УДК 004.8:004.056:681.518.5
DOI https://doi.org/10.32689/maup.it.2025.4.25

Андрій СУДИН
аспірант кафедри прикладної математики факультету прикладної математики та інформатики,
Львівський національний університет імені Івана Франка,
andrii.sudyn@lnu.edu.ua
***ORCID:** 0009-0006-8601-4682*

Лукаш СЦІСЛО
габілітований доктор інженерних наук,
завідувач кафедри автоматизації та комп'ютерної інженерії,
Краківський технологічний університет,
lukasz.scislo@pk.edu.pl
***ORCID:** 0000-0002-7728-9020*

Андрій ПЕРЕКРЕСТ
доктор технічних наук, професор,
завідувач кафедри комп'ютерної інженерії та електроніки,
Кременчуцький національний університет імені Михайла Остроградського,
pksg13@gmail.com
***ORCID:** 0000-0002-7728-9020*

Юрій ОНИЩЕНКО
кандидат наук з державного управління, доцент,
заступник директора інституту з освітньої та науково-дослідної діяльності
навчально-наукового інституту № 4 (підготовки фахівців з інформаційно-аналітичного забезпечення та кібербезпеки Національної поліції України),
доцент кафедри кібербезпеки та DATA-технологій,
Харківський національний університет внутрішніх справ, onischenko1980@gmail.com
ORCID: 0000-0002-7755-3071

ГІБРИДНА НЕЙРОМЕРЕЖЕВА МОДЕЛЬ ВИЯВЛЕННЯ ТА КІЛЬКІСНОЇ ОЦІНКИ РИЗИКУ ТАРГЕТОВАНИХ АТАК НА SCADA/ICS СИСТЕМИ ОБ'ЄКТІВ КРИТИЧНОЇ ІНФРАСТРУКТУРИ

Анотація. *Кількісна оцінка ризику таргетованих атак на SCADA/ICS системи об'єктів критичної інфраструктури є ключовою передумовою для формування обґрунтованих рішень щодо підвищення кіберстійкості, оптимізації механізмів захисту та мінімізації потенційних техногенних і соціально-економічних наслідків.*

Мета. *Розробка адаптивної гібридної нейромережевої моделі для виявлення й кількісної оцінки ризику таргетованих атак на SCADA/ICS системи об'єктів критичної інфраструктури.*

Методологія. *Запропонована архітектура поєднує багатомодальну інтеграцію процесних і мережевих сигналів через спеціалізовані енкодери, крос-модальний уваговий ф'юз із прототипною регуляризацією для підвищення локальної інтерпретованості та механізми обробки нерівномірної й частково відсутньої телеметрії (варіаційна автоімпутація, латентні звичайні диференціальні рівняння або трансформерні підходи з масками). Запропонований комбінований критерій детекції поєднує реконструкційну, прогнозну та контрастивну складові з адаптивними предиктивними компонентами для підвищення чутливості до «низько-повільних» сценаріїв атак. Для кількісної оцінки ризику введено калібрований ймовірнісний скор та функцію очікуваних збитків, що надало змогу формалізувати порогову політику реагування (моніторинг, ізоляція, автоматичні контрзаходи) у вигляді багаторівневої стратегії. Для забезпечення адаптивності до дрейфу та нових конфігурацій застосовано інкрементальне навчання з обмеженим буфером, MAML-подібну ініціалізацію і доменно-адверсаріальну регуляризацію, механізми XAI (внутрішня увага, прототипи, інтегровані градієнти та SHAP-подібні апроксимації) забезпечують логічні трасування причинно-наслідкових сценаріїв і підтримку судово-технічних висновків. Експериментальна валідація проведена на мульти-модальному датасеті, сформованому шляхом поєднання публічних SCADA/ICS наборів із модельованими траєкторіями та атакованими сценаріями. При цьому оцінювання включало ROC-AUC і F1-метрика для детекції аномалій, RMSE для прогнозної складової та економічно орієнтовані метрики очікуваного збитку і каліброваного ризику.*

Наукова новизна. *Розроблення адаптивної інтерпретованої нейромережевої моделі, що вперше поєднує багатомодальну інтеграцію мережевих і процесних сигналів SCADA/ICS, стійке виявлення «низько-повільних» таргетованих атак в умовах обмеженої телеметрії та формалізовану кількісну оцінку ризику з прогнозуванням наслідків для фізичних процесів критичної інфраструктури.*

Висновки. *Результати демонструють підвищену стабільність виявлення «низько-повільних» атак, кореляцію прогнозних відхилень із підвищенням RMSE у фазах атак для формалізованого вибору операційних порогів.*

Ключові слова: *SCADA/ICS, нейромережева модель, таргетована атака, регуляризація, об'єкт критичної інфраструктури.*

Andrii SUDYN, Łukasz ŚCISŁO, Andrii PEREKREST, Yurii ONYSHCHENKO. HYBRID NEURAL NETWORK MODEL FOR DETECTING AND QUANTIFYING THE TARGETED ATTACKS RISK ON SCADA/ICS SYSTEMS OF CRITICAL INFRASTRUCTURE FACILITIES

Abstract. *Quantifying the targeted attacks risk on SCADA/ICS systems of critical infrastructure facilities is a key prerequisite for making informed decisions to increase cyber resilience, optimize protection mechanisms, and minimize potential man-made and socio-economic consequences.*

Objective. *Development of the adaptive hybrid neural network model development for the detection and quantification of the targeted attacks risk on SCADA/ICS systems of critical infrastructure facilities. The proposed architecture combines multimodal integration of process and network signals through specialized encoders, cross-modal attentional fuse with prototypical regularization to increase local interpretability, and mechanisms for processing uneven and partially missing telemetry (variational auto-imputation, latent ordinary differential equations, or transformer approaches with masks). The proposed combined detection criterion combines reconstruction, prediction, and contrast components with adaptive predictive components to increase sensitivity to "low-to-slow" attack scenarios. For quantitative risk assessment, a calibrated probabilistic score and an expected loss function were introduced, which made it possible to formalize the threshold response policy (monitoring, isolation, and automatic countermeasures) as a multi-level strategy. Incremental learning with a limited buffer, MAML-like initialization, and domain-adversarial regularization were used to ensure adaptability to drift and new configurations. XAI mechanisms (internal attention, prototypes, integrated gradients, and SHAP-like approximations) provide logical tracing of cause-and-effect scenarios and support for forensic conclusions. Experimental validation was carried out on a multi-modal dataset formed by combining public SCADA/ICS sets with simulated trajectories and attacked scenarios. The evaluation included ROC-AUC and F1-metric for anomaly detection, RMSE for the predictive component, and economically oriented metrics of expected loss and calibrated risk.*

Scientific novelty. *Development of an adaptive interpreted neural network model that, for the first time, combines multimodal integration of SCADA/ICS network and process signals, robust detection of "low-slow" targeted attacks under limited telemetry conditions, and formalized quantitative risk assessment with prediction of consequences for physical processes of critical infrastructure.*

Conclusion. *The results demonstrate increased stability in detecting "low-slow" attacks and correlation of predictive deviations with increasing RMSE in attack phases for formalized selection of operational thresholds.*

Key words: *SCADA/ICS, neural network model, targeted attack, regularization, critical infrastructure object.*

Вступ. SCADA/ICS системи керування критичною інфраструктурою забезпечують безперервність життєво важливих послуг (енергетика, водопостачання, транспорт, промислові процеси) і характеризуються високою міжсекторною взаємодією, значною кількістю застарілих компонентів і адаптацією стандартних IT-протоколів до операційних мереж [5]. Інтеграція з корпоративними мережами та віддаленим доступом, а також підвищена складність ланцюгів постачання програмного забезпечення істотно розширюють поверхню атаки, що призводить до зростання частоти й витонченості таргетованих атак, спрямованих на порушення фізичних процесів і створення катастрофічних соціально-економічних наслідків [3].

Наявні підходи виявлення здебільшого орієнтовані на сигнатурні чи статистичні методи, які неадекватно працюють у разі «низько-повільні» таргетовані атаки, багатокомпонентних сценаріїв з відкладеним впливом або при недостатності телеметрії й різнорідності протоколів [6]. При цьому зазначається, що на теперішній час відсутні інтегровані механізми кількісної оцінки ризику, що поєднують детекцію аномалій із прогнозом наслідків для безпеки процесу. У зв'язку з цим виникає потреба в адаптивних, інтерпретованих моделях, здатних у режимі близькому до реального часу корелювати мережеві й технологічні індикатори, враховувати контекст інфраструктури та надавати кількісні метрики ризику для прийняття управлінських рішень. Отже, актуальність розробки нейромережевої моделі виявлення та кількісної оцінки ризику таргетованих атак на SCADA/ICS системи об'єктів критичної інфраструктури зумовлена необхідністю підвищити точність і своєчасність виявлення складних та «low-and-slow» атак, забезпечити кореляцію мережевих і технологічних індикаторів, інтегрувати прогнозування наслідків для фізичних процесів і надати кількісні метрики, що підвищують обґрунтованість управлінських рішень для мінімізації соціально-економічних втрат.

Стан дослідження проблеми. У дослідженнях, наприклад, М. А. Умер, Х. Н. Джунеджо, М. Т. Джілані, Ф. П. Матур [9], присвячених кіберзахисту SCADA/ICS, домінують теми класифікації вторгнень і виявлення аномалій із застосуванням статистичних методів, класичних алгоритмів машинного навчання та підходів на основі фізично орієнтованих детекторів. При цьому вони підкреслюють різноманітність підходів (керовані, напівкеровані та некеровані) і звертають увагу на потребу інтеграції інформації мережевого та процесного рівнів для підвищення надійності детекції.

Ключові дослідження відповідних інцидентів, зокрема, М. Ассанте і Р. Лі [2], визначають характерні вектори таргетованих атак проти критичної інфраструктури та запропонували таксономії супротивницьких кампаній (наприклад, аналіз Stuxnet та дослідження атак на енергетичну інфраструктуру [1]), що сформували базу для моделювання атакових ланцюжків і розробки процедур реагування.

В останні роки з'явилися прикладні дослідження нейромережевих рішень для SCADA/PLC, в яких широкого застосування набули автокодери, LSTM, CNN та гібридні архітектури. Результати досліджень Л. Ройтера, О. Юнга, Дж. Магін [10], а також М. Закарія, С. У. Аміна, Ф. С. Алрайс, М. Хелал і З. І. Хан [8] показали покращення точності виявлення аномалій у тестових наборах і стендах. Водночас з'являються нові перспективні підходи (наприклад, DeepFM та інші deep-learning схеми, наприклад, дослідження Д. Я. Квірумбай, Д. Фернандес Іглесіас, Ф. Новоа [7]) для роботи з високорозмірними й нелінійними ознаками мережевого та процесного трафіку. У [4] паралельно підкреслюється роль стендів і публічних датасетів для відтворюваності експериментів.

Незважаючи на прогрес, залишаються відкриті питання, для вирішення яких потрібна спеціалізована нейромережева модель виявлення та кількісної оцінки ризику таргетованих атак на SCADA/ICS:

– як забезпечити стійку детекцію «низько-повільних» атак за обмеженою та зашумленою телеметрією;

– як об'єднати мережеві й фізичні сигнали в єдиному інтерпретованому представленні та перетворити виявлені аномалії у кількісні метрики ризику з прогностичною цінністю;

– як підвищити інтерпретованість і пояснюваність нейромережевих рішень для оперативних рішень і судово-технічного аналізу;

– як адаптувати моделі до різнорідних і застарілих протоколів, обмежених датасетів і змінних конфігурацій інфраструктури;

– які процедури валідації й оцінки ризику гарантуватимуть переносимість результатів із лабораторних стендів у реальний експлуатаційний контекст.

Саме для усунення цих прогалин доцільною є розробка адаптивної, контекст-чутливої нейромережевої моделі виявлення та кількісної оцінки ризику таргетованих атак.

Метою дослідження є розробка адаптивної інтерпретованої нейромережевої моделі для виявлення та кількісної оцінки ризику таргетованих атак на SCADA/ICS системи об'єктів критичної інфраструктури, яка забезпечує стійку детекцію «низько-повільної» атаки при обмеженій і зашумленій телеметрії, інтегрує мережеві та фізичні сигнали в єдине контекст-чутливе представлення, формалізує прогнозні метрики ризику для підтримки оперативних і управлінських рішень, гарантує інтерпретованість висновків та переносимість між гетерогенними конфігураціями й обмеженими наборами даних.

Для досягнення установленої мети треба виконати такі **завдання дослідження:**

1. Розробити адаптивну гібридну нейромережеву архітектуру для виявлення аномалій і таргетованих атак у SCADA/ICS, що забезпечує стійкість до «низько-повільних» сценаріїв та ефективну роботу за обмеженої, зашумленої телеметрії.

2. Запровадити методи багатомодальної інтеграції та контекстного представлення мережевих і фізичних сигналів для кореляції індикаторів на різних рівнях системи та підвищення точності детекції.

3. Формалізувати кількісну модель ризику, що включає набір прогнозних метрик і ймовірнісних оцінок наслідків для фізичних процесів, а також реалізувати механізми інтерпретації результатів (XAI) для підтримки оперативних і судово-технічних рішень.

4. Провести експериментальну валідацію й оцінку переносимості моделі на репрезентативних наборах даних, розробивши процедури адаптації та оцінки надійності в гетерогенних конфігураціях SCADA/ICS.

Наукова новизна отриманих результатів полягає у розробленні адаптивної інтерпретованої нейромережевої моделі, що вперше поєднує багатомодальну інтеграцію мережевих і процесних сигналів SCADA/ICS, стійке виявлення «низько-повільних» таргетованих атак в умовах обмеженої телеметрії та формалізовану кількісну оцінку ризику з прогнозуванням наслідків для фізичних процесів критичної інфраструктури.

Виклад основного матеріалу. Приймається, що у дискретному часі спостереження надходять у вигляді двох потоків, до яких належать технологічні (процесні) сигнали $x_t^{(p)} \in \mathbb{R}^{d_p}$ (з показниками й вимірюваннями сенсорів, актюаторів, фізичних величин) та мережеві сигнали $x_t^{(n)} \in \mathbb{R}^{d_n}$ (пакетні ознаки, статистики потоків, логування протоколів). Наявність пропущених або обмежених вимірювань моделюється маскою спостережень $m_t \in \{0,1\}^d$ (об'єднаний розмір $d = d_p + d_n$), а шум – адитивним шумом ε_t. При цьому позначимо локальний контекст інфраструктури (конфігурація обладнання, режим роботи) як $c \in \mathcal{C}$. Атака задається вектором впливу a_t (прихований процес), а її метою є створення небажаних відхилені у фізичному стані $y_t \in \mathbb{R}^k$ (процесні величини, що відображають стан безпеки).

ISSN 2786-5460 (Print), ISSN 2786-5479 (Online)

Пропонована гібридна архітектура складається з трьох логічних блоків, до яких належать модулі обробки модальностей (процесу та мережі), блок мульти-модального ф'юзу та детекторно-прогнозний блок з механізмами інтерпретації та оцінки ризику (рис. 1).

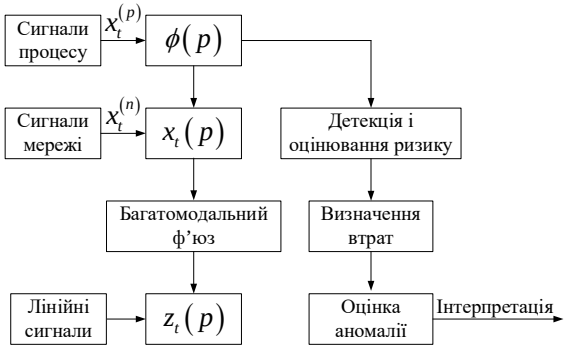

Рис. 1. Структурна схема пропонованої гібридної архітектури нейромережевої моделі

Кожний вхід проходить через спеціалізований енкодер:

$$z_t^{(p)} = \phi_p\left(x_{t-\tau:t}^{(p)}, m_{t-\tau:t}^{(p)}, \theta_p\right), z_t^{(n)} = \phi_n\left(x_{t-\tau:t}^{(n)}, m_{t-\tau:t}^{(n)}, \theta_p\right), \quad (1)$$

де $\phi_{\cdot}$ – параметризовані нейромережеві перетворення (у цьому дослідження застосована мультишарова LSTM для обробки нерівномірної телеметрії) з можливістю вбудованого врахування масок *m* для обробки пропущених вимірювань. Для стійкого виявлення «низько-повільних» атак застосовується багаторівнева тимчасова обробка, в рамках якої короткострокові та довготермінові контекстні вікна поєднуються через часові фільтри, а також через гістограми інтегрованих статистик (у цьому дослідженні застосовується статистика CUSUM) як детерміновані індикації.

Ф'юз представлень здійснюється через крос-модальний уваговий механізм як:

$$\alpha_t = \mathrm{softmax}\left(W_a\left[z_t^{(p)}; z_t^{(n)}\right] + b_a\right), \quad z_t = \mathrm{Fuse}\left(z_t^{(p)}, z_t^{(n)}\right) = \alpha_t^{(p)} \odot z_t^{(p)} + \alpha_t^{(n)} \odot z_t^{(n)}, \quad (2)$$

де ваги уваги α_t забезпечують інтерпретованість внеску кожної модальності в остаточне представлення. Для підвищення локальної інтерпретованості можна додатково використовувати структуру «прототипів» $P = \{p_j\}$ у латентному просторі і вводити прототип-регуляризатор, який примушує представлення до кластеризації навколо зрозумілих прототипів (навчання прототипів).

Пропонований комбінований підхід поєднує реконструкційний автокодер, предиктивну компоненту для фізичних величин і контрастивний компонент для підвищення роздільної здатності представлень. У цьому контексті реконструкційна складова визначається як:

$$\hat{x}_t = \psi_{\text{зап}}(z_t; \theta_{\text{зап}}), L_{\text{зап}} = \mathbb{E}[\|(x_t - \hat{x}_t) \odot m_t\|^2], \quad (3)$$

а прогнозна складова фізичних процесів – як:

$$\hat{y}_{t+1:t+H} = \psi_{\text{прогн.}}(z_t, c; \theta_{\text{прогн.}}), L_{\text{прогн.}} = \mathbb{E}\left[\sum_{\tau=1}^{H} \ell(y_{t+\tau}, \hat{y}_{t+\tau})\right], \quad (4)$$

де ℓ є квадратичною функцією помилки або негативним лог-ймовірнісним виходом (для стохастичних прогнозів). При цьому зазначається, що контрастивний елемент підвищує чутливість до *z* відмінностей між нормальними і атакувальними траєкторіями, тобто:

$$L_{конр} = -\mathbb{E}\left[\log\left(\frac{\exp\left(\frac{\mathrm{sim}\left(z_t, z_{t'}^{+}\right)}{\tau}\right)}{\sum_{z^-}\exp\left(\frac{\mathrm{sim}\left(z_t, z^-\right)}{\tau}\right)}\right)\right]. \quad (5)$$

Аномалійний скорінг поєднує реконструкційну помилку, прогнозну помилку та відхилення уваги й описується виразом:

$$s_t = \lambda_r \cdot \frac{\|(x_t - \hat{x}_t) \odot m_t\|^2}{\sigma_r} + \lambda_p \cdot \frac{\sum_{\tau=1}^{H} \ell(y_{t+\tau}, \hat{y}_{t+\tau})}{\sigma_p} + \lambda_a \cdot \mathrm{KL}(\alpha_t \| \bar{\alpha}), \quad (6)$$

де $\sigma_{\cdot}$ – нормалізуючі скалярні оцінки (наприклад, міри дисперсії на навчальній вибірці), $\bar{\alpha}$ – базова розподіл уваги у нормі, а ваги $\lambda_{\cdot}$ налаштовуються валідацією. При цьому варто наголосити на те,

що пропонована композиція чутлива до повільних, накопичуваних відхилень (через предиктивну та EWMA-компоненти) і до нетипових кореляцій між модальностями.

Для кількісної оцінки невизначеності застосовується варіаційний підхід (наприклад, MC-dropout). Нехай модель видає параметри умовного розподілу прогнозу $p_\theta(y_{t+1:t} + H \mid z_t)$ (наприклад, гаусові параметри $\mu_{t+\tau}$, $\Sigma_{t+\tau}$). При цьому ризик у конкретний момент визначається як очікуваний збиток за умови можливості атаки, помножений на оцінку імовірності наявності активної шкідливої кампанії:

$$R_t = \Pr(\mathcal{A}_t|\mathcal{D}_t) \cdot \mathbb{E}\left[L\left(y_{t+1:t+H}\right)|\mathcal{D}_t, \mathcal{A}_t \right] \tag{7}$$

де $\Pr(\mathcal{A}_t|\mathcal{D}_t)$ – ймовірність активної атаки, оцінена через відношення аномалійного скору s_t та калібрований перетворювач (наприклад, логістична калібрування або байєсова тональна модель):

$$\Pr(\mathcal{A}_t|\mathcal{D}_t) = \sigma(\gamma_0 + \gamma_1 \cdot s_t), \tag{8}$$

і очікуваний збиток визначається як інтеграл від розподілу прогнозованих фізичних величин щодо функції шкоди L, що формалізується через порогові перевищення та економічну оцінку:

$$\mathbb{E}\left[L\left(y_{t+1:t+H}\right)\right] = \int L\left(y_{t+1:t+H}\right) \cdot p_0\left(\left(y_{t+1:t+H}\right)|\mathcal{D}_t, \mathcal{A}_t\right) dy. \tag{9}$$

Для практичної реалізації пропонованої моделі застосовується оцінка через сценарну суму, яка подається як:

$$\mathbb{E}[L] = \frac{1}{M} \cdot \sum_{i=1}^{M} L\left(y_{t+1:t+H}^{(i)}\right), \tag{10}$$

де $y^{(i)}$ – вибірки з апроксимаційного розподілу (MC-семплювання або ансамблі).

Функція шкоди L формалізується як сума важкостей за перевищенням безпечних меж для кожної фізичної змінної. Отже,

$$L\left(y_{t+1:t+H}\right) = \sum_{\tau=1}^{H} \sum_{k=1}^{k} \omega_k \cdot \left[\max\left(0, y_{t+\tau}^{(k)} - T_{безп}^{k}\right)\right]^{\beta} \tag{11}$$

де $T_{безп}^{k}$ – безпечні пороги для кожної компоненти, ω_k – економічні ваги, $\beta \geq 1$ – ступінь нелінійності шкоди.

На основі викладеного відзначається, що параметри моделі оптимізуються за компромісним функціоналом:

$$L(\Theta) = \lambda_{зап.} \cdot L_{зап.} + \lambda_{прогн.} \cdot L_{прогн.} + \lambda_{контр.} \cdot L_{контр.} + \lambda_{опц.} \cdot L_{опц.} + \lambda_{пр.} \cdot L_{пр.} + \Omega(\Theta), \tag{12}$$

де $L_{опц.}$ – (опціональна) керована крос-ентропійна (або бінарна) втрата при наявності етикеток атак, $L_{пр.}$ – регуляризатор на прототипи (наприклад, відстань представлень до найближчого прототипу), а Ω – стандартні регуляризатори (у цій моделі застосовується L2-регуляризація). Окрім того, для інтерпретованості вводяться додаткові штрафи, які зменшують ентропію уваги та сприяють розрідженим поясненням:

$$L_{XAI} = \lambda_{ентр.} \cdot \sum_{t} H\left(\alpha_t\right) + \lambda_{L2} \cdot \alpha_{t1}. \tag{13}$$

Адаптивність до дрейфу та нових конфігурацій досягається за допомогою кількох механізмів, зокрема, інкрементального навчання з обмеженим буфером повторення (збереженням ключових нормальних сценаріїв), мета-навчання (MAML-подібний підхід) для швидкого перенавчання на нових конфігураціях та доменно-адверсаріальної регуляризації для зменшення залежності від специфічних датасетів. Для випадків обмеженої телеметрії застосовується варіаційна автоімпутація (наприклад, імпутер, що базується на варіаційному автоенкодері) і латентні звичайні диференціальні рівняння, що дозволяють інтерполювати нерегулярні часові точки. Формально, при новому контексті c' оновлення може виконуватись через байєсове онлайн оновлення параметрів або через адаптивний градієнтний крок на невеликому наборі локальних зразків:

$$\theta \leftarrow \theta - \eta_t \cdot \nabla_\theta \cdot L_{лок.}\left(\theta; \mathcal{D}_{лок.}\right) \tag{14}$$

де η_t – адаптивний крок (наприклад, оптимізатор ADAM зі зниженням ваг) і $\mathcal{D}_{лок.}$ – локальний буфер.

Механізми інтерпретації включають внутрішню увагу (наприклад, міжмодальна увага) з виводом ваг α_t, прототипну інтерпретацію (з поверненням найближчих прототипів та прикладів), інтегровані градієнти або SHAP-подібні апроксимації для локальної важливості ознак, а також модуль побудови причинно-наслідкових сценаріїв: на основі прогнозних семплів модель повертає ймовірнісні сценарії розвитку фізичного стану з описом ключових ознак (які сенсори, які мережеві індикатори спричинили прогнозне відхилення). Для формальної віддачі пояснення модель генерує короткий логічний «звіт-трейс» у вигляді набору тверджень щодо високої ймовірності ін'єкції команд у пристрій A, а саме, мережеві сесії з атиповою частотою (порт X), одночасне позитивне відхилення сенсора p_3 від прогнозу є більшим δ. У цьому контексті прогнозом є перевищення порогу $T^{(p_3)}$ через H кроків з ймовірністю q.

Процедура прийняття рішень та порогова політика базується на тому, що на практиці формується багаторівнева політика реагування:

1) при $R_t < R_{низьке}$ – моніторинг;
2) при $R_{низьке} \le R_t < R_{високе}$ – оперативне опитування або ізоляція потоків;
3) при $R_t \ge R_{високе}$ – автоматичні контрзаходи (ініціація аварійного відключення, перемикання на резервні режими).

При цьому зазначається, що порогові значення калібруються на основі допустимого ризику та економічних оцінок.

Таким чином, запропонована модель формально поєднує гібридне представлення з урахуванням масок і нерівномірної телеметрії (латентні звичайні диференціальні рівняння або трансформери з масками), крос-модальний уваговий ф'юз з прототипною інтерпретацією, комбінований критерій детекції, адаптований на виявлення «низько-повільних» рухів (через предиктивні та EWMA-компоненти) та формалізовану й калібровану ймовірнісну оцінку ризику, що поєднує ймовірність атаки та очікуваний фізичний збиток.

Проведено обчислювальний експеримент для валідації запропонованої гібридної нейромережевої моделі детекції та кількісної оцінки ризику таргетованих атак. Для експерименту було сформовано мульти-модальний датасет, отриманий шляхом поєднання репрезентативних публічних наборів даних SCADA/ICS (процесні телеметричні виміри та мережеві логи) та синтетичних траєкторій і атак, згенерованих на лабораторному стенді імітації технологічного процесу. Сценарії атак включали як класичні ін'єкції команд і DoS-сценарії, так і «низько-повільні» таргетовані кампанії з поступовим накопиченням впливу. Для підвищення реалістичності до даних було додано адитивний шум, випадкові пропуски вимірювань і варіанти конфігурацій мережі. Із сирих записів виконано екстракцію багатомодальних ознак (статистики пакетного трафіку, тимчасові вікна процесних змін, CUSUM-статистика, гістограми інтегрованих показників), нормалізацію, імпутацію пропусків і формування масок спостережень для навчання моделей. Дані було розбито на навчальну, валідаційну та тестову вибірки із збереженням часової послідовності (time-aware split). Також застосовано k-fold крос-валідацію для оцінки стабільності результатів при різних конфігураціях. Оцінювання моделі виконано за набором метрик: ROC-AUC і F1 для задачі класифікації аномалій, RMSE для прогнозної складової фізичних величин, а також економічно-орієнтовані метрики очікуваного збитку і калібрований ризик R_t, які використовувалися для верифікації порогових політик реагування.

Отримані результати демонструють ключові характеристики роботи запропонованого детектора та прогнозної компоненти. Зокрема, динаміка RMSE, обчислена у ковзному вікні (рис. 2), вказує на значне погіршення якості прогнозу в проміжку моделюваної атакувальної активності, що підкреслює чутливість прогнозного модуля до цільових впливів. При цьому функція очікуваних збитків як залежність від порога класифікації (рис. 3) має виражений U-подібний вигляд із однозначно визначеним мінімумом, що дозволяє вибрати оптимальний поріг мінімізації економічних витрат системи реагування.

Рис. 2. Діаграма динаміки помилки прогнозу за часом

Рис. 3. Діаграма функції очікуваних збитків від порога прийняття рішень

Відповідно до рис. 2 динаміка кореня середньоквадратичної помилки (RMSE) за часом демонструє стабільно низький рівень похибки в нормальному режимі та виразне і тривале збільшення помилки в інтервалі, ідентифікованому як період атак, що свідчить про деградацію прогнозної точності під впливом цільованих впливів і підтверджує чутливість прогнозного модуля до аномальних сценаріїв; така тимчасова структура помилок дозволяє локалізувати і корелювати порушення з відомими інцидентами, а також виправдовує використання адаптивних порогів спрацьовування для режимів підвищеного ризику. При цьому, відповідно до рис. 3, функція очікуваних збитків як залежність від порога класифікації має виражений U-подібний профіль із чітко визначеним мінімумом, що відображає класичний компроміс між витратами на пропущені інциденти (FN) та витратами на хибні тривоги (FP); положення мінімуму залежить від заданих економічних параметрів і може бути використане для формалізованого вибору операційного порога, причому оцінка чутливості оптимуму до варіації параметрів витрат та невизначеності даних є обов'язковим кроком перед впровадженням у виробничу політику реагування.

Висновки. Розроблено адаптивну гібридну нейромережеву модель, яка поєднує багатомодальну інтеграцію мережевих і процесних сигналів, крос-модальний уваговий ф'юз із прототипною регуляризацією та комбінований критерій детекції на основі реконструкційної, прогнозної та контрастивної складових. Запропонована архітектура спеціально орієнтована на виявлення «низько-повільних» таргетованих атак в умовах обмеженої й зашумленої телеметрії за рахунок багаторівневої тимчасової обробки, механізмів імпутації нерегулярних вимірювань та адаптивних стратегій перенавчання (інкрементальне оновлення, MAML-подібні підходи). Для кількісної оцінки ризику введено ймовірнісну модель, що комбінує калібрований аномалійний скор із очікуваним збитком, і формалізовано багаторівневу політику реагування на основі порогів ризику.

Експериментальна валідація на синтезованому мульти-модальному корпусі (поєднання публічних датасетів SCADA/ICS та лабораторних сценаріїв з доданим шумом і пропусками) підтвердила працездатність підходу: прогнозний модуль демонструє явне збільшення RMSE у відрізках атак, що робить його чутливим індикатором впливу на фізичні процеси; функція очікуваних збитків від порога має U-подібний профіль, що дозволяє формалізовано обирати операційний поріг мінімізації економічних втрат. При цьому застосування часового розподілу і k-fold валідації підтвердило стабільність результатів за різних конфігураціях.

Список використаних джерел:

1. Прокопович-Ткаченко Д. І., Звєрєв В. П., Козаченко І. М. Кіберзагрози та методи захисту фізичної інфраструктури промислового інтернету речей (ІІОТ). *Вчені записки ТНУ імені В.І. Вернадського. Серія: Технічні науки*. 2025. Том 36 (75), № 1. С. 218–225. doi: 10.32782/2663-5941/2025.1.2/32.

2. Assante M. J., Lee R. M. The Industrial Control System Cyber Kill Chain. 2015. 22 p. URL: https://icscsi.org/library/Documents/White_Papers/SANS%20-%20ICS%20Cyber%20Kill%20Chain.pdf

3. Cherdantseva Y., Burnap P., Nadjm-Tehrani S., Jones K. A Configurable Dependency Model of a SCADA System for Goal-Oriented Risk Assessment. *Applied Sciences*. 2022. Vol. 12, no. 10. 4880. doi: 10.3390/app12104880.

4. Ikotun A. M., Ezugwu A. E., Abualigah L., Abuhaija B., Heming J. K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*. 2023. Vol. 622. P. 178–210. doi: 10.1016/j.ins.2022.11.139.

5. Mesbah M., Elsayed M. S., Jurcut A. D., Azer M. Analysis of ICS and SCADA Systems Attacks Using Honeypots. *Future Internet*. 2023. Vol. 15, no. 7. 241. doi: 10.3390/fi15070241.

6. Okur C., Dener M. Symmetrical Resilience: Detection of Cyberattacks for SCADA Systems Used in IIoT in Big Data Environments. *Symmetry*. 2025. Vol. 17, no. 4. 480. doi: 10.3390/sym17040480.

7. Quirumbay Yagual D., Fernández Iglesias D., Nóvoa F. J. A Hybrid Deep Learning-Based Architecture for Network Traffic Anomaly Detection via EFMS-Enhanced KMeans Clustering and CNN-GRU Models. *Applied Sciences*. 2025. Vol. 15, no. 20. 10889. doi: 10.3390/app152010889.

8. Reuter L., Jung O., Magin J. Neural network based anomaly detection for SCADA systems. 23rd Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN), Paris, France, 24–27 February 2020, pp. 194–201.

9. Umer M. A., Junejo K. N., Jilani M. T., Mathur A. P. Machine learning for intrusion detection in industrial control systems: Applications, challenges, and recommendations. *International Journal of Critical Infrastructure Protection*. 2022. Vol. 38. 100516. doi: 10.1016/j.ijcip.2022.100516.

10. Zakariah M., Amin S. U., Alrayes F. S., Helal M., Khan Z. I. SCADA intrusion detection using deep factorization machines. *Scientific Reports*. 2025. Vol. 15. 39753. doi: 10.1038/s41598-025-20625-2.

Дата надходження статті: 28.11.2025
Дата прийняття статті: 10.12.2025
Опубліковано: 30.12.2025