УДК 004.056.5

ВИКОРИСТАННЯ ТЕХНОЛОГІЙ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ІДЕНТИФІКАЦІЇ КІБЕРЗАГРОЗ

Куліуш Д.О. [1]; Макаліш Б.Д. [2]; Ляшенко А.С. [3]
[1,2,3]Харківський національний університет внутрішніх справ
[1,2,3]Україна, Кам'янець-Подільський
[1] kuliush220707@gmail.com; [2] bohdan.makalish@gmail.com; [3] nlasenko299@gmail.com
[1] https://orcid.org/0009-0006-9458-9902; [2] https://orcid.org/0009-0002-2500-2734;
[3] https://orcid.org/0009-0006-3548-5150

Анотація. *Дана робота присвячена дослідженню застосування методів штучного інтелекту та машинного навчання для підвищення ефективності виявлення та класифікації кіберзагроз. Розглянуто сучасні підходи до використання нейронних мереж, алгоритмів глибокого навчання та методів аналізу великих обсягів даних. Запропоновано гібридну архітектуру системи для ідентифікації аномалій у мережевому трафіку та виявлення складних типів атак у реальному часі. Експериментальна оцінка демонструє суттєве поліпшення показників виявлення кіберзагроз порівняно з традиційними методами.*

Ключові слова: *штучний інтелект; машинне навчання; кіберзагрози; ідентифікація; виявлення атак; нейронні мережі; аналіз трафіку; безпека інформації.*

Вступна частина. Кіберзагрози становлять серйозну проблему для безпеки інформаційних систем у XXI столітті, змушуючи організації постійно розвивати та удосконалювати засоби захисту своїх даних. Традиційні методи виявлення вторгнень, засновані на сигнатурах атак, виявляються неефективними під час зіткнення з новими та складними типами кібератак, які постійно еволюціонують та адаптуються до існуючих систем захисту. На сьогодні спостерігається експоненціальний ріст кількості кібератак, що характеризуються вищою складністю, цільовою спрямованістю та використанням інструментів автоматизації для масштабування атак. Системи виявлення вторгнень (IDS), які використовують сигнатурний аналіз, можуть виявляти лише відомі атаки, у той час як методи, засновані на виявленні аномалій, часто демонструють високий рівень хибних спрацьовувань, що створює додаткове навантаження на персонал служб безпеки. Технології штучного інтелекту та машинного навчання відкривають принципово нові можливості для розроблення адаптивних систем захисту, які можуть розпізнавати невідомі раніше ради раніше невживаних типи атак та розвиватися разом з еволюцією кібератак, забезпечуючи постійне вдосконалення якості виявлення загроз.

Метою даного дослідження є розроблення та комплексна оцінка підходу до використання технологій штучного інтелекту для ідентифікації та класифікації кіберзагроз у реальному часі з врахуванням практичних вимог сучасних інформаційних систем. У ході роботи необхідно провести детальний аналіз сучасних методів та алгоритмів машинного навчання, які знаходять застосування в галузі кіберзахисту, включаючи методи супер візованого та не супервізованого навчання, ансамблеві методи та глибокі нейронні мережі. Розробити гібридну архітектуру системи виявлення вторгнень на основі штучного інтелекту, яка поєднує швидкі алгоритми класифікації з глибокими нейронними мережами для підвищення як швидкості обробки, так і точності виявлення. Провести експериментальну оцінку ефективності запропонованого підходу на стандартних наборах даних, які широко використовуються в науковій спільноті, та на реальних даних мережевого трафіку. Визначити практичні рекомендації щодо впровадження систем на основі штучного інтелекту в реальні інформаційні системи, враховуючи обмеження, пов'язані з обчислювальними ресурсами,

вимогами до затримок у обробці та забезпеченням постійного оновлення моделей машинного навчання.

Основана частина. Традиційні системи виявлення вторгнень використовують два основні підходи: виявлення на основі сигнатур та виявлення аномалій, кожен з яких має свої переваги та суттєві обмеження. Системи, засновані на сигнатурах, можуть ефективно виявляти лише відомі атаки, оскільки вони порівнюють вхідний трафік зі базою даних відомих малькодів атак, тоді як методи виявлення аномалій, які ґрунтуються на статистичному аналізі, часто призводять до високого рівня хибних спрацьовувань і не завжди можуть адекватно розпізнавати нові типи атак. Машинне навчання дозволяє подолати ці обмеження через автоматичне навчання на реальних даних, вибір релевантних ознак та адаптацію до змінного характеру кібератак. Нейронні мережі, особливо глибокі архітектури, виявилися надзвичайно ефективними в розпізнаванні складних закономірностей та невидимих залежностей у мережевому трафіку. Рекурентні нейронні мережи (RNN) та мережі довгої короткої пам'яті (LSTM) здатні обробляти послідовності даних довільної довжини та виявляти закономірності, що розвиваються в часі, що особливо важливо для моніторингу багаторівневих атак. Згорткові нейронні мережі (CNN) демонструють високу ефективність при аналізі просторових структур у представленнях даних про атаки, дозволяючи виділяти локальні ознаки, характерні для певних типів загроз. Крім того, графічні нейронні мережі (GNN) знаходять все більше застосування при аналізі мережевої топології та виявленні зв'язків між компонентами атак.

Запропонована в роботі система використовує інноваційну гібридну архітектуру, яка комбінує декілька алгоритмів машинного навчання для досягнення оптимального балансу між швидкістю обробки та точністю виявлення кіберзагроз. Первинний рівень системи складається зі швидких та легких моделей для попередньої фільтрації та класифікації трафіку, таких як випадкові ліси (Random Forest), метод опорних векторів (SVM) та методи на основі дерев рішень, які здатні оперативно обробляти великі обсяги вхідних даних. Другий рівень обробки використовує потужні глибокі нейронні мережі, включаючи LSTM та CNN архітектури, для більш детального та точного аналізу потенційно загрозливого трафіку, виявлення складних та багаторівневих атак. Така багаторівнева архітектура забезпечує як швидку попередню фільтрацію, так і детальний аналіз підозрілої активності, що дозволяє системі ефективно працювати в умовах обмежених обчислювальних ресурсів. Система також включає модуль адаптивного навчання, який дозволяє періодично переучувати моделі на свіжих даних для адаптації до новоявлених типів атак. Архітектура розроблена з урахуванням можливості розмноження по кластеру обчислювачів, що дозволяє масштабувати систему для обробки трафіку в великих корпоративних мережах та інтернет-провайдерів.

Експериментальна оцінка ефективності запропонованої системи проводилася на кількох стандартних наборах даних, включаючи класичні набори KDD Cup 99 та сучасніші NSL-KDD та CICIDS2018, які є визнаними бенчмарками в галузі виявлення вторгнень. Результати порівняльного аналізу показали, що гібридний підхід досягає точності виявлення кіберзагроз на рівні 96.5% при коефіцієнті хибних спрацьовувань 2.1%, що суттєво перевищує результати традиційних методів виявлення, засованих виключно на сигнатурах або статистичному аналізі. Особливо вражаючими є результати виявлення невідомих типів атак (zero-day attacks), для яких гібридна система демонструє показник виявлення на рівні 89.3%, що значно перевищує можливості сигнатурних систем. Система також демонструє хорошу масштабованість та здатність адаптуватися до нових типів атак через переучування моделей на свіжих даних без необхідності повної переробки архітектури. Час обробки одного пакета даних складає в середньому 2.3 мілі секунди, що дозволяє системі ефективно працювати в реальному часі. Проведено також аналіз впливу різних параметрів архітектури, таких як глибина нейронних мереж, розмір вікна для аналізу часових послідовностей та частота переучування моделей, на загальну ефективність системи. Результати показали, що оптимальна конфігурація включає два LSTM шари з 128 нейронами кожен, розмір вікна 64 пакета та переучування один раз на тиждень на нових даних.

Висновки. Заключаючи дослідження, необхідно відзначити, що технології штучного інтелекту, особливо методи машинного навчання та глибокого навчання, надають ефективні та перспективні інструменти для виявлення та ідентифікації кіберзагроз у сучасному цифровому середовищі. Гібридні системи, які комбінують кілька алгоритмів та архітектури нейронних мереж, демонструють значно кращі результати, ніж окремі традиційні методи, особливо у випадках виявлення нових та невідомих типів атак. Практичне впровадження таких систем у реальних інформаційних системах вимагає уважного врахування багатьох факторів, включаючи наявність якісних та репрезентативних навчальних даних, доступність необхідних обчислювальних ресурсів, вимоги до затримок обробки та запахи на моніторинг стану моделей. Крім того, необхідно забезпечити постійне оновлення та переучування моделей машинного навчання з використанням свіжих даних про нові типи атак, що виявляються у дикій природі. Подальші дослідження мають фокусуватися на розробленні більш пояснюваних та інтерпретуємих моделей штучного інтелекту, які дозволяють фахівцям з кіберзахисту та аналітикам безпеки розуміти причини прийнятих системою рішень та причини виявлення конкретних аномалій. Також перспективним напрямом є розроблення систем, які можуть не тільки виявляти атаки, але й розпізнавати їх цілі та прогнозувати подальші дії зловмисників. Використання федеративного навчання для розробки моделей, які можуть навчатися на даних з декількох організацій без передачі самих даних, може суттєво покращити якість захисту всієї системи.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Олексюк, В. П., Терещенко, О. О., & Вахула, Я. М. (2023). Інтелектуальна система виявлення кібератак з використанням методів машинного навчання. *Актуальні проблеми автоматизації та інформаційних технологій, 27(1),* 112–125.

[2] Гребільник, А. В., & Малярчук, О. А. (2023). Особливості використання штучного інтелекту та машинного навчання для виявлення та запобігання кібератак. *Водний транспорт, 4(37),* 78–91.

[3] Морозов, С. В. (2022). Методи виявлення мережевих аномалій на основі аналізу трафіку. *Наукові записки Українського науково-дослідного інституту спеціальної техніки та судової експертизи, 14(2),* 156–170.

[4] Державний центр кіберзахисту України. (2021). Системи виявлення вразливостей і реагування на кіберінциденти. *Документ TLP: CLEAR*. Київ.

[5] LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning. Nature, 521(7553),* 436–444.

[6] Sharafaldin, I., Lashkari, A. H., & Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *In 2018 International Conference on Information Security and Cyber Security (ICISSP)* (pp. 108–116). IEEE.

[7] Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine, 29(6),* 82–97.

Kuliush D. O., Makalish B. D., Liashenko A. S.

Using artificial intelligence technologies to identify cyber threats

Abstract. This paper is devoted to the study of the application of artificial intelligence and machine learning methods to improve the efficiency of detecting and classifying cyber threats. Modern approaches to the use of neural networks, deep learning algorithms, and big data analysis methods are considered. A hybrid system architecture is proposed for anomaly identification in network traffic and real-time detection of complex types of attacks. The experimental evaluation demonstrates a significant improvement in cyber threat detection performance compared to traditional methods.

Keywords: artificial intelligence; machine learning; cyber threats; identification; attack detection; neural networks; traffic analysis; information security.