DOI: 10.25140/2411-5363-2024-1(35)-135-145
УДК 004.8:004.056

Артем Волокита[1], Микита Меленчуков[2]

[1]кандидат технічних наук, доцент кафедри обчислювальної техніки,
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» (Київ, Україна)
E-mail: artem.volokita@kpi.ua. **ORCID**: https://orcid.org/0000-0001-9069-5544

[2]аспірант кафедри обчислювальної техніки,
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» (Київ, Україна)
E-mail: melenchukov.nikita@gmail.com. **ORCID**: https://orcid.org/0009-0005-6615-4306

НЕЙРОННІ МЕРЕЖІ У ВИЯВЛЕННІ АТАК НА РОЗПОДІЛЕНІ СИСТЕМИ

Сучасні виклики до обробки великих обсягів інформації вирішуються за допомогою складних розподілених систем, які своєю чергою потребують кіберзахисту, що дозволяє керувати ризиками безпеки, такими як заволодіння інформацією, шпигунства, зниження продуктивності систем та ін. У цій статті зроблено огляд деяких засобів виявлення кібератак, які зокрема застосовують машинне навчання, наведені їхні переваги, недоліки, методи роботи, вразливості та підходи до їх захисту.

Аналіз атак проти засобів захисту на основі машинного навчання, які використовують підхід детекції аномалій, показав що існують слабкі місця, що потребують додаткового захисту, наприклад, розподілені в часі атаки можуть адаптуватись до допустимих діапазонів відхилення показників мережі. Виконано огляд механізмів забезпечення стійкості систем захисту до таких впливів, а саме додавання різноманітних шумів під час навчання, зменшення інтервалів значень параметрів системи, варіації донавчання моделі на оманливих даних, використання спеціальних класифікаторів.

Ключові слова: *системи виявлення атак; нейронні мережі; розподілені системи; детекція аномалій; машинне навчання.*

Табл.: 5. Бібл.: 33.

Актуальність теми дослідження. З постійним технологічним розвитком людства стають дедалі більше поширенішими розподілені системи, що здатні обробляти великі об'єми даних та виконувати складні задачі. Разом із цим розвиваються і методи атак на ці системи, метою яких може бути заволодіння даними, захоплення контролю над системою, вплив на продуктивність системи тощо. У цій статті зроблено огляд деяких засобів виявлення кібератак, які зокрема застосовують машинне навчання, наведені їхні переваги, недоліки й методи роботи.

Постановка проблеми. Атаки на розподіленні системи здатні заподіяти великої шкоди, наприклад, заволодіння конфіденційною інформацією, отримання доступу до керування, зниження ефективності роботи системи. Застосування систем виявлення атак (СВА) є ефективним підходом для детекції атак на розподілені системи. У зв'язку з різноманітністю засобів захисту за швидкодією, точністю, варіацією типів атак, методами роботи із вхідними даними та різноманітністю цих даних, виникає необхідність в огляді наявних видів засобів виявлення атак, їхніх слабких та сильних сторін, способів роботи.

Аналіз останніх досліджень і публікацій. Системам виявлення атак присвячені статті [1]-[18]. У статті [1] класифіковано такі системи за методами та підходами детекції і розглянуто утиліти з відкритим кодом Snort та ClamAV з метою порівняння їх принципів роботи. Автори статті [2] розглянули застосування класифікатора на основі SVM та k-середніх з метою покращення точності виявлення кібератак та зниження кількості неправильних позитивних спрацювань тривог. Було оглянуто можливі види атак, яким допомагає запобігти обране рішення, способи пришвидшити перевірку трафіку, та проаналізовано отримані результати, описано переваги та недоліки застосованого методу. У статті [3] розглянуто виявлення аномальної поведінки у розподіленій системі за допомогою використання згорткових нейронних мереж (CNN) та Random Forest. Автори пропонують свій алгоритм TR-IDS для попередньої підготовки, обробки та аналізу даних що базується на моніторингу даних інтернет пакетів та їх заголовків. Також надано схему розробленої Text-CNN моделі що використовується із метою виокремлення ознак для подальшої класифікації трафіку за допомогою Random Forest. В межах статті [4] розглянуто використання CNN моделі для виявлення атак на основі даних системи (логи). Навчання

моделі було виконано на базі датасетів NGIDS-DS та ADFA-LD. Основа частина статті присвячена питанню аналізу дуже великого обсягу інформації з метою виявлення атаки, опису критеріїв відбору даних для перевірки згортковою мережею, проєктуванню самої моделі CNN та подальшого аналізу отриманих результатів.

Виділення недосліджених частин загальної проблеми. У результаті аналізу досліджень і публікацій показано, що існує необхідність об'єднання даних різних систем виявлення атак, які використовуються на різних шарах кіберзахисту. Є потреба в дослідженні доцільності використання моделей машинного навчання, нейронних мереж, базуючись на доступних вхідних даних і вимогах до точності та швидкодії.

Мета дослідження. Метою роботи є огляд систем виявлення атак, порівняння переваг та недоліків методів детекції атак на різних рівнях розподілених систем, а також аналіз підходів до використання різноманітних моделей машинного навчання для покращення точності та ефективності ідентифікації кібератак.

Виклад основного матеріалу. Система виявлення атак (intrusion detection system, IDS) – це програмна чи апаратна система що націлена на виявлення кібератак з метою підтримки безпеки комп'ютерної системи [1]. Завдяки різним підходам до моніторингу IDS дозволяє виявляти шкідливий трафік чи використання комп'ютерних ресурсів, щоб ідентифікувати атаку. У сучасному світі для захисту розподілених мереж застосовують комплексні підходи що поєднують у собі застосування шифрування, управління ідентифікацією та доступом, моніторингу та детекції, наборів інструментів, що допомагають відповідати стандартам безпеки.

Для впровадження детекції атак на розподілені системи із використанням штучного інтелекту аналізують різні параметри таких систем. Це можуть бути певні параметри в пакетах даних, тому при аналізі заголовків, протоколів, кодувань, часу відправлення, відправника або отримувача застосовують метод опорних векторів (SVM), метод k-середніх, алгоритм нечітких С-середніх [2]. Таку модель автори називають багатоядерним алгоритмом k-середніх із неповним ядром (MKKM-IC). Запропоноване рішення було протестовано на трьох датасетах – NSL-KDD, UNSW, AWID. У цих датасетах дані містять атрибути протоколів пакетів що передавались у мережі. На їх основі моделі нейронних мереж навчають виявляти DDoS атаки, ін'єкції, заволодіння доступом до адміністрування, імітації. Порівнюючи точності (precision) виявлення кібератак на розподілену систему найкраще себе показали моделі MKKM-IC та Змішана Модель Гаусса(GMM) із результатами 71-88 %, а найгірше алгоритм пікових щільностей та k-середніх, із точністю 55-72 %. Результати експериментів статті [2] відображено у таблиці 1.

Таблиця 1 – Порівняння точності виявлення атак нейронними мережами на різних датасетах

Алгоритм	Precision(%) для NSL-KDD датасету	Precision(%) для UNSW датасету	Precision(%) для AWID датасету
MKKM-IC	81,65	77,27	88,24
Density peaks	68,18	60,18	72,73
Алгоритм k-середніх	59,13	55,56	72,16
Змішана модель Гаусса(GMM)	76,19	71,17	85,71

Джерело: таблиця складена на основі даних [2].

У статті Мін та інших [3] для аналізу пакетів даних застосовують глибокі нейронні мережі, наприклад CNN, RNN, LSTM. Нейромережею CNN аналізувались дані, отримані поєднанням вмісту пакетів, та виокремлювались ознаки на основі яких, за допомогою моделі random-forest, виявлялись атаки. Точність (accuracy) виявлення атак такого підходу досягає 99,13 %.

У іншому дослідженні виявлення атак на цьому рівні, автори [5] пропонують застосовувати для навчання датасети, що містять дані HTTP сесій – KF-ISAC, CSIC-2010, CICIDS2017. Використовують комбінацію довгої короткочасної пам'яті (LSTM) і згорткової мережі (CNN), CNN-LSTN та окремо глибоку нейронну мережу (DNN). Поєднання таких

підходів дозволило виявляти кібератаки із точністю (precision) 92,49 % для CNN-LSTM, 88,32 % для LSTM-CNN, та 93 % для DNN. Автори наголошують на необхідності постійного продовження донавчання моделі на реальному трафіку для зниження рівня помилкових позитивних ідентифікацій загроз. До недоліків такої системи слід віднести проблемність роботи із зашифрованими даними, аналіз персональних даних без деперсоніфікації, необхідність обробки великих обсягів інформації, що значно впливає на продуктивність.

При моніторингу потокових атак використовують як SVM, random forest так і глибокі нейронні мережі. Все залежить від способу обробки вхідних даних. Для перевірки моделей використовують такі датасети як KDD99 чи NSL-KDD. Виокремлення трафіку у групи дозволяє покращити рівень виявлення [5]. Потокові дані складно одразу використовувати у нейронній мережі, тому застосовують різні методи для підготовки та виокремлення даних за ознаками. Розробники NSL-KDD пропонують декілька підходів до тестування СВА – на повному датасеті чи на датасеті без дій що зустрічаються дуже часто. Такий відбір можна робити і за допомогою SVM, дерева рішень, наївного алгоритму Баєса чи за допомогою K-середніх [6]. При виборі нейронних мереж важливим є не тільки точність виявлення атак, а і швидкодія моделі, яка іноді є критичною. Автори статті [7] для навчання пропонують використовувати датасети CICIDS-2017 та CSE-CICIDS2018 що складаються із перехоплених пакетів із мережевого трафіку. Для подальшої обробки інформації у контексті мережевих потоків, виконується виділення ознак, за якими можна аналізувати саме потокову діяльність (таблиця 2).

Таблиця 2 – Виділення ознак потоку для подальшого аналізу нейронними мережами

Ознаки	**Трактування**
Fl-dur	Тривалість потоку
Fl-iat-max	Максимальний час між потоками
Tot-fw-pk	Об’єднання пакетів у напрямку передачі
Tot-l-fw-pkt	Загальний розмір пакета що відправляється
Tot-bw-pk	Загальна кількість пакетів що приймаються
Fw-pkt-1-min	Найменший розмір пакета що відправляється
Fw-pkt-1-avg	Середній розмір пакета що відправляється
Fw-iat-min	Довжина найменшого проміжку часу між надсиланням двох пакетів
Bw-iat-tot	Загальний час прийняття всіх пакетів
Bw-iat-avg	Середній час прийняття пакетів
Bw-iat-std	Середній час надсилання на прийняття двох послідовних пакетів
Bw-iat-max	Найбільший період очікування між прийняттям двох пакетів
Bw-iat-min	Найменший період очікування між прийняттям двох пакетів

Джерело: дані взяті на основі інформації викладеної у статті [7].

Наступним кроком після виділення ознак потоку, порівнюються точності(accuracy) моделей DBN (95 %), DNN (90,25 %), LSTM (96,2 %), CNN (96 %) та власного рішення HCRNN (97,75 %) утвореного поєднанням CNN, RNN та DL. Автори статті наголошують на необхідності для нейронної мережі бути пристосованою до обробки великих обсягів інформації та важливому впливу попередньої обробки інформації з метою виокремлення ознак на кінцевий рівень точності виявлення загроз.

Використання правил для виявлення атак на систему може давати велику кількість помилкових тривог на відсутні атаки. Поліпшити цю ситуацію та знизити рівень неправильних спрацювань СВА можна за допомогою комбінування правил та нейронних мереж [8]. Досягнути таких результатів можна застосовуючи KNN, CNN, DNN. Також застосовують поєднання детекції на основі логів та аналізу системних викликів за допомогою CNN моделі [4].

Огляд сучасних новітніх досягнень доступних на ринку. Сучасні популярні рішення для роботи із розподіленими системами пропонують власні набори інструментів для виявлення атак. Пропозиції для моніторингу та детекції таких компаній як Oracle, AWS, Azure, DigitalOcean, Cisco, IBM мають схожості та особливості. СВА що продаються

цими компаніями (табл. 3), являють собою нашарування декількох різних систем які забезпечують спостереження і виявлення на різних рівнях. Це може бути моніторинг на основі керування реакцією на підозрілі події чи інциденти, або ж аналіз певних груп трафіку.

Таблиця 3 – Засоби для моніторингу та детекції атак на розподілені системи

Компанія	**Назва продукту**	**Особливості**	**Спільні можливості**
Cisco	Cisco Stealthwatch, Cisco Identity Services Engine, Cisco Umbrella, Cisco Threat Response, Cisco Advanced Malware Protection, Cisco Talos Intelligence Group	Виявлення загроз для кінцевих пристроїв	Система моніторингу
IBM	IBM QRadar, IBM Resilient Incident Response Platform, IBM Cloud Pak for Security, IBM X-Force Threat Management, IBM Trusteer, IBM Cloud Activity Tracker, IBM Security MaaS360	Керування реакціями на події та інциденти, обмін інформацією про загрози	
Oracle	OCI Monitoring, OCI Logging, OCI Security Monitoring and Analytics	Детекція загроз для хмарних середовищ	
Amazon	Amazon CloudWatch, AWS X-Ray, Amazon CloudTrail, Amazon GuardDuty, Amazon Inspector, AWS Security Hub, Amazon Macie	Система аналізу та виявлення потенційних проблем у безпеці. Система виявлення шахрайства.	
Azure	Azure Monitor, Azure Security Center, Azure Sentinel, Azure Network Watcher, Azure Resource Graph, Azure Service Health	Система управління подіями та інцидентами, рекомендації з безпеки для хмарних середовищ	
DigitalOcean	DigitalOcean Monitoring and Alerts	Система аудиту хмарних середовищ.	

Джерело: розроблено авторами.

Основні відмінності між наявними рішеннями компаній що розглядаються, полягають у різних цілях на яких фокусуються системи захисту, для Cisco це мережева безпека, для Oracle – безпека баз даних, IBM спеціалізується на рішеннях що аналізують та обробляють дані за допомогою штучного інтелекту. Azure та AWS пропонують широкий спектр різноманітних служб для покращення захищеності системи. Крім цього рішення кожної з вищезгаданих компаній націлене на використання всередині екосистеми та мають різний рівень складності інтеграції зі сторонніми продуктами.

Система що пропонується компанією IBM так і називається IDS. Існує можливість написання власних правил задаючи порогові значення параметрів для виявлення таких небажаних втручань як атака отруєння адреси, перенаправлення ICMP повідомлення, атака пошкодженими пакетами, атаку ping-of-death, TCP ACK storm та інші [9]. Навіть якщо не створити власні правила для виявлення атак, IBM IDS містить попередній набір правил для детекції поширених видів кібератак. IBM використовує машинне навчання для виявлення аномалій що допомагає помітити атаку на систему.

Microsoft Azure дозволяє збирати такі дані про роботу системи як логи про налаштування і трафік системи, записувати інтернет-трафік [10]. Подальший аналіз з метою виявити атаку можна робити вбудованим Azure Firewall у комбінації із Threat Intelligence використовуючи базу шкідливих IP та доменів. Для детекції атак за допомогою машинного навчання Azure пропонує використовувати детектор аномалій. Його можна налаштувати для спостереження за динамікою зміни від однієї до 300 змінних. Такий ефект досягається застосуванням графової нейронної мережі(GNN) із шаром уваги. Крім цього Azure дозволяє використовувати сторонні рішення IDS для підвищення рівня безпеки системи.

Amazon AWS пропонує використовувати їх рішення Amazon GuardDuty [11] що поєднує у собі виявлення аномалій, моніторинг мережі, ідентифікація шкідливих файлів. Для визначення атаки сервіс використовує аналіз логів системи, DNS, віртуальної приватної хмари(VPC), звертає увагу на атипічну геолокацію чи час активності, незвичні виклики API. Amazon GuardDuty виявляє шкідливий трафік ідентифікуючи аномалії за допомогою машинного навчання (AIDS) а також використовуючи шаблони для виявлення атак (SIDS).

Рішення також дозволяє користувачу створювати власні правила для виділення кібератак і може бути поєднаним з іншими наявними рішеннями з AWS marketplace.

Своїм користувачам Digital Ocean рекомендує використовувати Suricata [12] – високопродуктивну, систему для аналізу мережі з відкритим сирцевим кодом. При налаштуванні Suricata користувач має можливість налаштувати шаблони за якими буде виконувати виявлення атак. Ці шаблони складаються з опису бажаної дії яку треба виконати у випадку коли шаблон підійшов, а також з таких ознаки як хост, IP адреса, порт, протокол, напрямок трафіку, регулярні висловлювання для детекції пакетів по вмісту. Також доступний стандартний набір правил для виявлення підозрілого трафіку по шаблону.

У Oracle, Intrusion Detection System входить до Oracle Session Border Controller (SBC) [13] і дозволяє ідентифікувати кібератаки за допомогою виявлення аномалій (AIDS). Детекція атак виконується зважаючи на оцінку надійності вузлів, кількість спрацьованих тривог на проміжку часу, цілісності пакета, кількості активних сесій, середньої тривалості зв'язку. Крім цього, налаштовуються діапазони допустимих значень параметрів сесії, наприклад значення максимальної вхідної/вихідної швидкості, мінімальний процент успішних відповідей, час до відновлення сесії у випадку її призупинення через порушення правил.

Компанія Cisco пропонує своїм клієнтам різні продукти для захисту свої розподілених систем що відрізняються один від одного своєю складністю, та підходом до роботи. Так, Cisco Umbrella поєднує у собі фаєрвол, моніторинг вразливих застосунків, захист на рівні DNS, можливість збирати логи та налаштовувати правила обробки трафіку. Іншим рішенням компанії є Cisco XDR. Його перевагою є відносно легке впровадження у порівнянні з іншими IDS рішеннями, а також можливість захистити систему одразу на 4-х рівнях – на рівні мережі, користувача чи кінцевої точки, хмари та на рівні застосунків і ідентифікації. Можливість збирати логи із декількох рівнів системи, та обробляти їх разом, зіставляючи події один з одним, дозволяє краще оцінювати процеси що відбуваються у мережі. Такі великі обсяги інформації Cisco XDR аналізує за допомогою машинного навчання.

Отже, СВА мають можливість аналізувати різні ознаки. Для поліпшення роботи СВА використовують машинне навчання. Атаки на розподілені системи виконуються на різних рівнях, тому для їх виявлення необхідно аналізувати різні дані для обробки яких оптимальними є різні моделі нейронних мереж. Здебільшого (IBM, Azure, AWS, Oracle) при виявленні атак за допомогою машинного навчання застосовують детекцію аномалій у системах. Такий підхід має свої переваги та недоліки (табл. 4).

Таблиця 4 – Переваги та недоліки підходу до детекції атак на розподілені системи на основі виявлення аномалій

Переваги підходу детекції аномалій	**Недоліки підходу детекції аномалій**
Завдяки порівнянню повсякденної роботи системи із незвичайними відхиленнями, дозволяє виявляти найновіші види атак, включаючи атаки що використовують вразливості нульового дня [14]	Вимогливість до обчислювальних ресурсів
Здатність до розширення [15]	Необхідність великого об'єму даних для забезпечення достатнього рівня детекції
Здатність до обробки великих обсягів даних [16]	Наявність як позитивних, так і негативних помилкових виявлень втручань
Можливість виявлення атак у реальному часі	Питання конфіденційності даних і етики
Можливість проведення постійного донавчання та покращення рівня детекції	Вразливість до змагального типу атак на систему. До них відносять атаки отруєння та атаки ухилення
Низький рівень помилкових позитивних виявлень атак із можливістю його подальшого зниження	
Здатність виявляти втручання маючи обмежений або неповний об'єм даних, за відсутності інформації про структуру мережі [15]	
Висока точність(accuracy) виявлення втручань [7], [14]-[16], [17]	

Джерело: розроблено авторами.

Попри те, що метод виявлення аномалій є найпоширенішим серед застосунків що використовують машинне навчання у системах СВА, цей підхід є вразливим до атак змагального типу, що охоплюють застосування оманливих, отруєних даних. Пошук нових варіацій вторгнень що належать до цього типу, та захистів від них постійно продовжується на сьогоднішній час. Перелік деяких рішень, разом зі статтями, у яких вони описані, наведений у таблиці 5.

Таблиця 5 – Наявні види атак змагального типу що застосовуються проти моделей що працюють на основі виявлення аномалій, та наявні способи захисту від них

Атаки змагального типу	**Спосіб адаптації аномальних детекторів до такого виду атак**
Атака за допомогою мапи помітності на основі якобіана (JSMA) [18]. При її здійсненні складається мапа чутливості вхідних параметрів, за допомогою якої визначається зміна яких параметрів дозволить змінити кінцеву класифікацію даних	Звуження допустимих діапазонів "безпечних" значень які можуть набувати параметри [19]
	Донавчання моделей на датасети що містять дані отримані при виконанні змагальних атак [20]
	Впровадження методу "Дистиляції", при якому відбувається розділення початкової нейронної мережі на дві, при такому підході друга мережа приймає ймовірнісний вектор класифікації та робить подальший висновок чи відбувається атака [21]
Метод швидкого градієнта (FGSM) - атаки при яких максимізується значення функції втрат із метою зміни результату класифікації. Для прикладу додавши до зображення панди непомітні для людського ока шуми, можна отримати нове зображення панди яке буде класифікуватись як зображення гібона [22]	Використання шару маскування градієнта(GCM) який буде модифікувати дані що проходять крізь нього, тим самим приховуючи градієнт [23]
	Донавчання моделей на датасети що містять дані отримані при виконанні змагальних атак [22]
	Використання автокодувальників для забезпечення виявлення оманливих даних без залучення окремих датасетів із даними змагальних атак [24]
DeepFool - для атаки ітеративним підходом знаходиться напрямок найменших змін вхідних даних з метою зміни кінцевої класифікації [25]	Використання Minimax захисту що застосовує генератор оманливих даних (GAN) та додатковий класифікатор для визначення підробок [26]
	Застосування аналізу Фур'є на вхідних зображеннях та мапах параметрів [27]
	Додавання шуму [28]
Атака Carlini & Wagner-a у якій знаходяться найменші необхідні модифікації для зміни класифікації моделі враховуючи умови обмеження змін вхідних параметрів класифікатора [29]	Стиснення вхідних даних, таке як розмиття зображення, зменшення глибини кольорів [30]
	Застосування генератора оманливих даних (GAN) для підготовки моделі до виявлення оманливих даних [31]
	Створення надійних та стійких класифікаторів за допомогою глибокого навчання на основі k-найближчих сусідів (DkNN) [32], [33]

Джерело: розроблено авторами.

Висновки. Огляд наявних експериментів із застосування ШІ з метою покращення точності детекції атаки за допомогою СВА, показав які методи виявлення вторгнень є ефективними, яка проблематика цієї теми, та у якому напрямку варто вести дослідження.

Застосування СВА у різних шарах системи, дозволяє виявляти такі атаки як SQL вставка, отримання прав адміністратора, отримання доступу до вузла, DDoS, Probe. Це можна зробити аналізуючи логи, або ж мережеві пакети у випадку SQL вставки, чи застосовуючи моніторинг інтернет-трафіку при DDoS атаці. З урахуванням тенденції зростання кількості видів атак на розподілені системи та збільшення обсягів інформації що обробляється, використання машинного навчання у СВА стає дедалі привабливішим. Штучний інтелект можна залучати до відбору необхідних ознак для аналізу, виявлення втручань, зниження рівня неправильних детекцій атак, покращення точності їх розпізнавання. Можливість донавчати моделі на нових даних є суттєвою перевагою через малу кількість наявних тестових датасетів на тему атак на розподілені системи.

Визначено що детекція аномалій є найпоширенішим підходом до виявлення атак із використанням машинного навчання на ринку комерційних продуктів. Аналіз ефективних атак проти моделей що базуються на детекції аномалій, показав де у працюючих на основі такого ж принципу СВА, доцільно шукати слабкі місця, як можна зробити системи виявлення безпечнішими. Підвищити успішність втручання у розподілену систему можна за допомогою знаходження допустимих діапазонів відхилення показників мережі та адаптувавши атаку до них. Дізнатись який є простір для змін поведінки користувача аби не бути класифікованим як загроза можна за допомогою використання методів JSMA, FGSM, DeepFool, Carlini & Wagner-a. Зробити СВА що функціонує на основі детекції аномалій стійкою до таких впливів, можна по-різному. Наприклад додаванням різноманітних шумів під час навчання, зменшенням інтервалів допустимих значень параметрів системи, варіаціями донавчання СВА на оманливих даних, використовуванням спеціальних класифікаторів таких як DkNN.

Зважаючи на активний розвиток теми атак на системи аномального виявлення (наведені методи втручання та захист від них описані у 2015-2023 роках), обраний напрямок є перспективним для подальших досліджень. Це може бути інтеграція у СВА алгоритмів захисту від атак змагального типу які застосовуються у моделях що розпізнають зображення. Також варто спробувати оптимізувати швидкодію та ефективність втручань і їх запобіжників, наприклад реалізація атаки Carlini & Wagner-a вимагає суттєвих додаткових обчислювальних можливостей для її здійснення, а різноманітні варіації додавання шумів [23], [28], [30] не є ефективними проти всіх видів атак змагального типу.

Список використаних джерел

1. Intrusion detection system: A comprehensive review / H.-J. Liao, C.-H. R. Lin, Y.-C. Lin, K.-Y. Tung // Journal of Network and Computer Applications. – 2013. – Т. 36, № 1. – С. 16-24. DOI: https://doi.org/10.1016/j.jnca.2012.09.004.

2. A multiple-kernel clustering based intrusion detection scheme for 5G and IoT networks / N. Hu, Z. Tian, X. Du, H. Lu // International Journal of Machine Learning and Cybernetics. – 2021. – № 12(11). DOI: https://doi.org/10.1007/s13042-020-01253-w.

3. TR-IDS: Anomaly-Based Intrusion Detection through Text-Convolutional Neural Network and Random Forest / E. Min, J. Long, Q. Liu, J. Cui, W. Chen // Security and Communication Networks. – 2018. – С. 1-9. DOI: https://doi.org/10.1155/2018/4943509.

4. Tran, N. N. An Approach for Host-Based Intrusion Detection System Design Using Convolutional Neural Network / N. N. Tran, R. Sarker, J. Hu // Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. – Cham, 2018. – Pp. 116-126. DOI: https://doi.org/10.1007/978-3-319-90775-8_10.

5. Kim A. AI-IDS: Application of Deep Learning to Real-Time Web Intrusion Detection / A. Kim, M. Park, D. Hoon Lee // IEEE Access. – 2020. – Vol. 8. – Pp. 70245-70261. DOI: https://doi.org/10.1109/access.2020.2986882.

6. Intrusion detection based on K-Means clustering and Naïve Bayes classification / Z. Muda, W. Yassin, M. N. Sulaiman, N. I. Udzir // 2011 7th International Conference on IT in Asia (CITA) (Sarawak, Malaysia, 12-13. 07. 2011 p.). – 2011. DOI: https://doi.org/10.1109/cita.2011.5999520.

7. Khan M. A. HCRNNIDS: Hybrid Convolutional Recurrent Neural Network-Based Network Intrusion Detection System / M. A. Khan // Processes. – 2021. – Vol. 9, № 5. – Pp. 834. DOI: https://doi.org/10.3390/pr9050834.

8. Use of machine learning in big data analytics for insider threat detection / M. Jay Mayhew, M. Atighetchi, A. Adler, R. Greenstadt // MILCOM 2015 - 2015 IEEE Military Communications Conference (Tampa, FL, USA, 26-28 October 2015). – 2015. DOI: https://doi.org/10.1109/milcom.2015.7357562.

9. IBM. i Version 7.2 Security Intrusion detection [Electronic resource]. – Access mode: https://www.ibm.com/docs/en/ssw_ibm_i_72/rzaub/rzaubpdf.pdf.

10. Baldwin M. Security Control: Network Security [Electronic resource]. – Access mode: https://learn.microsoft.com/en-us/security/benchmark/azure/security-control-network-security.

11. Amazon. Amazon GuardDuty features [Electronic resource]. – Access mode: https://aws.amazon.com/guardduty/features/.

12. Camisso J. Understanding Suricata Signatures [Electronic resource]. – Access mode: https://www.digitalocean.com/community/tutorials/understanding-suricata-signatures.

13. Oracle. Oracle intrusion detection system [Electronic resource]. – Access mode: https://docs.oracle.com/cd/E95618_01/html/sbc_scz810_security/GUID-73A32803-097C-496F-9709-7B51CF54382B.htm#Intrusion-Detection-System.

14. Utilising Deep Learning Techniques for Effective Zero-Day Attack Detection / H. Hanan, R. Atkinson, C. Tachtatzis, J.-N.Colin, E. Bayne, X. Bellekens // Electronics. – 2020. – Vol. 9, № 10. – C. 1684. DOI: https://doi.org/10.3390/electronics9101684.

15. Network intrusion detection system for DDoS attacks in ICS using deep autoencoders / I. Ortega-Fernandez, M. Sestelo, J. C. Burguillo, C. Piñón-Blanco // Wireless Networks. – 2023. – January. DOI: https://doi.org/10.1007/s11276-022-03214-3.

16. Intrusion detection model using machine learning algorithm on Big Data environment / S. M. Othman, F. M. Ba-Alwi, N. T. Alsohybe, A. Y. Al-Hashida // Journal of Big Data. – 2018. – Vol. 5, № 1. DOI: https://doi.org/10.1186/s40537-018-0145-4.

17. CNN-Based Network Intrusion Detection against Denial-of-Service Attacks / J. Kim, J. Kim, H. Kim, M. Shim, E. Choi // Electronics. – 2020. – Vol. 9, № 6. – P. 916. DOI: https://doi.org/10.3390/electronics9060916.

18. Model Evasion Attack on Intrusion Detection Systems using Adversarial Machine Learning / Md. A. Ayub; W. A. Johnson; D. A. Talbert; A. Siraj // 2020 54th Annual Conference on Information Sciences and Systems (CISS) (Princeton, NJ, USA, 18-20 March 2020). – 2020. DOI: https://doi.org/10.1109/ciss48834.2020.1570617116.

19. Evasion Attacks against Machine Learning at Test Time / B. Biggio, I. Corona, D. Maiorca, B. Nelson, N. Srndic, P. Laskov, G. Giacinto, F. Roli // Joint European Conference on Machine Learning and Knowledge Discovery in Databases. – 2013. – Pp. 387-402. DOI: https://doi.org/10.1007/978-3-642-40994-3_25.

20. Secure Kernel Machines against Evasion Attacks / Paolo Russu,Ambra Demontis, Battista Biggio, Giorgio Fumera, Fabio Roli // CCS'16: 2016 ACM SIGSAC Conference on Computer and Communications Security (Vienna Austria. – New York, 2016). – 2016. DOI: https://doi.org/10.1145/2996758.2996771.

21. Distillation as a Defense to Adversarial Perturbations Against Deep Neural Networks / N. Papernot, P. McDaniel, X. Wu, S. Jha, A. Swami // 2016 IEEE Symposium on Security and Privacy (SP) (San Jose, 22-26. 05. 2016). – 2016. DOI: https://doi.org/10.1109/sp.2016.41.

22. Goodfellow, I. J. Goodfellow, I. J. Explaining and Harnessing Adversarial Examples [Electronic resource] / I. J. Goodfellow, J. Shlens, C. Szegedy. – Access mode: https://arxiv.org/abs/1412.6572.

23. Gradient Concealment: Free Lunch for Defending Adversarial Attacks [Electronic resource] / S. Pei, J. Sun, X. Zhang, G. Meng. – Access mode: https://arxiv.org/abs/2205.10617.

24. Meng, D. MagNet: A Two-Pronged Defense against Adversarial Examples / D. Meng, H. Chen // CCS '17: 2017 ACM SIGSAC Conference on Computer and Communications Security (Dallas Texas USA). – New York, 2017. DOI: https://doi.org/10.1145/3133956.3134057.

25. Moosavi-Dezfooli, S.-M. DeepFool: A Simple and Accurate Method to Fool Deep Neural Networks / S.-M.Moosavi-Dezfooli, A. Fawzi, P. Frossard // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Las Vegas, 27-30. 06. 2016). – 2016. – DOI: https://doi.org/10.1109/cvpr.2016.282.

26. Lindqvist, B. Minimax defense against gradient-based adversarial attacks [Electronic resource] / B. Lindqvist, R. Izmailov. – Access mode: https://arxiv.org/abs/2002.01256.

27. SpectralDefense: Detecting Adversarial Attacks on CNNs in the Fourier Domain [Electronic resource] / P. Harder, F.-J. Pfreundt, M. Keuper, J. Keuper. – Access mode: https://arxiv.org/abs/2103.03000.

28. Kwon, H. AdvGuard: Fortifying Deep Neural Networks against Optimized Adversarial Example Attack / H. Kwon, J. Lee // IEEE Access. – 2020. – P. 1. DOI: https://doi.org/10.1109/access.2020.3042839.

29. Carlini N. Towards Evaluating the Robustness of Neural Networks / N. Carlini, D. Wagner // 2017 IEEE Symposium on Security and Privacy (SP) (San Jose, 22- 26. 05. 2017). – 2017. DOI: https://doi.org/10.1109/sp.2017.49.

30. Xu, W. Feature Squeezing Mitigates and Detects Carlini/Wagner Adversarial Examples [Electronic resource] / W. Xu, D. Evans, Y. Qi. – Access mode: https://arxiv.org/abs/1705.10686.

31. Samangouei, P. Defense-GAN: Protecting Classifiers Against Adversarial Attacks Using Generative Models [Electronic resource] / P. Samangouei, M. Kabkab, R.Chellappa. – Access mode: https://arxiv.org/abs/1805.06605.

32. Deep k-Nearest Neighbors: Towards Confident, Interpretable and Robust Deep Learning [Electronic resource] / Nicolas Papernot, Patrick McDaniel – Access mode: https://arxiv.org/abs/1803.04765.

33. Adversarial-Aware Deep Learning System Based on a Secondary Classical Machine Learning Verification Approach / M. Alkhowaiter, H. Kholidy, M. A. Alyami, A. Alghamdi, C. Zou // Sensors. – 2023. – Vol. 23, № 14. – P. 6287. DOI: https://doi.org/10.3390/s23146287.

References

1. Liao, H.J., Lin, C.H.R., Lin, Y.C., & Tung, K.Y. (2013). Intrusion detection system: A comprehensive review. *Journal of Network and Computer Applications, 36(1),* 16-24.

2. Hu, N., Tian, Z., Lu, H., Du, X., & Guizani, M. (2021). A multiple-kernel clustering based intrusion detection scheme for 5G and IoT networks. I*nternational Journal of Machine Learning and Cybernetics,* 1-16.

3. Min, E., Long, J., Liu, Q., Cui, J., & Chen, W. (2018). *TR-IDS: Anomaly-based intrusion detection through text-convolutional neural network and random forest*. Security and Communication Networks.

4. Tran, N.N., Sarker, R., & Hu, J. (2018). An approach for host-based intrusion detection system design using convolutional neural network. In *Mobile Networks and Management*: 9th International Conference, MONAMI 2017, Melbourne, Australia, December 13-15, 2017, Proceedings 9 (pp. 116-126). Springer International Publishing.

5. Kim, A., Park, M., & Lee, D. H. (2020). AI-IDS: Application of deep learning to real-time Web intrusion detection. *IEEE Access, 8,* 70245-70261.

6. Yassin, W., Udzir, N.I., Muda, Z., & Sulaiman, M.N. (2013). Anomaly-based intrusion detection through k-means clustering and naives bayes classification.

7. Khan, M.A. (2021). HCRNNIDS: Hybrid convolutional recurrent neural network-based network intrusion detection system. *Processes, 9(5),* 834.

8. Mayhew, M., Atighetchi, M., Adler, A., & Greenstadt, R. (2015, October). Use of machine learning in big data analytics for insider threat detection. In *MILCOM 2015-2015 IEEE Military Communications Conference* (pp. 915-922). IEEE.

9. IBM. (2013). Security Intrusion detection. IBM i Version 7.2. https://www.ibm.com/docs/en/ssw_ibm_i_72/rzaub/rzaubpdf.pdf.

10. Microsoft & Baldwin, M. (2022). Security Control: Network Security. learn.microsoft.com. https://learn.microsoft.com/en-us/security/benchmark/azure/security-control-network-security.

11. Amazon. (2024). Amazon GuardDuty features. aws.amazon.com. https://aws.amazon.com/guardduty/features/.

12. Digitalocean & Camisso, J. (2021). Understanding Suricata Signature. www.digitalocean.com. https://www.digitalocean.com/community/tutorials/understanding-suricata-signatures.

13. Oracle. (2024). Oracle Intrusion Detection System. docs.oracle.com. https://docs.oracle.com/cd/E95618_01/html/sbc_scz810_security/GUID-73A32803-097C-496F-9709-7B51CF54382B.htm#Intrusion-Detection-System.

14. Hindy, H., Atkinson, R., Tachtatzis, C., Colin, J. N., Bayne, E., & Bellekens, X. (2020). Utilising deep learning techniques for effective zero-day attack detection. *Electronics, 9(10),* 1684.

15. Ortega-Fernandez, I., Sestelo, M., Burguillo, J. C., & Pinon-Blanco, C. (2023). Network intrusion detection system for DDoS attacks in ICS using deep autoencoders. *Wireless Networks*, 1-17.

16. Othman, S.M., Ba-Alwi, F.M., Alsohybe, N.T., & Al-Hashida, A.Y. (2018). Intrusion detection model using machine learning algorithm on Big Data environment. *Journal of big data, 5(1),* 1-12.

17. Kim, J., Kim, J., Kim, H., Shim, M., & Choi, E. (2020). CNN-based network intrusion detection against denial-of-service attacks. *Electronics, 9(6),* 916.

18. Ayub, M.A., Johnson, W.A., Talbert, D.A., & Siraj, A. (2020, March). Model evasion attack on intrusion detection systems using adversarial machine learning. In *2020 54th annual conference on information sciences and systems (CISS)* (pp. 1-6). IEEE.

19. Biggio, B., Corona, I., Maiorca, D., Nelson, B., Šrndić, N., Laskov, P., ... & Roli, F. (2013). Evasion attacks against machine learning at test time. In *Machine Learning and Knowledge Discovery in Databases* : European Conference, ECML PKDD 2013, Prague, Czech Republic, September 23-27, 2013, Proceedings, Part III 13 (pp. 387-402). Springer Berlin Heidelberg.

20. Russu, P., Demontis, A., Biggio, B., Fumera, G., & Roli, F. (2016, October). Secure kernel machines against evasion attacks. In *Proceedings of the 2016 ACM workshop on artificial intelligence and security* (pp. 59-69).

21 Papernot, N., McDaniel, P., Wu, X., Jha, S., & Swami, A. (2016, May). Distillation as a defense to adversarial perturbations against deep neural networks. In *2016 IEEE symposium on security and privacy (SP)* (pp. 582-597). IEEE.

22. Goodfellow, I. J., Shlens, J., & Szegedy, C. (2014). Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572.

23. Pei, S., Sun, J., Zhang, X., & Meng, G. (2022). *Gradient Concealment: Free Lunch for Defending Adversarial Attacks.* arXiv preprint arXiv:2205.10617.

24. Meng, D., & Chen, H. (2017, October). Magnet: a two-pronged defense against adversarial examples. In *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security* (pp. 135-147).

25. Moosavi-Dezfooli, S.M., Fawzi, A., & Frossard, P. (2016). Deepfool: a simple and accurate method to fool deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2574-2582).

26. Lindqvist, B., & Izmailov, R. (2020). Minimax defense against gradient-based adversarial attacks. arXiv preprint arXiv:2002.01256.

27. Harder, P., Pfreundt, F. J., Keuper, M., & Keuper, J. (2021, July). Spectraldefense: Detecting adversarial attacks on cnns in the fourier domain. In *2021 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.

28. Kwon, H., & Lee, J. (2020). AdvGuard: fortifying deep neural networks against optimized adversarial example attack. IEEE Access.

29. Carlini, N., & Wagner, D. (2017, May). Towards evaluating the robustness of neural networks. In *2017 ieee symposium on security and privacy* (sp) (pp. 39-57). Ieee.

30. Xu, W., Evans, D., & Qi, Y. (2017). Feature squeezing mitigates and detects carlini/wagner adversarial examples. arXiv preprint arXiv:1705.10686.

31. Samangouei, P., Kabkab, M., & Chellappa, R. (2018). Defense-gan: Protecting classifiers against adversarial attacks using generative models. arXiv preprint arXiv:1805.06605.

32. Papernot, N., & McDaniel, P. (2018). Deep k-nearest neighbors: Towards confident, interpretable and robust deep learning. arXiv preprint arXiv:1803.04765.

33. Alkhowaiter, M., Kholidy, H., Alyami, M. A., Alghamdi, A., & Zou, C. (2023). Adversarial-Aware Deep Learning System Based on a Secondary Classical Machine Learning Verification Approach. *Sensors, 23(14),* 6287.

Отримано 01.02.2024

UDC 004.8:004.056

Artem Volokyta[1], Mykyta Melenchukov[2]

[1]PhD in Technical Sciences, Associate Professor of Department of Computer Engineering
National Technical University of Ukraine "Ihor Sikorskyi Kyiv Polytechnic Institute" (Kyiv, Ukraine)
E-mail: artem.volokita@kpi.ua. **ORCID:** http://orcid.org/0000-0001-9069-5544

[2]PhD student of Department of Computer Engineering
National Technical University of Ukraine "Kyiv Polytechnic Institute named after Igor Sikorsky" (Kyiv, Ukraine)
E-mail: melenchukov.nikita@gmail.com. **ORCID**: https://orcid.org/0009-0005-6615-4306

NEURAL NETWORKS IN DETECTING ATTACKS ON DISTRIBUTED SYSTEMS

Modern challenges in processing vast amounts of data are solved with the help of complex distributed systems, which in turn require cyber protection, that has the instruments for managing security risks such as information acquisition, espionage, reduction of system productivity, etc. This article provides an overview of some approaches to detecting cyberattacks, which in particular use machine learning. Their advantages, disadvantages, work methods, vulnerabilities, and approaches to their protection are given. Approaches to using various machine learning models for pre-processing input data, which is subsequently analyzed by intrusion detectors, and ways of improving the accuracy and effectiveness of cyberattack identification were also investigated.

As a result of the analysis of research, it is shown that there is a need to combine data from various attack detection systems used at different layers of cyber defense. The use of attack detection systems in different layers of the system allows the detection of such attacks as SQL insertion, obtaining administrator rights, acquiring access to the node, DDoS, and Probe. This can be done by analyzing logs, or network packets in case of SQL insertion, or by monitoring Internet traffic during a DDoS attack. Taking into account the growing variety of attacks on distributed systems and the increase in the amount of information being processed, the use of machine learning in attack detection systems is becoming an increasingly attractive direction for study. Artificial intelligence can be involved in selecting the necessary features for analysis, detecting interventions, reducing the level of false attack detections, and improving the accuracy of their recognition. The ability to retrain the model on new data is a significant advantage due to the small number of available test datasets dedicated to attacks on distributed systems. There is a need to investigate the feasibility of using certain machine learning models and neural networks, based on available input data and requirements for accuracy and speed.

It has been determined that anomaly detection is the most common approach to recognizing attacks using machine learning in the commercial product market. Analysis of attacks against machine learning-based defenses that use an anomaly detection approach has shown that there are weaknesses that can be minimized with additional protection; for example, time-distributed attacks can adapt to acceptable ranges of deviation of network indicators. An overview of the mechanisms for ensuring the resistance of protection systems to such influences, including the addition of various noises during training, range reduction of system parameter values, variations in retraining the model on misleading data, and the use of special classifiers, was performed.

Keywords: *intrusion detection systems; neural networks; distributed systems; anomaly detection; machine learning.*

Table: 5. References: 33.

Волокита А., Меленчуков М. Нейронні мережі у виявленні атак на розподілені системи. *Технічні науки та технології.* 2024. № 1(35). С. 135-145.