DOI 10.31891/2307-5732-2023-321-3-251-257
УДК 004.896

БОБРОВНІКОВА КІРА
Хмельницький національний університет
https://orcid.org/0000-0002-1046-893X
e-mail: bobrovnikova.kira@gmail.com
ГУРМАН ІВАН
Хмельницький національний університет
https://orcid.org/0000-0002-2282-3484
e-mail: devastator167384@gmail.com
ПОПОВ ЮРІЙ
Хмельницький національний університет
e-mail: f.societyy98@gmail.com
БОЙЧУК ЯРОСЛАВ
Хмельницький національний університет
e-mail: Boichuk.yaroslav@gmail.com
КАЧУР ВОЛОДИМИР
Хмельницький національний університет
e-mail: erevanskiyshpek@gmail.com

ВИЯВЛЕННЯ КІБЕРАТАК В ІНФРАСТРУКТУРІ ІНТЕРНЕТУ РЕЧЕЙ НА ОСНОВІ МАШИННОГО НАВЧАННЯ

Зростаючий попит на пристрої Інтернету речей призводить до прискорення темпів їх виробництва. Прагнучи прискорити випуск нового пристрою на ринок та зменшити його собівартість, виробники дуже часто нехтують дотриманням вимог кібербезпеки стосовно цих пристроїв. Відсутність оновлень безпеки та прозорості щодо стану безпеки пристроїв Інтернету речей, а також небезпечне розгортання в мережі перетворює пристрої Інтернету речей на об'єкт атак кіберзлочинців. Щоквартальні звіти компаній, пов'язаних з забезпеченням кібербезпеки, свідчать про низький рівень безпеки інфраструктури Інтернету речей. Враховуючи широке використання пристроїв Інтернету речей не лише в приватному секторі, а й на об'єктах різного призначення, включаючи об'єкти критичної інфраструктури, безпека цих пристроїв та інфраструктури Інтернету речей набуває важливого значення.

На сьогоднішній день відомо багато різних методів виявлення кібератак на інфраструктуру Інтернету речей. Перевагами застосування методів машинного навчання в порівнянні з сигнатурним аналізом є вища точність виявлення та менша кількість хибних спрацювань, можливість виявлення аномалій та нових ознак атак. Проте ці методи мають і певні недоліки. Серед них необхідність в додаткових апаратних ресурсах та більш низька швидкість обробки даних. В роботі представлено огляд сучасних методів, спрямованих на виявлення кібератак та аномалій в мережах Інтернету речей із застосуванням методів машинного навчання. Основними недоліками відомих методів є неспроможність виявлення та адаптивного реагування на атаки нульового дня та мультивекторні атаки. Останній недолік є найбільш критичним, про що свідчить постійне зростання кількості кібератак на інфраструктуру Інтернету речей. Загальним обмеженням для більшості відомих підходів є потреба в значних обсягах обчислювальних ресурсів та значний час відгуку систем виявлення кібератак.

Ключові слова: Інтернет речей (IoT), машинне навчання, виявлення аномалій, виявлення атак, виявлення вторгнень.

HURMAN IVAN, BOBROVNIKOVA KIRA, POPOV YURY, BOYCHUK YAROSLAV, KACHUR VOLODYMYR
Khmelnytskyi National University

MACHINE LEARNING BASED METHODS FOR CYBERATACS DETECTION IN THE INTERNET OF THINGS INFRASTRUCTURE

The growing demand for IoT devices is accelerating the pace of their production. In an effort to accelerate the launch of a new device and reduce its cost, manufacturers often neglect to comply with cybersecurity requirements for these devices. The lack of security updates and transparency regarding the security status of IoT devices, as well as unsafe deployment on the Internet, makes IoT devices the target of cybercrime attacks. Quarterly reports from cybersecurity companies show a low level of security of the Internet of Things infrastructure. Considering the widespread use of IoT devices not only in the private sector but also in objects for various purposes, including critical infrastructure objects, the security of these devices and the IoT infrastructure becomes more important.

Nowadays, there are many different methods of detecting cyberattacks on the Internet of Things infrastructure. Advantages of applying the machine-based methods in comparison with signature analysis are the higher detection accuracy and fewer false positive, the possibility of detecting both anomalies and new features of attacks. However, these methods also have certain disadvantages. Among them there is the need for additional hardware resources and lower data processing speeds. The paper presents an overview of modern methods aimed at detecting cyberattacks and anomalies in the Internet of Things using machine learning methods. The main disadvantages of the known methods are the inability to detect and adaptively respond to zero-day attacks and multi-vector attacks. The latter shortcoming is the most critical, as evidenced by the constantly increasing number of cyber attacks on the Internet of Things infrastructure. A common limitation for most known approaches is the need for significant computing resources and the significant response time of cyberattack detection systems.

Keywords: Internet of Things (IoT), machine learning, anomaly detection, attacks detection, intrusions detection.

Постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями

Сьогодні Інтернет речей (IoT) є невід'ємною частиною сучасного суспільства. IoT інфраструктура

стає невід'ємною складовою інфраструктури сучасних мегаполісів. Віртуальна інфраструктура повсюдно контролює фізичні об'єкти різного призначення, включаючи об'єкти критичної інфраструктури: від приватних помешкань та об'єктів різного призначення, включаючи виробництва, до транспортних магістралей, дамб та електростанцій.

На сьогоднішній день Інтернет речей являє собою слабко пов'язані між собою розрізнені мережі фізичних об'єктів, кожна з яких розгорнута для розв'язку специфічних задач. Фізичні об'єкти Інтернету речей містять вбудовані технології, що дозволяють здійснювати взаємодію з зовнішнім середовищем, передавати дані про свій стан та приймати дані ззовні. Таким чином, інфраструктура Інтернету речей складається з мережі передачі даних між множиною фізичних пристроїв, які оснащені вбудованими засобами та технологіями для взаємодії в автоматичному режимі між собою та/або з зовнішнім середовищем (такими як давачі та виконавчі механізми), а також програмним забезпеченням. Для підключення пристроїв Інтернету речей можуть бути використані як провідні, так і безпровідні технології, що надає можливість передачі і обміну даними між зовнішнім середовищем і іншими пристроями Інтернету речей за допомогою використання стандартних протоколів зв'язку.

Зростаючий попит на різноманітні пристрої Інтернету речей для автоматизації житлових приміщень, інфраструктури розумних міст, медицини та сільського господарства призводить до прискорення темпів їх виробництва. Ці пристрої забезпечують нові послуги та уможливлюють автономну підтримку функціонування та комунікацій в різних галузях.

Прагнучи якнайшвидше випустити новий пристрій на ринок та зменшити його собівартість, виробники дуже часто розробляють ці пристрої без урахування останніх вимог кібербезпеки, спрощують або взагалі не впроваджують жодних функцій безпеки та захисту. Іншими критичними факторами є відсутність оновлень безпеки для пристроїв Інтернету речей, відсутність прозорості щодо стану їх безпеки, а також небезпечне розгортання з можливістю безпосереднього доступу до пристроїв Інтернету речей через Інтернет. Все це перетворює пристрої Інтернету речей на найслабшу ланку, що відкриває можливості для злому та компрометації захищеної інфраструктури мереж. Це, в свою чергу, призводить до здійснення кібератак навіть на ті сфери, які раніше не представляли ризиків для кібербезпеки [1, 2]. Крім того, набувають поширення розумні пристрої Інтернету речей, які регулярно збирають та використовують конфіденційну інформацію про своїх власників, що також робить їх бажаною ціллю для кіберзлочинців.

Вразливими ланками в інфраструктурі Інтернету речей можуть бути як пристрої Інтернету речей, які зазвичай є основним засобом ініціювання атак, так і канали, що з'єднують компоненти інфраструктури Інтернету речей між собою [3]. Вразливими для кібератак можуть бути незахищені за замовчуванням налаштування пристроїв Інтернету речей та застарі компоненти, неправильне конфігурування цих пристроїв, а також протоколи, що використовуються в інфраструктурі Інтернету речей [4]. Вразливості веб-додатків та програмного забезпечення пристроїв Інтернету речей можуть надати кіберзлочинцям можливість компрометації систем для надсилання зловмисних оновлень або викрадення облікових даних користувача.

В багатьох мережах Інтернету речей вже є скомпрометовані або вразливі до злому пристрої Інтернету речей, на які можуть бути спрямовані різноманітні кібератаки, або які самі можуть стати джерелом кібератаки на інші пристрої в мережі Інтернет. При цьому кібератаки на інфраструктуру Інтернету речей можуть бути спрямовані як на конфіденційність, так і на доступність або продуктивність мережі Інтернету речей. Таким чином, будь-який побутовий розумний пристрій, наприклад кавомашина, може бути скомпрометований і використаний в якості джерела кібератаки, що дозволить кіберзлочинцям впливати на критично важливі системи мережі за рахунок моніторингу систем Інтернету речей та збору даних в скомпрометованій мережі [1].

Сотні тисяч окремих незахищених пристроїв Інтернету речей, кожен з яких має невелику обчислювальну потужність, можуть бути інфіковані зловмисним програмним забезпеченням та об'єднані злочинцями в єдину зловмисну мережу. Постійне підключення до мережі Інтернет та низький рівень або повна відсутність функцій безпеки перетворюють незахищені пристрої Інтернету речей на зручний інструмент для організації потужних кібератак. Мережі інфікованих пристроїв найчастіше використовуються для організації потужних DDoS-атак або в якості вузлів виходу VPN. При цьому обсяг DDoS-трафіку, генерований мережею інфікованих пристроїв Інтернету речей, зазвичай набагато потужніший, ніж обсяг трафіку зловмисних мереж, сформованих з персональних комп'ютерів [5, 6].

Іншим способом використання інфікованих пристроїв Інтернету речей є криптомайнінг. Обмежена ємність батарей на смартфонах не дозволяє використовувати такі інфіковані пристрої для монетизації, тому з цією метою найчастіше використовуються інфіковані смарт-телевізори, приставки тощо. При цьому будь-які розумні пристрої Інтернету речей, підключені до мережі Інтернет, наприклад такі як лічильники води, електрики і газу, потенційно можуть бути об'єктами інтересу кіберзлочинців. Мотивація кіберзлочинців для здійснення атак може бути різна: розваги, викрадення конфіденційної інформації або інформації, що є комерційною таємницею, помста, вимагання або шантаж з метою отримання фінансової вигоди, або навіть терористичні акти з політичною або іншою метою.

Згідно з прогнозами Groupe Speciale Mobile Association [7], кількість використовуваних у всьому світі пристроїв Інтернету речей до 2025 р. досягне майже 25 мільярдів (що вдвічі перевищує кількість підключених пристроїв на сьогоднішній день). Це призведе до зростання ризику кібератак, спрямованих на

ці пристрої. Важливою проблемою на сьогоднішній день також є неможливість встановлення будь-яких захисних або моніторингових рішень на пристрої Інтернету речей, що ускладнює попередження зловмисної активності в інфраструктурі Інтернету речей. Враховуючи стрімку інтеграцію мережі Інтернет через платформу Інтернета речей в різні сфери людської діяльності, включаючи об'єкти критичної інфраструктури, захист інфраструктури Інтернету речей від кібератак набуває важливого значення.

Метою роботи є огляд сучасних методів, спрямованих на виявлення кібератак та аномалій в мережах Інтернету речей із застосуванням методів машинного навчання

Методи на основі машинного навчання для виявлення кібератак в інфраструктурі Інтернету речей

На сьогоднішній день відомо багато підходів, спрямованих на виявлення кібератак в інфраструктурі Інтернету речей, і одним з перспективних напрямків є методи на основі машинного навчання (табл. 1).

Запропонований в роботі [8] метод використовує хмарні технології та парадигму програмно-визначених мереж (SDN) для виявлення та пом'якшення DDoS-атак в бездротових мережах Інтернету речей (табл. 1). В цьому підході використано дворівневу децентралізовану SDN. Кожен домен підмережі містить локальний контролер, при чому в хмарному середовищі розташовано універсальний контролер, підключений до локальних контролерів. Весь трафік мережі контролюється локальними контролерами, які збирають трафік і витягують з нього множину ознак, що можуть вказувати на наявність DDoS-атак. З цією метою було використано 155 ознак, вилучених за допомогою функції switched port analyzer (SPAN) комутатора Cisco Nexus, наприклад: frame.interface_id, frame.time_epoch, frame.len, radiotap.pad, radiotap.length, wlan.fc.frag, wlan.duration, wlan.frag, data.len. Вилучені з трафіку ознаки слугують для модуля виявлення DDoS, який працює на всіх локальних контролерах. Модуль пом'якшення DDoS-атак також розгорнутий у локальних контролерах. З метою пом'якшення DDoS-атак запропоновано окремі стратегії для рухомих та нерухомих пристроїв бездротового Інтернету речей.

З метою виявлення DDoS-атак використано часткове навчання та машину екстремального навчання, extreme learning machine, ELM – нейронну мережу прямого поширення. Особливістю функціонування машин екстремального навчання є вибір початкових параметрів випадковим чином і включення простих матричних операцій, що надає можливість скоротити час навчання. Таким чином, машини екстремального навчання можуть бути використані в режимі реального часу, оскільки будь-яке перенавчання буде досить швидким і не порушить роботу додатків.

В роботі [10] представлено систему виявлення вторгнень в інфраструктуру Інтернету речей на основі глибокого навчання (DL-IDS). Згідно запропонованого підходу, трафік середовища Інтернету речей піддається попередній обробці для усунення невизначеностей та нормалізації набору даних, що полягає в усуненні надмірності та замінах відсутніх значень. З цією метою здійснюється вимірювання подібності даних у наборі даних з використанням відстані Мінковського для обчислення відстані між кожною парою даних, після чого повторювані та надлишкові дані видаляються з набору даних і передаються на наступний етап попередньої обробки. На наступному етапі відсутні значення атрибутів у даних замінюються обчисленими значеннями найближчого сусіда, щоб уникнути зміщення результату класифікації в бік більш частих записів. З цією метою визначаються K найближчих сусідів за Евклідовою відстанню, і відсутнє значення замінюється середнім значенням для одержаних даних.

З метою вибору найбільш важливих ознак трафіку, які можуть свідчити про факт вторгнення в середовище Інтернету речей, використано алгоритм оптимізації spider monkey (SMO). З метою виявлення вторгнень застосовано stacked–deep polynomial network (SDPN), що надає можливість класифікувати вхідні дані як нормальні або аномальні. Аномальні дані можуть вказувати на факт вторгнення, такий як атака на відмову в обслуговуванні (DoS), атака user–to–root (U2R), атаку probe, атака remote–to–local (R2L).

В [11] проведено комплексне дослідження ефективності та перспектив використання класифікаторів, заснованих на машинному навчанні, для систем виявлення вторгнень на основі аномалій (IDS) в інфраструктурі Інтернету речей. В якості основної атаки для аналізу було обрано найбільш поширений та небезпечний тип атаки – атаку на відмову в обслуговуванні (DoS). Було проаналізовано ефективність застосування як ансамблів класифікаторів, так і одиночних класифікаторів (табл. 1). З цієї метою було використано популярні набори даних (табл. 1).

Продуктивність усіх класифікаторів було виміряно з точки зору точності, специфічності, чутливості, частоти помилкових спрацьовувань та площі під кривою робочої характеристики приймача. Для статистичного аналізу значущих відмінностей між досліджуваними класифікаторами використано тести Фрідмана і Немен'ї. Крім того, проаналізовано час відгуку класифікаторів на спеціальному обладнанні IoT та обговорено методологію вибору найкращого класифікатора відповідно до вимог системи виявлення вторгнень. На основі одержаних результатів продуктивності та статистичних тестів було зроблено висновок, що такі класифікатори як дерева класифікації та регресії (classification trees, regression trees), а також extreme gradient boosting показують найкращий компроміс між проаналізованими показниками ефективності класифікації та часом відгуку, тому є прийнятним вибором для створення IDS для середовища IoT на основі аномалій (табл. 1).

В роботі [14] запропоновано фреймворк для виявлення трафіку атак в мережах Інтернету речей. Метод включає чотири кроки: (1) новий підхід CorrAUC до вибору ознак, що несуть достатню для виявлення трафіка атак інформацію; (2) на основі CorrAUC розроблено новий алгоритм CorrAUC вибору ознак, який базується на техніці обгортки для точної фільтрації найбільш ефективних ознак для вибраного алгоритму машинного навчання та складається з оцінки атрибутів кореляції (Correlation Attribute Evaluation,

CAE) і поєднується з метрикою Area Under Roc Curve (AUC); (3) застосування інтегрованих методу мультикритеріального аналізу рішень TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution, multi-criteria decision analysis method) та Shannon Entropy; (4) створення бієктивного м'якого набору для перевірки вибраних ознак для ідентифікації трафіку атак у мережі IoT. Ефективність запропонованого підходу була оцінена за допомогою чотирьох алгоритмів машинного навчання (табл. 1). Аналіз експериментальних результатів показав, що запропонований метод може досягати в середньому більше 96% ефективності для різних алгоритмів машинного навчання (табл. 1).

У [16] запропоновано систему AD-IoT для виявлення кібертатак на вузли туманних обчислень в інфраструктурі розумного міста, засновану на алгоритмі машинного навчання Random Forest. Запропоноване рішення може ефективно виявляти скомпрометовані пристрої IoT у розподілених туманних вузлах. Згідно цього підходу, з метою визначення нормальної та аномальної поведінки здійснюється моніторинг мережного трафіку, який проходить через кожен туманний вузол. Після виявлення атак на рівні туману система повинна повідомити хмарні служби безпеки про аналіз та оновлення системи. Результати проведених експериментів показали, що AD-IoT надає можливість досягти прийнятних результатів для виявлення атак (табл. 1).

В [17] представлено фреймворк для виявлення та захисту від аномальної активності в бездротових сенсорних мережах IoT (Wireless sensor networks, WSN). Зазначено, що на формування бот-мереж впливають специфічні для інфраструктури IoT особливості, такі як недостатня обчислювальна потужність, обмеження живлення та висока щільність вузлів IoT. Було проаналізовано два типи найбільш поширених атак на пристрої Інтернету речей. Перший тип – атаки Bashlite, спрямовані на фреймворки Linux, за яких здійснюється передача даних через відкритий telnet. Другий тип – атаки Mirai, спрямовані на знаходження слабких місць в гаджетах IoT, які можуть бути атаковані через їх IP- та Mac-адреси, після чого на зламані пристрої завантажується зловмисне програмне забезпечення. Запропоновано систему для виявлення та захисту від аномальної активності. В якості даних для аналізу цих двох типів атак було використано три стандартні набори доброякісних (нормальних) даних і (аномальних чи шкідливих) даних, зібраних з трьох пристроїв IoT (табл. 1). Дані, які використовуються для аналізу, розглядаються як великі дані. Таким чином, при обробці цих даних виникають наступні проблеми: великий обсяг даних, їх різноманітність та неструктурованість, нестача даних та необхідність високопродуктивної обробки. Тому при попередній обробці таких даних виключаються повторювані дані, після чого обчислюються мінімальне, максимальне, середнє та стандартне відхилення значень кожного атрибута. Після масштабування даних було одержано ознаки в діапазоні від 0 до 1. З метою оцінки кореляції ознак та рівня їх залежності було використано коефіцієнт Пірсона. Для аналізу одержаного набору даних було використано програмне забезпечення WEKA, а в якості методів машинного навчання було застосовано чотири класифікатори (табл. 1). Експериментальні результати показали, що поєднання алгоритмів Random Forest і Decision Tree може забезпечити достатньо високий рівень точності виявлення аномалій та атак на пристрої Інтернету речей.

В роботі [19] було проаналізовано трафік бот-мереж в середовищі IoT з використанням трьох класифікаторів машинного навчання (табл. 1). Було класифіковано дані для кожної атаки в кожній бот-мережі для дев'яти пристроїв. Для кожного класифікатора було обчислено такі показники як Accuracy, True Positive, False Positive, False Negative, True Negative, Precision, Recall, F1-score. За результатами експериментальних досліджень зазначено, що хоча було досягнуто високої точності виявлення (близько 99%), загалом застосування в якості класифікатора Random Forest дає найкращі результати, а застосування Support Vector Machine – найнижчі. Проте одержані високі показники F1-score демонструють надійність всіх трьох класифікаторів. Недоліком підходу є те, що для аналізу було використано всі наявні ознаки в наборах даних.

В роботі [20] проведено дослідження дванадцяти алгоритмів машинного навчання з точки зору їх здатності виявляти аномальну поведінку в мережі Інтернету речей. Оцінка проводиться за трьома загальнодоступними наборами даних (табл. 1). Експериментальні дослідження було проведено за допомогою ALICE high-performance computing facility at the University of Leicester. На основі проведених експериментальних досліджень здійснено комплексний аналіз застосування алгоритмів машинного навчання для виявлення аномальної поведінки в мережах Інтернету речей. Результати оцінки підтверджують, що алгоритм Random Forest досягає найкращої продуктивності з точки зору Accuracy, Precision, Recall, F1-Score and Receiver Operating Characteristic (ROC) curves для всіх застосованих наборів даних. Зауважено, що інші алгоритми машинного навчання працюють з близькою до Random Forest ефективністю, і що рішення стосовно вибору алгоритму машинного навчання залежить від даних, які підлягають аналізу.

В роботі [23] представлено систему виявлення вторгнень на основі аномалій, Також досліджено ефективність застосування різних алгоритмів машинного навчання для виявлення аномалій у використаному наборі даних про вторгнення в мережу Інтернету речей у реальному часі (табл. 1). Проведені експерименти показали найвищу точність класифікації для K-nearest Neighbours (табл. 1). Зазначено, що однією з важливих проблем, пов'язаних з безпекою в мережах Інтернету речей, є те, що більшість таких пристроїв мають обмежене живлення і обчислювальні можливості. Тому шифрування та аутентифікацію важко застосувати для захисту від кібератак. Виходячи з цього, виявлення мережних вторгнень на основі аномалій відіграє важливу роль у захисті мереж Інтернету речей від різних шкідливих дій. Перевага підходу полягає в тому, що коли відбувається атака нульового дня, сигнатура атаки не буде розпізнана, проте подальша поведінка мережі буде відхилятися від нормальних моделей трафіку, що дозволить IDS виявити аномалію.

Таблиця 1

Ефективність, використані методи машинного навчання та джерела наборів даних сучасних методів виявлення кібератак в інфраструктурі Інтернету речей

Автори	Мета	Застосовані методи	Множина даних	Результат
1	2	3	4	5
Ravi, N., Shalinie, S. M. [8] 2020	Виявлення та пом'якшення DDoS-атак	Extreme learning machines, ELM з частковим навчанням	UNB-ISCX [9]	Точність виявлення (Accuracy) DDoS-атак на рівні 96,28%
Otoum, Y., Liu, D., Nayak, A. [10] 2019	Виявлення вторгнень типу DoS, user–to–root (U2R), probe, remote–to–local (R2L)	Stacked–deep polynomial network	NSL-KDD [9]	Виявлення вторгнень на рівні: Accuracy (99.02%), Precision (0.9938), Recall (0.9829), F1-score (0.9883)
Verma, A., Ranga, V. [11] 2020	Дослідження ефективності та перспектив використання класифікаторів, заснованих на машинному навчанні, для IDS на основі аномалій на прикладі виявлення DoS-атак	Ансамблі класифікаторів: Random Forest, AdaBoost, Gradient Boosted Machine, Extreme Gradient Boosting, Extremely Randomized Trees; одиночні класифікатори: Classification and Regression Trees, Multi-layer Perceptron	CIDDS-001[12], UNSW-NB15 [13], NSL-KDD [9]	Classification Trees, Regression Trees, а також Extreme Gradient Boosting показують найкращий результат з рівнем Accuracy до 96,7%, Specificity до 96,2%, Sensitivity до 97,3%, та прийнятним часом відгуку
Shafiq, M., Tian, Z., Bashir, A. K., Du, X., Guizani, M. [14] 2020	Виявлення трафіку атак	Decision Tree (C4.5), Support Vector Machine, Naive Bayes, Random Forest	Bot-IoT Data Set [15]	Для Decision Tree C4.5 та Random Forest Specificity становить 98,95% і 99,99% відповідно, для Naive Bayes і SVM – 98,44% і 98,48% відповідно.
Alrashdi, I., Alqazzaz, A., Aloufi, E., Alharthi, R., Zohdy, M., Ming, H. [16] 2019	Виявлення аномалій та атак	Random Forest	UNSW-NB15 [13]	Виявлення атак на рівні: Precision – 0.79, Recall – 0.97, F1-score – 0.86
Aysa, M. H., Ibrahim, A. A., Mohammed, A. H. [17] 2020	Виявлення аномалій та атак на пристрої IoT	Decision Tree (J-48), Linear Support Vector Machine, Neural Network (Back- propagation), Random Forest	Нормальні та аномальні дані, зібрані з пристроїв Інтернету речей, з UCI Machine Learning Repository [18]	Поєднання алгоритмів Random Forest і Decision Tree може забезпечити високий рівень Accuracy
Bagui, S., Wang, X., Bagui, S. [19] 2021	Виявлення вторгнень	Logistic Regression, Random Forest, Support Vector Machine	UCI Machine Learning Repository [18]	Досягнуто високої точності виявлення (близько 99%)

Продовження табл. 1

1	2	3	4	5
Elmrabit, N., Zhou, F., Li, F., Zhou, H. [20] 2020	Виявлення аномальної активності, яка може вказувати на наявність атак	Logistic Regression, Naive Bayes, Decision tree, Simple Recurrent Neural Network, Gated Recurrent Units, Convolutional Neural Network and Long short-Term Memory, Convolutional Neural Network, Long short-Term Memory, Random Forest, Adaptive boosting, Deep Neural Network, K-nearest Neighbours	CICIDS-2017 [21], UNSW-NB15 [13], ICS Cyberattack [22]	Random Forest (RF) algorithm надає кращу продуктивність на рівні до 99.9 % для CICIDS-2017
Liu, Z., Thapa, N., Shaver, A., Roy, K., Yuan, X., Khorsandroo, S. [23] 2020	Підвищення рівня безпеки мережі Інтернету речей шляхом застосування кількох методів машинного навчання на наборі даних про вторгнення в мережу IoT	Logistic Regression, Support Vector Machine, K-nearest Neighbours, Random Forest, Extreme Gradient Boosting	IoT Network Intrusion Dataset [24]	Accuracy на рівні 99% із застосуванням K-nearest Neighbours, тоді як час виконання класифікації становить в середньому 2 хв. Accuracy на рівні 97% при застосуванні Extreme Gradient Boosting, час виконання класифікації – 10,8 с.

Висновки

Огляд звітів компаній, пов'язаних з забезпеченням кібербезпеки та виробників антивірусного програмного забезпечення, а також літературних джерел показує, що проблема виявлення кібератак в інфраструктурі Інтернету речей є надзвичайно актуальною. В роботі проведено короткий огляд підходів виявлення атак в інфраструктурі Інтернету речей на основі машинного навчання. Хоча відомі методи демонструють високий рівень ефективності, тим не менше їм властиві спільні недоліки та обмеження. Основними недоліками відомих методів є високий рівень хибних спрацювань, неспроможність виявлення та адаптивного реагування на атаки нульового дня та мультивекторні атаки. Останній недолік є найбільш критичним, про що свідчить постійне зростання кількості кібератак на інфраструктуру Інтернету речей. Загальним обмеженням для більшості відомих підходів є потреба в значних обсягах обчислювальних ресурсів та значний час відгуку систем виявлення, що є неприпустимим для роботи в режимі реального часу. Таким чином, все ще існує необхідність у розробленні нових методів виявлення атак в інфраструктурі Інтернету речей, які б усували недоліки відомих підходів та надавали можливість виявлення та адаптивного реагування на ще невідомі загрози, такі як атаки нульового дня та мультивекторні атаки.

References

1. Trend Micro. Inside the Smart Home: IoT Device Threats and Attack Scenarios. URL: https://www.trendmicro.com/vinfo/it/security/news/internet-of-things/inside-the-smart-home-iot-device-threats-and-attack-scenarios (accessed on November 1, 2021).

2. Check point software technologies LTD. Cyber security report 2021. URL: https://www.checkpoint.com/pages/cyber-security-report-2021/ (accessed on November 1, 2021).

3. OWASP Internet of Things. URL: https://owasp.org/www-project-internet-of-things/#tab=IoT_Attack_Surface_Areas (accessed on November 1, 2021).

4. Nozomi Networks Labs. What IT Needs to Know about OT/IoT Security Threats in 2020. URL: https://www.nozominetworks.com/blog/what-it-needs-to-know-about-ot-io-security-threats-in-2020/ (accessed on November 1, 2021).

5. McAfee Labs Threats Report. URL: https://www.mcafee.com/enterprise/en-us/lp/threats-reports/oct-2021.html (accessed on November 1, 2021).

6. Securelist. New trends in the world of IoT threats. URL: https://securelist.com/new-trends-in-the-world-

of-iot-threats/87991/ (accessed on November 1, 2021).

7. Global System for Mobile Communications. URL: https://www.gsma.com/ (accessed on November 1, 2021).

8. Ravi, N., & Shalinie, S. M. Learning-driven detection and mitigation of DDoS attack in IoT via SDN-cloud architecture. IEEE Internet of Things Journal, 2020, 7(4), pp. 3559-3570.

9. UNB. Canadian Institute for Cybersecurity. Datasets. URL: https://www.unb.ca/cic/datasets/index.html (accessed on November 1, 2021).

10. Otoum, Y., Liu, D., & Nayak, A. DL‐IDS: a deep learning-based intrusion detection framework for securing IoT. Transactions on Emerging Telecommunications Technologies, 2019, e3803.

11. Verma, A., & Ranga, V. Machine learning based intrusion detection systems for IoT applications. Wireless Personal Communications, 2020, 111(4), pp. 2287-2310.

12. Hochschule Coburg. CIDDS – Coburg Intrusion Detection Data Sets. URL: https://www.hs-coburg.de/forschung/forschungsprojekte-oeffentlich/informationstechnologie/cidds-coburg-intrusion-detection-data-sets.html (accessed on November 1, 2021).

13. UNSW Sydney. The UNSW-NB15 Dataset. URL: https://research.unsw.edu.au/projects/unsw-nb15-dataset (accessed on November 1, 2021).

14. Shafiq, M., Tian, Z., Bashir, A. K., Du, X., & Guizani, M. Corrauc: a malicious bot-iot traffic detection method in iot network using machine learning techniques. IEEE Internet of Things Journal, 2020.

15. UNSW Sydney. Bot-IoT Data Set. https://research.unsw.edu.au/projects/bot-iot-dataset (accessed on November 1, 2021).

16. Alrashdi, I., Alqazzaz, A., Aloufi, E., Alharthi, R., Zohdy, M., & Ming, H. Ad-iot: Anomaly detection of iot cyberattacks in smart city using machine learning. In 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), IEEE, 2019, pp. 0305-0310.

17. Aysa, M. H., Ibrahim, A. A., & Mohammed, A. H. IoT ddos attack detection using machine learning. In 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), IEEE, 2020, pp. 1-7.

18. UCI Machine Learning Repository. URL: https://archive.ics.uci.edu/ml/datasets.php (accessed on November 1, 2021).

19. Bagui, S., Wang, X., & Bagui, S. Machine Learning Based Intrusion Detection for IoT Botnet. International Journal of Machine Learning and Computing, 11(6), 2021.

20. Elmrabit, N., Zhou, F., Li, F., & Zhou, H. Evaluation of machine learning algorithms for anomaly detection. In 2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security), IEEE, 2020, pp. 1-8.

21. UNB. Canadian Institute for Cybersecurity. Intrusion Detection Evaluation Dataset (CIC-IDS2017). URL: https://www.unb.ca/cic/datasets/ids-2017.html.

22. Industrial Control System (ICS) Cyber Attack Datasets. URL; ttps://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets (accessed on November 1, 2021).

23. Liu, Z., Thapa, N., Shaver, A., Roy, K., Yuan, X., & Khorsandroo, S. Anomaly detection on iot network intrusion using machine learning. In 2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD), IEEE, 2020, pp. 1-5.

24. IEEE DataPort. IoT network intrusion dataset. URL: https://ieee-dataport.org/open-access/iot-network-intrusion-dataset.