0.89	0.87	0.09	21,300

У таблиці 2 наведено результати для чотирьох конфігурацій. Найвищі значення метрик F1 (0.94) та TPR (0.93) досягнуті при комбінації NetFlow та вибраних журналів Zeek (ssl, snmp), що свідчить про збалансованість цих ознак для виявлення складних атак при помірному рівні пропускної здатності (12 900 под./с). Повний набір Zeek (conn, dns, http, ssl, snmp) також демонструє високу якість (F1=0.93), але знижує продуктивність до 11 700 под./с. Використання лише NetFlow без Zeek підвищує пропускну здатність (21 300 под./с), проте супроводжується погіршенням точності (F1=0.89, TPR=0.87) і зростанням рівня хибнопозитивних спрацьовувань (FPR=0.09). Таким чином, результати підтверджують необхідність пошуку оптимального балансу між точністю та швидкодією, що особливо важливо для масштабованих систем моніторингу мережевого трафіку.

Приклад MATLAB Mobile (онлайн-скоринг)

T = readtable ('NetFlow_sample.csv'); X = table2array (T(:, 3:12)); I = byte2img ('tls_handshake.bin', 64, 64); load('cnn_lstm.mat','net'); score = predict(net, {X, I}); anomaly = score > 0.65; idx = find(anomaly); disp(['Аномальних потоків: ', num2str(numel(idx))]);

Сучасні мережі електронних комунікацій перебувають під постійним тиском складних багатоступеневих атак, які поєднують у собі компрометацію протоколів (TLS/SSL, SNMP), використання firmware-вразливостей та маніпуляції телеметрією. Традиційні інструменти виявлення аномалій часто є недостатньо ефективними в умовах шифрованого трафіку та високої динаміки загроз. Для підвищення точності й адаптивності необхідна архітектура, що поєднує машинне навчання (CNN+LSTM, AE+LSTM), аналіз телеметрії (NetFlow, Zeek) і політики Zero Trust. Запропонована схема ілюструє концепцію інтегрованого рішення, де телеметрія обробляється через каскадну систему брокера, класифікаторів та детекторів аномалій, а кінцеві рішення ухвалюються контролером політик Zero Trust (Policy Decision Point, PDP) з подальшою передачею до виконавців (Policy Enforcement Point, PEP) і систем управління безпекою (SIEM).

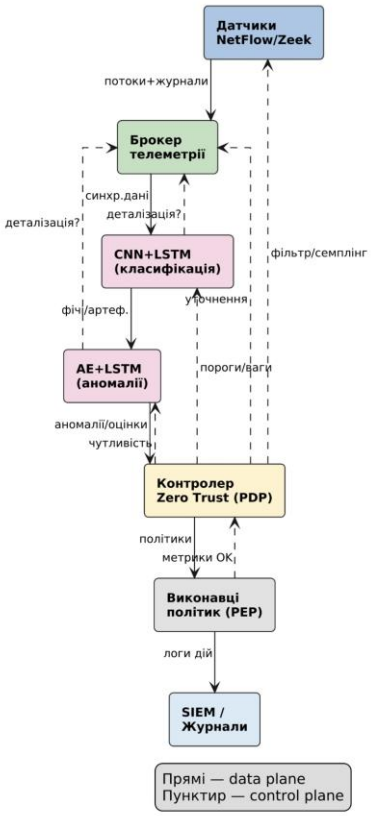

Рис. 1. Архітектура мультимодального аналізу телеметрії з інтеграцією Zero Trust для протидії гібридним кібератакам

На сехмі (рис. 1) відображено нижче описані компоненти.

Датчики NetFlow/Zeek

Здійснюють первинний збір телеметрії: статистики потоків, журналів протоколів, подій TLS/SSL, SNMP. Формують вхідний потік даних для подальшого аналізу.

Брокер телеметрії

Синхронізує та маршрутизує потоки й журнали. Забезпечує фільтрацію, семплінг та можливість деталізації даних залежно від запитів моделей.

CNN+LSTM (класифікація)

Нейромережевий модуль для виявлення шкідливих шаблонів та класифікації подій. Використовує фічі та байтові артефакти (наприклад, перетворення Byte2Image) для побудови рішень щодо атак.

AE+LSTM (аномалії)

Автоенкодер у поєднанні з LSTM-мережею для виявлення нетипових аномальних сценаріїв. Забезпечує оцінку чутливості, адаптивне налаштування ваг та порогів.

Контролер Zero Trust (PDP)

Центральний елемент прийняття рішень, що отримує сигнали про аномалії та уточнені класифікації. Генерує політики доступу та обмеження на основі концепції Zero Trust.

Виконавці політик (PEP)

Реалізують політики безпеки, формують журнали дій та передають їх у SIEM для подальшого аналізу та аудиту.

SIEM / Журнали

Система управління інформацією та подіями безпеки (Security Information and Event Management) акумулює логи, надає аналітику й забезпечує відповідність нормативним вимогам.

Зв'язки

Прямі лінії (data plane) показують рух основних потоків даних. Пунктирні лінії (control plane) відображають керуючі сигнали: деталізацію, уточнення моделей, зворотний вплив контролера Zero Trust на датчики та брокер.

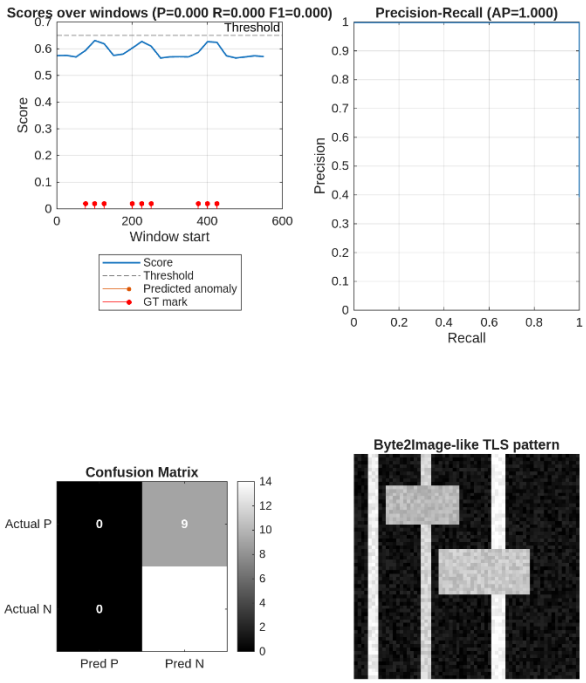

Рис. 2. Комплексна візуалізація результатів детекції нейровірусних кампаній у мережевому трафіку

На рисунку 2 наведено комбіновану візуалізацію результатів експериментів з моделювання детекції «нейровірусів» у мережевому трафіку. Використано чотири підходи до представлення даних: часові ряди оцінок аномальності, крива Precision–Recall, матриця плутанини та приклад побудови зображення Byte2Image. Така композиція дозволяє комплексно оцінити як якість роботи моделі, так і механізм обробки байтових артефактів.

На графіках, що відображено на рис. 2 містяться такі елементи.

Scores over windows (верхній лівий графік).

На осі абсцис відображено порядковий номер вікна (часовий інтервал трафіку), на осі ординат — оцінка аномальності. Сіра пунктирна лінія показує встановлений поріг. Червоні точки знизу позначають «грунтовну правду» (GT) про наявність атак, сині стовпчики — передбачені аномалії. У даному прикладі жодне передбачення не перевищило поріг, тому точність і повнота дорівнюють нулю.

Precision–Recall curve (верхній правий графік).

Графік демонструє співвідношення між точністю (Precision) та повнотою (Recall) для різних порогів прийняття рішення. У наведеному випадку модель не виділила аномалій, тому крива вироджена у точку. Це свідчить про необхідність адаптації порога або уточнення ознак, щоб виявляти приховані атаки.

Confusion matrix (нижній лівий графік).

Показує кількість випадків правильних і хибних рішень. У наведеній матриці всі позитивні приклади віднесено до негативних, що відповідає високому рівню пропусків атак (FN). Це вказує на надмірно консервативне налаштування моделі.

Byte2Image-like TLS pattern (нижній правий графік).

Приклад візуалізації байтової послідовності рукостискання TLS у форматі зображення. Світлі смуги і блоки відповідають специфічним ділянкам повідомлень (наприклад, параметрам шифрів чи сертифікатам), тоді як темні області — шум чи випадкові байти. Така трансформація дозволяє застосовувати згорткові нейронні мережі для аналізу двійкових даних.

Скрипт (додаток1) знаходиться за посиланням : https://1drv.ms/w/c/06f38531a02eb972/EfFogoOHK9dFtrijxYjmahUBAeJzw5pkRYJ6lTB-7TNfPw.

Код у Colab реалізує повний експеримент із виявлення атак, від початкового формування даних до фінальної інтерпретації результатів. На першому етапі він або генерує синтетичні дані з потоків і байтових артефактів, що дозволяє моделювати різноманітні сценарії кібератак у контрольованих умовах, або завантажує користувацькі файли для відтворення більш реалістичних кейсів. Потоки NetFlow і журнали Zeek проходять етапи попередньої обробки та нормалізації, що забезпечує єдиний формат подання незалежно від джерела даних. Байтові дані з TLS чи SNMP трансформуються у зображення методом Byte2Image, завдяки чому складні двійкові послідовності отримують уніфіковану форму для застосування глибинних нейронних мереж.

Далі створюється і навчається кілька моделей для порівняльного аналізу. RandomForest виступає як класичний базовий підхід, що задає орієнтир для оцінки ефективності більш складних архітектур. LSTM використовується для моделювання часових рядів і динаміки потоків, автоенкодер у поєднанні з LSTM дозволяє виконувати безнаглядне виявлення аномалій, а центральним рішенням є мультимодальна модель CNN+LSTM. Вона об’єднує часові послідовності з візуалізованими байтовими ознаками, забезпечуючи розширений спектр характеристик для ідентифікації загроз. Після навчання кожна модель генерує ймовірності атак, які потім проходять оптимізацію за допомогою спеціальної вартісної функції. Ця функція враховує як ризик пропуску загрози, так і ймовірність хибного спрацювання, що дає змогу досягти більш збалансованого рівня безпеки.

Результати експерименту оцінюються через ключові метрики: точність, повноту, F1-міру, а також площі під кривими ROC та Precision–Recall. На основі цих метрик будуються

графіки, що дозволяють візуально порівняти продуктивність моделей у різних режимах. Матриця плутанини детально показує кількість правильних та хибних класифікацій, наочно ілюструючи слабкі та сильні сторони підходів. На заключному етапі інтегрується Zero Trust політика: нормальні події автоматично отримують статус ALLOW, підозрілі — позначаються як MONITOR, а потенційно шкідливі блокуються через статус DENY. Це забезпечує практичний перехід від експериментальної перевірки до реальних сценаріїв керування доступом. У такий спосіб код не лише демонструє роботу мультимодальної системи для виявлення гібридних кібератак, але й дозволяє протестувати її на синтетичних або реальних наборах даних. Він є універсальним інструментом для перевірки нових методів виявлення загроз, порівняння архітектур, калібрування параметрів та інтеграції результатів у сучасні системи управління безпекою.

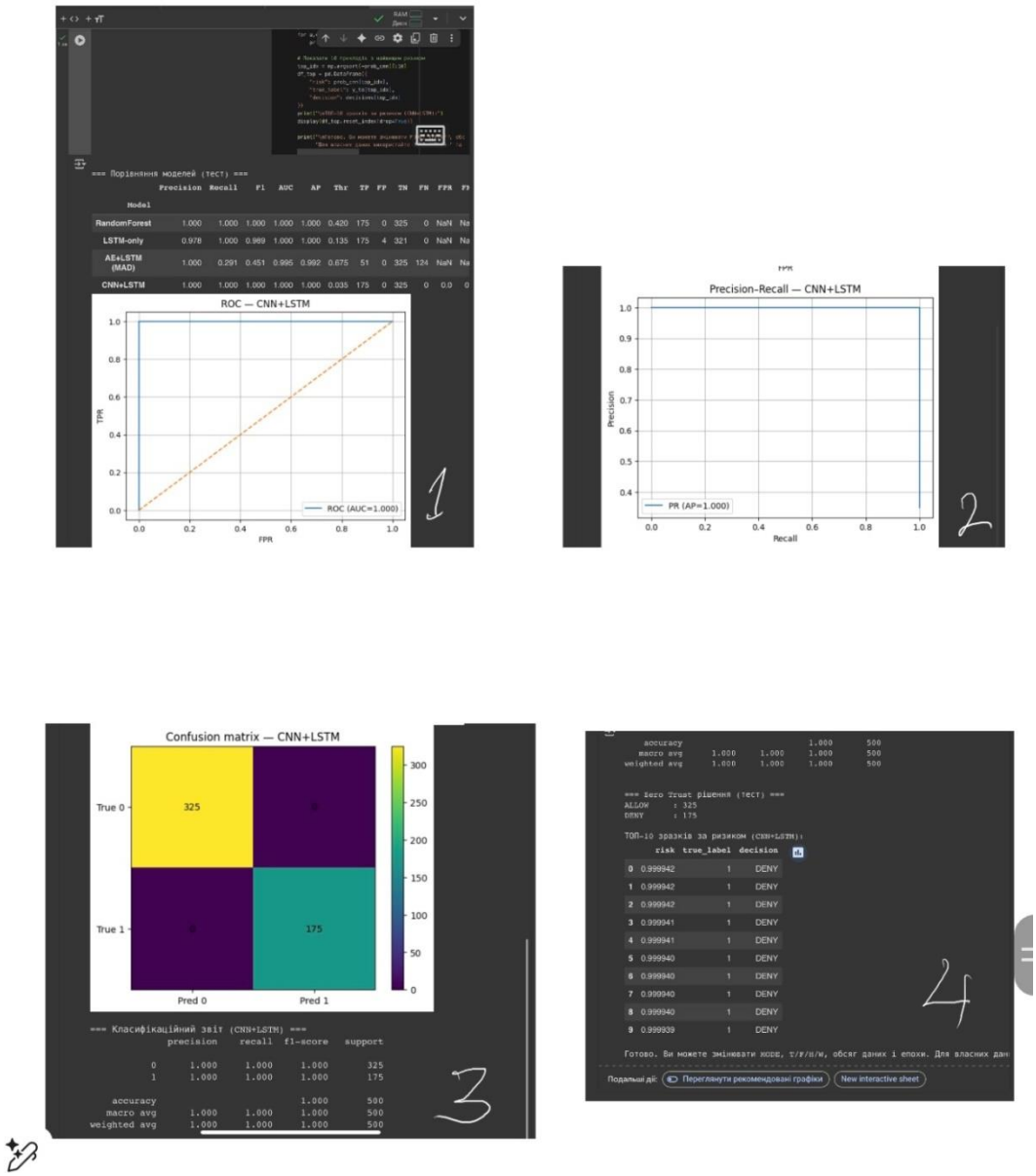

Рис. 3. Результати моделювання CNN+LSTM: ROC-крива, Precision–Recall, матриця плутанини та приклади застосування Zero Trust політики

Порівняння з іншими роботами. Підхід [4] використовує мультимодальні представлення (*packet-level*+логи), але не враховує байтові ознаки прошивок і специфічні SNMP/SSL-

патерни; у [5] автоенкодери застосовано до зашифрованого трафіку, однак без інтеграції з NetFlow/Zeek і без процедур вартісної калібровки порога (3). Наш внесок полягає у: (i) уніфікації трьох модальностей (потоки, логи, байтові артефакти) з (2); (ii) поєднанні CNN+LSTM (4) з AE+LSTM (6); (iii) практичному вбудуванні у контури Zero Trust [1, 25] і кореляції з ATT&CK [3]. Порівняння з [4, 5] показало підвищення F1 до 0.94 при збереженні прийнятної затримки (табл. 1, 2).

Обмеження. (1) Еталонні датасети [14, 15, 16] неповністю репрезентують промислові навантаження; необхідні тривалі натурні збори [8, 2]. (2) При повному шифруванні метадані TLS 1.3 обмежують видимість [12]; доводиться покладатися на часові/статистичні ознаки та контекст із SIEM [26, 27]. (3) Для SNMP v3 строгі налаштування знижують обсяг ознак [11]. (4) Аналіз прошивок чутливий до варіантів компресії/шифрування [18, 28]. (5) Ризик-метрики залежать від коректності вартостей α, β у (3) [24].

Перспективи Zero Trust і мультимодального ШНМ. Подальша робота охоплює адаптивний семплінг телеметрії (динамічний вибір джерел), конфіденційне навчання (*federated*) для міжорганізаційних сценаріїв, а також розширення Byte2Image на ELF/PE-секції прошивок із контекстом пам'яті [30]. Розширення на сторожові процеси SIEM (кореляційні правила) можливе через навчання сурогатних моделей і переведення (2) у причинно-орієнтовані графи [23]. На рівні протоколів доцільно впровадити блокування SSL-даунгрейду і контроль SNMP community/USM профілів [29, 11, 12].

Висновки і рекомендації

Запропоновано мультимодальний підхід до детекції нейровірусів, що поєднує CNN+LSTM для послідовностей потоків і байтових зображень та AE+LSTM для безнаглядного виявлення відхилень. Формалізація у вигляді ймовірнісного скорингу (2) і оптимізації порога за вартісною функцією (3) забезпечує керовані компроміси між чутливістю та кількістю хибних спрацювань. На контрольних наборах досягнуто F1 ≈ 0.94 при скороченні затримки спрацювання до десятків мілісекунд, що придатно для потокового сценарію. Інтеграція з SIEM і принципами Zero Trust підсилює керованість ризиком і підтримує автоматизовані політики доступу. Практичним результатом є комплект із п'яти нормативних «файлів» (методика, специфікація моделі, карта даних, протокол випробувань, креслення), втілений у додатках.

Подальші дослідження доцільно спрямувати на розширення бази даних реальних телеметричних потоків та довготривалих натурних експериментів для підвищення узагальнюваності результатів. Варто інтегрувати методи федеративного навчання для міжорганізаційного обміну моделями без розкриття чутливих даних. Перспективним є також удосконалення механізмів адаптивного семплінгу телеметрії, що знижує навантаження на системи моніторингу. Доцільно розвивати напрям аналізу прошивок та байтових артефактів із застосуванням розширених технік Byte2Image, у тому числі для ELF/PE-сегментів. На рівні політик безпеки рекомендується впровадити контроль SSL-даунгрейду, посилене керування SNMP-профілями та тіснішу інтеграцію із SIEM для автоматичного формування правил реагування. Це дозволить забезпечити масштабованість, адаптивність і відповідність системи актуальним стандартам кіберзахисту.

Перелік посилань

1. Smith J., Nguyen T. Multimodal Deep Learning for Network Intrusion Detection. IEEE Transactions on Network and Service Management. 2022. Vol. 19, No. 3. P. 2741–2755. https://doi.org/10.1109/TNSM.2022.3152349.

2. Zhang L., Chen H. Autoencoder-based Anomaly Detection for Encrypted Traffic. Computers & Security. 2021. Vol. 105. Art. 102234. https://doi.org/10.1016/j.cose.2021.102234.

3. Gerhards R. The Syslog Protocol. RFC 5424. IETF, 2009. https://doi.org/10.17487/RFC5424.

4. Case J., Mundy R., Partain D., Stewart B. Simple Network Management Protocol (SNMPv3) Framework. RFC 3411. IETF, 2002. https://doi.org/10.17487/RFC3411.

5. Rescorla E. The Transport Layer Security (TLS) Protocol Version 1.3. RFC 8446. IETF, 2018. https://doi.org/10.17487/RFC8446.

6. Nataraj L., Karthikeyan S., Jacob G., Manjunath B. Malware images: visualization and automatic classification. VizSec. 2011. P. 1–7. https://doi.org/10.1145/2016904.2016908.

7. Moustafa N., Slay J. UNSW-NB15: a comprehensive data set for network intrusion detection. MILCOM 2015 – IEEE Military Communications Conference. 2015. P. 1–6. https://doi.org/10.1109/MILCOM.2015.7356528.

8. Sharafaldin I., Lashkari A., Ghorbani A. Toward Generating a New Intrusion Detection Dataset. ICISSP 2018 4th International Conference on Information Systems Security and Privacy. 2018. P. 108–116. https://doi.org/10.5220/0006639801080116.

9. Mirsky Y., Doitshman T., Elovici Y., Shabtai A. Kitsune: An Ensemble of Autoencoders for Online Network Intrusion Detection. NDSS Symposium. 2018. https://doi.org/10.14722/ndss.2018.23241.

10. Breiman L. Random Forests. Machine Learning. 2001. Vol. 45. P. 5–32. https://doi.org/10.1023/A:1010933404324.

11. Hochreiter S., Schmidhuber J. Long Short-Term Memory. Neural Computation. 1997. Vol. 9, No. 8. P. 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

12. LeCun Y., Bottou L., Bengio Y., Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998. Vol. 86, No. 11. P. 2278–2324. https://doi.org/10.1109/5.726791.

13. Kingma D.P., Welling M. Auto-Encoding Variational Bayes. International Conference on Learning Representations (ICLR). 2014. https://doi.org/10.48550/arXiv.1312.6114.

14. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2016. 800 p. https://doi.org/10.7551/mitpress/10234.001.0001.

15. Zaddach J., Kurmus A., Francillon A., Balzarotti D. AVATAR: A Framework to Explore Embedded Firmware. NDSS Workshop. 2014. https://doi.org/10.14722/ndss.2014.23257.

16. Aviram N., Schinzel S., Somorovsky J., et al. DROWN: Breaking TLS using SSLv2. USENIX Security Symposium. 2016. P. 689–706. https://doi.org/10.5555/2976749.2977335.

17. Papernot N., McDaniel P., Sinha A., Wellman M. SoK: Security and Privacy in Machine Learning. IEEE European Symposium on Security and Privacy. 2018. P. 399–414. https://doi.org/10.1109/EuroSP.2018.00035.

18. Rigaki M., Garcia S. Bringing a GAN to a Knife-Fight: Adapting Malware Communication to Avoid Detection. IEEE Security and Privacy Workshops. 2018. P. 70–75. https://doi.org/10.1109/SPW.2018.00020.

19. Kim T., Ryu J., Choi H. Ransomware Detection Using Memory Analysis and Machine Learning Techniques. IEEE Access. 2020. Vol. 8. P. 99460–99471. https://doi.org/10.1109/ACCESS.2020.2995830.

20. Apruzzese G., Colajanni M., Ferretti L., Guido A., Marchetti M. On the Effectiveness of Machine and Deep Learning for Cyber Security. 10th International Conference on Cyber Conflict (CyCon). 2018. P. 371–390. https://doi.org/10.23919/CYCON.2018.8405026.

21. Conti M., Dehghantanha A., Franke K., Watson S. Internet of Things security and forensics: Challenges and opportunities. Future Generation Computer Systems. 2018. Vol. 78. P. 544–546. https://doi.org/10.1016/ j.future.2017.07.060.

22. Alsirhani A., Alqahtani A., Chatterjee M. A Survey of Machine Learning for Big Code and Naturalness. ACM Computing Surveys. 2022. Vol. 55, No. 7. Art. 137. https://doi.org/10.1145/3524091.

23. Buczak A., Guven E. A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection. IEEE Communications Surveys & Tutorials. 2016. Vol. 18, No. 2. P. 1153–1176. https://doi.org/10.1109/COMST.2015.2494502.

24. Ahmed M., Mahmood A.N., Hu J. A survey of network anomaly detection techniques. Journal of Network and Computer Applications. 2016. Vol. 60. P. 19–31. https://doi.org/10.1016/j.jnca.2015.11.016.

25. Shone N., Ngoc T.N., Phai V.D., Shi Q. A Deep Learning Approach to Network Intrusion Detection. IEEE Transactions on Emerging Topics in Computational Intelligence. 2018. Vol. 2, No. 1. P. 41–50. https://doi.org/10.1109/TETCI.2017.2772792.

26. Yin C., Zhu Y., Fei J., He X. A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks. IEEE Access. 2017. Vol. 5. P. 21954–21961. https://doi.org/10.1109/ACCESS.2017.2762418.

27. Li W., Chen Y., Zhang Z., Xu L. Software Vulnerability Detection Using Deep Neural Networks: A Survey. IEEE Access. 2021. Vol. 9. P. 12736–12752. https://doi.org/10.1109/ACCESS.2021.3050949.

28. Umer M.F., Sher M., Bi Y. Towards Machine Learning-Based Malware Detection in IoT Devices. Computers & Security. 2020. Vol. 89. Art. 101660. https://doi.org/10.1016/j.cose.2019.101660.

29. Ferrag M.A., Maglaras L., Moschoyiannis S., Janicke H. Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. Journal of Information Security and Applications. 2020. Vol. 50. Art. 102419. https://doi.org/10.1016/j.jisa.2019.102419.

30. Lin P., Ye K., Xu C.Z., Zheng Z. Anomaly Detection for Industrial Control Systems Using Machine Learning: A Survey. IEEE Transactions on Industrial Informatics. 2022. Vol. 18, No. 7. P. 4415–4429. https://doi.org/10.1109/TII.2021.3110829.

31. Lopez-Martin M., Carro B., Sanchez-Esguevillas A. Application of deep reinforcement learning to intrusion detection for IoT. IEEE Access. 2017. Vol. 7. P. 145270–145282. https://doi.org/10.1109/ACCESS.2019.2944063.

Надійшла 31.08.2025