АНАЛІЗ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ АТАК НУЛЬОВОГО ДНЯ

Вергелес А.Д., Лембей К.Д.

Кафедра інфокомунікаційної інженерії ім. В.В. Поповського,
Харківський національний університет радіоелектроніки,
Україна

E-mail: anastasiia.verheles@nure.ua ,
kyrylo.lembei@nure.ua

Abstract

Zero-day attacks pose a critical security challenge, as they exploit vulnerabilities that have not yet been documented, rendering signature-based intrusion detection systems ineffective. This work reviews modern machine-learning approaches designed to detect previously unseen attacks, with a focus on anomaly-based, deep-learning, generative, transfer-learning, and zero-shot paradigms. Behavioural models such as autoencoders and One-Class SVM demonstrate strong performance on high-volume anomalies but face difficulties with low-rate or weakly expressed attacks, while hybrid CNN–LSTM architectures offer higher accuracy at the cost of computational complexity. Generative models, particularly GAN-based frameworks, support the synthesis of rare or novel attack patterns, improving model robustness, whereas transfer learning enables effective adaptation under limited availability of labelled data. Special attention is given to Zero-Shot Learning, where classifiers infer malicious behaviour through semantic attributes rather than explicit examples, evaluated using the Z-DR metric for unseen-class detection. Overall, the analysis highlights the complementary strengths of traditional, deep, generative, and zero-shot methods and underscores the need for proactive detection systems capable of identifying emerging threats before their signatures become known.

Zero-day атаки становлять особливу загрозу, оскільки використовують вразливості, які ще не були описані чи задокументовані. Bilge і Dumitraş [1] показали, що такі атаки здатні залишатися непоміченими тривалий час (рисунок 1), що підтверджує обмеженість сигнатурних рішень і потребу у більш гнучких методах. Згідно з оглядом Guo Y. [2], саме машинне навчання стало базою сучасних підходів до виявлення невідомих типів шкідливої активності.

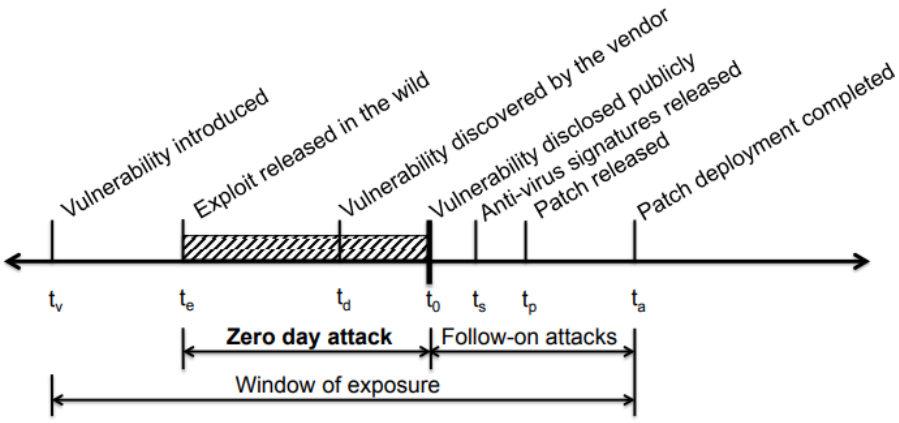

Рис. 1. Часовий проміжок середньої тривалості непомічених zero-day атак [1]

Поведінкові методи здебільшого спираються на аналіз аномалій. Автоенкодери формують модель «нормальної» поведінки мережі та сигналізують про відхилення через зростання помилки реконструкції. Як зазначається в сучасних оглядових працях з anomaly-based виявлення атак [3], такі моделі впевнено визначають масивні аномальні навантаження, включно з DDoS, але втрачають ефекти-

вність у випадку повільних low-rate атак, де зміни в трафіку мінімальні. One-Class SVM, як альтернатива, добре працює в умовах обмежених даних, проте слабко масштабується та поступається глибоким нейронним мережам при складних багатовимірних залежностях.

Моделі глибокого навчання дозволили розглядати трафік як часову послідовність. LSTM і Bi-LSTM ефективно враховують контекст подій, а CNN автоматично виділяють ключові ознаки. Дослідження, опубліковані в Scientific Reports [4], демонструють, що гібридні моделі CNN–LSTM можуть досягати дуже високої точності, хоча це потребує значних обчислювальних ресурсів. Схожу тенденцію спостерігають і автори, що використовують ансамблі LSTM, GRU та автоенкодерів: точність зростає, але впровадження в реальному часі ускладнюється ресурсомісткістю глибоких ансамблевих моделей [4].

Генеративні моделі відкрили іншу лінію досліджень. GAN-моделі [3], використовуються для генерування штучних прикладів атак, що допомагає компенсувати нестачу реальних зразків. Поєднання генеративних архітектур з автоенкодерами підсилює здатність моделей розрізняти нові або рідкісні загрози. Водночас такі підходи залежать від стабільності навчання GAN, що робить результати менш передбачуваними.

Класичні методи машинного навчання, такі як Random Forest, SVM та Decision Trees, продовжують застосовуватися завдяки простоті та інтерпретованості моделей. Проте, як зазначено в ряді робіт, дуже високі показники точності на статичних датасетах погано переносяться на реальні мережеві сценарії, де атаки більш різноманітні та менш чітко виражені, що підтверджується у порівняльних дослідженнях класичних та глибоких моделей [5].

Однією з ключових перешкод у дослідженні атак нульового дня є нестача мічених даних. У цьому контексті Transfer Learning показує значно кращі результати порівняно з повним навчанням з нуля [2], особливо коли різні типи атак мають подібні структурні властивості. Підходи на основі Manifold Alignment узгоджують структури даних між різними доменами, що дозволяє моделі працювати з новими типами атак. Інкрементальне перенесення знань поступово збагачує існуючу модель новою інформацією без її повного перенавчання. Завдяки цим підходам моделі можуть адаптуватися до нових даних, зберігаючи раніше набуті знання. Проте валідація таких підходів залишається складним завданням, що також наголошується в сучасних оглядах загроз нульового дня [3].

Особливу увагу привертає Zero-Shot Learning, де моделі працюють із семантичними описами поведінки атаки, а не з реальними зразками. У низці робіт із Zero-Shot Learning для виявлення мережевих атак [6] пропонуються метрики на кшталтзапропонували метрику Z-DR для оцінки здатності моделей ідентифікувати «невидимі» атаки. Хоча потенціал напряму є високим, ефективність значною мірою залежить від якості формування семантичних атрибутів, що все ще перебуває у фазі активних досліджень.

Систематизація переваг та обмежень основних методів машинного навчання для виявлення zero-day атак представлено в таблиці 1.

Таблиця 1. Порівняння методів

Метод	Переваги	Обмеження
1	2	3
Автоенкодери	Формують модель «нормальної» поведінки; добре визначають масивні аномалії.	Втрачають ефективність при low-rate атаках.
One-Class SVM	Добре працює за умови обмежених даних.	Погано масштабується; поступається глибоким мережам у складних задачах.
LSTM / Bi-LSTM	Враховують часовий контекст мережевих подій.	Потребують значних обчислювальних ресурсів.
Гібридні CNN-LSTM	Демонструють високий показник точності у дослідженнях.	Складне впровадження в реальному часі, потребують багато обчислювальних ресурсів.
Ансамблі LSTM/GRU/AE	Мають вищу точність порівняно з одиничними моделями.	Складне впровадження, потребують багато обчислювальних ресурсів.

Продовження таблиці 1

1	2	3
GAN	Генерація реалістичних синтетичних атак для рідкісних класів; допомагають компенсувати нестачу реальних зразків.	Потреба у великих обсягах тренувальних зразків.
Класичні ML (RF, SVM, DT)	Простота та інтерпретованість; високі результати на статичних датасетах.	Погано переносяться на реальні мережеві сценарії.
Transfer Learning	Показує кращі результати за навчання з нуля; добре працює при подібних типах атак.	Валідація перенесених моделей складна.
Zero-Shot Learning	Дозволяє працювати без зразків атак; базується на семантичних описах; оцінюється через Z-DR.	Ефективність залежить від якості семантичних атрибутів.

У висновку зазначимо, що на сьогодні не існує універсального рішення для виявлення атак нульового дня. Вибір методу визначається балансом між факторами точності виявлення, обчислювальними витратами та доступністю навчальних даних. Для високонавантажених мереж із обмеженими ресурсами доцільними є класичні методи або One-Class SVM. Критичні системи, де пріоритетом є максимальна точність, потребують гібридних глибоких архітектур. У випадку дефіциту мічених даних ефективними стають Transfer Learning та генеративні моделі. Подальші дослідження мають зосередитися на розробці ресурсоефективних гібридних рішень та покращенні стабільності навчання GAN-архітектур , що дозволить підвищити захищеність критичних систем від невідомих загроз.

Література

1. Bilge, L., Dumitraş, T. (2012), “Before We Knew It: An Empirical Study of Zero-Day Attacks in the Real World”, Proceedings of the 2012 ACM Conference on Computer and Communications Security (CCS’12), Association for Computing Machinery, New York, NY, USA, 16–18 Octobre , P. 833 – 844. DOI**:** https://doi.org/10.1145/2382196.2382284.

2. Guo, Y. (2023), “A review of machine-learning-based zero-day attack detection: Challenges and future directions”, Computer Communications, Vol. 198. P. 175 – 185. DOI**:** https://doi.org/10.1016/j.comcom.2022.11.001.

3. Mohamed, A., et al. (2025), “Emerging AI threats in cybercrime: a review of zero-day attacks via machine, deep, and federated learning”, Knowledge and Information Systems, Vol. 67. P. 10951 – 10987. DOI**:** https://doi.org/10.1007/s10115-025-02556-6.

4. Priyanshu Sinha, Dinesh Sahu, Shiv Prakash, Tiansheng Yang, Rajkumar Singh Rathore & Vivek Kumar Pandey (2025), “A high performance hybrid LSTM CNN secure architecture for IoT environments using deep learning”, Scientific Reports, Sci Rep 15, 9684 (2025). DOI**:** https://doi.org/10.1038/s41598-025-94500-5.

5. Ali, S.; Rehman, S.U.; Imran, A.; Adeem, G.; Iqbal, Z.; Kim, K.-I. (2022), “Comparative Evaluation of AI-Based Techniques for Zero-Day Attacks Detection”, Electronics 11, no. 23: 3934. DOI**:** https://doi.org/10.3390/electronics11233934.

6. Sarhan, M., Layeghy, S., Gallagher, M. et al. (2023), “From zero-shot machine learning to zero-day attack detection”, International Journal of Information Security, Vol. 22, P. 947 – 959. DOI**:** https://doi.org/10.1007/s10207-023-00676-0.