Форма № 24

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Запорізька політехніка»

Факультет інформаційної безпеки та електронних комунікацій
(повне найменування факультету)

Кафедра інформаційної безпеки та наноелектроніки
(повне найменування кафедри)

Пояснювальна записка

до дипломного проєкту (роботи)
магістр
(ступінь вищої освіти)

на тему Аналіз можливостей використання штучного інтелекту у сфері захисту інформації
(назва теми)

Виконав(ла): студент(ка) IIм курсу,
групи БКз-813м
Спеціальності 125 Кібербезпека та захист інформації
(код і найменування спеціальності)
Освітня програма (спеціалізація)
Безпека інформаційних і комунікаційних систем
КУЛІЄВА В.С.
(ПРІЗВИЩЕ та ініціали)

Керівник ЛІЗУНОВ С.І.
(ПРІЗВИЩЕ та ініціали)

Рецензент САМОЙЛИК С.С.
(ПРІЗВИЩЕ та ініціали)

2024

Форма № 25

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Запорізька політехніка»

Факультет інформаційної безпеки та електронних комунікацій
Кафедра інформаційної безпеки та наноелектроніки
Ступінь вищої освіти магістр
Спеціальність 125 Кібербезпека та захист інформації
(код і найменування)
Освітня програма (спеціалізація) Безпека інформаційних і комунікаційних систем
(назва освітньої програми (спеціалізації))

ЗАТВЕРДЖУЮ
Завідувач кафедри ІБтаН
Андрій КОРОТУН
«____» _____2024року

З А В Д А Н Н Я
НА ДИПЛОМНИЙ ПРОЄКТ (РОБОТУ) СТУДЕНТА(КИ)

КУЛІЄВОЇ Валерії Сергіївни
(ПРІЗВИЩЕ, ім'я, по батькові)

1. Тема проєкту (роботи) Аналіз можливостей використання штучного інтелекту у сфері захисту інформації
Analysis of the possibilities of using artificial intelligence in the field of information protection

керівник проєкту (роботи) к.т.н., доцент ЛІЗУНОВ Сергій Іванович,
(науковий ступінь, вчене звання, ПРІЗВИЩЕ, ім'я, по батькові)
затверджені наказом закладу вищої освіти від «05» грудня 2024 року №507
2. Строк подання студентом проєкту (роботи) 10.12.2024
3. Вихідні дані до проєкту (роботи) технології штучного інтелекту

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити)
Вплив штучного інтелекту на організацію кібербезпеки; протидія загрозам методами машинного навчання та глибинного навчання

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, кількість слайдів, плакатів)
Презентація доповіді (в MS PowerPoint), 10 слайдів.

6. Консультанти розділів проєкту (роботи)

Розділ	ПРІЗВИЩЕ, ініціали та посада консультанта	Підпис, дата	
		завдання видав	прийняв виконане завдання
1 – 2	ЛІЗУНОВ С.І., доцент кафедри ІБтаН	02.09.2024	05.12.2024
Нормоконтроль	КОРОЛЬКОВ Р. Ю., доцент кафедри ІБтаН		09.12.2024

7. Дата видачі завдання «02» вересня 2024року.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів дипломного проєкту (роботи)	Строк виконання етапів проєкту (роботи)	Примітка
1.	Збір та аналіз інформації про використання штучного інтелекту для захисту інформації	02.09.24 – 16.09.24	виконано
2.	Систематизація літературних даних	17.09.24 – 23.09.24	виконано
3.	Складання і затвердження наукового завдання	24.09.24 – 29.09.24	виконано
4.	Формування та уточнення наукового завдання	30.09.24 – 05.10.24	виконано
5.	Опис підходів на основі машинного навчання до розв’язання проблем кібербезпеки	06.10.24 – 15.10.24	виконано
6.	Опис та аналіз прикладів використання глибинного навчання для потреб кібербезпеки	16.10.24 – 13.11.24	виконано
7.	Оформлення графічної частини	14.11.24 – 19.11.24	виконано
8.	Оформлення ПЗ	20.11.24 – 30.11.24	виконано

Студент(ка) _____ Валерія КУЛІЄВА
(підпис) (Ім’я ПРІЗВИЩЕ)

Керівник проєкту (роботи) _____ Сергій ЛІЗУНОВ
(підпис) (Ім’я ПРІЗВИЩЕ)

АНОТАЦІЯ

Пояснювальна записка до магістерської роботи: 86 с., 8 рис., 97 джерел.

КІБЕРАТАКИ, КІБЕРБЕЗПЕКА, ГЛИБОКЕ НАВЧАННЯ, МАШИННЕ НАВЧАННЯ, ШТУЧНИЙ ІНТЕЛЕКТ

Мета роботи: дослідження можливостей штучного інтелекту у сфері кіберзахисту.

Об’єкт та предмет дослідження: об’єктом дослідження є технології штучного інтелекту; предметом дослідження є їх використання для захисту інформації.

Методи дослідження: описово-аналітичний.

Результати: результатом дослідження є визначення рішень машинного та глибокого навчання, найбільш придатних до використання у сфері кіберзахисту.

Рекомендації щодо впровадження: робота носить прикладний характер, її результати можуть бути використані для вирішення проблем кіберзахисту з використанням методів штучного інтелекту.

Практична цінність: визначено показники ефективності методів машинного навчання при використанні для захисту інформації.

ABSTRACT

Explanatory note to the master's thesis: 86 p., 8 figure, 97 sources.

CYBERATTACKS, CYBERSECURITY, DEEP LEARNING, MACHINE LEARNING, ARTIFICIAL INTELLIGENCE

The purpose of the work: study of the capabilities of artificial intelligence in the field of cyber security.

Object and subject of research: the object of the research is artificial intelligence technologies; the subject of the research is their use for information protection.

Research methods: descriptive-analytical.

Results: the result of the research is the identification of machine and deep learning solutions that are most suitable for use in the field of cyber security.

Recommendations for implementation: the work is of an applied nature, its results can be used to solve cyber security problems using artificial intelligence methods.

Practical value: indicators of the effectiveness of machine learning methods when used for information protection are determined.

ЗМІСТ

ВСТУП

Технологічна революція призвела до швидкого розвитку та прийняття нових і вдосконалених технологій. Однак це також призвело до швидкої еволюції кіберзагроз і атак. Ці атаки стають все більш частими, численними та вражаючими. Щоб протистояти цим загрозам, що постійно розвиваються, необхідно мати передові та безпечні заходи кібербезпеки та захисні механізми [1].

Кібербезпека захищає інформаційні та комунікаційні системи, що виходять в Інтернет, від зловмисних атак і загроз. Четверта промислова революція та промисловий Інтернет речей (IoT) розширили сферу кібербезпеки від безпеки мережі та додатків до безпеки інфраструктури, хмари та інформації, зробивши її багатовимірною [2]. Кібербезпека охоплює різні взаємопов'язані компоненти та технології в кіберпросторі, а не обмежується безпекою системи. В організаційному контексті кібербезпека передбачає захист усіх відповідних вимірів кіберпростору одночасно [3].

Концепція «штучного інтелекту» виникла в 1956 році і з тих пір перетворилася на практичні рішення, які використовуються в різних сферах. Використання машинного навчання в кібербезпеці починається з 1990-х років із розробкою систем виявлення аномалій (ADS) і систем виявлення вторгнень (IDS), хоча прогресу заважали обмеження даних і обчислень. Сьогодні штучний інтелект є невід'ємною частиною кібербезпеки, виходячи за межі корпоративного використання. Він може симулювати людський інтелект і поведінку, що призводить до автоматизації кібербезпеки, що перевищує можливості людини, та може виявити порушення безпеки в мережі за лічені секунди [4]. Пандемія COVID-19 прискорила цифрову трансформацію, зробивши бізнес залежним від таких технологій, як штучний інтелект, машинне та глибоке навчання (ML та DL) і великі дані (BD). Проте, це призвело і до сплеску кіберзлочинів, що поставило під загрозу окремих

осіб і відомі організації. Прогнозується, що до 2025 року кіберзлочини можуть коштувати 10,5 трильйонів доларів США. Через залежність від цих технологій підприємства стикаються з операційними ризиками та ризиками безперервності. Варто вивчати використання штучного інтелекту в кібербезпеці, щоб організації могли зрозуміти можливості штучного інтелекту в просторі кібербезпеки на благо своєї діяльності.

Досягнення в галузі великих даних та інформатики призвели до появи машинного навчання, найпоширенішого типу ШІ в організаційній кібербезпеці. Машинне навчання передбачає здатність машини навчатися та адаптуватися через набуття досвіду. Воно вважається підмножиною штучного інтелекту, і зосереджено на реалізації певних типів систем, які можуть навчатися на існуючих даних, щоб ідентифікувати закономірності та самостійно приймати рішення [5].

Величезні обсяги даних, які генерують організації, надають можливості для широкого спектру додатків ML у кіберпросторі, включаючи розвідку загроз, виявлення аномалій та автоматизацію завдань, пов’язаних із кібербезпекою. Цей зв’язок між ШІ та кібербезпекою називається кібер-ШІ.

В звязку з цим, дослідження впливу ШІ на кібербезпеку з організаційної точки зору є актуальною задачею.

1 ВПЛИВ ШТУЧНОГО ІНТЕЛЕКТУ НА ОРГАНІЗАЦІЮ КІБЕРБЕЗПЕКИ

1.1 Постановка проблеми

Впровадження штучного інтелекту, зокрема рішень ML у сфері кібербезпеки, можна простежити з кінця 1980-х років, коли була впроваджена перша система виявлення аномалій (ADS) [6]. У 1990 -х роках вона була замінена системою виявлення вторгнень (IDS). Через відсутність структурованих і чистих даних у поєднанні з обмеженнями обчислювальної потужності їх розвиток було відкладено на деякий час. На сьогодні штучний інтелект розвинувся і революціонізував можливості сучасних технологій у сфері кібербезпеки [7]. Впровадження рішень на основі штучного інтелекту в організаційну кібербезпеку стало необхідним.

Оскільки цифрова трансформація неухильно просувається в останні роки, зростає залежність всіх сфер людської діяльності від Інтернету та інформаційно-комунікаційних технологій (ІКТ) [8]. Це змусило компанії визнати величезний потенціал і значення сучасних технологій, таких як AI, ML і Big Data [9]. Проте, широке впровадження ІКТ також призвело до збільшення кількості кіберзлочинів, загроз і вразливостей, спрямованих як на окремих осіб, так і на відомі організації [4]. З лютого 2020 року спостерігався значний сплеск кіберзлочинів, і враховуючи залежність організацій від цих технологій, кіберзагрози та атаки можуть мати жахливі наслідки для їх діяльності та безперервності бізнесу [9].

В останні роки проведяться інтенсивні дослідження з метою оцінки впливу ШІ на різні технологічні середовища. Проводилося комплексне дослідження впливу ШІ на людські аспекти кібербезпеки на підприємствах. Отримані дані свідчать про те, що штучний інтелект наразі покращує людські здібності, і передбачають потенційну трансформацію, коли ШІ розвиватиметься в напрямку автономності.

В роботі [10], зосереджено увагу на кібератаках, керованих штучним інтелектом, щоб оцінити особливості таких атак для розробки заходів кібербезпеки.

Використання можливостей штучного інтелекту, що розвиваються, може стати значною перевагою для підприємств у захисті від кібератак.

Загалом дослідження, проведені на сьогоднішній день для визначення впливу штучного інтелекту на кібербезпеку, є досить мізерними через швидкий розвиток ІКТ.

Тому оцінимо вплив технологій, керованих штучним інтелектом, на організаційну кібербезпеку, включаючи їхні позитивні та негативні наслідки, а також визначимо ефективність кібер-ШІ порівняно з традиційними заходами кібербезпеки.

1.2 Результати пошуку та огляду літератури

За результатами пошуку та огляду дослідницьких статтей за 2018 – 2023 роки встановлено, що дослідження впливу штучного інтелекту на організаційну кібербезпеку почалися поступово в 2019 році, де було знайдено лише 1 дослідження (рис. 1.1). З 2020 по 2021 рік відбулося поступове збільшення з 8 статей до 13 статей; у 2022 році відбулося різке зростання, коли кількість публікацій зросла більше ніж удвічі до 28. Розподіл, зображений на рис. 1.1 показує зростаючий інтерес науковців до розуміння впливу штучного інтелекту на організаційну кібербезпеку, що може бути пов’язано з постійним прогресом у використанні штучного інтелекту та відповідних методів організаційної кібербезпеки.

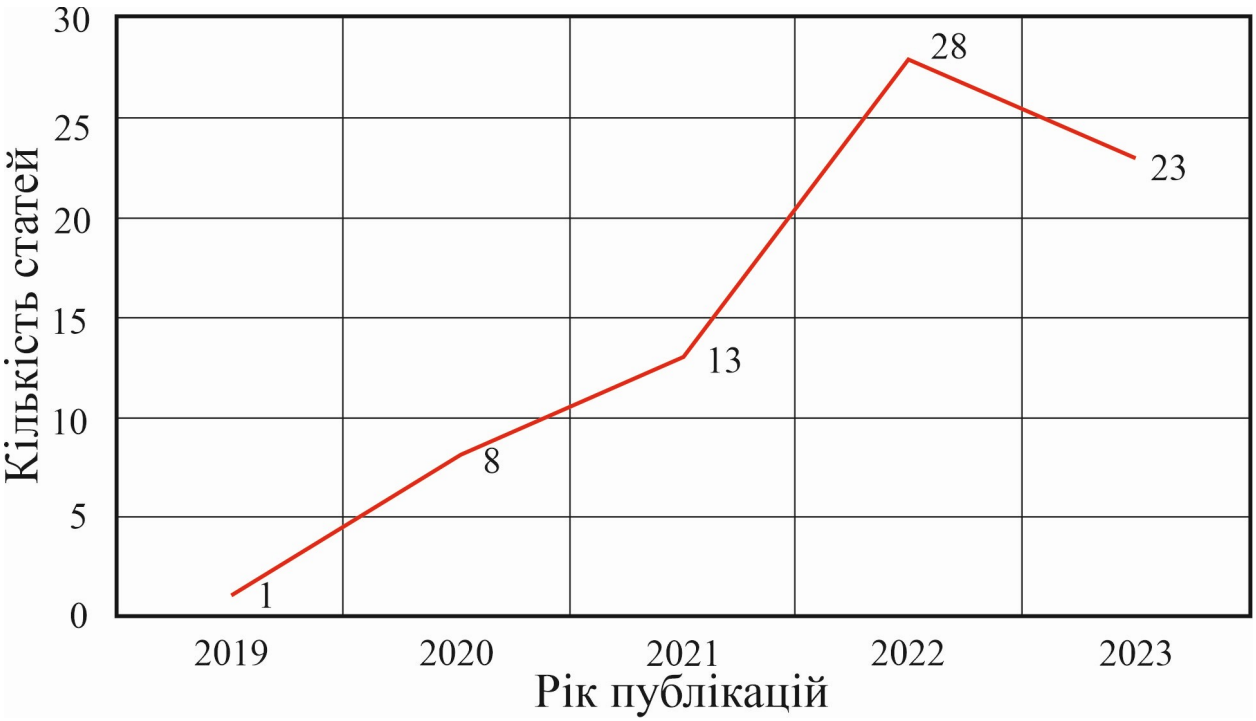

Рисунок 1.1 – Розповсюдження статтей

Схематичне зображення знайдених робіт наведено на рис. 1.2 узагальнює підтеми, визначені у зв'язку з питаннями дослідження. Розгляд зосереджено навколо впливу штучного інтелекту на кібербезпеку з організаційної точки зору, що є основною темою, поясненою у зв'язку з поставленими питаннями.

На рис. 1.2 надано візуальне представлення результатів дослідження. Воно відображає результати, пов'язані з Q1, які згруповані разом, і окремо демонструє висновки для Q2 і Q3. З огляду на те, що було проаналізовано обмежену кількість статей, можна відзначити, що тут висвітлено лише деякі підтеми щодо позитивного чи негативного впливу ШІ на кібербезпеку з організаційної точки зору та порівняння ШІ з традиційними підходами.

Можливості штучного інтелекту лежать у сфері управління вразливістю, особливо у виявленні вторгнень. Він також відіграє вирішальну роль у зміцненні безпеки організаційних мереж і систем від кіберзагроз. Позитивний вплив штучного інтелекту на кібербезпеку охоплює різні аспекти, зокрема передбачення кіберінцидентів і допомогу у відновленні

даних, що зрештою сприяє конкурентоспроможності організації . Ці теми підкреслюють певні тенденції в поширених програмах ШІ, включаючи виявлення вторгнень, покращені заходи безпеки та ідентифікацію шкідливого програмного забезпечення.

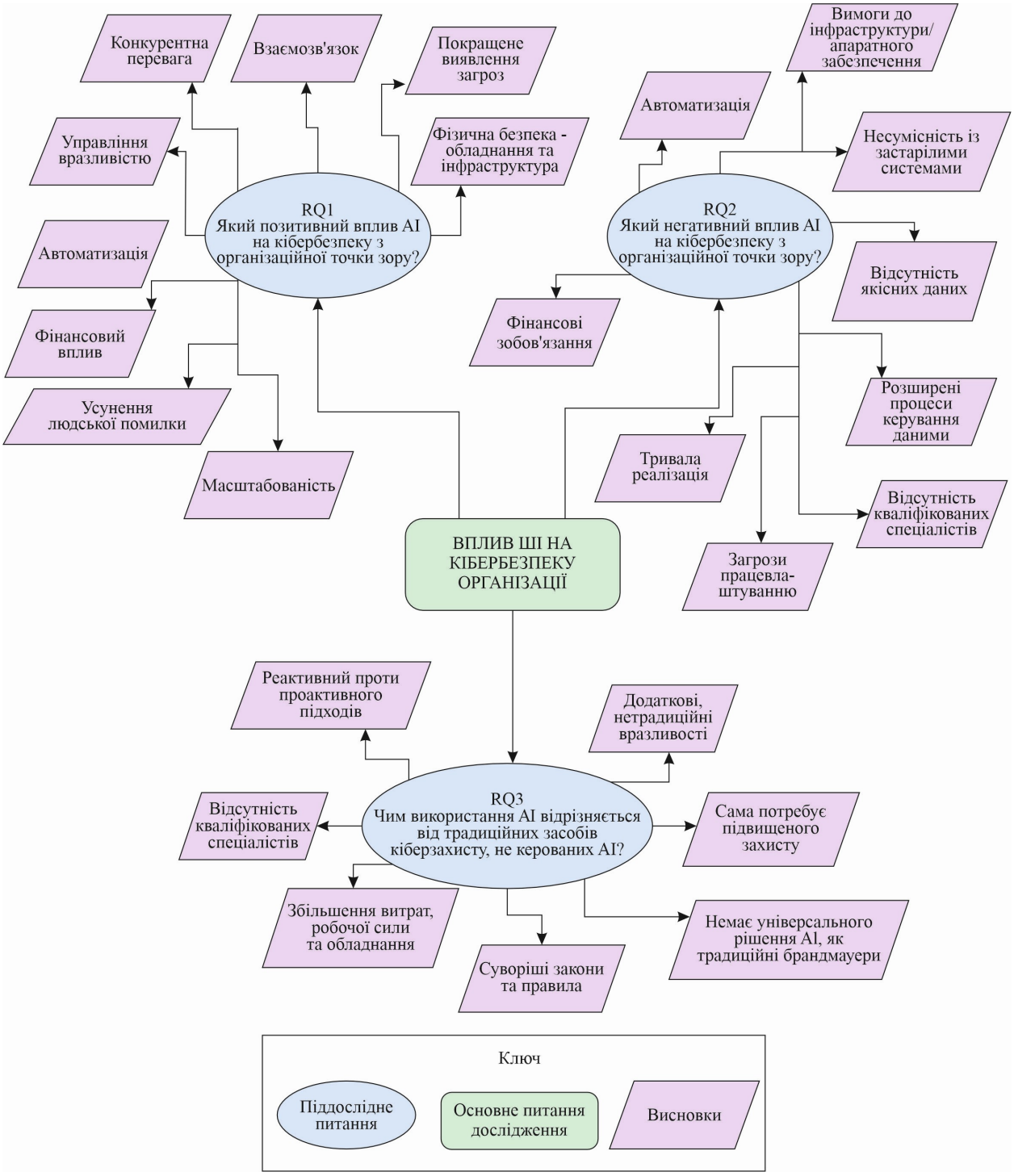

Рисунок 1.2 – Тематична карта

Згідно з результатами, більшість виявлених негативних наслідків були віднесені до категорії унікальних та до категорії «інші» . Крім того, в літературі висвітлено тривожну проблему вразливості систем ШІ до експлуатації, що завдає значної шкоди кібербезпеці організації.

Дефіцит кваліфікованих спеціалістів і відсутність ефективних методів штучного інтелекту для створення прогностичних пояснень сприймаються як фактори, які мають шкідливий вплив на кібербезпеку організацій . Крім того, визнається, що агресивні атаки сприяють проблемам кібербезпеки, оскільки хакери наполегливо намагаються застосувати оманливі тактики, які можуть перехитрити системи безпеки.

На основі результатів експериментів, проведених у деяких розглянутих літературних джерелах, порівнюючи підходи штучного інтелекту та інших підходів, стає зрозумілим, що методи на основі штучного інтелекту перевершують інші методи з точки зору ефективності та точності, пропонуючи підвищену безпеку та покращені можливості для виявлення вторгнень.

Представимо результати огляду літератури, щоб всебічно обговорити загальний вплив рішень на основі ШІ на кібербезпеку організації. У цьому аналізі розглядаються як позитивні, так і негативні впливи. Крім того, щоб отримати повне розуміння загального впливу, було необхідно порівняти ШІ з традиційними методами кіберзахисту для ширшого бачення. Таким чином, також розглядаються відмінності між штучним інтелектом і традиційними підходами до організаційної кібербезпеки, які не пов’язані зі штучним інтелектом.

Існуючі дослідження показують, що зростання кіберзагроз і атак змусило організації прийняти технології на основі ШІ для захисту своїх цифрових активів. У той час як початкова рушійна сила для цього впровадження необхідна для впровадження штучного інтелекту в організаційних налаштуваннях, а його застосування для кіберзахисту призводить до значних конкурентних переваг для організацій. Крім того,

вважається, що це принесе революційні зміни в сучасний кіберзахист і змінить його масштаби. Представимо стислий огляд основних підтем, отриманих на основі аналізу літератури, пов’язаних із позитивним ефектом використання штучного інтелекту в контексті виявлення зловмисного програмного забезпечення, а також у виявленні інших інцидентів вторгнення в мережу або систему. Крім того, сприятливий вплив штучного інтелекту поширюється на сферу адміністрування кібербезпеки, оптимізуючи операційні процедури та підвищуючи загальну зручність.

Розглядаючи людей як найслабшу ланку в ланцюжку безпеки, запровадження автоматизації завдань, керованої штучним інтелектом, усуне вразливі місця, пов’язані з людськими помилками в роботі. Це надзвичайно важливо, оскільки, згідно з [11], людська помилка є основною причиною порушень кібербезпеки. Незалежно від того, чи це помилка, заснована на прийнятті рішення, помилка, пов’язана з навичками, чи помилка у виконанні завдання, навмисна чи ні, усунення потенціалу людської помилки є першим кроком у створенні добре захищеного кіберсередовища. Цієї мети можна досягти за допомогою вдосконалених методів виявлення загроз, коли використання неконтрольованих систем виявлення вторгнень (IDS) дає змогу ідентифікувати навіть найменші загрози чи атаки до того, як вони відбудуться. У минулому це завдання займало години, але за допомогою штучного інтелекту його можна було ефективно виконати за секунди.

Ця автоматизація поширюється на керування вразливими місцями та прийняття рішень, де конкретна гілка штучного інтелекту, нейронні мережі та обробка природної мови (NLP), допомагає в управлінні та визначенні пріоритетів як відомих, так і невідомих загроз. Це досягається шляхом аналізу існуючих загроз, нещодавно виявлених загроз, помилкових спрацьовувань, базових показників поведінки мережі, систем і серверів. Шляхом ідентифікації шаблонів даних і виявлення ненормальної поведінки система може визначити потенційні ризики. Отже, ця автоматизація дозволяє організаціям прийняти проактивний підхід до розпізнавання, передбачення та

усунення знайомих і незнайомих загроз, а не покладатися виключно на реагування після кіберзлому. Як зазначено в [12], автоматизація задач кібербезпеки зменшує потребу в людському втручанні, мінімізує людську взаємодію, а згодом зменшує ймовірність людських помилок протягом усього життєвого циклу безпеки.

З точки зору організаційної кібербезпеки, це не лише передове програмне забезпечення та захисні рішення. Захист фізичної безпеки та життєво важливих компонентів апаратного забезпечення та інфраструктури має вирішальне значення для досягнення організацією комплексного та зрілого кіберзахисту. Багатогранний характер штучного інтелекту (ШІ) може мати позитивний вплив на безпеку обладнання та інфраструктури шляхом оптимізації та моніторингу центрів обробки даних, серверів і процесорів, відповідальних за цей захист. Рішення на основі штучного інтелекту використовують методи машинного навчання для моніторингу таких аспектів, як температура обладнання, системи охолодження, енергоспоживання та резервне живлення. Аналізуючи ці дані разом із історичною інформацією, ці рішення підвищують продуктивність апаратного забезпечення та загальну ефективність інфраструктури. Крім того, впровадження штучного інтелекту допомагає мінімізувати фінансовий тягар витрат на технічне обслуговування обладнання та інфраструктури, необхідних для захисту організації. Це досягається шляхом інтелектуального сповіщення організацій про заплановане технічне обслуговування або прогнозування потенційних збоїв окремих апаратних компонентів, уможливлюючи проактивну заміну до повної поломки. Зрештою, інтеграція технологій штучного інтелекту з обслуговуванням обладнання та інфраструктури може забезпечити фінансову економію для організацій і зменшити загальне енергоспоживання апаратних компонентів.

Іншим позитивним впливом використання кібер-ШІ, помітним у розглянутій літературі, була масштабованість і взаємозв’язок рішень ШІ на більш просунутому рівні. Захист мережі за допомогою систем аналізу мережі

(NAS) і систем захисту мережі (NPS) на основі штучного інтелекту може гарантувати безпеку та доступність комп'ютерних мереж в організації не лише для одного комп'ютера, а й для всієї системи комп'ютерної мережі одночасно. Ці рішення штучного інтелекту можна розгортати на кожному етапі життєвого циклу безпеки, таким чином створюючи більш повне, комплексне та взаємопов'язане рішення.

Хоча впровадження штучного інтелекту в організаційну кібербезпеку здатне досягати ефективності, що перевищує людські можливості, є кілька недоліків, пов'язаних з його впровадженням, особливо на організаційному рівні. Зростання впровадження штучного інтелекту призвело до зростання кількості атак, що підвищило загрозу системам кіберзахисту. Наявність або відсутність відповідних правил і стандартів може перешкоджати впровадженню ШІ в організації. Ці негативні наслідки перешкоджають або затримують широке визнання рішень ШІ як основного підходу до кібербезпеки.

Відповідно до сучасних уявлень, однією з головних перешкод для широкого впровадження ШІ в кіберсфері є його вплив на вимоги до інфраструктури та обладнання. Для ефективного впровадження рішень на основі ШІ на організаційному рівні необхідні значні обчислювальні потужності, можливості обробки та пам'ять. Крім того, більші та досконаліші моделі штучного інтелекту вимагають сучасних центральних процесорів (CPU), які можуть працювати в десять разів швидше, ніж традиційні процесори, що призводить до значних витрат на впровадження. Інша проблема полягає в проблемах сумісності, викликаних постійним використанням застарілих систем, мов програмування та загальної технологічної інфраструктури в багатьох організаціях. Ці застарілі системи не можуть належним чином підтримувати вимоги штучного інтелекту та техніки машинного навчання (ML). Наприклад, аналізу величезних обсягів складних даних, критичного кроку в успішному розгортанні штучного інтелекту та машинного навчання, перешкоджає відсутність

масштабованості, яку пропонують застарілі бази даних і застарілі системи. По суті, впровадження рішень штучного інтелекту в організаціях не є простим завданням, оскільки воно часто вимагає повного перегляду технологічної інфраструктури.

Література постійно висвітлює повторювану тему недостатньої доступності високоякісних, безпомилкових і очищених даних. Рішення штучного інтелекту покладаються на великі набори даних для навчання моделей і отримання точних результатів. Як наслідок, отримання великої кількості даних має важливе значення для ефективного навчання моделей ШІ. Крім того, впровадження кібер-штучного рішення потребує більш складного процесу управління організаційними даними через різноманітність обсягів і типів даних, що зберігаються, швидкість, з якою дані накопичуються, необхідність підтримувати конфіденційність даних і постійну потребу в додаткових даних. Цей аспект особливо важливий, оскільки наслідки рішень ШІ залежать виключно від якості наборів даних, які використовуються для навчання моделей.

Немає універсального рішення кібер-ШІ, яке б відповідало всім ситуаціям, оскільки більшість систем ШІ потрібно певним чином налаштовувати для конкретних організацій. Хоча деякі рішення можна застосовувати на практиці відносно швидко, було помічено, що час, необхідний для впровадження більшості рішень Cyber-AI на організаційному рівні, негативно впливає на їх впровадження. Цю затримку можна пояснити складністю самого сучасного ШІ, і навіть найпростіше рішення ШІ може зайняти місяці або навіть роки, щоб повністю впровадити в організації. Цей подовжений період впровадження в основному пов’язаний з реструктуризацією апаратного забезпечення, отриманням необхідних даних для навчання та тестування моделей, а також наданням достатнього часу для того, щоб моделі осягнули та засвоїли унікальні мережеві та поведінкові моделі, характерні для організації.

Окрім тривалого часу впровадження, було виявлено, що впровадження Cyber-AI є проблемою через його міждисциплінарний характер, вимагаючи ряду спеціалізованих фахівців, таких як спеціалісти з обробки даних, аналітики даних, експерти зі штучного інтелекту, спеціалісти з машинного навчання, розробники, спеціалісти з кібербезпеки та менеджери проектів, кожен з яких має різний рівень технічної експертизи. В [13] також зазначено, що ця велика вимога до кваліфікованого персоналу створює труднощі для організацій, враховуючи поточну нестачу кваліфікованих і досвідчених професіоналів у цих спеціалізованих галузях, які можуть ефективно впроваджувати та керувати рішеннями кібер-ШІ на організаційному рівні. Крім того, організації часто стикаються зі значним фінансовим тягарем при наймі цих дефіцитних спеціалістів. З іншого боку, запровадження кібер-штучних рішень приносить автоматизацію організації, що загрожує багатьом пов’язаним з кіберпрофесіями роботам через його здатність пропонувати швидші, точніші та надійніші рішення кібербезпеки. Хоча малоймовірно, що штучний інтелект повністю замінить потребу в працівниках відзначимо, що лише його впровадження в кібербезпеку створює загрозу для певних ролей, робіт і завдань, пов’язаних з кібербезпекою, оскільки вони можуть стати застарілими із впровадженням рішень кібер-ШІ.

Використання систем штучного інтелекту дає хакерам автономію для використання атак на основі штучного інтелекту, які можуть уникнути захисних заходів на основі штучного інтелекту, що потенційно може призвести до порушення конфіденційності. Ці атаки, керовані штучним інтелектом, мають здатність розвиватися швидше, ніж самі захисні інструменти, завдяки техніці під назвою нейронний фаззинг. Нейронний фаззинг використовує нейронні мережі для виявлення вразливостей у цільових системах, дозволяючи зловмисникам навчатися на існуючих захисних інструментах ШІ. Ненадійність генеративного ШІ також викликає занепокоєння в деяких організаціях через велику кількість помилкових спрацьовувань.

Для того, щоб повністю оцінити вплив штучного інтелекту на кібербезпеку організації, потрібно провести порівняння між традиційними підходами до кібербезпеки та підходами, керованими штучним інтелектом. Хоча існує обмежена кількість літератури, яка безпосередньо досліджує різницю між штучним інтелектом і традиційними засобами кіберзахисту, не керованими штучним інтелектом, було виявлено, що, крім того, традиційний реактивний підхід до кібербезпеки, мета якого полягає в тому, щоб дочекатися атаки, а потім щоб нейтралізувати її, штучний інтелект забезпечує проактивний підхід не лише до реагування, але й до передбачення та вирішення проблем. Це стало можливим завдяки аналізу та прогнозуванню загроз на основі штучного інтелекту, які можуть вивчати минулий досвід і доступні дані для розпізнавання аномальної поведінки. Методи штучного інтелекту продемонстрували кращу продуктивність на відміну від альтернативних методів, про що свідчить ряд досліджень. У проведених експериментах рішення штучного інтелекту незмінно перевершували інші підходи у виявленні вторгнень завдяки їх підвищеній точності і підвищеним можливостям безпеки, що вказує на взаємопов’язаний характер штучного інтелекту з іншими захисними компонентами та створює додаткові, нетрадиційні вразливості в ланцюжку безпеки. Як зазначено в [14], якщо будь-який взаємопов’язаний компонент системи штучного інтелекту скомпрометовано, цілком імовірно, що буде скомпрометовано всю систему. Наприклад, якщо набори навчальних даних, які використовуються для навчання моделей, скомпрометовані, це може вплинути на результат «навчання», і, отже, вся модель може не працювати належним чином. Таким чином, незважаючи на те, що AI забезпечує кращий захист, системи AI вимагають підвищеного захисту протягом усього життєвого циклу розробки безпеки на відміну від традиційних підходів.

Крім того, було виявлено, що різниця між цими двома підходами також випливає з того факту, що не існує універсального рішення, такого як традиційні брандмауери, антивірусне програмне забезпечення або програмне

забезпечення для захисту від зловмисних програм. Кожне рішення штучного інтелекту має бути адаптоване до конкретної організації з використанням внутрішніх і зовнішніх даних організації. Це вимагає не лише більших фінансових і людських зобов'язань, але й більших зобов'язань щодо апаратного забезпечення та інфраструктури для впровадження порівняно з традиційними підходами. Було очевидно, що в той час як традиційні інструменти мережевої безпеки, такі як брандмауери та антивірусне програмне забезпечення, є універсальними, сьогодні вони розглядаються як прикордонний захист. Системи аналізу мережі (NAS), керовані штучним інтелектом, замінили ці традиційні механізми захисту мережі, оскільки вони просто набагато швидші та ефективніші порівняно з традиційними підходами.

Нарешті, впровадження рішень штучного інтелекту на організаційному рівні регулюється більш суворими законами та правилами порівняно з традиційними підходами до кібербезпеки. Незважаючи на те, що штучний інтелект в основному використовується для захисних цілей в організаційній кібербезпеці, деякі уряди та регулюючі органи регулюють високоризикові програми ШІ, щоб забезпечити відповідальне використання такої потужної технології.

2 ЗАГРОЗИ КІБЕРБЕЗПЕЦІ ТА ЗАСОБИ ПРОТИДІЇ ЗА ДОПОМОГОЮ МАШИННОГО ТА ГЛИБОКОГО НАВЧАННЯ

2.1 Дослідницькі виклики та проблеми кібербезпеки

Кібербезпека є динамічною сферою, де виклики будуть постійно зростати, і професіонали або окремі люди повинні бути готові протистояти цим викликам. На рис. 2.1 наведено основні сектори, які постраждали від кібератак у 2021 році. Електронний бізнес, онлайн-транзакції більш вразливі до кіберзагроз. Це призводить до втрати конфіденційної інформації, шкоди репутації та навіть відповідальності за судові позови. Кіберзагрози можуть мати будь-яку форму залежно від мотивів зловмисника, наприклад кіберзлочинність, кібервійна, кібертероризм і кібершпигунство. Кіберзлочинність призводить до різноманітних злочинних дій, що завдає значних фінансових збитків підприємствам і окремим особам. Через кібершпигунство величезна кількість даних, конфіденційної інформації та інтелектуальної власності витягується з веб-сайтів уряду та приватного сектору задля економічної вигоди чи політичних причин. Статистика показала, що 11% кібератак відбувається через шпигунство.

Аерокосмічний і оборонний сектори стикаються з кіберзагрозами з наміром викрасти інтелектуальну власність і оборонні секрети. У кібервійні кібер-зловмисники відстежують, проникають і підривають захист інших країн, щоб порушити їх критичну інфраструктуру. Кібератаки на захист мають каскадні наслідки та порушують систему національної безпеки. Кібератаки здійснюються таємно, щоб послабити або завдати удару противнику для досягнення політичних цілей. Ворог невидимий, а жертва не знає, як і де реагувати. Жодних доказів своєї причетності до цих атак зловмисник не залишив. Нападників називають недержавними зловмисниками. До недержавних суб’єктів належать злочинні організації, сценарні діти, хактивісти, шахраї та хакери. Майбутня війна буде «кібертероризмом» або безконтактною війною, у якій немає «фізичних» або

«кінетичних» дій через кордони, які постійно зростають. У кібертероризмі кіберпростір свідомо використовується для планування терористичних атак. Останнім часом терористи використовують кіберпростір для своєї комунікації, для командування та контролю, для промивання мізків невинних людей, а також для цілей навчання та фінансування. Забезпечення кібербезпеки в системі оборони є складним питанням, яке вимагає багатовимірних, багаторівневих ініціатив і відповідей. Визначення кібертероризму таке: незаконне вторгнення та неминучість атаки на обчислювальні вузли, мережі та критично важливі дані, коли це робиться з метою зловживання чи примушення уряду, державних службовців чи обраних людей, з наголосом на соціально-політичних цілях. Інші терміни для кібертероризму: кіберджихад, електронний джихад, електронний джихад та інтернет-джихад.

Системи супутникового зв'язку, системи навігації та системи спостереження Землі часто становлять загрозу від кібератак. Кібер-зловмисники можуть використовувати механізми програмного забезпечення, підсилювачі, передавачі та керовані антени, щоб заважати або генерувати супутникові сигнали. Уразливості в системах супутникового зв’язку є критично важливими, оскільки вони можуть порушити системи запуску, телеметрію, стеження, командування та зв’язок. Для захисту цих систем космічного базування необхідно вживати заходів постійного моніторингу та захисту.

Безліч кіберзагроз загрожує сектору охорони здоров’я. У багатьох країнах конфіденційність даних у сфері охорони здоров’я викликає більше занепокоєння. Кіберзагрози для сектору охорони здоров’я можуть виникати через зловмисне програмне забезпечення, яке ставить під загрозу роботу системи, або через DDOS-атаки, втрачаючи конфіденційність пацієнтів або порушуючи роботу засобів, доступних пацієнтам. Кіберзагрози в секторі охорони здоров’я мають наслідки, окрім фінансових втрат і порушення конфіденційності. Наприклад, програмне забезпечення-вимагач для лікарень

викрадає дані пацієнтів і ставить під загрозу їх життя. Повідомляється, що понад 18 мільйонів даних пацієнтів постраждали від атак програм-вимагачів. Цей сектор більш схильний до кібератак, оскільки особиста інформація пацієнта та медичні дані були скомпрометовані кібератаками. Дані про хвороби можуть бути використані для шантажу конфіденційної інформації пацієнтів, такої як результати діагностики, тяжкість, типи лікування та захворювання, які надсилаються маркетинговим компаніям для реклами та рекомендацій їхніх продуктів. У системі електронних медичних записів (EHR) зберігаються записи пацієнтів і медичних пристроїв.

Рисунок 2.1 – Середня тижнева кількість атак на організацію за галузями (2021)

Кіберзловмисник може скомпрометувати систему EHR та пристрої, підключені до системи EHR, і може здійснити кібератаки.

Кібератаки в IoT (Інтернет речей), де пристрої, підключені до мережі, більш сприйнятливі, коли зловмисники намагаються захопити IP-адресу, порт програми, DNS-сервер та IP-адресу сервера. Пристрої IoT чутливі до

кібератак, оскільки більшість вузлів IoT постійно підключені та обмінюються даними через Інтернет. Поточний розвиток портативних пристроїв і пристроїв Інтернету речей ще більше посилив наслідки атак шкідливих програм. Як результат, ризики експоненціально більші для пристроїв Інтернету речей. Захист пристрою IoT ускладнюється масштабом і обсягом даних, що генеруються та збираються. Піратство програмного забезпечення та атаки зловмисного програмного забезпечення піддають ризику організації та певні операційні можливості. Ці атаки на IoT із зростанням усюдисущих пристроїв збільшували кількість загроз безпеці. Алгоритми кібербезпеки допомагають захищати хости, можуть захищатися від цих програм і даних і відновлюватися після збою контрольованим і вимірюваним способом.

Хоча існує багато доступних механізмів виявлення кібератак, швидке вдосконалення навичок хакерства та збільшення кількості кібератак вимагає нових систем виявлення кібератак. Останнім часом набули розповсюдження різні цифрові транзакції, такі як онлайн-транзакції електронного бізнесу, онлайн-банківські транзакції, онлайн-транзакції фондового ринку та онлайн-ведення карток пацієнтів. Усі ці онлайн-транзакції схильні до кібератак, коли зловмисник інтерпретує та підслуховує важливу інформацію, пов’язану з бізнес-транзакцією. Подібним чином під час онлайн-банківських транзакцій зловмисники захоплюють облікові дані користувача, облікові дані про позику тощо, щоб знати фінансовий стан клієнтів. Він повторно використовує облікові дані для входу та контролю фінансового стану для створення загроз безпеці.

Атаки на енергосистему країни вважаються вразливими, оскільки важлива інформація про конструкцію ядерного реактора та роботу реакторів може бути передана іншим країнам. Кібератака на енергомережу може призвести до припинення електропостачання та раптового припинення роботи реактора [15]. Основною загрозою для АЕС є кібердиверсія. Кіберсаботаж може фізично вивести з ладу ядерне обладнання, занести віруси/зловмисне програмне забезпечення на електростанцію та призвести до

ядерного вибуху. Деякими прикладами кібератак на атомні електростанції є атака комп'ютерного хробака Stuxnet на атомні електростанції Ірану, кібератака на атомну електростанцію Куданкулам в Індії та злом енергомережі України [16].

Нові покоління кіберфізичних систем (CPS), що складаються з програмного забезпечення та фізичних частин [17] є більш уразливими до загроз і легко порушують цілісність цих систем. Крім того, датчики цих систем можуть бути зламані хакерами, а помилкові дані можуть проникнути в систему, щоб контролер працював зі зловмисними даними. Зловмисник може навіть скомпрометувати приводи, щоб вони не функціонували належним чином [18].

Під час шахрайства з технічною підтримкою кіберзловмисники використовують тактику страху, щоб переконати людей заплатити за завищені «довідкові послуги» для діагностики технічних проблем комп'ютерного обладнання та програмного забезпечення.

Онлайн-ігри стали способом розваги для молоді, але це також створює можливість для ворогів запроваджувати кібератаки. Найпоширеніші загрози кібербезпеці в онлайн-іграх включають розкриття особистої інформації, місцезнаходження, IP-адреси пристроїв, флуд, злом, проблеми з обслуговуванням сервера тощо.

Безпека та конфіденційність даних громадян є головними проблемами розумного міста. Автори роботи [19] обговорили основні проблеми та виклики кібербезпеки в розумних містах. Проблеми кібербезпеки класифікуються з трьох точок зору: управління, технологічна та соціально-економічна. Процес управління використовує інструменти інформаційно-комунікаційних технологій (ІКТ) та Інтернет для надання інформації та державних послуг. Розумні міста мають забезпечувати конфіденційність громадян, надаючи контрольні показники конфіденційності та безпеки для забезпечення безпеки та конфіденційності.

В останні роки спостерігається бум соціальних мереж, які стають все більш популярними. Мільярди людей використовують такі соціальні мережі, як Instagram, Facebook, Twitter, YouTube, LinkedIn тощо, щоб спілкуватися та взаємодіяти один з одним. Однак цільовий спам, фішинг, наклеп, видавання себе за іншу особу, кіберзалякування та підроблені облікові записи є одними із найпоширеніших загроз у соціальній кібербезпеці [20]. Крім того, кібер-зловмисники надсилають користувачам Twitter і Facebook фальшиві повідомлення про скарги щодо авторських прав, які містять шкідливі посилання, і натискання таких посилань може пошкодити пристрої або програмне забезпечення на пристрої.

У наш час багато пошукових систем ранжують веб-сторінки, щоб дати релевантні результати пошуку, коли користувач запитує через пошукову оптимізацію (SEO) [21]. Хоча багато організацій використовують спеціальні методи для розміщення своїх сторінок у результатах пошуку, кіберзлочинці можуть використовувати SEO-отруєння для створення шкідливих веб-сайтів і використовувати тактику оптимізації пошукових систем, щоб вони з’являлися переважно в результатах пошуку. Цей тип атаки також називають пошуковим отруєнням. Блокчейн і криптовалюта поширюються і викликають більше інтересу, ніж будь-коли. Крипто-транзакції є цифровими, і суб’єкти господарювання повинні застосовувати відповідні заходи кібербезпеки для захисту від порушень безпеки, крадіжки особистих даних та інших потенційних загроз. Захищене керування сховищем ключів і недоторкане обчислення є критично важливими для пристроїв Blockchain [22].

Оскільки хмарні обчислення залежать від Інтернету, використання хмарних технологій надзвичайно зростає і стало конкурентною потребою; забезпечення безпеки архітектури є серйозною проблемою. Деякі основні загрози для архітектури could – DoS-атаки, внутрішні ризики, викрадення облікового запису, порушення даних, неправильна конфігурація та зниження видимості інфраструктури.

Внутрішня атака є ще однією загрозою безпеці будь-якої організації. Тут зловмисниками можуть бути поточні або колишні співробітники, ділові партнери, посадові особи, консультанти або рада директорів. Незадоволені співробітники можуть вирішити цілеспрямовано завдати шкоди організації. Співробітники зі зловмисними намірами можуть розкрити секрети організації стороннім особам, оскільки вони можуть знати структуру мережі, уразливість і коди доступу. Хоча необачні користувачі можуть і не мати наміру завдати шкоди, вони мають доступ до інформації організації та конфіденційних даних, які вони випадково розкривають. Кібербезпека є найбільшою проблемою з інсайдерськими атаками, де інсайдерське шахрайство має бути виявлено та виправлено. Інсайдерські атаки становлять серйозну загрозу для CPS, як-от Smart Cities та їх компонентів.

Зв’язок 5G швидко розвивається, забезпечуючи високу швидкість і оперативність для технології бездротового зв’язку. Але нові технології пов’язані з невідомими ризиками, для яких фахівці з кібербезпеки повинні знайти рішення для потенційних загроз. Мережі 5G відіграють важливу роль у розумних містах, автентифікації особи, онлайн-банкінгу тощо. Кібербезпека має важливе значення для захисту транзакцій, пом’якшення крадіжек особистих даних, захисту даних та ідентифікації користувачів, а також додаткових механізмів інтелектуального контролю доступу (IAC) [23].

Відзначається, що кібернетична загроза є глобальною проблемою, і багато країн страждають від неї. На рис. 2.2 показано країни, які постраждали від великих кібератак. У всьому світі зростає потреба у боротьбі з кібератаками та створення механізмів захисту.

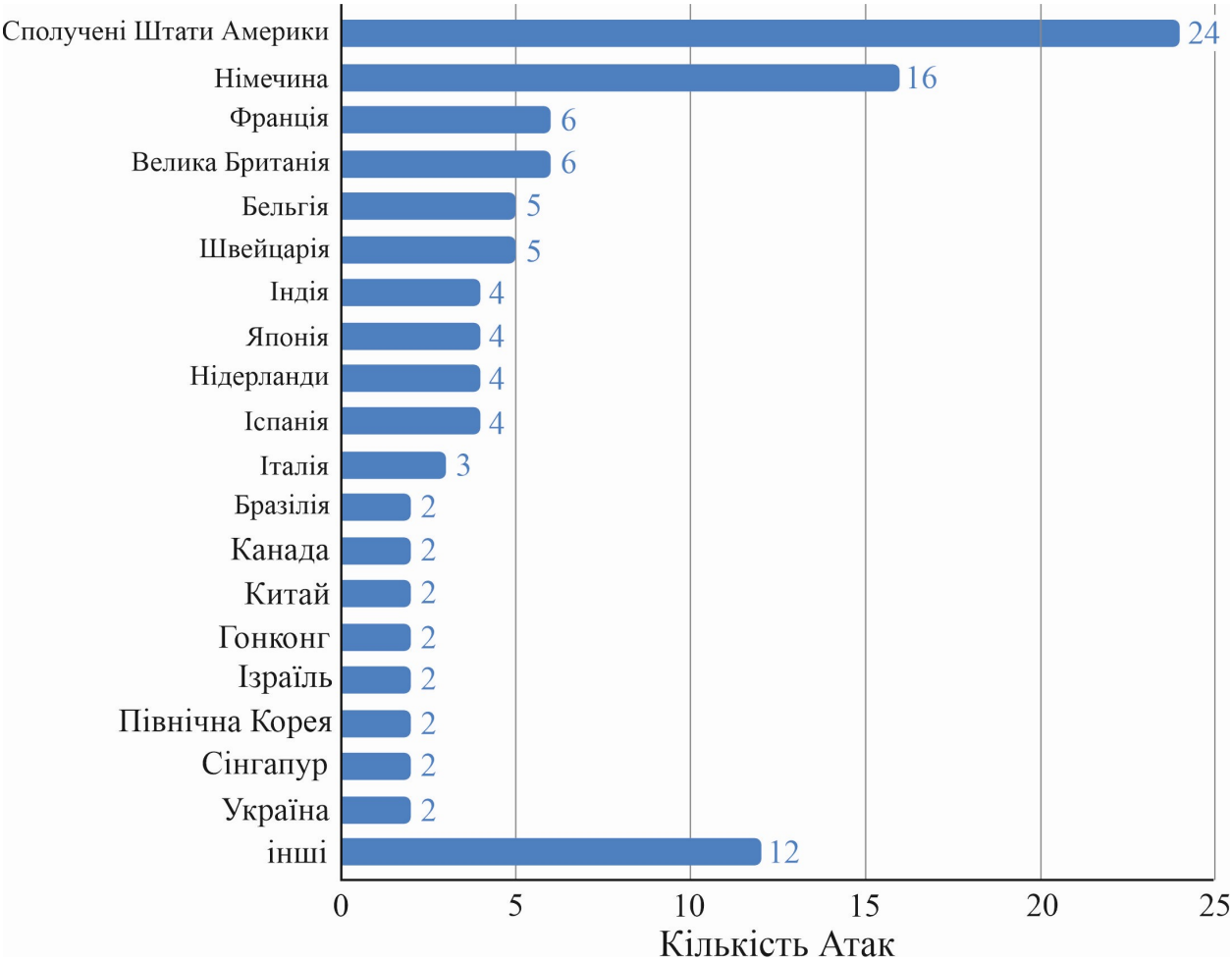

Рисунок 2.2 – Країни, які постраждали від великих кібератак у січні 2022 року

2.2 Домени в кібербезпеці

Обговоримо деякі сфери кібербезпеки. Для цих сфер немає жорстких меж, оскільки вони постійно розвиваються. Основними визначеними доменами є соціальний домен, інформаційний домен, фізичний домен і когнітивний домен. Фізичний домен включає захист системи/настільного комп'ютера та периферійних апаратних компонентів від кіберкрадіжок. Інформаційна сфера фокусується на конфіденційності, цілісності та доступності даних. Інформаційний домен пропонує стратегії для захисту

програм, даних, комп'ютерів і мереж від несанкціонованого доступу або атак. Модель інформаційної безпеки призначена для політики організації, яка забезпечує безпеку даних. Сприйняття даних, аналіз і те, як дані використовуються для прийняття рішень, пояснюють когнітивну сферу. Соціальна сфера стосується норм, етики та політики організації та широкого соціального ландшафту. Технічна реалізація різних форм безпеки включає безпеку додатків, безпеку інформації, управління вразливістю, безпеку мережі, хмарну безпеку, криптографію, безпеку критичної інфраструктури тощо.

Рисунок 2.3 – Різні підходи до кібербезпеки

Для визначення наслідків кібератак використовуються різні методи прогнозування. Таксономія методів передбачення та прогнозування атак у кібербезпеці пояснюється більш детально на рис. 2.3. У методі проекції атаки та методі розпізнавання намірів прогнозуються наміри зловмисника та його наступний хід. Майбутні кібератаки передбачено у передбаченні вторгнень. Прогнозування кібератак на всю мережу виконується в прогнозуванні ситуації безпеки мережі. Існують різні підходи до формулювання загроз кібербезпеці у вигляді моделей, таких як байєсовські мережі, марковські моделі, графіки атак, або безперервні моделі, такі як сірі моделі, часові ряди

тощо. Проблеми кібербезпеки також можна вирішити за допомогою машинного навчання, глибокого багаторівневого репрезентативного навчання та підходів до відкритих знань.

Системи виявлення вторгнень (IDS). Сучасні мережеві підприємства потребують високотехнологічних технологій для захисту організацій. Системи виявлення вторгнень використовуються як засоби безпеки для виявлення можливих вторгнень у мережу або хост [24]. Зловмисник усередині або за межами організації може ініціювати аномальну діяльність, щоб порушити роботу мережі. IDS захищають систему, надаючи автентифікацію користувача та забезпечуючи захисний доступ від неавторизованих користувачів, щоб отримати більше системних привілеїв або зловживати своїми привілеями. Це також гарантує запобігання втраті конфіденційності даних. IDS можуть розрізняти зловмисні та доброякісні дії [25]. Залежно від функціональних можливостей IDS можуть бути мережевими, хостовими та розподіленими.

Методи виявлення, системи виявлення вторгнень можуть працювати на основі правил (також називаються аномаліями), на основі сигнатур (також називаються на основі неправильного використання) під час аналізу та виявлення атак, а також бути гібридними [26].

У системі на основі правил нормальні стани поведінки системи зберігаються в базі даних. Поведінка програми постійно контролюється, якщо будь-які відхилення, окрім цих указаних правил, позначаються аварійними сигналами. Детектор зловмисного програмного забезпечення має збирач даних, який збирає інформацію про програмний інтерпретатор і збіг даних. Програмний інтерпретатор перетворює дані на корисне представлення, засоби зіставлення даних порівнюють інтерпретовані дані з поведінкою програми [27]. Більшість дослідників класифікує виявлення аномалій наступним чином [28]:

1) Аномалії точок – це точка даних, яка розглядається як ненормальна під час порівняння з іншими даними.

2) Контекстуальні аномалії – аномалія ґрунтується на певному контексті.

3) Колективні аномалії – це набір точок даних як набір даних, який розглядається як аномальний.

Продуктивність IDS на основі аномалій краща для невідомих і складних атак, ніж атаки на основі сигнатур [29]. Він добре виявляє невідомі типи атак (також звані експлойтами нульового дня) [30]. Основний виклик с виявлення аномалії полягає у відокремленні нормальної та відхиленої поведінки. Рішення для виявлення аномалій не є стандартними для програм. Це важко передбачити, оскільки шкідливі дії постійно розвиваються. Крім того, уся невидима поведінка розглядається як аномалія, що підвищує частоту помилкових тривог.

У системах на основі сигнатур шаблони атак зазвичай зберігаються в сховищі даних. Ці шаблони атак порівнюються в мережі IDS. У атаках на основі сигнатур уже відомі типи атак можна виявити з високою точністю, і вони не створюють помилкових тривог. Зазвичай IDS на основі сигнатур забезпечує вищу ефективність виявлення порівняно з аномаліями для відомих типів атак. Недолік полягає в тому, що він може ідентифікувати атаки, згадані лише в базі даних. Адміністратор повинен дуже часто оновлювати правила бази даних і підписи. Вилучення різних підписів вимагає багато часу та зусиль. Він не надає точних результатів для атак нульового дня та вірусів, які мають поліморфну поведінку. *«Нульовий день»* відноситься до нещодавно реалізованих уразливостей безпеки, які зловмисники можуть використовувати для атаки на системи. Іншими словами, постачальник або розробник має «нуль днів», щоб це виправити [31].

Інший спосіб виявлення вторгнення відомий як гібридний метод. Цей метод поєднує в собі переваги виявлення аномалій і неправильного використання. Це збільшує частоту ідентифікації вторгнень і мінімізує частоту помилкових спрацьовувань для невідомих типів атак. Більшість

методів ML/DL є гібридним виявленням вторгнень [32]. У гібридних атаках невідомі типи атак ідентифікуються шляхом виявлення аномалій, а відомі атаки виявляються шляхом виявлення зловживання. Гібридне виявлення поділяється на 2 категорії [30] - послідовне і паралельне виявлення. У першому випадку спочатку використовується виявлення неправильного використання або аномалії. В останньому підході кілька детекторів застосовуються паралельно, щоб отримати кілька вихідних сигналів для прийняття рішення. Повну класифікацію IDS представлено на рис. 2.4.

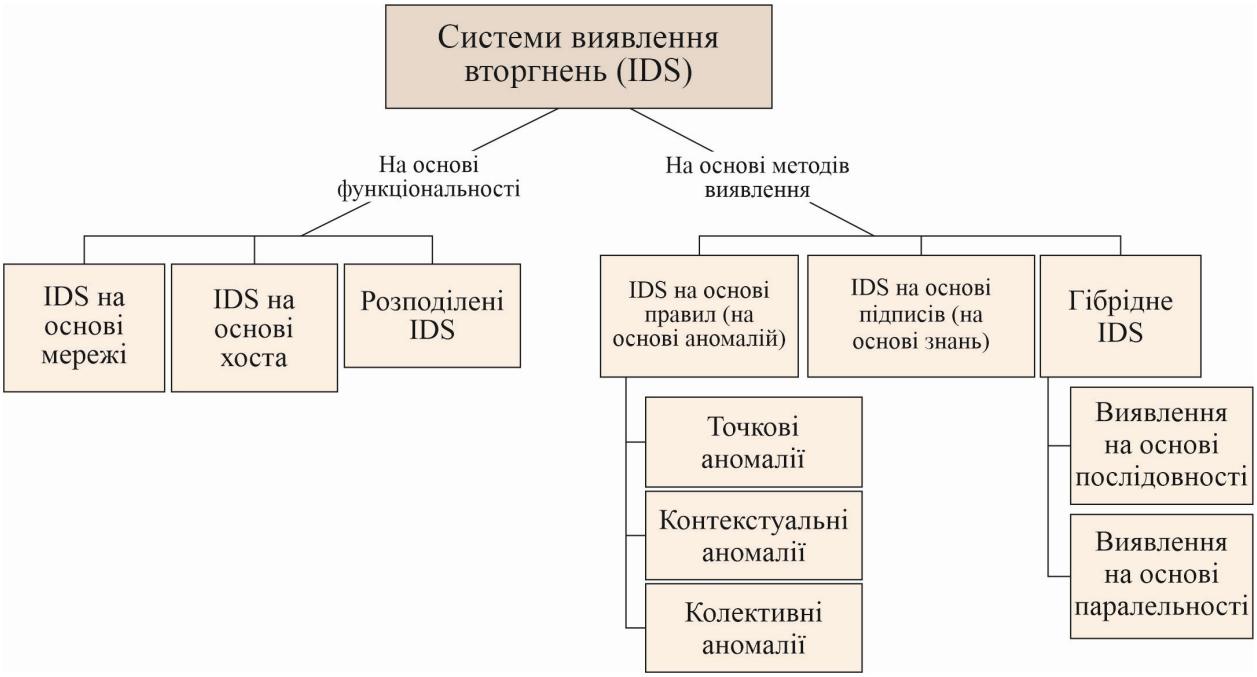

Рисунок 2.4 – Система виявлення вторгнень (IDS)

Система запобігання вторгненням (IPS) — це механізм захисту для взаємопов’язаних пристроїв, який постійно спостерігає за зловмисною активністю в мережі. Він вживає відповідних заходів, щоб запобігти таким діям, блокуючи, видаляючи або повідомляючи про них. Система запобігання вторгненням (IPS) повинна бути системою на основі сигнатур або статистичних аномалій [33]. За допомогою систем запобігання вторгненням можна контролювати доступ до IT-мережі та захищати її від зловживання та атак.

Деякі системи IDS/IPS з відкритим кодом: OSSEC (Open-Source Security), SNORT, Suricata, Zeek, Samhain, Fail2ban, Security Onion, Bro-IDS, Kismet, OpenWIPS -ng, Sagan тощо.

Типи кіберзагроз. Світ кібербезпеки не стоїть на місці. Кіберзагрози змінюються з великою швидкістю. Крім того, тактика захисту кібербезпеки та методи атак змінюються та вдосконалюються щодня. Деякі з основних кіберзагроз наведено на рис. 2.5.

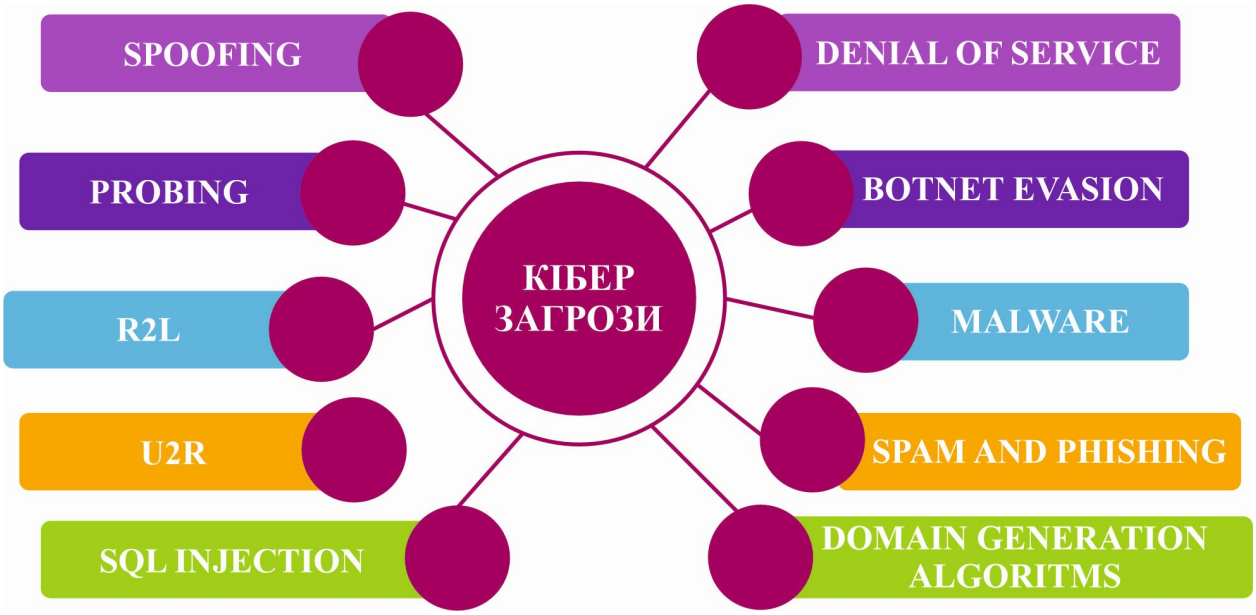

Рисунок 2.5 – Типи загроз кібербезпеці

Кіберзагрози загалом поділяються на наступні категорії.

А. Відмова в обслуговуванні (DoS).

Як правило, банківський сектор, державні організації, комерційні програми, медіакомпанії тощо вразливі до атак DoS. В епізоді відмови в обслуговуванні хакер заповнює системи, мережі або сервери небажаними запитами. Це призводить до того, що ресурси та пропускна здатність сервера виснажуються трафіком зловмисників. В результаті відмови в обслуговуванні система не може виконати законні запити від законних користувачів.

Припустімо, що зловмисники використовують кілька скомпрометованих пристроїв для здійснення DoS-атак; атака добре відома як

розподілена відмова в обслуговуванні (DDoS). За словами кількох дослідників у літературі, доведено, що DDoS має кілька наслідків. Мотивація запровадження DDoS-атак полягає в тому, щоб порушити трафік цільової служби або служби з метою отримання фінансової вигоди, економічного зростання, помсти, ідеологічних переконань, інтелектуального виклику, кібервійни тощо. DDoS-атаки поділяються на різні типи залежно від цілей атаки [34].

1. DDoS-флудове вторгнення на рівні мережі/транспорту

Зловмисник робить недоступними ресурси пропускної здатності мережі. Ця атака затоплення далі класифікується на різні типи наступним чином:

а) Flooding атаки

Під час лавинних атак хакер переповнює смугу пропускання цільової мережі, надсилаючи хибні запити, переважно з пакетами ICMP або UDP, що зрештою порушує роботу законного користувача.

Такі атаки можна ініціювати за допомогою ботнетів.

б) Атаки Flooding з використанням протоколу

Атаки на протокол використовують можливості обробки ресурсів мережевої інфраструктури, таких як брандмауери, сервери та балансувальники навантаження. Вони націлені на протоколи рівня 3 і рівня 4 із шкідливими запитами на підключення.

в) Атаки затоплення на основі відображення

Тут зловмисник підробляє IP-адресу цілі та передає запит на пристрої, які надають послугу. Сервер відповідає та відповідає на IP-адресу цілі. Для цього зловмисник в основному використовує UDP або TCP у деяких випадках, таким чином маючи той самий протокол, що й «Reflection» в обох напрямках.

г) Атаки затоплення на основі посилення

Під час атак із посиленням зловмисники надсилають «пусковий пакет» до пристроїв-рефлекторів, встановлюючи IP-адресу джерела як IP-адресу

цілі. Він, у свою чергу, переповнює машину жертви тригерними пакетами. Зловмисник може надсилати мільйони таких запитів до вразливих служб, тим самим генеруючи значно більшу кількість відповідей, ніж оригінальний запит, і значно збільшуючи розмір і пропускну здатність, виділену цілі.

2. Атака на ресурси системи (атака SYNFlooding)

Ця атака використовує процес рукостискання TCP, необхідний для запуску з'єднання TCP. При цьому зловмисник передає SYN-повідомлення на сервер, на що він відповідає підтвердженням. Оскільки запити є фальшивими, сервер чекає, поки клієнт завершить механізм рукостискання, і повторно передає SYN + ACK безперервно до часу очікування. Зрештою сервер змушений тримати відкритими багато напіввідкритих з'єднань, які з часом перевантажують такі ресурси, як час ЦП, пам'ять та інші ресурси пристрою, часто до точки, коли сервер виходить з ладу.

3. DDoS-атаки на рівні програми

Ці складні DDoS-атаки використовують слабкі місця на прикладному рівні. Він відкриває з'єднання, ініціює процеси та виконує транзакції, які виснажують обмежені ресурси, такі як дисковий простір і доступна пам'ять. DDoS-атаки на рівні додатків поділяються на категорії:

а) Атаки затоплення з відображенням/посиленням: це подібно до атаки на рівні мережі/транспорту

б) Атаки HTTP flooding. Нижче наведено чотири різновиди цього типу атак

I Атаки затоплення сесії

Він виснажує ресурси сервера, надсилаючи на сервер велику кількість запитів на підключення до сеансу, наприклад: атака HTTP get/post flooding.

II Запит Flooding Attack

Під час атаки із затопленням запитів зловмисники надсилають сеанси, що містять більше запитів, ніж зазвичай, що призводить до затоплення DDoS-атак, що відмовляє клієнту в обслуговуванні, наприклад: сеанс HTTP get flood/HTTP post flooding.

III Асиметричні атаки

Під час асиметричної атаки зловмисники передають сеанси, які включають запити з високим навантаженням із кількох HTTP-запитів, вбудованих в один пакет, наприклад: кілька HTTP get/post flood.

IV Атаки повільного запиту/відповіді

У повільній атаці запит/відповідь зловмисник надсилає часткові HTTP-запити, які постійно та швидко зростають, поступово оновлюються та не припиняються. Епізод зберігається, доки ці запити не займуть усі доступні сокети, а веб-сервер стане недоступним, наприклад: повільна атака Loris, атака фрагментації HTTP, повільна пост-атака, атака повільного читання.

Б. Ухилення від ботнетів

Ботнет-атака — це багатоетапна переважна кібератака, яка починається зі сканування мережевих пристроїв. Він заражає пристрої шкідливим програмним забезпеченням, наприклад вірусами [35]. Щоб збільшити масштаб своїх атак, зловмисники можуть отримати контроль над ботнетом без розуміння власника пристрою. Крім того, ботнет переповнює системи в мережах під час DDoS-атаки. Незважаючи на те, що фактичною ціллю ботнетів є комп’ютери, останніми роками противники все частіше націлюються на пристрої Інтернету речей (IoT). У 2016 році ботнет Mirai націлився на півмільйона пристроїв IoT з відкритими портами Telnet і використовував імена користувачів і паролі за замовчуванням для входу на ці пристрої та перетворення їх на зомбі [36]. Метою запуску ботнет-атаки є ініціювання шкідливих дій, таких як створення спаму, реєстрація ключів, порушення авторських прав тощо. Зазвичай боти використовують різноманітні інвазивні підходи, щоб отримати максимальну вигоду. Творець ботнетів широко відомий як Bot Masters, як правило, це особа або об’єднання людей, які мають намір запустити зловмисну діяльність.

Комунікації ботнетів класифікуються на:

1) Централізований ботнет (тобто клієнт-серверна модель)

2) Децентралізований ботнет (тобто одноранговий комунікаційна модель)

3) Гібридна модель

Наприклад, ботнети *Mirai, Muhstik, Toraii, Hakai, Trojan, Gagfyt, Okiru, Kenjiro, Hajime, IRCBot, Hide and Seek* є найпоширенішими атаками ботнетів [37].

В. Атака шкідливих програм

Термін «зловмисне програмне забезпечення» походить від слів «шкідливий» і «програмне забезпечення». Зловмисне програмне забезпечення широко використовується для позначення хробаків, програм-вимагачів, вірусів, шпигунського програмного забезпечення, рекламного ПЗ, троянських програм та інших типів шкідливого програмного забезпечення. Під час атаки зловмисного програмного забезпечення зловмисник порушує вразливі мережеві посилання, коли особа переходить за підозрілим посиланням або відкриває вкладення електронної пошти. Це призводить до встановлення в системі незахищеного або ненадійного програмного забезпечення. Останнім часом велика кількість нових шкідливих програм генерується за допомогою метаморфічних, поліморфних та різних методів уникнення [38]. Спочатку зловмисне програмне забезпечення перебуває в інкубаційному періоді, протягом якого воно буде тихо поширюватися в мережі, заражаючи хости. Під час інкубаційного періоду зловмисне програмне забезпечення не завдає шкоди жодній системі в мережі, і атака починається лише тоді, коли гарантовано, що достатньо систем заражено. У період розширення він поширює всю мережу, запускаючи/заражаючи ботів. Імовірність виявлення зловмисного програмного забезпечення та інкубаційний період є ключовими факторами, які визначають ступінь серйозності атаки зловмисного програмного забезпечення [39].

Зловмисне програмне забезпечення отримало різні назви залежно від його поведінки та призначення. Найпоширеніші типи зловмисного програмного забезпечення включають зловмисне програмне забезпечення,

Cryptojacking, шпигунське програмне забезпечення, рекламне програмне забезпечення, програмне забезпечення-вимагач, троянський кінь, хробаки, руткіти, Man-In- TheMiddle (MitM), бекдори, віруси, бот, Scareware, Man - InThe -Mobile (MitMo), тощо. Останніми атаками зловмисного програмного забезпечення були Shlayer, ZeuS, Agent Tesla, NanoCore, CoinMiner, Delf, Gh0st, Jupyter, Arechclient2, Mirai.

Загалом методи аналізу та виявлення атак зловмисного програмного забезпечення класифікуються на такі групи:

1) Динамічний

2) Статичний

3) Гібрид

Статичний аналіз є швидшим, оскільки вони можуть аналізувати код без запуску та мають справу з хибно-позитивними результатами. Методи, засновані на статичному аналізі, є обчислювально ефективними та безпечнішими. Статичний аналіз не передбачає більш точного прогнозування зловмисного програмного забезпечення, оскільки воно відображається лише для деяких шаблонів. Він може виявити найпоширеніші типи шкідливих програм. Однак він є неефективним для просунутих зловмисних програм, які використовують передові методи виявлення ухилення, такі як поліморфізм і обфускація (де докази зловмисної діяльності приховані).

Динамічний аналіз працює з виконуваним кодом і ефективний проти обфускації. Динамічний аналіз використовує характеристики зловмисного програмного забезпечення та його функції для визначення ступеня серйозності зловмисного програмного забезпечення. Крім того, поведінка функціональних можливостей шкідливого програмного забезпечення визначається після виконання коду в середовищі пісочниці. Динамічний аналіз може виявити будь-які невидимі зразки, оскільки файл аналізується в системах віртуального середовища для підвищення продуктивності. Коли файл буде виконано, динамічний аналіз досягає кращої точності та визначає

всі відповідні шаблони. Гібридні методи використовують переваги як статичних, так і динамічних методів.

Г. Спам і фішинг-атаки

Фішинг — це надсилання шахрайських повідомлень, які, здається, надходять із передбачуваного джерела, зазвичай електронною поштою. За допомогою фішингових атак зловмисники видають себе за надійних контактів і отримують конфіденційну інформацію від користувача. Щоб отримати важливу конфіденційну інформацію, як-от номери кредитних карток, PIN-коди та дані для входу користувачів, зловмисники здійснюють фішингові атаки. Під час фішингу на комп'ютер жертви встановлюється або облікові дані для входу, або зловмисне програмне забезпечення. Фішинг є поширеною кіберзагрозою в соціальних мережах, таких як Twitter, Facebook тощо. Фішингові листи переконують користувачів бездоганними словами та оригінальними логотипами. Фішингові посилання спрямовують на веб-сайти, заражені шкідливим програмним забезпеченням. Фішингові атаки використовують уразливі місця людини більше, ніж уразливості системи. Це змушує користувача вводити свої дані на підроблений веб-сайт, який нагадує законний веб-сайт.

Фішингові атаки поділяють на два основні типи: соціальна інженерія (тобто оманливий фішинг) і фішинг на основі шкідливого програмного забезпечення. Атаки соціальної інженерії зазвичай пов'язані з психологічними маніпуляціями користувачів, щоб вони зробили помилки або поділилися своєю конфіденційною інформацією. Під час фішингу на основі зловмисного програмного забезпечення на комп'ютері користувача запускається шкідливе програмне забезпечення для отримання конфіденційної інформації користувачів. Фішингові атаки на основі зловмисного програмного забезпечення включають DNS-фішинг, викрадення сеансу, фішинг ін'єкцій вмісту, реєстратори ключів, телефонний фішинг, маніпуляції посиланнями, реконфігурацію системи тощо. За допомогою фішингових атак можна встановити зловмисне програмне забезпечення на

машину жертви, яка може перетворити машину жертви на Ботнет і ботнети тепер можуть запускати DDoS або будь-який інший вид атаки.

Нижче наведено різні фішингові атаки.

1) Фішинг на основі алгоритму

Його вперше було виявлено в 1996 році, коли було розроблено алгоритм для генерації випадкових номерів кредитних карток, які збігаються з оригінальними номерами кредитних карток облікових записів America Online (AOL) (Tang і Mahmoud, 2021).

2) Оманливий фішинг

Під час оманливої атаки зловмисник використовує електронні листи або SMS-повідомлення, щоб надсилати шахрайські посилання та обманом змусити людей натиснути посилання. Веб-сайти за посиланнями викрадають і зберігають особисту інформацію жертви.

3) URL-фішинг

У цій атаці зловмисники використовують URL-адресу фішингової сторінки, щоб заразити ціль. Приховане посилання веде на сайт хакера. Коли жертва натискає URL-адресу, вона спрямовується на веб-сайт хакера, який викрадає інформацію жертви.

4) Отруєння файлів хостів

Файл хосту в операційних системах отруєний таким чином, що коли користувач запитує потрібний веб-сайт, він або перенаправляється на інший веб-сайт, або повертає помилку «Сторінку не знайдено». Коли він перенаправляється на підроблений веб-сайт, дані користувача викрадаються. Через отруєння хост-файлу змінюється спосіб, у який ОС розпізнає ім’я DNS.

5) Фішинг із впровадженням вмісту

Фішинг із впровадженням вмісту є поширеною вразливістю веб-безпеки. Уразливі веб-програми роблять фактичний вміст веб-сторінки підробним або зміненим. Фішинг із впровадженням вмісту відбувається, коли програма неправильно обробляє дані, надані користувачем, і зловмисник може надати вміст веб-додаткам.

6) Клонувальний фішинг

У Clone Phishing електронний лист, який надсилається до того, як містить будь-яке посилання, використовується для створення ідентичної копії електронного листа, але зі шкідливим посиланням. Цей новий електронний лист є лише копією оригіналу, але з підробленими посиланнями чи вкладеннями. Ця копія електронної пошти надсилається всім контактам із папки "Вхідні". Особа, яка отримує клоновану пошту, натискає на підроблені посилання, вважаючи, що це законна електронна пошта. Ця атака небезпечна, оскільки одержувачі ніколи не запідозрять електронний лист.

7) Китобійна атака

Китобійна атака завжди спрямована на високопоставлених керівників, таких як генеральні директори, технічні директори та виконавчі директори. Зловмисники зазвичай змушують жертву діяти, наприклад переказувати кошти. Важко знайти ці атаки, оскільки вони часто не використовують шкідливі URL-адреси чи збройні вкладення.

8) Фішинг

Зазвичай фішинг націлений на велику кількість одержувачів, але електронні листи з фішингом ретельно розроблені, щоб отримати дані у формі відповіді від конкретної особи. Хоча рівень ризику високий, фішинг має високий рівень успіху і став одним із основних аспектів, що впливають на безпеку мережі [40].

Фішинг електронної пошти та URL-адреси важко ідентифікувати, оскільки зловмисники часто змінюють свої стратегії. Деякі підходи до захисту від фішингових атак включають інструменти на стороні клієнта, автентифікацію, фільтри та класифікатори на стороні сервера, захист на рівні мережі, а також навчання користувачів [41].

Ґ. Алгоритм генерації домену (DGA)

Це тип атаки, під час якої зловмисники розробляють програмне забезпечення, яке генерує велику кількість псевдовипадкових доменних імен [42]. За допомогою цього DGA зловмисне програмне забезпечення

генеруватиме сотні чи тисячі доменних імен випадковим чином за короткий проміжок часу. Згенеровані доменні імена явно призначаються сайтам. Доменні імена, призначені для сайтів, отримають контроль від зловмисного програмного забезпечення та віддадуть свої вказівки. DGA поширені у зловмисному програмному забезпеченні, яке намагається встановити командний і контрольний зв’язок із ботмайстром та зараженою машиною. Оскільки доменні імена недовговічні, захисникам або аналітикам важко їх виявити. Використовуючи DGA, зловмисники можуть керувати веб-сайтами, які поширюють інфекцію, і розгортати командно-контрольні (C&C). Атака DGA складається з наступних фаз: зараження, C&C, бічного поширення та ексфільтрації даних. Атаки DGA можна загалом класифікувати на двійкові та сценарні, залежно від способу їх розгортання.

Д. Спуфінг

Спуфінг також називають атакою з уособленням. Під час спуфінгу зловмисник викрадає облікові дані користувача, щоб отримати неавторизований доступ до служб. Облікові дані користувача можна отримати шляхом підслуховування в мережі або викрасти з пристрою за допомогою фішингової атаки. Зловмисник пов'язує свою MAC-адресу з IP-адресою незахищеної мережі. Зловмиснику стає легко здійснити крадіжку або видалити дані в цій вразливій мережі. Спуфінг зазвичай можна класифікувати на ARP (Address Resolution Protocol), спуфінг IP, спуфінг DNS [43]. Під час підробки ARP зловмисник надсилає підроблене ARP-повідомлення в локальну мережу. Потім адреса контролю доступу до медіа (MAC) зловмисника приєднується до будь-якого з законних користувачів у локальній мережі. Завдяки цьому зловмисник зможе модифікувати, викрасти або навіть зупинити мережевий трафік. Підробка IP-адреси здійснюється шляхом зміни вихідної IP-адреси, за допомогою якої змінюється ідентифікація відправника.

Підробка DNS відбувається шляхом зміни записів DNS-сервера (який зіставляє доменні імена з IP-адресами). Тепер зловмисник може перенаправити конкретне доменне ім’я до шкідливої або зараженої системи.

Е. Зондування

Зловмисник використовує зондування, щоб дізнатися слабкі місця в системі та отримати доступ до неї. Хакери надсилають пакети сканування в систему та ефективно збирають інформацію та дані. Приклади атак включають Nmap, Satan, сканування портів, сканування IP-адрес, mscan тощо [26].

Є. Remote-to-Local (R2L)

Зловмисник під час атаки R2L визначає вразливість пристрою, надсилаючи пакети через мережу. Потім зловмисник отримує несанкціонований доступ до машини жертви [44]. Зазвичай атаки спричинені переповненням буфера (як у imap, named, sendmail), неправильно налаштованими політиками безпеки (як у ftpwrite) або троянами (xsnoop). Атаку R2L може бути складно виявити, оскільки вона включає функції як на рівні мережі, так і на рівні хоста.

Ж. User-to-Root (U2R)

Під час атаки U2R користувач отримує легальний доступ до облікового запису (цільової машини), за допомогою якого зловмисник намагається незаконно отримати права суперкористувача root, використовуючи сприйнятливість системи [45]. Прикладами цієї атаки є атаки Load Module, Eject, Buffer_overflow і Perl.

З. SQL ін'єкція

Атаки SQL-ін’єкції здебільшого атакують веб-додатки. Лазівки в базах даних веб-сайтів використовуються для компрометації бази даних. Завдяки цьому хакери можуть отримати доступ до конфіденційної інформації користувача на веб-сайті. Хакери можуть навіть змінювати, видаляти або оновлювати інформацію користувача на веб-сайті. Ці атаки дозволяють зловмисникам підробити ідентифікаційні дані, спричинити проблеми з

відмовою, знищити дані або зробити дані недоступними та навіть змінити налаштування адміністратора сервера веб-сайту. Зловмисники можуть отримати доступ до серверних даних на основі таких методів, як Out-of-band SQLi, In-band SQLi (класичний) і Inferential SQLi (сліпий).

Дані та набори даних відіграють вирішальну роль у дослідженнях кібербезпеки, щоб проводити дослідження та оцінювати дослідницьку діяльність у сфері кібербезпеки. Важливо визначити та використовувати відповідний набір даних для проведення дослідницьких експериментів, щоб оцінити значимість і ефективність запропонованих рішень кібербезпеки. Ефективність і впровадження моделей ML і DL залежать від розміру наборів даних, які використовуються для навчання методам ML і DL. Щоб побудувати ефективну IDS, потрібні відповідні гетерогенні та масивні набори даних для навчання запропонованих моделей та оцінки продуктивності запропонованої IDS [46]. Деякі з життєво важливих наборів даних за деякий час зображено на рис. 2.6.

У роботі [47] класифікували доступні публічні набори даних на 7 категорій. Класифіковані набори даних базуються на мережевому трафіку, інтернет-трафіку, електричній мережі, додатках Android для віртуальної приватної мережі, трафіку Інтернету речей і пристроях, підключених до Інтернету.

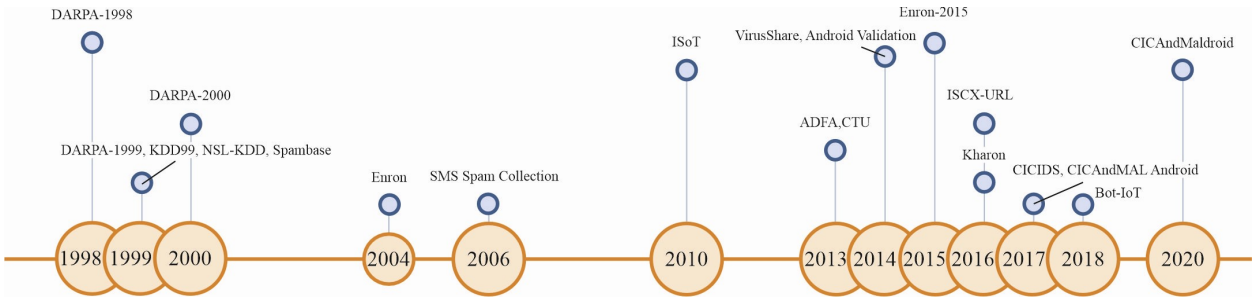

Рисунок 2.6 – Еволюція часто використовуваних наборів даних безпеки

У роботі [48] проілюстрували важливість підходів машинного навчання в системах виявлення вторгнень. Автори використовували дані на рівні пакетів, Netflow і публічні дані для оцінки алгоритму ML.

Лабораторії MIT Lincoln Labs розробили та підтримували кілька загальнодоступних наборів даних. Набори даних зазвичай називаються наборами даних DARPA, і їх набори даних доступні для загального користування. Загальнодоступні набори даних корисні дослідникам для проведення експериментів у сфері кібербезпеки. Різними наборами даних є DARPA 1998, DARPA 1999 та DARPA 2000.

Як правило, продуктивність систем виявлення вторгнень на основі ML і DL оцінюється за допомогою доступних наборів даних, таких як NSL-KDD, KDD CUP 99, UNB ISCX 2012, ADFA, UNSW-NB15, CSE-CIC-IDS 2018, CIC IDS 2017 [26]. DEFCON, CAIDAs, LBNL, CDX, KYOTO, TWENTE, UMASS і ADFA2013 є ще деякими структурами оцінки набору даних IDS.

2.3 Підходи на основі машинного навчання до вирішення проблем кібербезпеки

Алгоритми машинного навчання створюють моделі поведінки за допомогою математичних методів у величезних наборах даних і роблять неминучі прогнози за допомогою нового набору вхідних даних. Методи машинного навчання підходять для механізмів виявлення вторгнень. Машинне навчання (ML) дозволяє комп’ютеру навчатися без явного програмування. Зокрема, ML виконує категоризацію та регресію, встановлену на основі раніше вивчених функцій із набору екземплярів навчання. Стратегія складається з двох етапів: навчання та тестування.

Підходи машинного навчання зазвичай класифікуються як неконтрольовані, контрольовані та методи з підкріпленням.

Алгоритм/система навчається за допомогою набору позначених вхідних і вихідних даних у контрольованому алгоритмі навчання. Навчання виконується за допомогою набору функцій введення та правильного виведення, завдяки чому модель навчається з часом. Тобто навчальний набір даних має цільовий вектор. У той час як у неконтрольованому навчанні алгоритми навчаються на даних навчання, але без будь-якого доступного цільового вектора. У контрольованих методиках використовуються різні алгоритми та підходи до обчислень. Найпоширенішими методами навчання під наглядом є класифікація та регресія, встановлена на цільових мітках, які можуть бути дискретними або числовими. Навчання без нагляду включає зменшення розмірності, оцінку щільності та кластеризацію [49].

Неконтрольоване навчання не потребує позначених навчальних даних для виявлення зловмисної активності, вони найкраще підходять для кібербезпеки порівняно з контрольованим навчанням, для якого потрібні позначені навчальні дані. У навчанні з підкріпленням машина, навчена методом проб і помилок в інтерактивному середовищі з досвідом і прогнозованим результатом, оцінюється на основі позитивної чи негативної винагороди [28]. Основними методами навчання з підкріпленням є апроксимація функції значення та пошук політики.

Алгоритми ML ефективно визначають атаки нульового дня та незвичайні характеристики системи [50].

Підходи, які використовуються в системах виявлення загроз, це карти самоорганізації, опорні векторні машини (SVM), наївні байєсівські класифікатори (NB), байєсівські класифікатори, дерева рішень (DT), класифікатори нейронних мереж [51].

Дослідницький експеримент використовує Feedforward Fully Connected Deep Multi-Layer Neural Network і алгоритми Random Forest. Алгоритми ML застосовуються для ідентифікації вторгнень, аналізу зловмисного програмного забезпечення та виявлення спаму. DGA Detection, Network Intrusion Detection і Botnets зосереджені на виявленні вторгнень. Іншими

алгоритмами машинного навчання, які використовуються в кібербезпеці, є байєсівський підхід – байєсівські класифікатори та моделі Маркова. K Nearest Neighbor (KNN), Naive Bayesian classification, SVM і Neural Networks – це методи машинного навчання, які використовуються для фільтрації спаму.

Також було запропоновано змішаний неконтрольований метод для процесу виявлення аномалій, що поєднує кластерні методи, такі як One-Class SVM (OCSVM) і Subspace Clustering (SSC). SSC є розширенням традиційних підходів кластеризації. SVM – це контрольований підхід, який досліджує дані та визначає шаблони. OCSVM є розширенням моделі SVM і особливо підходить для немаркованих даних. Запропонований метод оцінюється з використанням відомого набору даних NSL-KDD [46].

Зловмисники використовують шкідливі веб-сайти, щоб отримати контроль над системою та впровадити зловмисне програмне забезпечення для збору даних користувачів або шкоди системі. Як правило, зловмисники постійно змінюють URL-адреси шкідливих веб-сайтів. Запропоновано метод класифікації URL-адрес веб-сайтів як зловмисних або доброякісних. Для класифікації автори використовували класифікатори машинного навчання, такі як Gradient Boosted Decision Trees, Random Forests і Deep Neural Networks. Для цих класифікаторів вони використовували функції ContentBased , Host-Based і Lexical із URL-адрес. Підкреслено дрейф веб-сайтів, щоб вирішити проблему активної природи шкідливих веб-сайтів. Веб-дрейфи спостерігаються шляхом зміни зв’язку між вхідними даними та цільовою змінною.

Під час аналізу зловмисного програмного забезпечення підходи ML використовуються для виявлення зловмисного програмного забезпечення та класифікації шкідливого програмного забезпечення за різними категоріями [52]. Під час виявлення зловмисного програмного забезпечення алгоритми класифікують програмне забезпечення як шкідливе або безпечне. Основна проблема шкідливих програм полягає в тому, що вони включають

метаморфічні, поліморфні та інші методи уникнення, які можуть змінити їх поведінку та створити новий тип шкідливих програм [38]. Ці методи обфускації використовуються хакерами проти традиційних методів на основі сигнатур. Представлено також методи виявлення шкідливих програм. Запропонований метод заснований на самоорганізуючих інкрементних нейронних мережах (SOINN) і бінарній візуалізації . Двійкові дані будь-якого файлу перетворюються на зображення, а шкідливий трафік аналізується та виявляється за допомогою SOINN. Перетворені зображення попередньо обробляються, а витягнуті функції передаються в SOINN для кластеризації та класифікації. Подібний процес відбувається на етапі тестування. Алгоритм досягає 74% загального рівня виявлення з помилковими спрацьовуваннями в 12% і помилково негативними результатами в 14%.

Для ефективного виявлення зловмисного програмного забезпечення автори [53] розробили структуру ідентифікації значущих дозволів (SigPID), яка використовує класифікатор SVM. Фреймворк SigPID отримує ефективні дозволи від програм і ефективно використовує витягнуті дані для виявлення зловмисного програмного забезпечення за допомогою контрольованих алгоритмів навчання. Для вилучення значних дозволів автори запропонували підхід багаторівневого скорочення даних (MLDP) із ранжуванням дозволів на основі підтримки (SPR), аналізом дозволів із правилами асоціації (PMAR) і ранжуванням дозволів із негативним коефіцієнтом (PRNR). Потім автори використали класифікатор SVM для класифікації зловмисного програмного забезпечення та безпечних програм. Запропонована структура досягає кращої точності та запам’ятовування при виявленні зловмисного програмного забезпечення, що є основною метою структури.

Класифікатор SVM є ще одним ефективним методом виявлення шкідливих програм. Авторами [37] доведено ефективність класифікаторів SVM у виявленні активності ботнету для домашнього середовища IoT. Використовувані показники ефективності: частота помилкових тривог, частота виявлення та точність тестування. Класифікатори, які

використовуються для виявлення активності ботнету: випадковий ліс, дерева рішень, двокласові нейронні мережі, багатокласові дерева рішень і багатокласові нейронні мережі. Автор дійшов висновку, що продуктивність класифікаторів зростала разом із збільшенням розміру набору даних, кількості та різноманітності шкідливих дій.

Системи виявлення вторгнень використовуються для моніторингу шкідливих дій у системі. Підхід IDS на основі ML включає три категорії, такі як класифікація даних, метод на основі аномалій і кластеризація даних. Класифікація даних – це стратегія контрольованого машинного навчання, у якій набір даних класифікується за різними типами атак. Відхилення від очікуваної поведінки визначаються за допомогою методу на основі аномалій, напівконтрольованої техніки машинного навчання. У кластеризації даних дані кластеризуються на основі шаблонів.

Адаптивний алгоритм Байєса (ABA), штучні нейронні мережі (ANN), KNN, DT і SVM – це методи машинного навчання, які вчені-дослідники широко використовують для виявлення вторгнень. Модель машинного навчання, радіальна функція SVM (RBF-SVM), призвела до найбільшого підвищення точності [50]. Пропонується адаптивна система виявлення вторгнень (Adaptive-IDS), яка називається адаптивно контрольована та кластеризована гібридна система виявлення вторгнень (ASCH-IDS) для класифікації зведених даних. Ця модель використовує методи машинного навчання, а саме випадковий класифікатор на основі лісу дерев як підсистему виявлення зловживань для виявлення відомих атак і вдосконалений класифікатор DBSCAN як підсистему виявлення аномалій для виявлення невідомих атак.

SVM використовувався у розробці системи виявлення вторгнень для запобігання можливим атакам, таким як U2R, DoS тощо. Запропонована методологія використовує SVM для класифікації шаблону зловмисного трафіку від типового шаблону трафіку, який виявляється нелінійним.

Хоча виявлення вторгнень використовує багато алгоритмів машинного навчання, кожен має свої переваги та недоліки. Кожен алгоритм працює по-різному під час різних атак. Ансамбль у машинному навчанні – це техніка, за якої кілька базових моделей машинного навчання поєднуються, щоб отримати оптимальну прогностичну модель. Ці ансамблеві моделі виявилися ефективними у виявленні кібератак. Розроблено унікальну структуру інтелектуальної системи виявлення вторгнень для класифікації багатьох атак на основі набору даних CIC-IDS 2018. Ця техніка ансамблю використовує змішаний режим підходу до вибору функцій із застосуванням випадкового лісу (RF) і аналізу основних компонентів (PCA) Інші алгоритми машинного навчання, які використовуються, це KNN, DT, Extra Trees, Light GBM, Gradient Boosting на основі гістограми (HBGB) і Extreme Gradient Boosting (XGB). Запропоновано метод ансамблю для ефективного вибору функцій і класифікації виявлення мережевих вторгнень для поточних загроз у хмарних обчисленнях. Цей запропонований підхід ґрунтується на техніці однофакторного ансамблевого вибору функцій із скороченими наборами функцій, вибраними з наборів даних про вторгнення, таких як набір даних реального часу Honeypot, Kyoto та NSLKDD.

В роботі [54] використовували техніку K-NN для контрольованого навчання та метод K-Means у класифікаторі KNN для неконтрольованого навчання, щоб підвищити продуктивність класифікатора вторгнень для атак U2R. Щоб досягти цього, автори впроваджують методи зважування ознак і неконтрольоване навчання в процесі KNN. Отримані результати показують, що запропонований підхід може ефективно класифікувати мережеві атаки та значно покращити класифікацію атак U2R.

В роботі [48] запропоновано підходи на основі ML і DL для виявлення атак кібервторгнень і зловживання, які застосовуються в дротових і бездротових мережах. Автор зосередився на виявленні неправильного використання, виявленні аномалій та гібридному виявленні для різних моделей ML і DL, таких як байєсівські мережі, еволюційні обчислення,

штучні нейронні мережі, кластеризація, дерева рішень, правила асоціації та нечіткі правила асоціації, послідовний шаблон Майнінг, індуктивне навчання, машина опорних векторів, приховані марковські моделі та наївний Байєс. Продуктивність цих моделей порівнюється з такими параметрами, як час на навчання моделі, класифікація неідентифікованих прикладів за допомогою навченої моделі ML, розуміння остаточних результатів (класифікація) і точність.

У роботі [55], автори детально описали стратегії ML і Deep Multilayered Representative Learning, які використовуються для виявлення мережевого вторгнення. У своєму дослідженні вони розглядали SVM, KNN, дерева рішень, мережі глибоких переконань (DBN), рекурентні нейронні мережі (RNN) і, нарешті, згорткові нейронні мережі (CNN). Вони висвітлили деякі проблеми, такі як відсутність контрольних наборів даних, нерегулярні показники оцінки та недостатнє вимірювання ефективності алгоритмів.

У дослідженнях [56] використовують підходи машинного навчання для виявлення розподілених кібератак. Робота зосереджена на виявленні зв’язку C&C (командування та управління) між сервером C&C та скомпрометованими ботами. Контакт C&C відбувається на етапі підготовки розподілених атак. Автори використовували 55 функцій для вибору трафіку C&C для раннього виявлення DDoS-атак. Вони використовували в основному PCA і SVM для вибору функцій. Для побудови класифікатора використовуються методи SVM і RF. Експеримент був зосереджений на зменшенні кількості використовуваних функцій і пошуку критичних функцій, необхідних для раннього виявлення зв’язку C&C. Дослідження прийшло до висновку, що, незважаючи на те, що під час виявлення використовується більше функцій, оскільки кількість сягає приблизно 40, ефективність виявлення буде не дуже високою.

У літературі доведено, що алгоритми машинного навчання найкраще підходять для фішингових атак, оскільки вони мають більшість спільних характеристик [57]. Багато результатів, заснованих на алгоритмах машинного

навчання, були представлені в публікаціях для запобігання фішинговим атакам. Однак існуючі рішення на основі ML мають більший час відгуку, високий рівень хибно-позитивних результатів і включають інформацію третіх сторін (неавтентифіковану). Запропонував рішення [58] для фішингових атак, яке виявляє фішингові атаки URL-адрес у реальному часі. Автори використали добре відомі алгоритми, такі як Random Forest, Spearman correlation і K, які найкраще підходять для виявлення фішингових атак. У запропонованій роботі використано дев'ять лексичних функцій для досягнення високої точності з випадковими лісами з дуже малим часом відповіді. Автори провели детальне дослідження часу відгуку, який включає час для вилучення функцій, підготовки набору даних, завантаження модулів і прогнозування результатів як дійсних або фішингових атак. Автори дійшли висновку, що алгоритм Random Forest має найвищий час відгуку, а SVM – мінімальний.

Ефективність інших класифікованих алгоритмів перевірена в [59]. Класифікація використовуються такі алгоритми: DT, K-NN, SVM, логістична регресія (LR), RF і ансамблеве навчання. Автори застосовували злиті класифікатори на основі алгоритмів на основі пріоритетів, таких як алгоритм пріоритету 1 (PA1) і алгоритм пріоритету 2 (PA2). Потім застосовується остаточне злиття на основі пріоритетів, отриманих у PA1 і PA2, щоб досягти точності 97%.

Прогнозування фішингу можна зробити за допомогою різних методів машинного навчання, таких як SVM, дерево рішень, випадковий ліс, наївний байєсівський метод, байєсівська класифікація, K-найближчий сусід і штучні нейронні мережі. Вибір функцій класифікується як функції вихідного коду, функції URL-адреси та функції зображення, і вони базуються на правилах [41,60]. Випадкові ліси та дерева рішень використовуються для виявлення фішингових атак. Набори даних збираються з Kaggle, а вибір функцій здійснюється за допомогою аналізу основних компонентів (PCA). Він визначає та класифікує компоненти набору даних. Дерева рішень

використовуються для категоризації веб-сайту, а для класифікації використовується випадковий ліс. Висока точність була досягнута завдяки Random Forest. Робота [61] пропонує виявлення фішингової електронної пошти на основі автентифікації (SPBA), яка використовує особливості особистості, стилометричні характеристики та ознаки статі, отримані з електронних листів того самого відправника, за допомогою якого створюється модель портрета особи відправника. Для автентифікації в якості класифікаторів використовуються KNN, SVM і Random Forest. Справжній портрет відправника потім порівнюється з портретом невпевненого електронного листа. Якщо його буде виявлено ідентичним, електронний лист розглядатиметься як звичайний, інакше електронний лист класифікуватиметься як фішинг від прихованого відправника. Це дослідження перевершує PHILFER і FSSPD щодо швидкості виявлення та точності.

У роботі [62] показали, що алгоритми ML використовуються в багатьох проблемах, тоді як алгоритми DL в основному використовуються для дослідження зловмисного програмного забезпечення, менше для виявлення вторгнень. Для виявлення спаму використовуються неконтрольовані алгоритми DL. Результати надали переконливі докази того, що методи ML мають недостатню ефективність для кібербезпеки. Відсутність стеження з боку людини може дозволити професійним зловмисникам проникнути, викрасти дані та навіть знищити підприємство. Автори дійшли висновку, що методи ML схильні до агресивних атак, алгоритми потребують постійного перенавчання, а параметри потребують ретельного налаштування. Зловмисник може здійснювати змагальні атаки на алгоритми машинного навчання під час періоду навчання або тестування (виведення) [50].

Змагальне машинне навчання – це метод ML, який призводить до збоїв у роботі машини, надаючи неправильні вхідні дані моделі під час навчання машини. Це змушує машину робити помилкові прогнози. Атака зловмисника може бути цілеспрямованою, коли націлена певна частина навчальної

вибірки, або це може бути випадкова атака, коли націлена будь-яка частина навчальної вибірки. В обох методах кінцевою метою є неправильна класифікація вихідного результату. Ефект суперництва може бути порушенням цілісності, порушенням доступності або порушенням конфіденційності на основі цілей супротивника. Цілеспрямована атака на нейронну мережу, яка призводить до неправильної класифікації, називається порушенням цілісності. Якщо цільова система недоступна для користувачів протягом певного періоду, це називається порушенням доступності. Порушення конфіденційності відбувається, якщо зловмисник успішно скомпрометував конфіденційну інформацію. Однак конкурентні приклади можна використати для підвищення продуктивності та надійності моделей машинного навчання.

Дослідження показали, що методи ML в IDS досягають підвищеної частоти виявлення, але меншої кількості помилкових позитивних результатів. Але також помічено, що алгоритми ML можуть неправильно класифікувати мережеві дані через руйнівне навчання [63,55].

Процес, який змушує алгоритми машинного навчання виконувати небажані дії/функції, називається машинною атакою. Змагальні машинні атаки класифікуються як [49]:

1) Отруєння (також відоме як причинний напад)

2) Атака ухилення і

3) Розвідувальна атака

Атака отруєння – це різновид суперницького вторгнення, під час якого супротивник маніпулює навчальним набором даних моделі машинного навчання. Під час атаки отруєння супротивник надає ретельно розроблені навчальні дані, які вводяться в систему на етапі навчання. Забруднені/отруєні набори даних призводять до неправильної поведінки моделі, що призводить до зниження продуктивності. Це вплине на точність системи. Атаки отруєння можуть бути двох типів: отруєння зі зміною характеристик (ярликів) і отруєння без зміни характеристик. У дослідницькому епізоді супротивник

вивчає модельний алгоритм і може маніпулювати параметрами системи, щоб досягти своїх цілей.

Ще одна широко відома атака - атака ухилення. Під час атаки Evasion шкідливі зразки ухиляються/неправильно класифікуються як дійсні протягом тестового часу. Атака ухилення спрямована на вивчені моделі під час фази тестування, створюючи вибрані супротивником результати. Завдяки атаці ухилення супротивник може пройти через процес тестування, змінюючи тестові зразки, і модель призводить до неправильних результатів. Атаки ухилення можна розділити на три типи. Атака «чорна скринька» є найбільш часто використовуваним типом атаки, коли зловмисник не матиме жодних знань про моделі ML/DL. У атаці «білого ящика» хакер має дозвіл на доступ до параметрів прототипу, тоді як у моделі «сірого ящика» зловмисник має помірні знання про модель [64,65]. Атаки на етапі тестування включають Deep Fool, Fast Gradient Sign Method (FGSM), метод на основі оптимізації, підхід Saliency Map Approach на основі Jacobian (JSMA) тощо. Техніки запобігання атакам на моделі ML класифікуються на чотири типи: механізми оцінки безпеки, контрагент на етапі навчання та тестування, безпека даних і конфіденційність. Деякі приклади захисних методів: змагальність, метод ансамблю, дедуплікація даних, безпечна дедуплікація даних, сенсибілізація даних, відмова від негативного впливу (RONI), шифрування на основі ідентифікації, дистиляція захисту, диференціальна конфіденційність, рішення на основі блокчейну, гомоморфне шифрування тощо.

Також пропонуються методи атаки чорної скриньки для моделей, які виявляють аномальний потік мережі за допомогою алгоритмів машинного навчання. Запропонований метод створення змагального прикладу «Чорної скриньки» використовує атаку «Білої скриньки» на замінну модель. Цільова модель і модель-замінник навчаються ідентично на наборі даних KDD99 і на наборі даних CSE-CICIDS2018. Зловмисник може здійснити атаку на замінну модель методом білого ящика. Ці створені змагальні приклади потім використовуються в цільовій моделі, щоб перевірити, чи можуть ці змагальні

приклади неправильно класифікувати цільову модель. Результати експерименту показали, що автори ефективно генерували змагальні приклади на основі мережевого потоку, який може ввести в оману моделі виявлення, засновані на машинному навчанні.

Загалом зловмисники використовують алгоритми змагального машинного навчання (AML), тому алгоритми машинного навчання неправильно класифікують доброякісну вибірку. Основна причина полягає в тому, щоб зробити модель машинного навчання несправною. Для цього противника використовуйте дані про отруту. Ці дані можуть бути призначені для використання певних вразливостей і компрометації результатів. Деякі з моделей AML – це Droid API Miner, Mystique, Pin droid і Droid Chameleon, які знижують рівень виявлення класифікації моделей машинного навчання. Змагальна класифікація може бути хибно негативною або хибно позитивною. У False positive зловмисник неправильно обчислює негативний екземпляр, щоб класифікувати його як позитивний. Навпаки, у хибно-негативному результаті доброякісні дані додаються зі зловмисним програмним забезпеченням, щоб воно могло обійти виявлення.

В [65] пропонується алгоритм оптимізації мурашиної колонії (ACO) для створення зразків отруйних шкідливих програм. У цьому підході спочатку застосовується алгоритм лінійної регресії, щоб вибрати екземпляри зловмисного програмного забезпечення, майже ідентичні доброякісним прикладам у навчальному наборі даних. Далі функція ACO використовується для пошуку вибіркових даних противника. Використане значення феромону ACO – це кількість змінених функцій. Алгоритм починається з однієї функції, а нові зразки створюються шляхом модифікації зразків зловмисного програмного забезпечення без атрибутів, наявних у законних програмах. Це повторюється за допомогою додаткових функцій. Оцінюється відстань між нещодавно згенерованою вибіркою та дискримінатором. Якщо воно знаходиться в межах зазначеного шкідливого програмного забезпечення та діапазону дискримінатора, щойно створений зразок додається до нещодавно

розроблених зразків; в іншому випадку відкиньте цей зразок. Значення ознак змінюються, а відстань перераховується. Це повторюється до максимальної ітерації, або класифікатор неправильно класифікує зразки шкідливих програм.

Доменні імена, згенеровані DGA, зазвичай виявляються шляхом вилучення особливостей трафіку DNS і статистичних характеристик мови доменних імен. Пізніше алгоритми ML аналізують витягнуті характеристики, щоб ідентифікувати та класифікувати доменні імена DGA. Автори [66] використовували глибоку нейронну мережу LSTM, щоб запропонувати модель виявлення доменних імен DGA. В роботі [67] представили модель обробки загроз DGA, оскільки звичайні підходи до контролю зловмисного програмного забезпечення (наприклад, чорний список) не можуть з ними впоратися. Стаття присвячена структурі машинного навчання, яка може ідентифікувати та виявляти атаки DGA. Він також пропонує техніку глибокого навчання (DNN) для організації великої кількості доменних імен. У цьому дослідженні представлено структуру машинного навчання з дворівневою моделлю та моделлю прогнозування. На першому рівні класифікації документ визначає Decision Tree-J48 як найкращий класифікатор серед NB, ANN, LR, SVM, RF і Gradient Boosting Tree (GBT) для класифікації доменів DGA. Алгоритм класифікації DT-J48 працював з високою точністю та мінімальним часом класифікації. Фреймворк використовує алгоритм DBSCAN для кластеризації другого рівня, яка є кластеризацією на основі щільності. Оскільки модель HMM добре працює завдяки швидкому часу виконання та підвищеній точності відповідності, її використовують для аналізу результатів кластеризації. У порівнянні з алгоритмом класифікації DT-J48 модель DNN краще працює для класифікації великих наборів даних. Дослідження екстраполює, що алгоритми глибокого навчання працюють краще порівняно з алгоритмами машинного навчання для класифікації великих наборів даних.

Сьогодні підходи ML сприйнятливі до змагальних випадків через Generative Adversarial Networks (GAN). Це неконтрольована техніка ML, яка поєднує в собі генератор і дискримінатор [26]. GAN створює серйозні проблеми для програм кібербезпеки, які є критично важливими для безпеки. Потрібна додаткова робота, щоб вивчити вплив змагальних прикладів у кібербезпеці. Генератор створює дані з випадкового розподілу, які можна легко прийняти за справжні дані, а сегрегатор (дискримінатор) відокремлює справжні дані від хибних. Вони вивчають розподіл даних за допомогою неконтрольованих методів. Генератором є згорточна нейронна мережа, а дискримінатором є згорточна нейронна мережа DE. Дані, створені генератором, відповідають розподілу ймовірностей навчальних даних, тоді як дискримінатор відрізняє дані навчання від згенерованих даних. Згенеровані зразки можуть підвищити ефективність виявлення. GAN можна використовувати для вирішення проблем із відсутніми даними, щоб створити негативні зразки, щоб задовольнити негативні зразки, необхідні для навчання глибоких мереж.

Також використовується метод грубої сили Black-Box для запуску вторгнення в системи, які працюють з машинним навчанням. Запропонований метод виявляє виявлення вторгнень у мережу (NIDS), оскільки методи ML є вразливими до змагальних прикладів. Метод грубої атаки (BFAM) оцінює стійкість класифікаторів машинного навчання у виявленні кібербезпеки. Він використовує показники достовірності з цільових класифікаторів для розробки змагальних прикладів, щоб BFAM можна було використовувати для інших змагальних вторгнень у кібербезпеку.

Технології ML можна ефективно використовувати для захисту від кібератак, крім того, системи на основі ML використовуються агресивно проти всіх типів атак. [68] вивчав різні моделі AI/ML для захисту кібербезпеки. Автори також перераховують неправильне використання ШІ/ML для загроз кібербезпеці. Як правило, моделі штучного інтелекту/ML,

фреймворки та інструменти доступні з відкритим кодом, хакери можуть легко адаптувати ці моделі для своєї вигоди. Змагальні моделі атак на основі AI/ML відрізняються швидкістю, автоматизацією, масштабом і витонченістю. На основі діяльності/дій, класифіковано кібератаки за допомогою ШІ/ML на зондування, сканування, підробку, затоплення, неправильне спрямування, виконання зловмисних процесів та обхід.

В [69] систематично пояснюється продуктивність алгоритмів машинного навчання по-різному для різних програм кібербезпеки. Автори дійшли висновку, що краще використовувати комбінацію моделей класифікації.

2.4 Рішення глибокого навчання для кібербезпеки

Глибоке навчання (DL) вважається підкатегорією ML, яка створює багатошарову нейронну мережу [70,71]. Алгоритми глибокого навчання довели, що вони можуть подолати обмеження алгоритмів машинного навчання. Алгоритми глибокого навчання виграють від традиційних алгоритмів машинного навчання, де функції високого рівня автоматично генеруються з наявних функцій.

Алгоритми DL знижують вимоги до розробки функцій і простору для функцій. Вони можуть ефективно працювати під контролем, без контролю та частково під контролем. Алгоритми DL обробляють величезні набори даних і можуть ефективно обробляти неструктуровані дані. Алгоритми DL відіграють важливу роль у вирішенні проблем у різних областях досліджень, таких як: обробка зображень, біоінформатика, ігри, розпізнавання мовлення, виявлення об’єктів, сегментація, класифікація, розпізнавання образів та зіставлення, автоматизація управління взаємовідносинами з клієнтами, система автоматизації транспортних засобів тощо.

Надійність, швидкість, точність і здатність обробки великого обсягу даних методів глибокого навчання привернули увагу дослідників в останні роки.

Алгоритми глибокого навчання (DL) ефективно виявляють передові загрози кібербезпеці. Очевидно, що методи DL можна використовувати для вирішення проблем кібербезпеки. Алгоритми глибокого навчання можуть ідентифікувати відомі та невідомі атаки, можуть керувати неповними, непослідовними та складними даними [72]. Досліджувалися різні алгоритми DL, а потім класифікувалися алгоритми DL на генеративні (неконтрольовані), дискримінаційні (контрольовані) та гібридні.

У роботі [73] запропонували структуру для моделей DL і дослідили використання багаторівневих моделей навчання для виявлення кількох проблем кібербезпеки, таких як вторгнення, зловмисне програмне забезпечення, спам-фішинг і пошкодження веб-сайту. Автори використовували генеративні моделі глибокого навчання замість дискримінаційних або гібридних підходів. Автори підкреслили переваги напівконтрольованого навчання для немаркованих даних.

Deep Boltzmann Machine (DBM), DBN, CNN, Restricted Boltzmann Machine (RBM), Deep Neural Network (DNN), Deep Reinforcement Learning (DRL), Generative Adversarial Network (GAN,) Stacked Autoencoder (SAE), LSTM, RNN, Deep Auto Encoder (DAE), глибока нейронна мережа прямого зв'язку та комбіновані механізми глибокого навчання корисні для виявлення кібератак [26].

У порівнянні з класичними методами машинного навчання, глибокі мережі можуть автоматично отримувати функції з даних, зменшуючи зусилля на попередню обробку вхідних даних і не покладаючись на функції, створені людиною. Це робить алгоритми глибокого навчання придатними для обробки в режимі реального часу. Але продуктивність алгоритму DL знижується, якщо алгоритми не забезпечені достатньою кількістю відповідних навчальних даних [74,75]. Існуючі методи машинного навчання

не масштабуються на величезний обсяг даних, тому виявлення кібератак у великих слабкозв'язаних пристроях є серйозним викликом. Помічено, що методи ML неефективні у виявленні внутрішніх атак або неідентифікованого шкідливого програмного забезпечення та дуже погано зберігають конфіденційність користувачів [45].

Методи DL можуть подолати недоліки моделей ML для існуючих рішень кібербезпеки. DL має потенціал для обробки складних шаблонів і створення міцних і надійних моделей. Методи DL є швидшими та точнішими в обробці, оскільки вони мають можливості самонавчання, які покращують швидкість обробки, а також точність додатків [76]. Методи DL підходять для виявлення зловмисного програмного забезпечення, виявлення мережевих вторгнень, атак DDoS, виявлення фішингу/спаму, виявлення аномалій поведінки, виявлення ботнетів і виявлення пошкодження веб-сайту [34].

Автори [67] використовували різні штучні нейрони мережеві методи, такі як CNN, LSTM і FCNN, для розробки системи штучного інтелекту, безпеки інформації та управління подіями (AI-SIEM). Запропонована модель може розрізняти істинні позитивні та хибні позитивні повідомлення. Ця модель дозволяє аналітикам кібербезпеки виявляти кіберзагрози та швидко захищатися від них. Автори зробили висновок, що AI-SIEM має значення для моделей виявлення мережевих вторгнень на основі навчання. Вони також дійшли висновку, що численні підходи до глибокого навчання можуть бути ефективно використані для покращення прогнозування загроз, щоб уникнути кібератак.

Дослідження [55] висвітлює відмінності між методами DL і ML, які використовуються для кібербезпеки. Алгоритми DL добре працюють, коли доступний великий обсяг даних, і для цього потрібні високопродуктивні машини з графічними процесорами, які не підходять для алгоритмів ML. У машинному навчанні вилучення функцій виконує експерт, тоді як у глибокому навчанні алгоритм намагається автоматично витягти функції. Продуктивність алгоритму ML оцінюється на основі точності витягнутих

ознак, чого немає в алгоритмах DL. Що стосується методів вирішення проблем, ML поділяє проблему на підпроблеми, а потім вирішує ці підпроблеми, тоді як алгоритми DL виконують наскрізне вирішення проблем. Період навчання довший у моделях DL, але час тестування набагато менший у порівнянні з алгоритмами ML. Алгоритми машинного навчання можуть працювати на будь-якому звичайному процесорі, але для запуску алгоритмів глибокого навчання потрібні високопродуктивні машини. Ручне виділення ознак виконується в підходах ML, тоді як алгоритми глибокого навчання автоматично витягують абстрактні та гнучкі функції шляхом узагальнення в класифікації. [77] у своїй роботі продемонстрували, що підхід глибокого навчання LSTM перевершує класифікатори машинного навчання, такі як J48, RF, KNN, NB, DT, і алгоритми для ефективного виявлення вторгнень грубою силою FTP і SSH.

У роботі [47] провели вичерпне дослідження систем виявлення вторгнень, наборів даних, а також порівняльний аналіз різних моделей DL. Автори використовували стратегії глибокого навчання, такі як DNN, RNN, RBM, CNN, DBN, DBM і DA, для виявлення таких вторгнень, як Brute Force, DoS, DDOS, SQL Injection і Botnet-атаки, і порівнювали їх з різними підходами машинного навчання, такими як RF, NB., SVM, ANN щодо глобального рівня виявлення. DBM, RNN і CNN це моделі DL, вбудовані для виявлення мережевих вторгнень. Автори [78] перерахували компоненти, задіяні в IDS для підвищення безпеки мережі. Компонентами IDS є збір даних, вибір функцій і механізм прийняття рішень. Третій компонент – критичний, коли зібрані дані класифікуються як безпечні або шкідливі на основі попередньої інформації.

Непросто знайти аномальні особливості у великих зразках мережевого трафіку. Нейронні мережеві автокодери з упередженим зв’язком найкраще підходять для виявлення мережевих аномалій, оскільки легко навчити вхідні дані та реконструювати вихідні дані [79]. Автоенкодери — це підхід до навчання нейронної мережі без вчителя. Автоенкодери зменшують

розмірність, стискаючи вхідні дані та перебудовуючи вихідні дані з їх представлення. Вони можуть виявити структуру в даних для створення стисненого представлення вхідних даних. У дослідженні [79] представили новий метод на основі Autoencoder, що складається з п’яти рівнів для виявлення аномального трафіку в мережі. Цей підхід перетворює вхідний набір даних у збалансовані набори даних щодо розміру та типів даних шляхом видалення викидів і уникнення упередженості при виявленні аномалій. У 5-рівневій архітектурі прихований рівень має оптимізовану кількість нейронів, а рівень латентного простору забезпечує найкращу продуктивність порівняно з іншими архітектурами.

Автоенкодери також можна використовувати для вивчення та вилучення функцій. Автори [80] використовували глибоке навчання функцій із багатоканальним доступом для виявлення вторгнень у систему. Фреймворк MINDFUL (MultI-chanNel Deep FeaUre Learning) використовує автоенкодери. Автоенкодери реалізовані [81] для виявлення атак нульового дня. Це дослідження намагається подолати недоліки виявлення атаки нульового дня на основі викидів, яке мають велику кількість хибнонегативних результатів. Автори побудували модель IDS, щоб зменшити частоту хибнонегативних результатів (тобто частоту пропусків) із високим рівнем запам’ятовування (тобто частоту справді позитивних). Автори використовували набори даних CICIDS2017 і NSL-KDD. Вони відзначили відмінний показник точності в порівнянні з однокласовою опорною векторною машиною (SVM).

Неконтрольований стековий автоматичний кодер (SAE) поєднується зі зваженим вибором функцій [82], щоб покращити процес вивчення функцій для IDS. Автори довели, що SAE є ефективним і цінним для механізмів вилучення ознак, кластеризації та класифікації. Результати перевірялися за допомогою набору даних Aegean Wi-Fi Intrusion Dataset (AWID), що складається з класів доброякісності, ін’єкції, імітації та затоплення. Автори дійшли висновку, що IDS, яка використовувала SAE як класифікатор,

призвела до низького рівня виявлення уособлення. Таким чином, SAE можна використовувати як класифікатор, а не засіб виділення ознак.

Вичерпне дослідження виявлення вторгнень на основі глибокого навчання запропоновано в [83]. У цій роботі використовується методологія Adaptively Supervised and Clustered Hybrid (ASCH-IDS). Ця модель виявлення вторгнень, Restricted Boltzmann-based Clustered IDS (RBC-IDS), призначена для критичних програм на основі бездротових сенсорних мереж. Результати показали, що IDS на основі ML бажана, якщо вона схожа на IDS на основі DL щодо точності, навчання та часу тестування для моніторингу критичної інфраструктури на основі WSN. В роботі [84] проведено серію експериментів з виявлення вторгнень за допомогою DBN. Завдяки цим експериментам стало можливим ідентифікувати невідомі атаки та після 50 ітерацій досягти 97,5% точності.

У роботі [85] розробили дві методи глибокого навчання, такі як пакетний градієнтний спуск і стохастичний градієнтний спуск, які порівнюються та перевіряються на методі повторної вибірки для кібербезпеки. Пакетний градієнтний спуск — це ітераційна техніка, яка використовує повні шаблони навчання введення для оптимізації функції витрат. У стохастичному градієнтному спуску шаблони навчання вибираються випадковим чином для оновлення вагових коефіцієнтів. Автор дійшов висновку, що Stochastic Gradient Descent забезпечує ефективний алгоритм оптимізації для кібербезпеки з хорошою продуктивністю та меншими обчислювальними витратами.

В роботі [46] описано модель виявлення вторгнень на основі DBN. Автор порівнює фундаментальні алгоритми, різні методи навчання та набори даних та інтерпретує результати різних досліджень, починаючи з 2016 року. Виявлення вторгнень на основі DBN використовує ADFA, NSL-KDD, UNSW-NB15 і KDD Cup 99. Структура на основі DBN-IDS складається з таких компонентів, як навчання попереднього процесора даних, класифікатор, оптимізатор і алгоритм тонкого налаштування.

Шкідливі дії виявляються з мережевого трафіку за допомогою виявлення аномалій. Для виявлення аномалій було запропоновано багато методів глибокого навчання. Щоб виявити аномалії, [86] розробили кластер підходів, заснованих на структурах варіаційного автокодувальника (VAE), повністю підключеної мережі (FCN) і LSTM Seq2Seq, і дійшли висновку, що методи глибокого навчання є правильним вибором для виявлення переконливих мережевих аномалій. Автори перевірили запропоновані архітектури з різними публічними наборами даних трафіку, включаючи IDS2017, UNSW-NB15, Kyoto-Honeypot і NSL-KDD. Під час попередньої обробки даних числові характеристики нормалізуються за допомогою z-показника, а категоріальні ознаки перетворюються на числові за допомогою одноразового кодування — попередньо оброблені дані надходять у підключену мережу для навчання. Автори розглядали ReLU як функцію активації в прихованих шарах. Рівень Softmax створює кінцевий результат із функцією перехресної ентропійної вартості, яка може бути нормальною або атакуючою. Далі тестуються два варіанти моделей VAE, такі як моделі VAE-Pure і VAEFCN. Вихідні дані та виявлені дані порівнюються для обчислення втрат. Модель LSTM-Seq2Seq базується на RNN, яка дає цільову послідовність та умовну ймовірність через кодер і декодер. Структура LSTM Seq2Seq показала багатообіцяючий результат 99% точності двійкової класифікації як на наборі даних NSL-KDD, так і на даних Honeypot Кіотського університету («Kyoto-Honeypot»). Результати SVM і RF демонструють меншу точність, якщо класифікувати їх за набором даних NSL KDD, і високу точність за набором даних UNSW-NB15.

Запропонували [87] дворівневу модель DL, яка діє як надійна система для виявлення аномалій і захисту від кібератак в архітектурі 5G для мобільної мережі. Контрольований або напівконтрольований метод навчання використовується на першому рівні для впровадження DBN або SAE, що працює на кожній RAN. Контрольована мережа LSTM Recurrent Network використовується на другому рівні для обмеження кібератак.

Для виявлення аномалій за допомогою багатовимірних вхідних даних було проведено дослідження за допомогою згорткових нейронних мереж (CNN) [28]. Хоча механізми глибокого навчання найкраще підходять для виявлення аномалій, проблеми, з якими доводиться стикатися, полягають у швидшому виявленні загроз, а профіль трафіку має формуватися автоматично. Профіль трафіку включає статистику потоку, таку як швидкість передачі, кількість пакетів, розмір потоку тощо. У CNN такі функції автоматично витягуються з профілю трафіку. В роботі [88], шаблони трафіку будуються шляхом дослідження початкових байтів перших кількох пакетів трафіку. Оскільки для виявлення аномалій використовуються лише перші кілька пакетів, швидкість виявлення загрози збільшується. Запропонована система автоматично використовує модуль CNN для визначення характеристик вихідних даних. Модель досягає 99,77% точності виявлення шкідливих дій і менше 1% FNR і FPR. Набір даних містить чотири класи атак DDoS: HTTP-флуд, ACK-флуд, UDP-флуд і SYN-флуд.

Підхід машинного навчання, який поєднує підходи глибокого навчання з навчанням з підкріпленням (RL), називається Deep Reinforcement Learning (DRL). Автори [89] використовували механізми DRL у своїй роботі, щоб запропонувати підхід до розробки, який захищає детектори ботнетів від агресивних атак. Нова стратегія використовує DRL для підвищення надійності детекторів. Детектори ботнетів використовують класифікатори Wide and Deep (WnD) і Random Forest (RF). Агенти в запропонованій моделі базуються на таких підходах глибокого підкріплення, як Double Deep Q-Network (2DQN) і Deep State-action-reward-state-action (Sarsa), які використовують методи поза політикою та згідно з політикою відповідно . На наступному етапі цей навчений агент DRL проводить атаку противника. Ці зразки можуть уникнути виявлення ботнету. Завдяки змагальному навчанню модель використовує зразки для посилення детекторів ботнетів.

Глибинні нейронні мережі найбільше підходять для алгоритмів генерації доменів, оскільки вони можуть ефективно класифікувати доменні

імена, як шкідливі так і доброякісні [75]. Для виявлення DGA автори досліджують переваги мічених даних для навчання класифікаторів DL. Для цього автори використовували RNN, LSTM, CNN та гібридні моделі CNN/RNN. В роботі [42] використовували такі архітектури RNN, як двонаправлений LSTM (Bi-LSTM), мережі довгострокової короткочасної пам'яті (LSTM) і Gated Recurrent Units (GRU), щоб обчислити продуктивність класифікатора DGA. Запропонований класифікатор DGA бере доменні імена із запитів DNS і не вимагає ручного створення функції. Без будь-якої контекстної інформації модель виконує багатокласову класифікацію, щоб визначити родину домену, до якої вона належить.

У порівнянні з алгоритмами ML алгоритми DL найбільш підходять для виявлення зловмисного програмного забезпечення [62]. Причина полягає в зменшенні продуктивності алгоритмів ML, коли розмір даних збільшується. Алгоритми DL покращують продуктивність, хоча розмір вхідних даних більший. Оскільки зловмисне програмне забезпечення розмножується разом із технологією, виявлення шкідливого програмного забезпечення має впоратися з проблемами масштабованості. У [38] запропонували гібридну масштабовану структуру глибокого навчання під назвою Scale Mal Net, яка обробляє великі зразки шкідливого програмного забезпечення. Модель збирає зразки шкідливих програм і застосовує їх для попередньої обробки розподіленим способом. Виконувані файли класифікуються на безпечні або шкідливі зразки за допомогою статичного та динамічного дослідження на першому етапі. Після цього йде другий етап, на якому виконувані файли зловмисного програмного забезпечення поділяються на сімейства. Автори дійшли висновку, що архітектури глибокого навчання перевершують класичні моделі машинного навчання.

Для класифікації шкідливих програм автори [73] запропонували структуру з гібридною глибокою нейронною мережею. Цей гібридний підхід поєднує кілька попередньо навчених мережевих моделей, і результати тестування довели, що запропонована структура може відокремлювати

зловмисне програмне забезпечення з підвищеною точністю, запам’ятовуваністю, точністю та оцінкою F1.

Запропоновано структуру для класифікації категорій шкідливих програм для Android [90]. Ця структура використовує динамічну класифікацію категорій зловмисного програмного забезпечення, а також застосовує частково контрольовані глибокі нейронні мережі. Результати експерименту показують, що показник F1 кращий і має частоту помилкових позитивних результатів на 2,76%, перевершуючи типові алгоритми машинного навчання. Вхідний рівень складався з 470 нейронів, а вихідний — з 5 нейронів. Функція sigmoid використовується для активації, а для оптимізації використовувався метод mini-Batch Gradient Descent.

Подібно до алгоритмів машинного навчання, методи глибокого навчання також зазнають впливу атак противника. Моделі глибокого навчання є крихкими під час агресивних атак. Змагальні атаки можуть бути атаками сірого ящика, білого ящика та чорного ящика. Запропоновано багато алгоритмів атаки для створення змагальної вибірки для цих моделей загроз. Деякі з алгоритмів атаки: Deep Fool, Fast Gradient Sign Method (FGSM), метод на основі оптимізації, Saliency Map Approach на основі якобіана (JSMA), алгоритм Broyden -Fletcher Goldfarb- Shanno (L-BFGS) з обмеженою пам’яттю, Basic Ітеративний метод (BIM)/проекційний градієнтний спуск (PGD), атаки Карліні та Вагнера (C&W) та атаки Distribution Ally Adversarial [91,92].

Нейронні мережі глибокого навчання, засновані на системах виявлення вторгнень, сприйнятливі до атак за сценаріями «білої скриньки» та бекдорів [93]. У цій галузі проведено багато дослідницької роботи. Однією з таких робіт є дослідження суперечливих прикладів, що впливають на інтерпретацію систем виявлення вторгнень за допомогою глибоких нейронних мереж (DNN) [94]. Автор ілюструє, що зловмисник може генерувати змагальні приклади, щоб ввести в оману модель DNN, навіть якщо внутрішня інформація моделей ізольована від супротивника. Ці

змагальні приклади генеруються та оцінюються в моделі чорної скриньки. Хоча тут немає доступу до внутрішніх деталей моделі, зловмисник все одно може ввести в оману класифікатор, щоб неправильно класифікувати вхідні дані атаки як звичайні вхідні дані.

Запропоновано [95] класифікатор машинного навчання для генерації та захисту від атак ухилення та причинних атак, поєднуючи дослідницьку атаку на основі DL. Спочатку зловмисник створює класифікатор, використовуючи пошукову атаку, засновану на глибокому навчанні (DL), подібну до оригінального класифікатора. З побудованого класифікатора зразки збираються та передаються до оригінального класифікатора. Щоб досягти атаки ухилення в навченому класифікаторі, зловмисник намагається ввести в оману алгоритм машинного навчання, надаючи неправильні вхідні дані, що призводить до неправильної мітки, і таким чином неправильно класифікуючи зразки. Для причинної атаки супротивник надає цільовому класифікатору помилкову інформацію про клас, таким чином знижуючи точність навченого класифікатора. Це дослідження, проведене авторами, продемонструвало, що атака ухилення збільшила помилку на етапі тестування, а причинна атака зросла так само під час фази навчання. Вони завершили роботу, забезпечивши агресивний механізм захисту з невеликими збуреннями, які показують, що помилка під час атаки ідентична помилці, коли атаки немає.

Пропонують [96] поєднання атак для створення прикладів зловмисного програмного забезпечення, що суперечить один одному. Для цього автор використовує кілька генеративних процедур і наборів маніпуляцій. Щоб перевірити надійність детекторів шкідливих програм, автор використовує 26 атак ухилення. Ці атаки ухилення класифікуються на градієнтні атаки, безградієнтні атаки, атаки з перенесенням, обфускацію та підходи змішаних атак.

Показники ефективності. Основними показниками, які оцінюють ефективність методів DL і ML, є матриця заплутаності, точність, коефіцієнт виявлення (DR) (також називається відкликанням або істинно позитивним

показником), хибно-позитивний рівень, справді негативний рівень, F1-Score, точність. Крива робочих характеристик приймача (ROC) і площа під ROC (AuC) також використовуються для оцінки ефективності класифікації.

У наборі даних випадкового розміру компонент може належати до двійкової або n- нарної класифікації. У бінарній класифікації елемент можна розглядати як нападний або доброякісний. Інвазія представлена як позитивна, а доброякісна категорія позначена як негативна [38]. Справжній позитивний (*TP*) — це компонент із позитивної категорії, який алгоритм розглядає як позитивний. Подібним чином, True Negative *(TN)* — це елемент з негативного класу, який правильно розглядається алгоритмом як негативний. Але в помилково позитивному *(FP)* елемент визначається як атака, хоча насправді це не так. Подібним чином у False Negatives (*FN*) алгоритм не може визначити атаку.

Точність вимірюється як частина правильно передбачених елементів:

$$Точність = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.1}$$

Точність (позитивне прогнозне значення) – це частина елементів, які правильно передбачені для загальної прогнозованої атаки. Це визначає кількість атак, класифікованих як позитивні:

$$Точність(позитивна) = \frac{TP}{TP + FP} \tag{2.2}$$

Коефіцієнт виявлення (DR) показує кількість ідентифікованих атак:

$$DR = \frac{TP}{TP + FN} \tag{2.3}$$

Рівень хибнопозитивних результатів (FPR) вказує на кількість нерозпізнаних вторгнень:

$$FRP = \frac{FP}{TN + FP} \tag{2.4}$$

Відкликання – це обчислений відсоток правильно класифікованих даних атаки до загальної кількості даних атаки в наданому наборі даних. Чим вищий коефіцієнт відкликання, тим краща продуктивність моделі машинного навчання:

$$Відкликання = \frac{TP}{TP + FN} \tag{2.5}$$

F1-Score/F1-Measure розраховується як гармонійне середнє значення точності та запам’ятовування. Збільшення показника F1-Score демонструє відмінну роботу алгоритму машинного навчання:

$$F1 - Score = 2X \frac{Точність(позитивна) \times Відкликання}{Точність(позитивна) + Відкликання} \tag{2.6}$$

Високе значення коефіцієнта помилкових негативних результатів може свідчити про те, що NIDS не вдалося ідентифікувати відомі або анонімні атаки. Навпаки, підвищена частота хибних спрацьовувань вказує на генерацію помилкових тривог, коли в мережі немає атаки [97].

Деякі з показників, які використовуються для створення AE, це загальна вартість часу (TTC), коефіцієнт виявлення змагальності (ADR) і коефіцієнт початкового виявлення (ODR) (Zhang *et al.* 2020). TTC – це загальний час, необхідний для створення набору AE. ODR визначає

ефективність виявлення цільових класифікаторів на відміну від реальних прикладів атак. ADR передбачає ефективність виявлення цільових класифікаторів на противагу агресивним атакам:

$$ORD = \frac{No.of\ \ right\ indentified\ orignal\ attack\ examples}{No.of\ \ all\ the\ orignal\ attack\ examples} \tag{2.7}$$

$$КПВ = \frac{К-ть\ реальних\ індетифікованих\ орігінальних\ прикладів\ атак}{К-ть\ усіх\ орігінальних\ прикладів\ атак} \tag{2.7}$$

ВИСНОВКИ

Розглянуто підходи до використання різних технологій штучного інтелекту (зокрема машинного та глибокого навчання) в кібербезпеці.

Запропоновано показники, які оцінюють ефективність методів машинного та глибокого навчання в кібербезпеці.

Встановлено, що архітектури глибокого навчання перевершують класичні моделі машинного навчання.

Доведено, що алгоритми глибокого навчання можуть подолати обмеження алгоритмів машинного навчання оскільки алгоритми глибокого навчання виграють від традиційних алгоритмів машинного навчання, де функції високого рівня автоматично генеруються з наявних функцій.

Показано, що у порівнянні з класичними методами машинного навчання, глибокі мережі можуть автоматично отримувати функції з даних, зменшуючи зусилля на попередню обробку вхідних даних і не покладаючись на функції, створені людиною що робить алгоритми глибокого навчання придатними для обробки в режимі реального часу.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Artificial intelligence for cybersecurity: a systematic mapping of literature / I. Wiafe et al. *IEEE access*. 2020. Vol. 8. P. 146598–146612. URL: https://doi.org/10.1109/access.2020.3013145.

2. Yu X., Guo H. A survey on iiot security. *2019 IEEE VTS asia pacific wireless communications symposium (APWCS)*, Singapore, 28–30 August 2019. 2019. URL: https://doi.org/10.1109/vts-apwcs.2019.8851679.

3. Li Y., Liu Q. A comprehensive review study of cyber-attacks and cyber security; Emerging trends and recent developments. *Energy reports*. 2021. URL: https://doi.org/10.1016/j.egyr.2021.08.126.

4. Artificial intelligence in cyber security: research advances, challenges, and opportunities / Z. Zhang et al. *Artificial intelligence review*. 2021. URL: https://doi.org/10.1007/s10462-021-09976-0.

5. Uniting cyber security and machine learning: advantages, challenges and future research / M. Wazid et al. *ICT express*. 2022. URL: https://doi.org/10.1016/j.icte.2022.04.007.

6. Machine learning methods for computer security / A. D. Joseph et al. *Dagstuhl manifestos,*. 2013. Vol. 3, no. 1. URL: https://doi.org/10.4230/DagMan.3.1.1.

7. Lo C. K. Systematic reviews on flipped learning in various education contexts. *Systematic reviews in educational research*. Wiesbaden, 2019. P. 129–143. URL: https://doi.org/10.1007/978-3-658-27602-7_8.

8. Cyber attacks in the era of COVID-19 and possible solution domains / I. C. Eian et al. 2020. (Preprint. 202009.0630). URL: https://www.preprints.org/manuscript/202009.0630.

9. Almeida F., Duarte Santos J., Augusto Monteiro J. The challenges and opportunities in the digitalization of companies in a post-covid-19 world. *IEEE*

engineering management review. 2020. Vol. 48, no. 3. P. 97–103. URL: https://doi.org/10.1109/emr.2020.3013206.

10. Artificial intelligence-based cyber security in the context of industry 4.0–A survey / A. J. G. de Azambuja et al. *Electronics*. 2023. Vol. 12, no. 8. P. 1920. URL: https://doi.org/10.3390/electronics12081920.

11. Effectiveness of artificial intelligence techniques against cyber security risks apply of IT industry / B. Alhayani et al. *Materials today: proceedings*. 2021. URL: https://doi.org/10.1016/j.matpr.2021.02.531.

12. Hariyanti E., Djunaidy A., Siahaan D. Information security vulnerability prediction based on business process model using machine learning approach. *Computers & security*. 2021. Vol. 110. P. 102422. URL: https://doi.org/10.1016/j.cose.2021.102422.

13. Arasada S. These four challenges in adopting machine learning can lower your ROI and sabotage success. *Forbes*. 2021. 31 August. URL: https://www.forbes.com/councils/forbestechcouncil/2021/08/31/these-four-challenges-in-adopting-machine-learning-can-lower-your-roi-and-sabotage-success/.

14. Mohammed I. A. Artificial intelligence for cybersecurity: a systematic mapping of literature. *ResearchGate*. URL: https://www.researchgate.net/publication/353887583_ARTIFICIAL_INTELLIGENCE_FOR_CYBERSECURITY_A_SYSTEMATIC_MAPPING_OF_LITERATURE.

15. Farooq Q., Shan L., Yuan H. Y. New approach towards cyber security for nuclear power control system. *2021 IEEE kansas power and energy conference (KPEC)*, Manhattan, KS, USA, 19–20 April 2021. 2021. URL: https://doi.org/10.1109/kpec51835.2021.9446235.

16. Kumar V., Gupta C. P. Cyber security issue in smart grid. *2021 IEEE 4th international conference on computing, power and communication technologies (GUCON)*, Kuala Lumpur, Malaysia, 24–26 September 2021. 2021. URL: https://doi.org/10.1109/gucon50781.2021.9573600.

17. Falsification of cyber-physical systems using deep reinforcement learning / Y. Yamagata et al. *IEEE transactions on software engineering*. 2020. P. 1. URL: https://doi.org/10.1109/tse.2020.2969178.

18. Cyber-Physical systems security–a survey / A. Humayed et al. *IEEE internet of things journal*. 2017. Vol. 4, no. 6. P. 1802–1831. URL: https://doi.org/10.1109/jiot.2017.2703172.

19. Cyber security issues and challenges for smart cities: a survey / B. Hamid et al. *2019 13th international conference on mathematics, actuarial science, computer science and statistics (MACS)*, Karachi, Pakistan, 14–15 December 2019. 2019. URL: https://doi.org/10.1109/macs48846.2019.9024768.

20. Thakur K., Hayajneh T., Tseng J. Cyber security in social media: challenges and the way forward. *IT professional*. 2019. Vol. 21, no. 2. P. 41–49. URL: https://doi.org/10.1109/mitp.2018.2881373.

21. The effect and technique in search engine optimization / A. Dramilio et al. *2020 international conference on information management and technology (icimtech)*, Bandung, Indonesia, 13–14 August 2020. 2020. URL: https://doi.org/10.1109/icimtech50083.2020.9211171.

22. Urien P. Innovative countermeasures to defeat cyber attacks against blockchain wallets. *2021 5th cyber security in networking conference (csnet)*, Abu Dhabi, United Arab Emirates, 12–14 October 2021. 2021. URL: https://doi.org/10.1109/csnet52717.2021.9614649.

23. Deep learning modalities for biometric alteration detection in 5G networks-based secure smart cities / A. Sedik et al. *IEEE access*. 2021. Vol. 9. P. 94780–94788. URL: https://doi.org/10.1109/access.2021.3088341.

24. A survey of deep learning methods for cyber security / D. Berman et al. *Information*. 2019. Vol. 10, no. 4. P. 122. URL: https://doi.org/10.3390/info10040122.

25. Deep learning techniques for cyber security intrusion detection : a detailed analysis / M. A. Ferrag et al. *6th international symposium for ICS &*

SCADA cyber security research 2019. 2019. URL: https://doi.org/10.14236/ewic/icscsr19.16.

26. A comprehensive survey of databases and deep learning methods for cybersecurity and intrusion detection systems / D. Gumusbas et al. *IEEE systems journal*. 2020. P. 1–15. URL: https://doi.org/10.1109/jsyst.2020.2992966.

27. Al-Janabi M., Altamimi A. M. A comparative analysis of machine learning techniques for classification and detection of malware. *2020 21st international arab conference on information technology (ACIT)*, 6 of October, Giza, 28–30 November 2020. 2020. URL: https://doi.org/10.1109/acit50332.2020.9300081.

28. Alabadi M., Celik Y. Anomaly detection for cyber-security based on convolution neural network : a survey. *2020 international congress on human-computer interaction, optimization and robotic applications (HORA)*, Ankara, Turkey, 26–28 June 2020. 2020. URL: https://doi.org/10.1109/hora49412.2020.9152899.

29. Utilising deep learning techniques for effective zero-day attack detection / H. Hindy et al. *Electronics*. 2020. Vol. 9, no. 10. P. 1684. URL: https://doi.org/10.3390/electronics9101684.

30. Rashid A., Siddique M. J., Ahmed S. M. Machine and deep learning based comparative analysis using hybrid approaches for intrusion detection system. *2020 3rd international conference on advancements in computational sciences (ICACS)*, Lahore, Pakistan, 17–19 February 2020. 2020. URL: https://doi.org/10.1109/icacs47775.2020.9055946.

31. When machine learning meets hardware cybersecurity: delving into accurate zero-day malware detection / Z. He et al. *2021 22nd international symposium on quality electronic design (ISQED)*, Santa Clara, CA, USA, 7–9 April 2021. 2021. URL: https://doi.org/10.1109/isqed51717.2021.9424330.

32. Mahdavifar S., Ghorbani A. A. Application of deep learning to cybersecurity: a survey. *Neurocomputing*. 2019. Vol. 347. P. 149–176. URL: https://doi.org/10.1016/j.neucom.2019.02.056.

33. Intrusion detection and prevention system using deep learning / A. Krishna et al. *2020 international conference on electronics and sustainable communication systems (ICESC)*, Coimbatore, India, 2–4 July 2020. 2020. URL: https://doi.org/10.1109/icesc48915.2020.9155711.

34. Chen Z. Deep learning for cybersecurity: a review. *2020 international conference on computing and data science (CDS)*, Stanford, CA, 1–2 August 2020. 2020. URL: https://doi.org/10.1109/cds49703.2020.00009.

35. A two-fold machine learning approach to prevent and detect iot botnet attacks / F. Hussain et al. *IEEE access*. 2021. Vol. 9. P. 163412–163430. URL: https://doi.org/10.1109/access.2021.3131014.

36. Kambourakis G., Kolias C., Stavrou A. The mirai botnet and the iot zombie armies. *2017 IEEE military communications conference (MILCOM)*, Baltimore, MD, 23–25 October 2017. 2017. URL: https://doi.org/10.1109/milcom.2017.8170867.

37. Identification of botnet activity in iot network traffic using machine learning / M. Hegde et al. *2020 international conference on intelligent data science technologies and applications (IDSTA)*, Valencia, Spain, 19–22 October 2020. 2020. URL: https://doi.org/10.1109/idsta50958.2020.9264143.

38. Robust intelligent malware detection using deep learning / R. Vinayakumar et al. *IEEE access*. 2019. Vol. 7. P. 46717–46738. URL: https://doi.org/10.1109/access.2019.2906934.

39. Xu S., Xia Y., Shen H.-L. Analysis of malware-induced cyber attacks in cyber-physical power systems. *IEEE transactions on circuits and systems II: express briefs*. 2020. Vol. 67, no. 12. P. 3482–3486. URL: https://doi.org/10.1109/tcsii.2020.2999875.

40. Detecting spear-phishing emails based on authentication / W. Xiujuan et al. *2019 IEEE 4th international conference on computer and communication systems (ICCCS)*, Singapore, 23–25 February 2019. 2019. URL: https://doi.org/10.1109/ccoms.2019.8821758.

41. Singh C., Meenu. Phishing website detection based on machine learning: a survey. *2020 6th international conference on advanced computing and communication systems (ICACCS)*, Coimbatore, India, 6–7 March 2020. 2020. URL: https://doi.org/10.1109/icaccs48705.2020.9074400.

42. Shahzad H., Sattar A. R., Skandaraniyam J. DGA domain detection using deep learning. *2021 IEEE 5th international conference on cryptography, security and privacy (CSP)*, Zhuhai, China, 8–10 January 2021. 2021. URL: https://doi.org/10.1109/csp51677.2021.9357591.

43. Cyber security issues and challenges for smart cities: a survey / B. Hamid et al. *2019 13th international conference on mathematics, actuarial science, computer science and statistics (MACS)*, Karachi, Pakistan, 14–15 December 2019. 2019. URL: https://doi.org/10.1109/macs48846.2019.9024768.

44. Detecting abnormal traffic in large-scale networks / M. S. Elsayed et al. *2020 international symposium on networks, computers and communications (ISNCC)*, Montreal, QC, 20–22 October 2020. 2020. URL: https://doi.org/10.1109/isncc49221.2020.9297358.

45. Sapre S., Islam K., Ahmadi P. A comprehensive data sampling analysis applied to the classification of rare iot network intrusion types. *2021 IEEE 18th annual consumer communications & networking conference (CCNC)*, Las Vegas, NV, USA, 9–12 January 2021. 2021. URL: https://doi.org/10.1109/ccnc49032.2021.9369617.

46. Sohn I. Deep belief network based intrusion detection techniques: a survey. *Expert systems with applications*. 2020. P. 114170. URL: https://doi.org/10.1016/j.eswa.2020.114170.

47. Deep learning for cyber security intrusion detection: approaches, datasets, and comparative study / M. A. Ferrag et al. *Journal of information security and applications*. 2020. Vol. 50. P. 102419. URL: https://doi.org/10.1016/j.jisa.2019.102419.

48. Buczak A. L., Guven E. A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE communications surveys &*

tutorials. 2016. Vol. 18, no. 2. P. 1153–1176. URL: https://doi.org/10.1109/comst.2015.2494502.

49. Machine learning for security and the internet of things: the good, the bad, and the ugly / F. Liang et al. *IEEE access*. 2019. Vol. 7. P. 158126–158147. URL: https://doi.org/10.1109/access.2019.2948912.

50. A review of various challenges in cybersecurity using Artificial Intelligence / H. Chaudhary et al. *2020 3rd international conference on intelligent sustainable systems (ICISS)*, Thoothukudi, India, 3–5 December 2020. 2020. URL: https://doi.org/10.1109/iciss49785.2020.9316003.

51. Kilincer I. F., Ertam F., Sengur A. Machine learning methods for cyber security intrusion detection: datasets and comparative study. *Computer networks*. 2021. Vol. 188. P. 107840. URL: https://doi.org/10.1016/j.comnet.2021.107840.

52. Large-Scale malicious software classification with fuzzified features and boosted fuzzy random forest / F. Li et al. *IEEE transactions on fuzzy systems*. 2020. P. 1. URL: https://doi.org/10.1109/tfuzz.2020.3016023.

53. Significant permission identification for machine-learning-based android malware detection / J. Li et al. *IEEE transactions on industrial informatics*. 2018. Vol. 14, no. 7. P. 3216–3225. URL: https://doi.org/10.1109/tii.2017.2789219.

54. A high-performance intrusion detection method based on combining supervised and unsupervised learning / H. Wang et al. *2018 IEEE smartworld, ubiquitous intelligence & computing, advanced & trusted computing, scalable computing & communications, cloud & big data computing, internet of people and smart city innovation (smartworld/scalcom/uic/atc/cbdcom/iop/sci)*, Guangzhou, China, 8–12 October 2018. 2018. URL: https://doi.org/10.1109/smartworld.2018.00304.

55. Machine learning and deep learning methods for cybersecurity / Y. Xin et al. *IEEE access*. 2018. Vol. 6. P. 35365–35381. URL: https://doi.org/10.1109/access.2018.2836950.

56. Feature selection for machine learning-based early detection of distributed cyber attacks / Y. Feng et al. *2018 IEEE 16th intl conf on dependable,*

autonomic and secure computing, 16th intl conf on pervasive intelligence and computing, 4th intl conf on big data intelligence and computing and cyber science and technology congress(dasc/picom/datacom/cyberscitech), Athens, 12–15 August 2018. 2018. URL: https://doi.org/10.1109/dasc/picom/datacom/cyberscitec.2018.00040.

57. Lakshmanarao A., Rao P. S. P., Krishna M. M. B. Phishing website detection using novel machine learning fusion approach. *2021 international conference on artificial intelligence and smart systems (ICAIS)*, Coimbatore, India, 25–27 March 2021. 2021. URL: https://doi.org/10.1109/icais50930.2021.9395810.

58. A novel approach for phishing URLs detection using lexical based machine learning in a real-time environment / B. B. Gupta et al. *Computer communications*. 2021. Vol. 175. P. 47–57. URL: https://doi.org/10.1016/j.comcom.2021.04.023.

59. Analysis of modern intrusion detection algorithms and developing a smart IDS / P. Iyer et al. *2021 international conference on intelligent technologies (CONIT)*, Hubli, India, 25–27 June 2021. 2021. URL: https://doi.org/10.1109/conit51480.2021.9498519.

60. Tang L., Mahmoud Q. H. A survey of machine learning-based solutions for phishing website detection. *Machine learning and knowledge extraction*. 2021. Vol. 3, no. 3. P. 672–694. URL: https://doi.org/10.3390/make3030034.

61. Detecting spear-phishing emails based on authentication / W. Xiujuan et al. *2019 IEEE 4th international conference on computer and communication systems (ICCCS)*, Singapore, 23–25 February 2019. 2019. URL: https://doi.org/10.1109/ccoms.2019.8821758.

62. On the effectiveness of machine and deep learning for cyber security / G. Apruzzese et al. *2018 10th international conference on cyber conflict (cycon)*, Tallinn, 29 May – 1 June 2018. 2018. URL: https://doi.org/10.23919/cycon.2018.8405026.

63. Sharma R. K., Kalita H. K., Borah P. Analysis of machine learning techniques based intrusion detection systems. *Proceedings of 3rd international*

conference on advanced computing, networking and informatics. New Delhi, 2015. P. 485–493. URL: https://doi.org/10.1007/978-81-322-2529-4_51.

64. Dixit P., Silakari S. Deep learning algorithms for cybersecurity applications: a technological and status review. *Computer science review*. 2021. Vol. 39. P. 100317. URL: https://doi.org/10.1016/j.cosrev.2020.100317.

65. Can machine learning model with static features be fooled: an adversarial machine learning approach / R. Taheri et al. *Cluster computing*. 2020. Vol. 23, no. 4. P. 3233–3253. URL: https://doi.org/10.1007/s10586-020-03083-5.

66. DGA-based botnet detection toward imbalanced multiclass learning / Y. Chen et al. *Tsinghua science and technology*. 2021. Vol. 26, no. 4. P. 387–402. URL: https://doi.org/10.26599/tst.2020.9010021.

67. A machine learning framework for domain generation algorithm-based malware detection / Y. Li et al. *IEEE access*. 2019. Vol. 7. P. 32765–32782. URL: https://doi.org/10.1109/access.2019.2891588.

68. AI and machine learning: a mixed blessing for cybersecurity / F. Kamoun et al. *2020 international symposium on networks, computers and communications (ISNCC)*, Montreal, QC, 20–22 October 2020. 2020. URL: https://doi.org/10.1109/isncc49221.2020.9297323.

69. Nguyen T. T. T., Armitage G. A survey of techniques for internet traffic classification using machine learning. *IEEE communications surveys & tutorials*. 2008. Vol. 10, no. 4. P. 56–76. URL: https://doi.org/10.1109/surv.2008.080406.

70. Martínez Torres J., Iglesias Comesaña C., García-Nieto P. J. Review: machine learning techniques applied to cybersecurity. *International journal of machine learning and cybernetics*. 2019. Vol. 10, no. 10. P. 2823–2836. URL: https://doi.org/10.1007/s13042-018-00906-1.

71. Hu W., Tan Y. Generating adversarial malware examples for black-box attacks based on GAN. 2017. (Preprint. 1702.05983). URL: https://doi.org/10.48550/arXiv.1702.05983.

72. Geluvaraj B., Satwik P. M., Ashok Kumar T. A. The future of cybersecurity: major role of artificial intelligence, machine learning, and deep

learning in cyberspace. *International conference on computer networks and communication technologies*. Singapore, 2018. P. 739–747. URL: https://doi.org/10.1007/978-981-10-8681-6_67.

73. Aslan O., Yilmaz A. A. A new malware classification framework based on deep learning algorithms. *IEEE access*. 2021. Vol. 9. P. 87936–87951. URL: https://doi.org/10.1109/access.2021.3089586.

74. Mahdavifar S., Ghorbani A. A. Application of deep learning to cybersecurity: a survey. *Neurocomputing*. 2019. Vol. 347. P. 149–176. URL: https://doi.org/10.1016/j.neucom.2019.02.056.

75. Weakly supervised deep learning for the detection of domain generation algorithms / B. Yu et al. *IEEE access*. 2019. Vol. 7. P. 51542–51556. URL: https://doi.org/10.1109/access.2019.2911522.

76. Imamverdiyev Y. N., Abdullayeva F. J. Deep learning in cybersecurity. *International journal of cyber warfare and terrorism*. 2020. Vol. 10, no. 2. P. 82–105. URL: https://doi.org/10.4018/ijcwt.2020040105.

77. SSH and FTP brute-force attacks detection in computer networks: LSTM and machine learning approaches / M. D. Hossain et al. *2020 5th international conference on computer and communication systems (ICCCS)*, Shanghai, China, 15–18 May 2020. 2020. URL: https://doi.org/10.1109/icccs49078.2020.9118459.

78. Karatas G., Demir O., Koray Sahingoz O. Deep learning in intrusion detection systems. *2018 international congress on big data, deep learning and fighting cyber terrorism (IBIGDELFT)*, ANKARA, Turkey, 3–4 December 2018. 2018. URL: https://doi.org/10.1109/ibigdelft.2018.8625278.

79. Improving performance of autoencoder-based network anomaly detection on NSL-KDD dataset / W. Xu et al. *IEEE access*. 2021. Vol. 9. P. 140136–140146. URL: https://doi.org/10.1109/access.2021.3116612.

80. Multi-Channel deep feature learning for intrusion detection / G. Andresini et al. *IEEE access*. 2020. Vol. 8. P. 53346–53359. URL: https://doi.org/10.1109/access.2020.2980937.

81. Utilising deep learning techniques for effective zero-day attack detection / H. Hindy et al. *Electronics*. 2020. Vol. 9, no. 10. P. 1684. URL: https://doi.org/10.3390/electronics9101684.

82. Kim K., Aminanto M. E. Deep learning in intrusion detection perspective: overview and further challenges. *2017 international workshop on big data and information security (IWBIS)*, Jakarta, 23–24 September 2017. 2017. URL: https://doi.org/10.1109/iwbis.2017.8275095.

83. Otoum S., Kantarci B., Mouftah H. T. On the feasibility of deep learning in sensor network intrusion detection. *IEEE networking letters*. 2019. Vol. 1, no. 2. P. 68–71. URL: https://doi.org/10.1109/lnet.2019.2901792.

84. Alom M. Z., Bontupalli V., Taha T. M. Intrusion detection using deep belief networks. *NAECON 2015 - IEEE national aerospace and electronics conference*, Dayton, OH, USA, 15–19 June 2015. 2015. URL: https://doi.org/10.1109/naecon.2015.7443094.

85. Djellali C., Adda M., Moutacalli M. T. A comparative study to deep learning for pattern recognition, by using online and batch learning; taking cybersecurity as a case. *ASONAM '19: international conference on advances in social networks analysis and mining*, Vancouver British Columbia Canada. New York, NY, USA, 2019. URL: https://doi.org/10.1145/3341161.3343533.

86. An empirical evaluation of deep learning for network anomaly detection / R. K. Malaiya et al. *2018 international conference on computing, networking and communications (ICNC)*, Maui, HI, 5–8 March 2018. 2018. URL: https://doi.org/10.1109/iccnc.2018.8390278.

87. A self-adaptive deep learning-based system for anomaly detection in 5G networks / L. Fernandez Maimo et al. *IEEE access*. 2018. Vol. 6. P. 7700–7712. URL: https://doi.org/10.1109/access.2018.2803446.

88. An unsupervised deep learning model for early network traffic anomaly detection / R.-H. Hwang et al. *IEEE access*. 2020. Vol. 8. P. 30387–30399. URL: https://doi.org/10.1109/access.2020.2973023.

89. Deep reinforcement adversarial learning against botnet evasion attacks / G. Apruzzese et al. *IEEE transactions on network and service management*. 2020. Vol. 17, no. 4. P. 1975–1987. URL: https://doi.org/10.1109/tnsm.2020.3031843.

90. Dynamic android malware category classification using semi-supervised deep learning / S. Mahdavifar et al. *2020 IEEE intl conf on dependable, autonomic and secure computing, intl conf on pervasive intelligence and computing, intl conf on cloud and big data computing, intl conf on cyber science and technology congress (dasc/picom/cbdcom/cyberscitech)*, Calgary, AB, Canada, 17–22 August 2020. 2020. URL: https://doi.org/10.1109/dasc-picom-cbdcom-cyberscitech49142.2020.00094.

91. Ren C., Xu Y. A fully data-driven method based on generative adversarial networks for power system dynamic security assessment with missing data. *IEEE transactions on power systems*. 2019. Vol. 34, no. 6. P. 5044–5052. URL: https://doi.org/10.1109/tpwrs.2019.2922671.

92. Li D., Li Q. Adversarial deep ensemble: evasion attacks and defenses for malware detection. *IEEE transactions on information forensics and security*. 2020. Vol. 15. P. 3886–3900. URL: https://doi.org/10.1109/tifs.2020.3003571.

93. Alrawashdeh K., Goldsmith S. Defending deep learning based anomaly detection systems against white-box adversarial examples and backdoor attacks. *2020 IEEE international symposium on technology and society (ISTAS)*, Tempe, AZ, USA, 12–15 November 2020. 2020. URL: https://doi.org/10.1109/istas50296.2020.9462227.

94. Adversarial examples against the deep learning based network intrusion detection systems / K. Yang et al. *MILCOM 2018 - IEEE military communications conference*, Los Angeles, CA, 29–31 October 2018. 2018. URL: https://doi.org/10.1109/milcom.2018.8599759.

95. Shi Y., Sagduyu Y. E. Evasion and causative attacks with adversarial deep learning. *2017 IEEE military communications conference (MILCOM)*, Baltimore, MD, 23–25 October 2017. 2017. URL: https://doi.org/10.1109/milcom.2017.8170807.

96. A framework for enhancing deep neural networks against adversarial malware / D. Li et al. *IEEE transactions on network science and engineering*. 2021. Vol. 8, no. 1. P. 736–750. URL: https://doi.org/10.1109/tnse.2021.3051354.

97. Kilincer I. F., Ertam F., Sengur A. Machine learning methods for cyber security intrusion detection: datasets and comparative study. *Computer networks*. 2021. Vol. 188. P. 107840. URL: https://doi.org/10.1016/j.comnet.2021.107840.