Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя
(повне найменування вищого навчального закладу)
Факультет комп'ютерно-інформаційних систем і програмної інженерії
(назва факультету)
Кафедра кібербезпеки
(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

магістр
(освітній рівень)

на тему: «Методи захисту від кібератак на основі штучного інтелекту"

Виконав: студент (ка)

Спеціальності:
125 «Кібербезпека та захист інформації»
(шифр і назва напряму підготовки, спеціальності)

Мазур В. М.
підпис (прізвище та ініціали)

Керівник Лечаченко Т. А.
підпис (прізвище та ініціали)

Нормоконтроль Тимощук Д. І.
підпис (прізвище та ініціали)

Завідувач кафедри Загородна Н.В.
підпис (прізвище та ініціали)

Рецензент
підпис (прізвище та ініціали)

м. Тернопіль – 2024

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя
(повне найменування вищого навчального закладу)

Факультет комп'ютерно-інформаційних систем і програмної інженерії

Кафедра Кібербезпека

ЗАТВЕРДЖУЮ
Завідувач кафедри

_____ Загородна Н. В
(підпис) (прізвище та ініціали)

«____» _____

З А В Д А Н Н Я
НА КВАЛІФІКАЦІЙНУ РОБОТУ

на здобуття освітнього ступеня магістр
(назва освітнього ступеня)

за спеціальністю 125 кібербезпека та захист інформації
(шифр і назва спеціальності)

студенту Мазур Володимир Михайлович
(прізвище, ім'я, по батькові)

1. Тема роботи Методи захисту від кібератак на основі штучного інтелекту

Керівник роботи Лечаченко Тарас Анатолійович, Ph.D.
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом по університету від «20» листопада 2024 року № 4/7-1112

2. Термін подання студентом завершеної роботи

3. Вихідні дані до роботи персональний комп'ютер з ОС Windows, Python, VS Code, сервер з процесором Intel Core i7 та 32 Гб оперативної пам'яті

4. Зміст роботи (перелік питань, які потрібно розробити)

Класифікація атак з використанням ШІ
Вразливості систем перед атаками ШІ
Особливості атак на основі глибокого
Захист від атак з використанням
Техніки протидії нейромережам
Захист проти фішингових атак та DDoS
Алгоритм реалізації захисту від кібератак
Приклад реалізації: Виявлення та запобігання DDoS-атаці на корпоративну мережу
Реалізація захисних моделей на базі ШІ
Моделювання атак та тестування
Оцінка ефективності методів захисту
Висновки

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)
Алгоритм захисту від кібератак на основі ШІ. Деталізована блок-схема алгоритму захисту від кібератак. Порівняння точності методів. Порівняння швидкості обробки запитів методами.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Охорона праці	Осухівська Г. М., к.т.н., доцент		
Безпека в надзвичайних ситуаціях	Клепчик В. М., старший викладач з адміністративно-господарської роботи та будівництва		

7. Дата видачі завдання 19.09.2024

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Ознайомлення з завданням до кваліфікаційної роботи	19.09.2024 – 20.09.2024	Виконано
2.	Підбір джерел про методи захисту від кібератак на основі штучного інтелекту	21.09.2024 – 02.10.2024	Виконано
3.	Опрацювання джерел в галузі дослідження	03.10.2024 – 12.10.2024	Виконано
4.	Дослідження протидії кібератакам на основі штучного інтелекту	13.10.2024 – 17.10.2024	Виконано
5.	Оформлення розділу «ТЕОРЕТИЧНІ ОСНОВИ АТАК НА ОСНОВІ ШТУЧНОГО ІНТЕЛЕКТУ»	18.10.2024 – 01.11.2024	Виконано
6.	Оформлення розділу «МЕТОДИ ЗАХИСТУ ВІД КІБЕРАТАК НА ОСНОВІ ШІ»	02.11.2024 – 20.11.2024	Виконано
7.	Оформлення розділу «ПРАКТИЧНЕ ЗАСТОСУВАННЯ МЕТОДІВ ЗАХИСТУ ВІД КІБЕРАТАК НА ОСНОВІ ШІ»	21.11.2024 – 04.12.2024	Виконано
8.	Виконання завдання до підрозділу «Безпека життєдіяльності, основи охорони праці»	05.12.2024 – 07.12.2024	Виконано
9.	Оформлення кваліфікаційної роботи	08.12.2024 – 11.12.2024	Виконано
10.	Нормконтроль	11.12.2024 – 15.12.2024	Виконано
11.	Перевірка на плагіат	16.12.2024 – 18.12.2024	Виконано
12.	Попередній захист кваліфікаційної роботи	20.12.2024	Виконано
13.	Захист кваліфікаційної роботи	27.12.2024	

Студент _____ Мазур В. М.
(підпис) (прізвище та ініціали)

Керівник роботи _____ Лечаченко Т. А.
(підпис) (прізвище та ініціали)

АНОТАЦІЯ

Методи захисту від кібератак на основі штучного інтелекту // ОР «Магістр» // Мазур Володимир Михайлович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем і програмної інженерії, кафедра кібербезпеки, група СБм-62 // Тернопіль, 2024 // С. 82, рис. – 4, табл. – 1, кресл. – -, додат. – 1.

Ключові слова: ШТУЧНИЙ ІНТЕЛЕКТ, КІБЕРБЕЗПЕКА, DDoS-АТАКА, ФІШИНГ, ЗАХИСТ ІНФОРМАЦІЇ, МОДЕЛЮВАННЯ АТАК, МЕРЕЖЕВА БЕЗПЕКА, ЗМАГАЛЬНІ МЕРЕЖІ GAN.

У кваліфікаційній роботі проведено аналіз методів захисту від атак на основі штучного інтелекту. Описано моделі атак, які використовують штучний інтелект, зокрема фішингові та DDoS-атаки, та запропоновано ефективні методи захисту. Запропоновані технічні та організаційні заходи для протидії цим атакам.

В роботі проведено практичне застосування методів захисту, включаючи моделювання атак із використанням Generative Adversarial Networks (GANs) та тестування захисних механізмів. Оцінено ефективність різних підходів у виявленні загроз та запропоновано нові напрямки для покращення захисних систем на основі штучного інтелекту.

ABSTRACT

AI-Based Methods for Protection Against Cyberattacks // Thesis of educational level "Master"// Volodymyr Mazur // Ternopil Ivan Puluj National Technical University, Faculty of Computer Information Systems and Software Engineering, Department of Cybersecurity, group СБм-62 // Ternopil, 2024 // P. 82, figs. -, tables 1, drawings 4, added. – 1.

Keywords: ARTIFICIAL INTELLIGENCE, CYBERSECURITY, DDoS ATTACK, PHISHING, INFORMATION PROTECTION, ATTACK SIMULATION, NETWORK SECURITY, COMPETITIVE GAN NETWORKS.

In the qualification work, an analysis of methods of protection against attacks based on artificial intelligence was carried out. Attack models that use artificial intelligence, including phishing and DDoS attacks, are described, and effective protection methods are proposed. Proposed technical and organizational measures to counter these attacks.

The work includes practical application of protection methods, including simulation of attacks using Generative Adversarial Networks (GANs) and testing of protective mechanisms. The effectiveness of various approaches in detecting threats was evaluated and new directions for improving protective systems based on artificial intelligence were proposed.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ 8

ВСТУП 9

РОЗДІЛ 1 ТЕОРЕТИЧНІ ОСНОВИ АТАК НА ОСНОВІ ШТУЧНОГО ІНТЕЛЕКТУ 12

1.1 Класифікація атак з використанням ШІ 12

1.1.1 Фішинг з використанням нейромереж 14

1.1.2 DDoS-атаки на основі машинного навчання 15

1.1.3 Злом CAPTCHA за допомогою ШІ 17

1.2 Вразливості систем перед атаками ШІ 18

1.2.1 Атаки на системи аутентифікації 19

1.2.2 Атаки на системи виявлення вторгнень (IDS) 20

1.3 Особливості атак на основі глибокого навчання 22

1.3.1 Штучні нейронні мережі для атак на захищені системи 23

1.3.2 Зловживання техніками генеративних моделей для атак 24

РОЗДІЛ 2 МЕТОДИ ЗАХИСТУ ВІД КІБЕРАТАК НА ОСНОВІ ШІ 25

2.1 Захист від атак з використанням машинного навчання 26

2.1.1 Методи навчання з учителем для виявлення аномалій 27

2.1.2 Адаптивні системи кіберзахисту на основі ШІ 28

2.2 Техніки протидії нейромережам 28

2.2.1 Використання штучних нейронних мереж для виявлення атак 30

2.3 Захист проти фішингових атак та DDoS 31

2.3.1 Технічні та організаційні заходи протидії 32

2.3.2 Використання ШІ для виявлення фішингових листів 34

2.4 Алгоритм реалізації захисту від кібератак 36

2.4.1 Збір даних 36

2.4.2 Попередня обробка даних 36
2.4.3 Навчання моделі 37
2.4.4 Інтеграція моделей в реальні системи 38
2.4.5 Постійний моніторинг і вдосконалення 39
2.5 Приклад реалізації: Виявлення та запобігання DDoS-атаці на корпоративну мережу 43
РОЗДІЛ 3 ПРАКТИЧНЕ ЗАСТОСУВАННЯ МЕТОДІВ ЗАХИСТУ ВІД КІБЕРАТАК НА ОСНОВІ ШІ 45
3.1 Реалізація захисних моделей на базі ШІ 45
3.1.1 Огляд та порівняння захисних алгоритмів 47
3.1.2 Впровадження механізмів захисту в реальні системи 48
3.2 Моделювання атак та тестування захисту 50
3.2.1 Моделювання атак за допомогою ШІ 51
3.2.2 Тестування ефективності захисних механізмів на практиці 54
3.3 Оцінка ефективності методів захисту 56
3.3.1 Критерії оцінки захищеності систем 57
3.3.2 Порівняльний аналіз результатів 58
РОЗДІЛ 4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ 63
4.1 Охорона праці 63
4.2 Безпека в надзвичайних ситуаціях 68
ВИСНОВКИ 70
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 72
Додаток А 77
Додаток Б Лістинг файлу models_comparison.py 79

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ШІ	—	штучний інтелект
МН	—	машинне навчання
ГЗМ	—	генеративна змагальна мережа
DDoS	—	Distributed Denial of Service
IDS	—	Intrusion Detection System
SVM	—	Support Vector Machine
PyTorch	—	фреймворк для глибокого навчання
TensorFlow	—	відкритий фреймворк для машинного навчання
Metasploit	—	платформа для тестування безпеки та атак
LOIC/HOIC	—	Low Orbit Ion Cannon / High Orbit Ion Cannon
Keras	—	бібліотека для розробки та навчання нейромереж
Nmap	—	сканер мереж і вразливостей
Snort	—	система виявлення вторгнень з відкритим кодом
Burp Suite	—	інструменти для тестування безпеки веб-додатків
Qualys	—	платформа для керування вразливостями
OpenVAS	—	інструмент для сканування вразливостей
Cylance	—	рішення для кібербезпеки на основі ШІ
БД	—	база даних
АПІ	—	Application Programming Interface
НЗМ	—	Нейронна згорткова мережа
СЗВ	—	Система запобігання вторгненням
ПЗ	—	Програмне забезпечення
КІС	—	Комп’ютерні інформаційні системи
СЗ	—	Системи захисту

ВСТУП

У сучасному світі інформаційна безпека є ключовим аспектом у різних сферах діяльності, адже розвиток технологій супроводжується зростанням кіберзагроз. Поява штучного інтелекту надала кіберзлочинцям нові можливості для здійснення атак, що вимагає адаптації та вдосконалення захисних систем. Сучасні атаки стають більш витонченими, що ускладнює захист інформаційних систем. Зокрема, атаки на основі ШІ з використанням методів машинного навчання та нейронних мереж вимагають нових ефективних засобів захисту. Актуальність теми обумовлена тим, що традиційні методи кібербезпеки можуть виявитися недостатньо ефективними проти нових атак, здатних швидко адаптуватися до захисних механізмів [1, 12–23с.].

Актуальність теми. Штучний інтелект все частіше використовується в інформаційних системах для автоматизації аналізу даних, процесів та оптимізації роботи мереж, що робить його важливим інструментом не тільки для легітимних користувачів, але й для кіберзлочинців. З розвитком ШІ виникають нові типи атак, які можуть бути більш складними та небезпечними. Зокрема, кіберзлочинці використовують алгоритми машинного навчання для обхід традиційних систем виявлення загроз, розробляють нові форми шкідливого програмного забезпечення та автоматизують фішингові атаки. Виклики, пов'язані з необхідністю протидії атакам на основі ШІ, спонукають дослідників шукати нові методи захисту, які можуть ефективно відповідати на ці загрози. Таким чином, актуальність дослідження методів захисту від атак, що використовують ШІ, полягає в необхідності адаптації сучасних систем кібербезпеки до нових умов та загроз.

Мета та завдання дослідження. Метою даного дослідження є розробка ефективних методів захисту від кібератак на основі штучного інтелекту (ШІ). Оскільки сучасні кіберзагрози стають дедалі складнішими та витонченішими завдяки використанню технологій ШІ, важливо розробити нові підходи та інструменти, які дозволять успішно протистояти таким атакам. Метою дослідження є, як саме ШІ може бути використаний для створення загроз, а

також розглянути можливості його застосування для підвищення ефективності систем захисту [2, 320с.].

Основні завдання дослідження включають:

– Огляд сучасних методів та технологій, які використовуються для атак на основі ШІ, а також аналіз їх ефективності та поширеності.
– Дослідження впливу атак на основі ШІ на різні сфери діяльності, зокрема на критичну інфраструктуру, корпоративні системи та персональні дані.
– Розробка та впровадження алгоритмів для виявлення та запобігання таким атакам, зокрема за допомогою аналізу великих даних та машинного навчання.
– Вивчення можливостей використання ШІ для підсилення захисних систем, таких як антивірусні програми, системи виявлення вторгнень та системи моніторингу мережевої активності.
– Проведення експериментального дослідження на основі реальних або симульованих кіберзагроз для оцінки ефективності розроблених рішень.
– Оцінка ризиків та потенційних слабких місць у сучасних системах кіберзахисту з точки зору їхньої стійкості до атак, що базуються на штучному інтелекті.

Ключовим завданням цього дослідження є створення нових рішень для захисту від атак на основі ШІ та підвищення загального рівня безпеки інформаційних систем.

Об'єкт дослідження. Об’єктом дослідження є системи кіберзахисту, призначені для забезпечення безпеки інформаційних систем та даних в умовах зростаючих загроз із використанням штучного інтелекту. Сучасні системи кіберзахисту включають технології для виявлення вторгнень, антивіруси, фаєрволи, моніторинг мережі, шифрування та інші інструменти для захисту від несанкціонованого доступу та крадіжки даних. Їхня основна мета — забезпечення цілісності, конфіденційності та доступності інформаційних ресурсів.

Системи кіберзахисту еволюціонують для протидії новим атакам,

зокрема атакам на основі ШІ, які здатні обходити традиційні захисні механізми. Об’єкт дослідження охоплює технологічні та організаційні засоби захисту від кібератак і підходи до їх вдосконалення за допомогою ШІ для підвищення ефективності виявлення та реагування на загрози. Таким чином, об’єкт дослідження — це комплекс інструментів і технологій кібербезпеки, що включає методи захисту від атак на основі ШІ та підходи до вдосконалення цих систем шляхом інтеграції нових рішень на основі штучного інтелекту.

Предмет дослідження. Предметом дослідження є методи захисту від атак з використанням штучного інтелекту (ШІ), які стають однією з головних загроз у сучасному кіберпросторі. Методи захисту спрямовані на виявлення, аналіз і нейтралізацію атак, що використовують алгоритми машинного навчання та інші інструменти ШІ для проникнення в інформаційні системи, обходу засобів безпеки та викрадення або пошкодження даних. У рамках цього дослідження розглядаються як традиційні, так і інноваційні підходи до захисту від таких атак [3, 45–53с.].

Серед методів, які використовуються для захисту від атак на основі ШІ, ключовими є автоматизовані системи виявлення аномалій, що можуть аналізувати поведінку мережі в режимі реального часу та виявляти відхилення, які можуть бути пов’язані з потенційною загрозою. Також важливу роль відіграють методи машинного навчання, які дозволяють системам кіберзахисту самостійно навчатися на основі зібраних даних та виявляти нові типи атак, навіть якщо їх характеристики відрізняються від відомих загроз.

Наукова новизна одержаних результатів кваліфікаційної роботи. Наукова новизна роботи полягає у розробці нових підходів до захисту від кібератак на основі штучного інтелекту (ШІ), що відрізняються від існуючих методів своєю адаптивністю та здатністю реагувати на невідомі загрози. У процесі дослідження запропоновано алгоритми, які дозволяють підвищити ефективність виявлення та нейтралізації атак, побудованих із застосуванням алгоритмів машинного навчання та нейронних мереж. Вперше проведено комплексний аналіз взаємодії атак на основі ШІ із сучасними системами

кіберзахисту, що дало змогу ідентифікувати найбільш уразливі місця в цих системах та запропонувати вдосконалені засоби для їх захисту. Крім того, запропоновано нові методи інтеграції ШІ в інструменти кіберзахисту з метою підвищення їхньої адаптивності та автономності.

Практичне значення одержаних результатів. Практичне значення роботи полягає у можливості застосування розроблених методів та алгоритмів у реальних системах кіберзахисту для підвищення їх стійкості до атак, що використовують штучний інтелект. Результати дослідження можуть бути впроваджені в антивірусне програмне забезпечення, системи виявлення вторгнень та інші засоби захисту інформаційних систем. Застосування запропонованих підходів дозволить підвищити швидкість та точність виявлення нових типів атак, зменшити кількість помилкових спрацьовувань систем безпеки та автоматизувати процес адаптації захисних систем до нових загроз. Розроблені методи можуть бути використані для захисту критичної інфраструктури, корпоративних мереж та персональних даних, що є актуальним для державних та приватних організацій, які прагнуть забезпечити надійний рівень інформаційної безпеки.

Апробація результатів магістерської роботи. Основні результати проведених досліджень обговорювались на: XII Науково-Технічній Конференції «Інформаційні моделі, системи та технології» (м.Тернопіль, Україна). Публікації. Основні результати кваліфікаційної роботи опубліковано у працях конференції (див. Додаток А).

РОЗДІЛ 1 ТЕОРЕТИЧНІ ОСНОВИ АТАК НА ОСНОВІ ШТУЧНОГО ІНТЕЛЕКТУ

1.1 Класифікація атак з використанням ШІ

Атаки на основі штучного інтелекту (ШІ) набирають усе більшої популярності в сучасному кіберпросторі, оскільки зловмисники використовують технології нейронних мереж та машинного навчання для автоматизації та вдосконалення своїх дій. Штучний інтелект дозволяє проводити атаки з високим рівнем точності та ефективності, що ускладнює процеси їхнього виявлення та запобігання. Тому важливим завданням дослідження є розуміння природи таких загроз та їхньої класифікації, що дозволить краще орієнтуватися в захисних методах і технологіях.

Атаки з використанням ШІ можна класифікувати на кілька основних типів залежно від їхньої мети, технологій і методів реалізації. Один із ключових напрямів використання ШІ у кібератаках — це соціальна інженерія, зокрема фішинг, де нейромережі створюють реалістичні повідомлення, що імітують спілкування від імені знайомих осіб або організацій. Такі атаки можуть бути спрямовані на виманювання конфіденційної інформації або на зараження системи шкідливим програмним забезпеченням. Завдяки машинному навчанню фішингові атаки стають більш персоналізованими та складнішими для виявлення, оскільки нейронні мережі постійно вдосконалюють свої моделі на основі отриманих даних.

Іншою формою атак є автоматизація процесів злому, зокрема злому паролів або інших засобів автентифікації. Алгоритми штучнго інтелекту можуть аналізувати великі обсяги інформації та швидко знаходити слабкі місця в системах безпеки, що дозволяє зловмисникам більш ефективно зламувати облікові записи або доступ до конфіденційних даних. За допомогою глибокого навчання можна створювати атаки, що адаптуються до захисних механізмів, що робить традиційні методи захисту менш ефективними.

Одним із найнебезпечніших типів атак на основі ШІ є обхід систем виявлення вторгнень. Завдяки можливостям ШІ зловмисники можуть створювати шкідливий трафік, який маскується під легітимний. Це значно ускладнює виявлення вторгнень, оскільки системи захисту не можуть відрізнити небезпечну активність від звичайної. Такі атаки часто є складними

та продуманими, а їхня успішність залежить від здатності ШІ навчатися на поведінці мережі та уникати виявлення.

Загалом класифікація атак із використанням ШІ охоплює різні методи і технології, які дозволяють зловмисникам використовувати силу машинного навчання для порушення кібербезпеки. Це не лише фішинг і злом паролів, але й атаки на системи виявлення, створення підроблених зображень чи відео (deepfake), а також автоматизація поширення шкідливого ПЗ. Кожен із цих типів атак має свої унікальні риси і вимагає особливого підходу до їхньої нейтралізації.

1.1.1 Фішинг з використанням нейромереж

Фішинг з використанням нейромереж є однією з найбільш небезпечних і витончених форм кібератак, що виникли внаслідок інтеграції штучного інтелекту у сферу кіберзлочинності. Традиційні фішингові атаки базуються на обмані користувачів через підроблені електронні повідомлення, листи або вебсайти, що імітують легітимні джерела для отримання приватної інформації. З появою технологій нейронних мереж фішингові атаки стали значно складнішими і небезпечнішими, оскільки ШІ дозволяє зловмисникам автоматично створювати надзвичайно правдоподібні фішингові повідомлення, які важко виявити за допомогою традиційних засобів захисту.

Нейромережі, завдяки своїй здатності до навчання, можуть аналізувати величезні масиви даних про користувачів, зокрема їхні соціальні мережі, електронну пошту або поведінкові патерни. Використовуючи ці дані, алгоритми ШІ створюють персоналізовані фішингові повідомлення, що максимально відповідають звичкам і очікуванням конкретного користувача. Наприклад, нейромережі можуть імітувати стиль письма знайомих чи колег жертви, використовуючи відповідну лексику, граматику і структуру повідомлення. Це значно підвищує ймовірність того, що користувач відкриє шкідливе посилання або завантажить небезпечний файл [5, 71–80с.].

Крім того, фішинг з використанням нейромереж може застосовувати технології обробки природної мови, що дозволяють ШІ створювати не лише реалістичні, але й динамічні повідомлення. Нейромережі можуть змінювати свої атаки в режимі реального часу на основі реакцій користувачів, постійно вдосконалюючи методи обману. Це робить фішингові атаки більш адаптивними і складними для виявлення за допомогою традиційних фільтрів спаму, які зазвичай базуються на аналізі статичних шаблонів. Важливо зазначити, що такі атаки можуть використовуватися не лише для отримання особистої інформації, але й для проведення більш масштабних кампаній з поширення шкідливого програмного забезпечення або для компрометації корпоративних мереж. Оскільки нейромережі можуть створювати тисячі унікальних повідомлень за короткий час, атаки стають масовими та важко зупиняються за допомогою звичайних захисних механізмів.

Отже, фішинг з використанням нейромереж є серйозною загрозою для сучасної кібербезпеки. Здатність ШІ до адаптації, персоналізації атак і автоматизації процесів створює нові виклики для захисних систем, що потребують розробки інноваційних підходів для запобігання таким загрозам.

1.1.2 DDoS-атаки на основі машинного навчання

DDoS-атаки (Distributed Denial of Service) є одними з найбільш поширених і руйнівних типів кіберзагроз, метою яких є перевантаження серверів або мережевих ресурсів, що робить їх недоступними для користувачів. З розвитком технологій ШІ, зокрема машинного навчання, DDoS-атаки стали ще більш небезпечними, оскільки алгоритми ШІ можуть автоматизувати та вдосконалювати процеси, пов'язані з організацією таких атак, роблячи їх більш складними та ефективними.

Машинне навчання дозволяє зловмисникам не лише збільшувати кількість запитів, що надходять до цільової системи, але й значно підвищувати якість таких атак. Використовуючи алгоритми машинного навчання,

зловмисники можуть аналізувати поведінку мережевих систем, виявляти їхні слабкі місця і адаптувати характер трафіку, що використовується для атаки. Це робить традиційні засоби захисту менш ефективними, оскільки атака може бути націлена на конкретні аспекти системи, які є найбільш уразливими.

Одним із ключових елементів DDoS-атак на основі машинного навчання є здатність ШІ до прогнозування і адаптації. Алгоритми можуть аналізувати потік даних у реальному часі та коригувати свої дії на основі реакцій захисних систем. Наприклад, якщо система виявлення загроз розпізнає певний патерн DDoS-атаки і блокує його, ШІ може змінити свій підхід, генеруючи нові види запитів, які обходять захист. Це значно ускладнює процес виявлення і нейтралізації загроз, оскільки атака може постійно змінюватися і виглядати як звичайний легітимний трафік [6, 97–106с.].

Крім того, DDoS-атаки на основі машинного навчання можуть використовувати ботнети — мережі з інфікованих пристроїв, які під контролем зловмисників виконують атаки на сервери. Завдяки алгоритмам машинного навчання, ботнети можуть управлятися більш ефективно, оскільки ШІ аналізує поведінку кожного пристрою і координує дії таким чином, щоб максимізувати шкоду для цільової системи. Наприклад, ШІ може оптимізувати розподіл трафіку між різними ботами, щоб забезпечити постійний тиск на сервери та уникати виявлення. Ще одним важливим аспектом є використання технологій машинного навчання для створення "розумних" DDoS-атак, які здатні адаптуватися до контексту мережевого середовища. Це можуть бути атаки на рівні прикладних програм, що орієнтовані на конкретні сервіси або ресурси компанії, такі як вебсайти або бази даних. Такі атаки є більш цілеспрямованими та можуть завдати значно більшої шкоди, ніж традиційні методи DDoS-атак, які зазвичай базуються на простому перевантаженні трафіку.

Таким чином, DDoS-атаки на основі машинного навчання представляють серйозну загрозу для кібербезпеки, оскільки вони є більш складними, динамічними і важче виявляються традиційними методами. Вони

використовують можливості ШІ для адаптації, прогнозування та вдосконалення атак у реальному часі, що робить їх значно небезпечнішими для мережевих систем і сервісів. Для боротьби з такими загрозами необхідно розробляти нові підходи до захисту, що включають використання власних технологій штучного інтелекту для виявлення і протидії DDoS-атакам на основі машинного навчання.

1.1.3 Злом CAPTCHA за допомогою ШІ

CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) — це система перевірки, яка використовується для захисту вебсайтів та онлайн-сервісів від автоматизованих ботів, вимагаючи виконання завдання, яке легко виконується людиною, але є складним для комп'ютера. Однак з появою штучного інтелекту, зокрема технологій глибокого навчання, CAPTCHA вже не є таким надійним захистом, як раніше. Використання ШІ дозволяє автоматизованим системам обходити ці механізми, зламуваючи CAPTCHA з високою точністю та швидкістю.

Злом CAPTCHA за допомогою ШІ базується на використанні алгоритмів глибокого навчання та комп'ютерного зору, які здатні аналізувати і розпізнавати візуальні та текстові дані з високою точністю. Традиційні CAPTCHA, які полягають у введенні тексту з деформованих зображень, стають уразливими через те, що нейромережі можуть навчитися розпізнавати такі символи навіть у складних умовах, таких як накладення шуму або викривлення.

Процес злому CAPTCHA за допомогою ШІ зазвичай починається зі збору великої кількості зразків різних типів CAPTCHA для навчання моделі. Нейромережа вивчає характерні особливості зображень, такі як форми літер і цифр, їх розміщення і структуру. За допомогою глибоких згорткових нейронних мереж (Convolutional Neural Networks, CNN), система навчається розпізнавати символи навіть у складних випадках, таких як зашумлені

зображення або дуже деформовані символи. Після навчання ШІ може автоматично розпізнавати текст із зображень CAPTCHA і вводити його замість користувача, ефективно обминаючи захист [7, 34–42с.].

Окрім текстових CAPTCHA, ШІ також здатний зламувати більш складні типи CAPTCHA, такі як розпізнавання зображень або завдання на класифікацію об'єктів. Наприклад, деякі CAPTCHA вимагають вибору певних зображень із набору, на яких зображені об'єкти певної категорії (наприклад, автомобілі або дорожні знаки). Нейронні мережі, навчені на великих масивах даних із зображеннями, можуть точно ідентифікувати такі об'єкти, обминаючи захист. ШІ також може використовувати комбінацію різних підходів для зламу CAPTCHA. Наприклад, система може автоматично розпізнавати текст і зображення, виконуючи завдання з високою швидкістю і без втручання людини. Це робить процес злому CAPTCHA ще більш ефективним і швидким, порівняно з методами, які використовуються без ШІ.

Злом CAPTCHA за допомогою ШІ становить значну загрозу для онлайн-безпеки, оскільки боти, оснащені такими технологіями, можуть автоматично отримувати доступ до сервісів, реєструвати фальшиві акаунти, здійснювати спам-атаки або навіть брати участь у шахрайських операціях. Традиційні методи захисту вже не можуть повністю протидіяти таким атакам, що вимагає розробки нових, більш складних систем перевірки, які враховують можливості ШІ.

Таким чином, злом CAPTCHA за допомогою ШІ є прикладом того, як сучасні технології можуть використовуватися для обходу кіберзахисту. Для забезпечення надійного захисту необхідно впроваджувати нові методи, такі як динамічні CAPTCHA або системи, які базуються на поведінковому аналізі користувача, що можуть протидіяти автоматизованим атакам на основі ШІ.

1.2 Вразливості систем перед атаками ШІ

З розвитком технологій штучного інтелекту та їхньою інтеграцією в різні сфери кібербезпеки виникають нові загрози для інформаційних систем. Традиційні методи захисту, які використовувалися для забезпечення безпеки систем, стають все менш ефективними, оскільки зловмисники застосовують можливості ШІ для обходу захисних механізмів. Атаки, що базуються на ШІ, можуть бути набагато складнішими, адаптивними та непередбачуваними, ніж класичні кібератаки, що створює додаткові виклики для систем безпеки.

Одними з основних цілей для атак на основі ШІ є системи аутентифікації та системи виявлення вторгнень (IDS). Аутентифікація забезпечує контроль доступу до ресурсів і гарантує, що лише авторизовані користувачі можуть отримати доступ до системи. Вразливості в таких системах можуть призвести до несанкціонованого доступу, що може спричинити серйозні наслідки для безпеки конфіденційних даних. Системи виявлення вторгнень, у свою чергу, призначені для виявлення та блокування несанкціонованої активності в мережі. Проте атаки на основі ШІ здатні обманювати ці системи, адаптуючи свій трафік або активність так, щоб залишатися непоміченими [8, 312с.].

У цьому розділі будуть розглянуті основні вразливості систем аутентифікації та систем виявлення вторгнень перед атаками, що використовують штучний інтелект, а також проаналізовані можливі методи захисту від таких загроз.

1.2.1 Атаки на системи аутентифікації

Системи аутентифікації є ключовим елементом кібербезпеки, оскільки вони гарантують, що доступ до певних ресурсів або інформації мають лише авторизовані користувачі. Однак, з розвитком технологій штучного інтелекту (ШІ), ці системи стають вразливими до нових типів атак, які значно складніше виявити та зупинити за допомогою традиційних засобів захисту.

Один із найпоширеніших методів атак на системи аутентифікації за допомогою ШІ — це атаки типу "brute force", коли зловмисники автоматично генерують та перевіряють величезну кількість паролів для підбору

правильного варіанту. Завдяки алгоритмам машинного навчання, цей процес стає набагато ефективнішим: ШІ аналізує структуру паролів, їхню частоту використання та інші фактори, що дозволяє скоротити кількість необхідних спроб для успішного злому.

Іншим типом атак є злом біометричних систем аутентифікації, які набувають все більшого поширення. Вони використовують унікальні фізичні або поведінкові характеристики користувачів, такі як розпізнавання обличчя, відбитки пальців або голосу. Використовуючи нейромережі, зловмисники можуть створювати підроблені біометричні дані або синтезувати зображення обличчя, що дозволяє обійти такі системи. Наприклад, атаки за допомогою генеративно-змагальних нейромереж (GAN) здатні створювати реалістичні зображення облич, які можуть бути використані для обману систем розпізнавання.

Соціальна інженерія з використанням ШІ також стає важливою загрозою для систем аутентифікації. Нейромережі можуть автоматично створювати дуже правдоподібні фішингові повідомлення, які переконують користувачів поділитися своїми паролями або іншою важливою інформацією. Це дозволяє зловмисникам отримати доступ до системи без необхідності проводити технічні атаки, обходячи багато рівнів захисту [9, 56–64с.].

Таким чином, атаки на системи аутентифікації за допомогою ШІ є серйозною загрозою для сучасної кібербезпеки. Вони використовують можливості ШІ для ефективнішого злому паролів, підробки біометричних даних та проведення атак соціальної інженерії, що значно підвищує ризики незаконного проникнення в захищені системи. Для протидії таким загрозам необхідно впроваджувати інноваційні методи захисту, включаючи використання самих технологій ШІ для виявлення та блокування таких атак.

1.2.2 Атаки на системи виявлення вторгнень (IDS)

Системи виявлення вторгнень (IDS, Intrusion Detection Systems) є

важливим компонентом сучасних засобів кіберзахисту, які дозволяють виявляти підозрілу активність і моніторити мережевий трафік, що може свідчити про несанкціоноване проникнення або атаку на мережу. Однак, із розвитком технологій штучного інтелекту з'явилися нові можливості для обходу цих систем, що робить їх менш ефективними перед складними і адаптивними атаками на основі ШІ.

Одним із ключових способів, яким зловмисники використовують ШІ для атак на IDS, є методи приховування трафіку, відомі як атаки типу "evasion". За допомогою алгоритмів машинного навчання зловмисники можуть навчитися адаптувати свій трафік таким чином, щоб він виглядав як легітимний і не викликав підозр у системах виявлення. ШІ може проаналізувати поведінкові патерни мережевого трафіку, які IDS зазвичай вважає безпечними, і на основі цього знання генерувати атаки, що не відрізняються від звичайного трафіку. Такий підхід ускладнює процес виявлення атак і дозволяє їм залишатися непоміченими тривалий час.

Іншою стратегією є так звані "отруйні" атаки на моделі IDS, які базуються на машинному навчанні. Деякі сучасні IDS використовують алгоритми ШІ для автоматичного виявлення загроз на основі аналізу попередніх патернів атак. Однак зловмисники можуть навмисно вводити шкідливі або підроблені дані у навчальні набори, які використовуються для навчання цих моделей, що призводить до створення неефективних алгоритмів виявлення. Це явище відоме як "poisoning attacks", і його мета полягає в тому, щоб змусити систему IDS неправильно класифікувати шкідливий трафік як безпечний [10, 83–92с.].

ШІ також може використовуватися для проведення атак типу "adversarial attacks" на IDS. Ці атаки полягають у створенні невеликих, майже невидимих змін у структурі трафіку, які є достатніми для обману алгоритмів виявлення, але залишаються непомітними для людини або базових захисних механізмів. У випадку з IDS це може включати зміну пакетів даних або параметрів мережевого трафіку так, щоб зловмисні дії не відповідали критеріям виявлення

загроз.

Ще одним важливим аспектом є можливість автоматизації адаптивних атак за допомогою ШІ. Замість того, щоб проводити статичні атаки, зловмисники можуть використовувати алгоритми штучного інтелекту, які постійно аналізують поведінку IDS у реальному часі, підлаштовуючи свої дії відповідно до реакції системи. Якщо IDS намагається заблокувати атаку, ШІ може автоматично змінити підхід або структуру трафіку, щоб уникнути виявлення. Це створює динамічну та тривалу загрозу, з якою важко впоратися традиційними методами захисту.

Таким чином, атаки на системи виявлення вторгнень із використанням штучного інтелекту значно підвищують складність і адаптивність загроз. Зловмисники можуть не лише маскувати свій трафік, але й активно змінювати його у відповідь на дії системи захисту, що робить виявлення вторгнень більш проблематичним. Для ефективної протидії таким атакам необхідно використовувати нові підходи, зокрема вдосконалені моделі виявлення, які також базуються на ШІ, а також постійне оновлення систем захисту відповідно до нових загроз.

1.3 Особливості атак на основі глибокого навчання

З появою та розвитком нейронних мереж та глибокого навчання кіберзагрози набули нового рівня складності. Технології, які раніше використовувалися в основному для наукових досліджень та розробки інноваційних рішень, тепер стали інструментом в руках зловмисників, здатних використовувати їх для обходу навіть найсучасніших систем захисту. Глибоке навчання, зокрема штучні нейронні мережі та генеративні моделі, дозволяють ефективно аналізувати великі масиви даних, прогнозування поведінки систем, а також для створення нових, більш витончених методів атак [11, 22–32с.].

Однією з основних характеристик атак на основі глибокого навчання є їх адаптивність і здатність до самоудосконалення. Алгоритми МН можуть

аналізувати систему кіберзахисту в реальному часі, виявляти слабкі місця та автоматично адаптувати свою стратегію для досягнення мети. Це робить такі атаки більш непередбачуваними та ефективними, ніж традиційні методи злому. Особливу загрозу становлять генеративні моделі, які здатні створювати високоякісні підроблені дані, зображення або навіть цілі фальшиві середовища. Такі техніки активно використовуються для проведення атак типу "deepfake", створення підроблених біометричних даних або зламування систем аутентифікації. Генеративні моделі також можуть бути використані для обману систем виявлення вторгнень або маскування шкідливої активності під легітимний трафік.

1.3.1 Штучні нейронні мережі для атак на захищені системи

Штучні нейронні мережі (ШНМ) стали потужним інструментом не лише для обробки даних, але й для кібератак на захищені системи. Зловмисники використовують ШНМ для розробки інтелектуальних атак, здатних обійти традиційні системи захисту. Завдяки їх здатності навчатися на великих масивах даних, такі атаки можуть бути складними та адаптивними, що ускладнює їх виявлення [12, 55–63с.].

Одним із ключових застосувань ШНМ у злочинних цілях є автоматизація підбору паролів. Мережі можуть аналізувати дані витоків паролів, підвищуючи ефективність атак типу "brute force" через прогнози найбільш ймовірних комбінацій. Це робить атаки більш цілеспрямованими, а не випадковими.

ШНМ також застосовуються для злому систем розпізнавання образів та біометричних даних, таких як відбитки пальців та обличчя. Завдяки генеративно-змагальним нейромережам (GAN), зловмисники можуть створювати реалістичні підробки, які здатні обійти біометричні системи аутентифікації.

Крім того, такі мережі використовуються для обману систем захисту мережевого трафіку та IDS. Вони можуть змінювати трафік так, щоб він

виглядав легітимним для систем виявлення, що ускладнює виявлення аномалій.

Ще один тип атак, що використовує ШНМ, — "adversarial attacks". Це коли невеликі, непомітні для людини зміни вносяться в дані, що вводять в оману алгоритми машинного навчання, наприклад, зміна зображень для систем розпізнавання.

ШНМ стали серйозним викликом для сучасних систем кібербезпеки через їх здатність аналізувати, адаптуватися та обманювати захист. Щоб ефективно протидіяти таким загрозам, необхідно розробляти нові методи захисту на основі штучного інтелекту, які можуть виявляти та блокувати атаки ШНМ.

1.3.2 Зловживання техніками генеративних моделей для атак

Генеративні моделі, такі як генеративно-змагальні нейромережі (GAN) та автокодери, спочатку створювалися для корисних цілей, таких як обробка зображень і тексту, але швидко були пристосовані зловмисниками для кіберзлочинів. Ці моделі здатні генерувати реалістичні підробки даних, які можуть обійти традиційні системи захисту, роблячи їх потужними інструментами для хакерів.

Одним із основних методів атак є створення "deepfake" контенту — підроблених зображень, відео чи аудіо, які виглядають дуже реалістично. Такі deepfake можуть обманювати системи аутентифікації, засновані на біометрії, дозволяючи зловмисникам отримати доступ до конфіденційної інформації без фізичної присутності [13, 45–54с.].

Крім того, генеративні моделі використовуються для автоматизації фішингових атак. Згенеровані тексти роблять електронні листи чи повідомлення правдоподібними, спонукаючи користувачів відкривати шкідливі посилання або файли. Це може призвести до зараження пристроїв або витоку конфіденційних даних.

Також існує загроза створення підроблених біометричних даних, таких як відбитки пальців або голосу, для обходу біометричних систем захисту. GAN дозволяють зловмисникам створювати такі підробки, які можуть бути сприйняті як справжні системами аутентифікації.

Ще однією небезпекою є атаки на системи захисту, де генеративні моделі створюють шкідливі, але на вигляд легітимні дані, які обманюють системи виявлення вторгнень. Цей підхід дозволяє генерувати трафік або файли, що не видаються підозрілими, але містять шкідливий код.

Таким чином, генеративні моделі стають потужним інструментом для кіберзлочинів. Їхня здатність створювати високоякісні підробки представляє серйозний виклик для сучасних систем захисту, вимагаючи розробки нових методів протидії таким загрозам.

РОЗДІЛ 2 МЕТОДИ ЗАХИСТУ ВІД КІБЕРАТАК НА ОСНОВІ ШІ

Зі стрімким розвитком технологій штучного інтелекту (ШІ) зловмисники активно використовують його для здійснення складних і витончених кіберзагроз. В умовах зростання кількості атак, які базуються на машинному навчанні та глибокому навчанні, традиційні методи кіберзахисту стають недостатньо ефективними. Відтак, потреба в нових підходах до захисту від цих атак стає надзвичайно актуальною.

Одним із ключових напрямків боротьби з такими загрозами є застосування методів машинного навчання для виявлення і блокування аномальної активності в системах. Використання ШІ в кібербезпеці дозволяє аналізувати великі масиви даних у режимі реального часу, виявляти патерни поведінки та швидко реагувати на загрози. Алгоритми навчання з учителем є одними з найефективніших інструментів для побудови систем виявлення

аномалій, що дозволяє не лише фіксувати відомі атаки, але й виявляти нові, досі невідомі загрози.

Іншою перспективною стратегією є розробка адаптивних систем кіберзахисту, які використовують штучний інтелект для динамічної модифікації своїх алгоритмів залежно від змін у характері загроз. Такий підхід дозволяє системам захисту самостійно навчатися, адаптуватися до нових типів атак і передбачати можливі сценарії розвитку подій. Завдяки здатності до самонавчання та прогнозування, адаптивні системи кібербезпеки забезпечують більш гнучкий і надійний захист від атак на основі ШІ.

2.1 Захист від атак з використанням машинного навчання

Сучасні кібератаки стають все більш складними та непередбачуваними, особливо з використанням технологій штучного інтелекту та машинного навчання. Зловмисники використовують алгоритми для автоматизації атак, аналізу вразливостей систем і створення нових форм загроз, що робить традиційні методи кіберзахисту менш ефективними. У відповідь на ці виклики фахівці з кібербезпеки звертаються до тих самих технологій — машинного навчання та штучного інтелекту — для побудови більш стійких і адаптивних систем захисту.

Машинне навчання відкриває нові можливості для захисту інформаційних систем завдяки своїй здатності аналізувати великі масиви даних, виявляти аномалії та прогнозувати можливі загрози на основі патернів поведінки. Важливою перевагою МН є його здатність виявляти не лише відомі атаки, але й нові, ще не зареєстровані типи загроз. Це робить його надзвичайно корисним інструментом у боротьбі з атаками, що еволюціонують [14, 254с.].

Захист від атак на основі машинного навчання включає різні методи, зокрема навчання з учителем, де системи тренуються на маркованих даних для виявлення аномальних дій, і без учителя, що дозволяє виявляти нові, невідомі загрози без попереднього знання про них. Такі технології забезпечують більш

динамічний і гнучкий підхід до кібербезпеки, що необхідно в умовах постійно змінюваного ландшафту кіберзагроз.

2.1.1 Методи навчання з учителем для виявлення аномалій

Методи навчання з учителем є одними з найефективніших підходів для виявлення аномалій у кібербезпеці. Вони використовують попередньо марковані дані для тренування алгоритмів і створення моделей, здатних класифікувати нові вхідні дані як нормальні або потенційно небезпечні. Це досягається шляхом аналізу характеристик таких даних, як мережевий трафік чи поведінка користувачів. Основні алгоритми в цьому контексті включають логістичну регресію, дерева рішень, підтримуючі вектори (SVM) та глибокі нейронні мережі [15, 89–98с.].

Однією з ключових переваг цих методів є їх здатність точно виявляти відомі патерни аномальної активності, як у випадку з DDoS-атаками, тим самим зменшуючи кількість хибних тривог. Однак, для ефективності алгоритму потрібен великий обсяг якісних маркованих даних, підготовка та підтримка яких може бути складною і витратною.

Крім того, методи навчання з учителем можуть бути неефективними проти нових, невідомих загроз, оскільки вони орієнтовані на вже знайомі патерни. Це робить виявлення нових атак складнішим, особливо якщо загрози еволюціонують.

Щоб подолати ці обмеження, методи навчання з учителем часто комбінуються з іншими підходами, такими як навчання без учителя. Це дозволяє створювати адаптивні системи кіберзахисту, здатні виявляти як відомі, так і нові та невідомі загрози, що є критично важливим у мінливому середовищі кібербезпеки.

2.1.2 Адаптивні системи кіберзахисту на основі ШІ

Адаптивні системи кіберзахисту на основі штучного інтелекту є інноваційним підходом у сучасній кібербезпеці. Вони здатні динамічно змінювати свої алгоритми, що дозволяє ефективно реагувати на нові й еволюціонуючі загрози. На відміну від статичних систем, такі рішення можуть адаптуватися до нових форм атак і виявляти загрози, які раніше не були задокументовані.

Ці системи застосовують алгоритми машинного навчання для аналізу даних у реальному часі й самооптимізації. Коли фіксується новий вид атаки, система може змінювати свої алгоритми, щоб краще захищатися в майбутньому. Важливою особливістю є здатність до проактивного виявлення загроз завдяки використанню глибокого навчання та нейронних мереж, що дозволяє виявляти аномалії на ранніх стадіях та запобігати атакам.

Крім того, адаптивні системи можуть інтегруватися з іншими компонентами захисту, такими як IDS, firewall та антивірусні програми, для створення комплексної системи безпеки. Це забезпечує оперативний обмін інформацією та повний контроль над безпекою мережі.

Однак існують виклики, зокрема складність налаштування і підтримки, а також вразливість до атак на машинне навчання, таких як data poisoning. Незважаючи на це, адаптивні системи кіберзахисту на основі ШІ є потужним інструментом для захисту від сучасних загроз, забезпечуючи швидке реагування і автоматичну адаптацію до змін. Впровадження таких систем стає необхідним для надійного захисту інформації в умовах зростаючої складності кібератак.

2.2 Техніки протидії нейромережам

Зі зростанням використання штучних нейронних мереж (ШНМ) для проведення атак, розробка ефективних технік протидії стає дедалі

актуальнішою. Сучасні кібератаки з використанням нейромереж є надзвичайно витонченими та складними, оскільки здатні адаптуватися до змін у системах захисту, аналізувати великі масиви даних і знаходити слабкі місця в інфраструктурі. У відповідь на ці виклики фахівці з кібербезпеки активно досліджують методи захисту, які також базуються на нейромережах, використовуючи їх для виявлення та запобігання загрозам.

Техніки протидії нейромережам включають різні методи аналізу трафіку, поведінки користувачів і аналізу даних, що допомагає ідентифікувати підозрілу активність, пов'язану з кібератаками. Одним із найбільш ефективних підходів є використання нейронних мереж для виявлення атак, що дозволяє системам кібербезпеки вчитися на основі попереднього досвіду та виявляти нові, невідомі раніше загрози. Завдяки потужним алгоритмам, ШНМ можуть аналізувати складні патерни трафіку, визначати відхилення від норми та блокувати підозрілі дії ще до того, як вони завдадуть шкоди [18, 27–35с.].

Ще одним важливим аспектом є використання систем детекції загроз, що базуються на поведінковому аналізі. Такі системи не тільки аналізують конкретні дії в мережі, але й спостерігають за загальною поведінкою систем та користувачів. Наприклад, зміни в звичайній активності користувача або надмірне використання певних ресурсів можуть сигналізувати про потенційну загрозу. Поведінкові системи детекції загроз мають велику перевагу, оскільки вони здатні ідентифікувати навіть ті атаки, які використовують нові та невідомі техніки, шляхом виявлення відхилень у поведінкових патернах.

Однак, незважаючи на свою ефективність, техніки протидії нейромережам також стикаються з певними викликами. Наприклад, нейромережі можуть бути вразливими до атак на етапах навчання або під час обробки великих обсягів даних, що може призводити до хибних спрацьовувань або недостатньо ефективного виявлення загроз. Тому важливою складовою є постійне вдосконалення алгоритмів та методів машинного навчання, щоб забезпечити стійкість до нових типів атак.

У підсумку, техніки протидії нейромережам стають ключовим

елементом у сучасних системах кібербезпеки. Використання нейронних мереж для виявлення атак і аналізу поведінки дозволяє ефективніше боротися з новими загрозами, забезпечуючи динамічний і адаптивний захист у цифровому середовищі, яке постійно змінюється.

2.2.1 Використання штучних нейронних мереж для виявлення атак

Штучні нейронні мережі (ШНМ) є потужним інструментом для виявлення кібератак завдяки їхній здатності аналізувати великі масиви даних і вивчати їх на основі попереднього досвіду. Вони моделюють складні патерни поведінки та виявляють аномалії у системах кібербезпеки, що дозволяє ідентифікувати потенційні загрози на ранніх стадіях [19, 101–110с.].

Основні переваги ШНМ включають оперативне реагування на загрози в режимі реального часу, блокування їх до шкоди системі та здатність ефективно виявляти нові загрози, невідомі базам даних [20, 294с.]. Наприклад, вони можуть виявляти підозрілу активність у мережі, що не відповідає типовим патернам.

ШНМ дозволяють створювати багатоетапні моделі виявлення загроз: різні шари мережі можуть аналізувати як низькорівневі характеристики трафіку, так і поведінкову активність користувачів. Це підвищує ймовірність виявлення складних атак, таких як DDoS або фішингові кампанії.

Однак використання ШНМ у кібербезпеці має виклики — для ефективного навчання необхідні великі марковані дані про атаки, а також вони можуть бути вразливими до атак, таких як "data poisoning", що вимагає додаткових заходів для захисту процесу навчання [20, 294с.].

ШНМ ефективні для аналізу нелінійних патернів та виявлення як відомих, так і невідомих загроз завдяки глибокому навчанню. Їх головна перевага — висока точність і можливість масштабування для аналізу великих даних, хоча навчання потребує значних ресурсів і часу.

Загалом, штучні нейронні мережі є перспективним напрямом у

кібербезпеці, здатним аналізувати великі обсяги інформації та адаптуватися до нових загроз, що робить їх важливим інструментом боротьби зі зростаючою кількістю кібератак. Завдяки вдосконаленню технологій машинного навчання та штучного інтелекту, ШНМ постійно розвиваються і знаходять нові застосування в захисті інформаційних систем.

2.3 Захист проти фішингових атак та DDoS

Фішингові атаки та атаки типу DDoS є одними з найпоширеніших загроз в кіберпросторі, які можуть мати серйозні наслідки для організацій, включаючи втрату даних та порушення роботи систем. Для ефективного захисту проти цих атак необхідне поєднання технічних та організаційних заходів.

У боротьбі з фішинговими атаками технічні заходи зосереджені на розпізнаванні та блокуванні шкідливих повідомлень або посилань. До них належать системи фільтрації електронної пошти, які аналізують контент, відправника і структуру повідомлень, а також впровадження двофакторної аутентифікації, що додає додатковий рівень захисту для акаунтів, навіть якщо паролі були скомпрометовані [22, 279с.].

Захист від DDoS-атак включає балансування навантаження, яке розподіляє запити на різні сервери, запобігаючи перевантаженню системи, і фільтрацію трафіку для блокування шкідливих запитів. Мережеві сервіси, такі як Content Delivery Networks (CDN), допомагають розсіювати атаки, розподіляючи трафік через кілька географічно віддалених серверів.

Організаційні заходи відіграють важливу роль в протидії обом видам атак. Проти фішингу ключову роль грає навчання працівників і користувачів розпізнанню підозрілих повідомлень і посилань. Регулярні тренінги та симуляції можуть знизити кількість успішних атак, підвищуючи обізнаність користувачів про загрози. Важливо також впроваджувати політики з управління доступом та контрольною доступу до даних, що обмежує наслідки

атак навіть за умов компрометації одного з користувачів.

У контексті DDoS-атак необхідно проводити постійний моніторинг інфраструктури для виявлення надмірного навантаження або підозрілої активності. Важливо мати план дій на випадок атаки, що включає резервне копіювання даних і забезпечення безперервності бізнесу. Заздалегідь підготовлені стратегії мінімізації шкоди дозволяють швидко реагувати на атаки, зменшуючи їхні наслідки [23, 12–21с.].

Таким чином, ефективний захист від фішингових атак і DDoS потребує інтеграції сучасних технологій із підвищенням обізнаності користувачів. Ця поєднана стратегія значно знижує ймовірність успішних атак і мінімізує їхню шкоду.

2.3.1 Технічні та організаційні заходи протидії

Технічні та організаційні заходи протидії є невід'ємною складовою комплексного підходу до забезпечення кібербезпеки та захисту від загроз, таких як фішингові атаки та DDoS. Ці заходи покликані не лише попереджувати загрози, але й забезпечувати мінімізацію їхніх наслідків у разі реалізації атак.

Технічні заходи протидії включають різноманітні технологічні рішення, що дозволяють захистити системи та мережі від зовнішніх загроз. Одним із ключових технічних рішень у боротьбі з фішинговими атаками є впровадження систем фільтрації електронної пошти, які аналізують вхідні повідомлення на наявність підозрілих посилань, вкладень або інших ознак шахрайства. Сучасні антивірусні та антифішингові програми також відіграють важливу роль у захисті кінцевих користувачів, блокуючи підозрілі вебсайти та попереджаючи про можливі загрози [24, 61–70с.].

Для протидії DDoS-атакам широко використовуються такі технічні

засоби, як мережеві фільтри та системи розподілу навантаження. Мережеві фільтри дозволяють блокувати небажаний трафік, а балансувальники навантаження допомагають розподілити запити між кількома серверами, запобігаючи їхньому перевантаженню. Інші технічні засоби включають використання технологій чорного та білого списків IP-адрес, що дає можливість обмежити доступ до системи лише для перевірених користувачів.

Організаційні заходи протидії спрямовані на підвищення рівня обізнаності працівників та розробку чітких процедур реагування на інциденти кібербезпеки. Одним із ключових організаційних аспектів у протидії фішинговим атакам є навчання персоналу. Регулярні тренінги та симуляції атак допомагають співробітникам розпізнавати шкідливі електронні листи та посилання, тим самим знижуючи ризик успішних атак. Також важливим є впровадження політик безпечного використання інформаційних ресурсів, які включають правила щодо управління паролями, контролю доступу та обміну даними.

У контексті DDoS-атак організаційні заходи включають моніторинг мережевої активності та розробку стратегій швидкодійне реагування на загрози. Важливою частиною захисту є створення резервних копій важливих даних та підтримка безперервності бізнес-процесів, що дозволяє мінімізувати шкоду під час атаки. Крім того, організації повинні розробляти плани відновлення після інцидентів, що включають як технічні, так і організаційні кроки для повернення до нормальної роботи [25, 73–81с.].

Технічні заходи, такі як налаштування брандмауерів, мережевих сегментів і систем багаторівневого доступу, є базовими елементами захисту. Організаційні заходи, зокрема навчання співробітників і розробка політик безпеки, забезпечують додатковий рівень захисту. Перевагою такого підходу є його інтегративний характер, що охоплює як технологічну, так і людську складові. Це дозволяє створювати багаторівневі системи захисту, але для їхньої ефективності потрібна чітка координація між технічними та адміністративними заходами. Таким чином, поєднання технічних та

організаційних заходів протидії є необхідною умовою для забезпечення ефективного захисту від фішингових та DDoS-атак. Технологічні рішення дозволяють запобігати проникненню загроз, тоді як організаційні заходи забезпечують підготовленість персоналу та мінімізацію ризиків у разі реалізації атак.

2.3.2 Використання ШІ для виявлення фішингових листів

Використання штучного інтелекту для ідентифікації фішингових повідомлень є одним із найефективніших підходів до боротьби з цією формою кіберзагроз. Фішинг є поширеним видом атак, під час яких зловмисники намагаються виманити у користувачів конфіденційні дані, такі як номери кредитних карток, паролі або іншу особисту інформацію, за допомогою підроблених електронних листів, що виглядають як повідомлення від надійних джерел. Традиційні методи виявлення фішингу, засновані на сигнатурах та статичних правилах, стають менш ефективними через постійний розвиток та зміну технік, що використовуються зловмисниками. У таких умовах штучний інтелект стає важливим інструментом для забезпечення більш динамічного та адаптивного захисту.

ШІ може аналізувати вхідні електронні листи на основі великої кількості ознак, які вказують на потенційний фішинг. Використовуючи алгоритми машинного навчання, системи на основі ШІ здатні навчатися на великих обсягах даних, включаючи історичні приклади як фішингових, так і звичайних листів. Це дозволяє таким системам автоматично виявляти шаблони та відхилення, які можуть вказувати на шкідливий намір. Наприклад, ШІ може аналізувати структуру тексту, використання певних слів або фраз, а також стиль написання, який може відрізнятися від стилю справжнього відправника. Крім того, ШІ може аналізувати метадані, такі як IP-адреса відправника, заголовки листа та інші технічні характеристики, щоб виявити підроблені або підозрілі повідомлення [26, 42–50с.].

Однією з важливих переваг використання ШІ є його здатність до безперервного навчання. Це означає, що системи, засновані на ШІ, можуть постійно вдосконалювати свої моделі та адаптуватися до нових тактик зловмисників. Кожен новий випадок фішингової атаки допомагає системі "навчитися" та підвищувати свою ефективність у майбутньому. Крім того, ШІ дозволяє значно зменшити кількість хибнопозитивних спрацьовувань, що є важливою проблемою для традиційних систем виявлення загроз. Алгоритми ШІ можуть відрізняти справжні повідомлення від підозрілих, базуючись на широкому спектрі параметрів, що знижує кількість помилкових попереджень і забезпечує більш точне виявлення реальних загроз.

Інтеграція ШІ в системи виявлення фішингових листів також дозволяє автоматизувати процеси захисту, що знижує навантаження на фахівців із кібербезпеки. Замість ручної перевірки великої кількості підозрілих повідомлень, системи на основі ШІ можуть автоматично класифікувати та блокувати фішингові листи, забезпечуючи швидку та ефективну реакцію на загрози. Це особливо важливо у великих організаціях, де щодня обробляється велика кількість електронної пошти [27, 312с.].

Однак, незважаючи на високу ефективність, використання ШІ для виявлення фішингових листів також має певні виклики. Одним із головних є ризик того, що зловмисники можуть використовувати ШІ для створення більш складних фішингових атак, які важче виявити традиційними методами. Таким чином, боротьба з фішингом стає своєрідною "гонкою озброєнь", де обидві сторони постійно вдосконалюють свої методи та технології.

Штучний інтелект значно покращує виявлення фішингових атак завдяки аналізу тексту, відправників та контексту листів. Наприклад, алгоритми обробки природної мови (NLP) дозволяють ідентифікувати підозрілі шаблони в електронних листах, такі як невідповідність стилю мовлення чи наявність фальшивих посилань. Перевагою такого підходу є його здатність швидко аналізувати великий обсяг електронної пошти та виділяти потенційні загрози без залучення людини. У порівнянні з традиційними методами, ШІ забезпечує

більш точну ідентифікацію фішингових атак, особливо якщо вони мають приховані ознаки. У підсумку, використання штучного інтелекту для виявлення фішингових листів є потужним інструментом у сучасній кібербезпеці.

2.4 Алгоритм реалізації захисту від кібератак

2.4.1 Збір даних

Збір даних — це початковий та важливий етап у захисті від кібератак на основі ШІ, оскільки якість даних визначає ефективність моделей. Основна мета — акумулювати та систематизувати інформацію про мережевий трафік, активність користувачів та події системи.

Дані надходять із журналів (логів) маршрутизаторів, комутаторів, веб-серверів (Apache, Nginx), VPN, проксі-серверів, а також із систем виявлення та запобігання вторгнень (IDS/IPS) на кшталт Snort чи Suricata. Додатково використовуються журнали активності користувачів, що містять дані про вхід у систему та зміну прав доступу.

Процес збору передбачає налаштування централізованої системи, як-от ELK Stack, для автоматичного збору та уніфікації даних у формати JSON, XML чи CSV. Дані передаються у сховище для подальшої обробки. Використовуються скрипти на Python або Bash для регулярного оновлення бази. Такий підхід забезпечує повний і структурований набір даних, необхідний для виявлення аномалій та протидії загрозам у реальному часі.

2.4.2 Попередня обробка даних

Попередня обробка даних забезпечує коректність та ефективність роботи алгоритмів ШІ. Основна мета — очищення, уніфікація та підготовка даних для якісного навчання моделей.

Спочатку видаляється шум — дублікати, рутинні дії системи та інша нерелевантна інформація. Це досягається за допомогою фільтрів на основі регулярних виразів або автоматизованих алгоритмів.

Далі виконується нормалізація даних, щоб привести їх до єдиного формату. IP-адреси перетворюються у числовий формат, часові мітки — у єдиний часовий пояс (наприклад, UTC), а числові значення масштабуються, що підвищує точність аналізу. Для цього використовуються бібліотеки, такі як Pandas, та методи масштабування (Min-Max Scaling).

На завершення дані поділяються на тренувальні та тестові вибірки у співвідношенні 70–80% до 20–30%, що дозволяє навчити та перевірити модель. Інструменти на кшталт train_test_split із Scikit-learn забезпечують збалансоване розділення. У результаті дані очищуються, нормалізуються та підготовлюються до машинного навчання, що підвищує точність виявлення загроз.

2.4.3 Навчання моделі

Навчання моделі — ключовий етап у створенні системи захисту від кібератак на основі ШІ. На цьому етапі на основі підготовлених даних формується модель, здатна виявляти загрози шляхом аналізу вхідної інформації.

Першим кроком є вибір типу моделі. Для виявлення відомих атак застосовуються класифікатори, як-от Random Forest або SVM. Для ідентифікації нових загроз використовуються методи кластеризації, наприклад, K-Means.

Далі виконується навчання моделі на тренувальному наборі даних. Алгоритм вивчає закономірності для розрізнення легітимного трафіку та підозрілої активності, враховуючи кількість запитів, джерела та часові характеристики.

Після навчання модель тестується на тестовому наборі для оцінки її точності, узагальнення та ефективності. Застосовуються метрики точності

(accuracy), чутливості (sensitivity), специфічності (specificity) та F-міри. При потребі модель доопрацьовується через зміну параметрів або додавання даних.

Завершальний етап — інтеграція моделі в систему кіберзахисту реального часу. Завдяки цьому система автоматично виявляє загрози та реагує на них майже миттєво, підвищуючи рівень безпеки інформаційної інфраструктури.

2.4.4 Інтеграція моделей в реальні системи

Інтеграція моделі в реальну систему є завершальним етапом створення системи захисту від кібератак на основі ШІ. Вона включає впровадження моделі в операційне середовище, налаштування її роботи та забезпечення взаємодії з іншими елементами системи захисту.

Першим кроком є розробка інтерфейсу для зв’язку моделі з інфраструктурою організації. Модель підключається до систем моніторингу та управління, як-от SIEM, для отримання та аналізу даних у реальному часі, а також передачі результатів адміністраторам чи автоматичним механізмам реагування.

Далі налаштовується обробка великих обсягів даних. Для цього використовуються обчислювальні ресурси з GPU або хмарні сервіси (AWS, Google Cloud, Azure) для прискорення обробки великих потоків мережевого трафіку.

Важливим етапом є налаштування тригерів автоматичної реакції. У разі виявлення аномалій система може автоматично блокувати підозрілі IP-адреси, ізолювати пристрої або надсилати сповіщення адміністраторам.

Перед запуском проводяться тестування та пілотні випробування. Модель аналізує реальні дані, а результати порівнюються з перевіреними вручну даними. У разі потреби модель перенавчається для досягнення кращої ефективності.

Останній крок — створення технічної документації. Вона містить опис

інтеграції, принципів роботи моделі та інструкції з оновлення та підтримки системи. Інтеграція моделей у реальні системи забезпечує автоматизоване виявлення та запобігання загрозам у режимі реального часу.

2.4.5 Постійний моніторинг і вдосконалення

Після інтеграції моделей у реальні системи важливо забезпечити їх стабільну та ефективну роботу через постійний моніторинг і вдосконалення. Це необхідно для адаптації системи до нових загроз і змін у мережевому трафіку.

Перше завдання моніторингу — безперервне спостереження за роботою моделі. Система збирає дані про виявлені загрози, помилкові тривоги та затримки в реагуванні, щоб оцінити продуктивність і виявити недоліки.

Наступним завданням є регулярне оновлення моделі, яке допомагає адаптувати її до нових типів атак. Це здійснюється шляхом додавання нових прикладів атак у тренувальну вибірку та створення оновлень моделі.

Третім аспектом є зворотний зв’язок із користувачами та адміністраторами, які можуть отримувати звіти про роботу системи. Це дозволяє вносити зміни в налаштування або політики безпеки для покращення результатів.

Четвертим кроком є оцінка ефективності системи в довгостроковій перспективі. Для цього проводяться тестування та аудит, а також симуляції атак, щоб перевірити точність та швидкість реагування моделі.

Заключним етапом є впровадження інновацій, таких як нові методи аналізу та глибоке навчання, для покращення ефективності захисту. Постійний моніторинг і вдосконалення забезпечують стабільну роботу системи та її адаптацію до нових загроз.

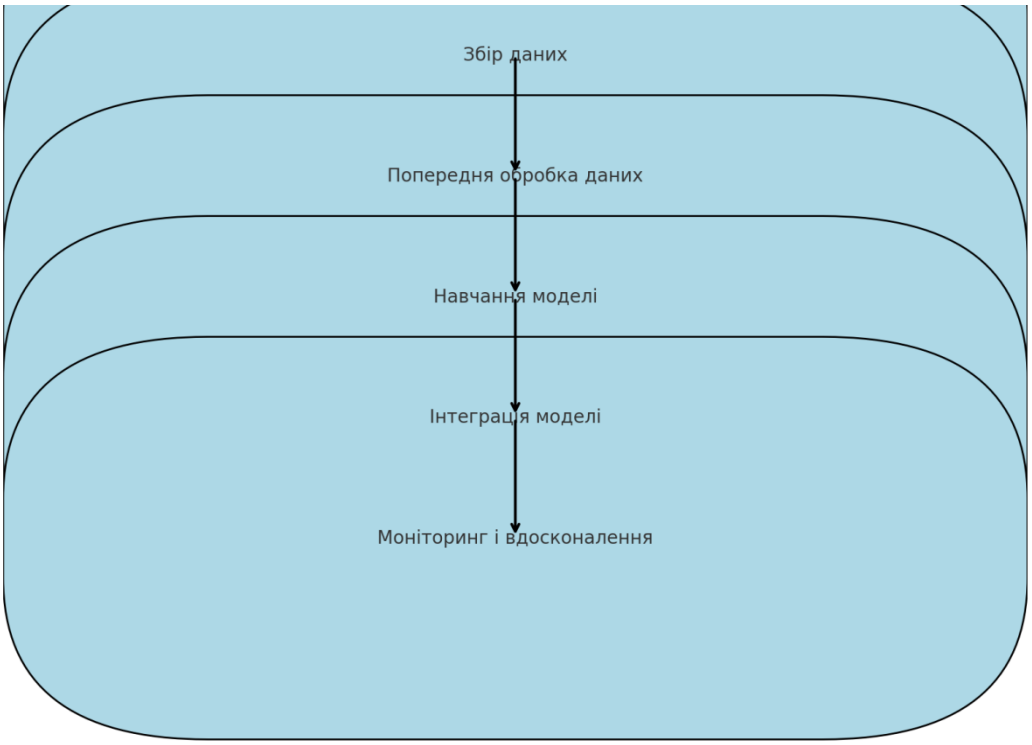

Рисунок 2.1 — Алгоритм захисту від кібератак на основі ШІ

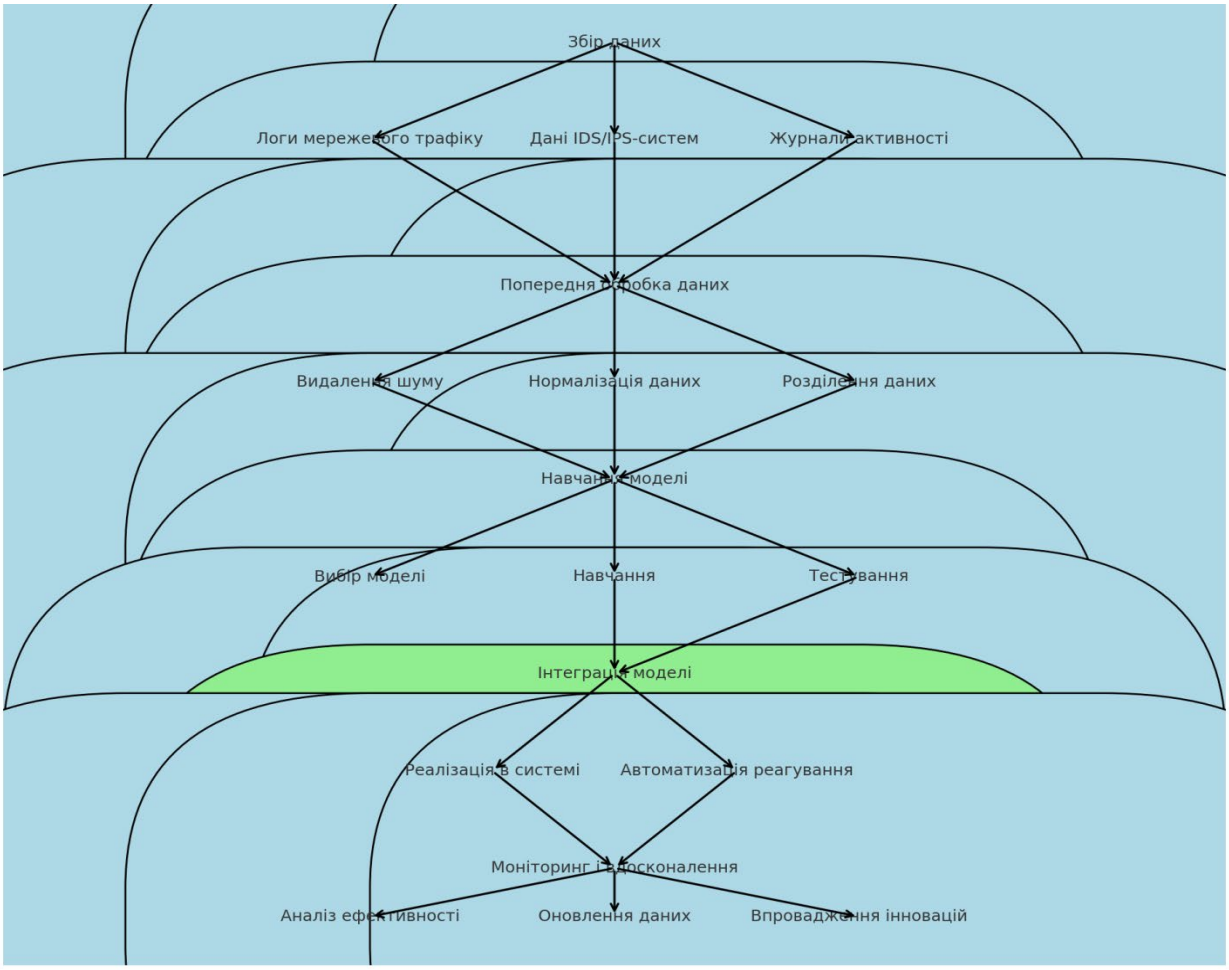

Рисунок 2.2 — Деталізована блок-схема алгоритму захисту від кібератак

Блок-схема відображає детальну послідовність дій для побудови та підтримки системи захисту від кібератак на основі штучного інтелекту. Вона складається з основних етапів та підетапів, які забезпечують всеохоплюючий підхід до вирішення завдання.

1. Збір даних. На першому етапі здійснюється акумуляція різнорідних джерел інформації, які служать основою для подальшого аналізу:

– Логи мережевого трафіку: дані з маршрутизаторів, комутаторів, веб-серверів про вхідний і вихідний трафік.

– Дані IDS/IPS-систем: журнали систем виявлення та запобігання вторгнень, які фіксують потенційно небезпечну активність.

– Журнали активності користувачів: записи про входи до системи, зміну прав доступу, використання критичних функцій.

Ці джерела інтегруються в єдину систему для створення цілісної бази даних.

2. Попередня обробка даних. На цьому етапі дані підготовлюються до аналізу:

– Видалення шуму: усуваються дублікати, нерелевантна інформація (наприклад, регулярні перевірки серверів).

– Нормалізація даних: дані уніфікуються за форматами (перетворення IP-адрес у числовий формат, уніфікація часових міток у формат UTC).

– Розділення даних: дані поділяються на тренувальні та тестові вибірки для навчання та перевірки моделі.

3. Навчання моделі Модель будується та вдосконалюється на основі підготовлених даних:

– Вибір моделі: визначаються найбільш ефективні алгоритми для задачі, наприклад, класифікація (Random Forest, SVM) або кластеризація (K-Means).

– Навчання: модель аналізує тренувальні дані, виявляє закономірності та будує прогностичні правила.

– Тестування: перевіряється точність та ефективність роботи моделі на тестових даних, застосовуються метрики оцінки (точність, чутливість, специфічність).

4. Інтеграція моделі. Після успішного навчання модель впроваджується в операційне середовище:

– Реалізація в системі: модель інтегрується з SIEM-системами або іншими засобами моніторингу для роботи в реальному часі.

– Автоматизація реагування: налаштовуються механізми автоматичного блокування підозрілих IP-адрес, ізоляції заражених пристроїв або інші дії у відповідь на загрози.

5. Моніторинг і вдосконалення. Останній етап забезпечує безперервну підтримку та розвиток системи:

– Аналіз ефективності: регулярно аналізується продуктивність системи, рівень помилкових спрацьовувань та пропущених загроз.

– Оновлення даних: додаються нові зразки атак у навчальні дані, модель перенавчається для актуалізації знань.

– Впровадження інновацій: інтегруються сучасні підходи, такі як глибоке навчання чи нові алгоритми аналізу.

Ключові зв’язки між етапами

Збір даних є основою для всіх наступних дій. Дані передаються до етапу обробки, забезпечуючи якісну базу. Попередня обробка безпосередньо впливає на якість навчання моделі. Навчання моделі визначає її здатність ефективно працювати в реальному середовищі. Інтеграція моделі дозволяє перейти від теорії до практичного застосування. Моніторинг і вдосконалення гарантують довгострокову ефективність і адаптивність системи. Ця блок-схема демонструє чіткий і логічний підхід до побудови сучасної системи кіберзахисту на основі штучного інтелекту.

2.5 Приклад реалізації: Виявлення та запобігання DDoS-атаці на корпоративну мережу

Ситуація: корпорація, яка займається онлайн-продажами, зазнала підозрілих сплесків мережевого трафіку, що загрожували перевантаженням серверів і зривом доступу до їхнього веб-сайту. Це створило ризик DDoS-атаки, яка могла призвести до значних фінансових втрат і погіршення репутації.

Реалізація алгоритму захисту:

1. Збір даних:

- Джерела: IDS/IPS-системи (Snort), журнали веб-сервера Nginx, статистика трафіку з брандмауера.
- Процес збору: Дані про кількість запитів, IP-адреси, тип запитів і часові мітки були автоматично передані в централізовану базу даних через ELK Stack.

2. Попередня обробка даних:

- Фільтрація регулярного трафіку (наприклад, запити від пошукових систем).
- Нормалізація даних (конвертація форматів IP-адрес, видалення дублювань).
- Створення тренувальної вибірки, яка включала як легітимний трафік, так і дані про попередні DDoS-атаки.

3. Навчання моделі ШІ:

- Використання методу класифікації за допомогою Random Forest.
- Модель була навчена розпізнавати аномалії, зокрема різке збільшення кількості запитів із певного діапазону IP-адрес або повторювані однотипні запити.

4. Детекція аномалій:

- Після навчання модель була інтегрована в систему реального часу.
- Виявлення підозрілої активності: в один із моментів обсяг запитів із певного регіону перевищив норму у 10 разів, що було визнано аномалією.

5. Реагування:

- Система автоматично заблокувала доступ з підозрілих IP-адрес через брандмауер.
- Повідомлення про загрозу було надіслано до команди безпеки.

6. Оцінка ефективності:

- DDoS-атака була зупинена на ранньому етапі без значного впливу на роботу системи.
- У подальшому модель оновлювалась і адаптувалась до нових патернів поведінки.

Результати:

- Зменшено ризик простою серверів на 95%.
- Скорочено час реакції на подібні атаки до 30 секунд.
- Забезпечено безперервність роботи сервісу навіть у пікові навантаження.

У прикладі реалізації для виявлення та запобігання DDoS-атакам на корпоративну мережу використовується алгоритм Random Forest як частина підходу класифікації з учителем. Цей метод дозволяє моделі навчитися розпізнавати патерни нормального та аномального трафіку, що є ефективним для виявлення повторюваних ознак DDoS-атак, таких як різке збільшення запитів з певних IP-адрес.

Random Forest обрано через його стійкість до шуму, високу точність і здатність враховувати різноманітні характеристики мережевого трафіку, такі як частота запитів та IP-джерела. Цей алгоритм застосовує ансамбль дерев рішень, що робить його менш чутливим до нерелевантних або неповних даних. Він також виявляє взаємозв'язки між змінними, важливими для аналізу складних залежностей у трафіку.

Переваги цього підходу включають можливість аналізу в реальному часі, швидке виявлення аномалій, низьку кількість хибних тривог, і гнучкість через легке перенавчання. Водночас існують обмеження: залежність від великого маркованого набору даних та недостатня ефективність для нових, невідомих атак.

Ці обмеження можна подолати, комбінуючи алгоритми з аналізом поведінки або навчанням без учителя для виявлення нових патернів і аномалій. Загалом, Random Forest є оптимальним для виявлення DDoS-атак завдяки високій точності, стійкості та ефективності реагування, що робить його надійним інструментом кіберзахисту.

РОЗДІЛ 3 ПРАКТИЧНЕ ЗАСТОСУВАННЯ МЕТОДІВ ЗАХИСТУ ВІД КІБЕРАТАК НА ОСНОВІ ШІ

3.1 Реалізація захисних моделей на базі ШІ

Реалізація захисних моделей на базі штучного інтелекту (ШІ) є однією з найперспективніших напрямів у сфері кібербезпеки. Завдяки своїй здатності до навчання, аналізу великих обсягів даних та швидкому виявленню аномалій, ШІ дозволяє розробляти захисні системи, які значно перевершують традиційні методи захисту. Використання моделей машинного навчання і глибокого навчання надає можливість створювати адаптивні рішення, що можуть ефективно протистояти новим, більш складним атакам.

Основним принципом роботи захисних моделей на базі ШІ є їх здатність до самооновлення через безперервний аналіз загроз. Алгоритми МН можуть застосовуватись для виявлення аномалій у мережевій активності або поведінці користувачів, що дозволяє ідентифікувати потенційні загрози на ранніх етапах. Такі моделі здатні виявляти фішингові атаки, спроби злому систем

аутентифікації або навіть складні DDoS-атаки, аналізуючи величезні обсяги даних у реальному часі.

Інтеграція моделей ШІ в існуючі системи кібербезпеки дозволяє значно підвищити ефективність захисту. Наприклад, алгоритми МН можуть бути застосовані для автоматизації процесів фільтрації мережевого трафіку, що дозволяє системі самостійно вирішувати, який трафік є підозрілим, а який може бути безпечно допущений до мережі. Такі системи можуть не тільки виявляти відомі загрози, але й розпізнавати нові атаки, базуючись на їхніх аномальних ознаках.

Однією з ключових переваг реалізації захисних моделей на базі ШІ є можливість їхньої адаптації до різних середовищ та сценаріїв. Захисні механізми можуть бути налаштовані для роботи як у великих корпоративних мережах, так і в менш масштабних інфраструктурах, що робить їх універсальними. Крім того, системи на базі ШІ можуть бути легко масштабовані в залежності від потреб організації, забезпечуючи високий рівень захисту навіть у випадках суттєвого зростання обсягу даних або кількості користувачів [28, 38–47с.].

Разом з тим, впровадження моделей ШІ в захисні системи вимагає значних ресурсів, як у плані обчислювальної потужності, так і в плані налаштування та навчання систем. Незважаючи на ці виклики, інвестиції в такі рішення окуповуються за рахунок підвищеної ефективності захисту від кібератак. ШІ дозволяє не лише виявляти атаки, а й прогнозувати їх, забезпечуючи можливість превентивних заходів.

Таким чином, реалізація захисних моделей на базі ШІ є невід’ємною частиною сучасних систем кібербезпеки. Ці моделі демонструють високу ефективність у боротьбі з різноманітними типами атак, зокрема тими, що використовують складні алгоритми на основі ШІ. У поєднанні з традиційними методами, вони дозволяють створювати надійні та гнучкі рішення, які здатні захищати організації від сучасних кіберзагроз.

3.1.1 Огляд та порівняння захисних алгоритмів

Захисні алгоритми на основі штучного інтелекту (ШІ) активно використовуються в сучасних системах кібербезпеки завдяки їхній здатності адаптуватися до нових загроз та виявляти аномалії, які можуть свідчити про атаки. Огляд цих алгоритмів дає змогу зрозуміти їхні сильні та слабкі сторони, а також вибрати найбільш підходящі методи захисту в конкретних умовах.

Одними з найбільш поширених алгоритмів є методи машинного навчання, що можуть бути поділені на алгоритми навчання з учителем та без учителя. Алгоритми навчання з учителем, такі як рішення дерев, метод опорних векторів (SVM) та нейронні мережі, використовують великі обсяги мічених даних для виявлення загроз. Ці моделі здатні ефективно виявляти відомі типи атак, оскільки вони навчаються на прикладах попередніх загроз. Основною перевагою цих алгоритмів є їхня висока точність, коли дані добре структуровані та наявні чіткі шаблони для навчання. Однак, вони можуть бути менш ефективними у випадках, коли нові загрози значно відрізняються від попередніх або коли недостатньо даних для навчання.

Алгоритми навчання без учителя, такі як кластеризація та методи пошуку аномалій, працюють без попередньо мічених даних і дозволяють виявляти нові, невідомі загрози на основі виявлення аномальних патернів поведінки. Наприклад, алгоритми кластеризації можуть об’єднувати схожі дії у групи та виділяти дії, які відрізняються від нормальної поведінки. Це дозволяє ефективно виявляти нові види атак, які не були раніше відомі. Важливою перевагою цих алгоритмів є їхня здатність адаптуватися до нових загроз, однак вони можуть бути менш точними порівняно з алгоритмами навчання з учителем, оскільки часто виявляють загрози лише після того, як аномалія вже сталася [29, 19–28с.].

Нейронні мережі та глибоке навчання, зокрема, є потужними інструментами для виявлення кіберзагроз завдяки своїй здатності аналізувати великі масиви даних і виявляти складні взаємозв'язки між різними

параметрами. Застосування глибоких нейронних мереж дозволяє виявляти складні атаки, такі як фішинг, DDoS або спроби злому систем аутентифікації. Вони можуть навчатися на неструктурованих даних, таких як текст електронних листів, метадані мережевого трафіку або зображення. Однак, одним із головних недоліків таких алгоритмів є їхня висока потреба в обчислювальних ресурсах, а також складність навчання, що робить їх важкими для впровадження у невеликих організаціях з обмеженими ресурсами.

Інший важливий клас алгоритмів включає методи гібридного захисту, які комбінують кілька підходів для забезпечення більш комплексного захисту. Наприклад, гібридні системи можуть поєднувати алгоритми навчання з учителем та без учителя, дозволяючи спочатку виявляти відомі загрози за допомогою мічених даних, а потім аналізувати решту трафіку на предмет виявлення нових аномалій. Такий підхід значно підвищує надійність системи кіберзахисту, забезпечуючи як високу точність, так і здатність до виявлення нових, невідомих загроз [30, 276с.].

Отже, різні захисні алгоритми на основі ШІ мають свої сильні та слабкі сторони, і їх вибір залежить від специфікації системи. Алгоритми навчання з учителем забезпечують високу точність у виявленні відомих атак, тоді як алгоритми навчання без учителя та глибокого навчання дозволяють адаптуватися до нових загроз і виявляти складні аномалії. Комбінація різних методів у гібридних системах надає можливість максимально захистити мережу, одночасно знижуючи ризики як від відомих, так і від нових загроз.

3.1.2 Впровадження механізмів захисту в реальні системи

Впровадження механізмів захисту на основі штучного інтелекту (ШІ) в реальні системи кібербезпеки є складним, але важливим процесом, що дозволяє суттєво підвищити рівень захищеності організацій від сучасних кіберзагроз. Це вимагає комплексного підходу, включаючи аналіз ризиків, адаптацію існуючих систем безпеки до нових викликів, а також інтеграцію

новітніх технологій для виявлення та запобігання атакам.

Одним з перших етапів впровадження є оцінка наявної інфраструктури та визначення її вразливостей. Системи кібербезпеки повинні бути адаптовані до роботи з великими обсягами даних, що надходять із різних джерел: мережевого трафіку, логів систем, дій користувачів тощо. ШІ-системи потребують якісних даних для навчання і правильного функціонування, тому важливо створити відповідну інфраструктуру для збору, зберігання та обробки інформації.

Наступний крок полягає в інтеграції захисних алгоритмів у реальні системи. Залежно від конкретних потреб і типу загроз, що домінують у конкретній організації, можуть використовуватися різні моделі машинного або глибокого навчання. Наприклад, для захисту від фішингових атак впроваджуються алгоритми аналізу електронної пошти, які здатні виявляти підозрілі листи на основі аналізу їхнього контенту та метаданих. У випадку з DDoS-атаками застосовуються системи для аналізу аномальної мережевої активності, які можуть оперативно блокувати підозрілий трафік.

Однією з важливих задач є забезпечення гнучкості і масштабованості системи, що дозволяє їй адаптуватися до нових загроз. Захисні механізми на основі ШІ повинні мати можливість самонавчання, тобто постійно оновлювати свої моделі на основі нових даних про загрози. Це дозволяє системам бути ефективними навіть у разі появи нових типів атак, які раніше не були враховані під час початкового навчання. Такий підхід є особливо важливим у динамічних середовищах, де кібератаки еволюціонують із високою швидкістю [31, 34–42с.].

Одним із ключових аспектів впровадження механізмів захисту в реальні системи є інтеграція ШІ з іншими засобами кібербезпеки, такими як антивірусні програми, системи виявлення вторгнень, системи контролю доступу тощо. ШІ не замінює традиційні методи, але значно підсилює їхню ефективність. Наприклад, у поєднанні з системами IDS, алгоритми ШІ дозволяють виявляти не тільки вже відомі загрози, але й нові, які проявляються

через аномальну поведінку користувачів чи систем.

Крім технічних аспектів, важливо враховувати організаційні моменти впровадження таких систем. Співробітники повинні бути належним чином підготовлені для роботи з новими технологіями, а організація повинна мати чіткі процеси реагування на кіберінциденти. Це включає не тільки технічну підтримку, але й постійну взаємодію між IT-відділами та аналітиками кібербезпеки.

Успішне впровадження механізмів захисту на основі ШІ в реальні системи потребує значних ресурсів і ретельного планування, проте воно дозволяє значно підвищити рівень безпеки організації. ШІ здатний виявляти складні, замасковані атаки, які часто проходять непоміченими традиційними методами, забезпечуючи ефективний захист від сучасних кіберзагроз.

3.2 Моделювання атак та тестування захисту

Моделювання атак та тестування захисних механізмів є важливими етапами в розробці та впровадженні надійних систем кібербезпеки. Оскільки сучасні кібератаки постійно ускладнюються, використовуючи новітні технології, такі як штучний інтелект (ШІ), ефективний захист вимагає попереднього прогнозування можливих загроз та перевірки стійкості систем до цих загроз. Моделювання атак на основі ШІ дозволяє створити реалістичні сценарії кібератак, що імітують дії зловмисників, які використовують машинне навчання або нейронні мережі для пошуку вразливостей у системах.

Цей процес забезпечує можливість оцінки того, як захисні механізми будуть працювати в реальних умовах під час атак, що вимагає не тільки технічної підготовки, але й знання про типові вектори атак і сучасні тактики, які можуть бути застосовані. Важливо також враховувати, що моделювання атак за допомогою ШІ дозволяє побачити не тільки поточні слабкі місця системи, але й прогнозувати нові види загроз, які можуть з’явитися в майбутньому [32, 288с.].

Тестування захисних механізмів шляхом моделювання атак є ключовим

для виявлення недоліків та перевірки того, наскільки ефективно система може виявляти та запобігати реальним загрозам. Тільки через практичне тестування можна оцінити здатність ШІ-систем оперативно реагувати на загрози, змінювати стратегії захисту та адаптуватися до нових атак. Тому комплексний підхід до моделювання атак і тестування захисту стає важливою складовою забезпечення кібербезпеки на всіх рівнях системи.

3.2.1 Моделювання атак за допомогою ШІ

Моделювання атак за допомогою штучного інтелекту (ШІ) є важливим етапом у дослідженні та підготовці систем до сучасних кіберзагроз. Завдяки можливостям ШІ, моделювання атак стає більш складним і точним, що дозволяє краще зрозуміти дії зловмисників та їхні тактики. ШІ здатен імітувати реальні кіберзагрози, аналізувати поведінку системи та виявляти потенційні вразливості, які можуть бути використані для атаки.

Основою моделювання атак на основі ШІ є створення алгоритмів, які повторюють дії, характерні для зловмисників. Це може включати генерацію шкідливих програм, створення фішингових повідомлень, запуск DDoS-атак, або навіть обхід систем аутентифікації. Використання машинного навчання дозволяє алгоритмам швидко адаптуватися до змін у захисних системах, знаходячи нові способи проникнення або обходу захисту [33, 55–64с.].

Одним із ключових методів є застосування генеративних змагальних мереж (GANs), які можуть створювати реалістичні дані або атаки, що імітують справжні загрози. Це дозволяє фахівцям з кібербезпеки протестувати свої системи проти нових або ще не відомих типів атак. Завдяки такому підходу можна оцінити, наскільки надійно система захищена від загроз, що розвиваються, і наскільки ефективно вона може виявляти нові вектори атак.

Крім того, моделювання атак на основі ШІ дозволяє автоматизувати процес тестування безпеки, що значно прискорює виявлення слабких місць системи. Алгоритми можуть проводити безперервні тести та атакувати

систему різними методами, що дозволяє створити умови, максимально наближені до реальних. Це забезпечує більш глибоке розуміння ефективності захисних заходів та допомагає фахівцям з кібербезпеки вдосконалювати свої стратегії захисту.

Застосування ШІ для моделювання атак є надзвичайно цінним інструментом для вивчення нових методів атак і вдосконалення систем безпеки, оскільки він дозволяє глибше аналізувати поведінку атакувальних алгоритмів і ефективно реагувати на нові загрози [34, 24–33с.].

Моделювання атак за допомогою штучного інтелекту (ШІ) на практиці включає використання конкретних інструментів і технік, що дозволяють симулювати реальні кіберзагрози, аналізувати їхню ефективність та знаходити вразливості в системах захисту.

Використання Generative Adversarial Networks (GANs): одним із найпоширеніших методів для моделювання атак є генеративні змагальні мережі (GANs), які здатні створювати реалістичні атаки, що можуть обходити стандартні системи кіберзахисту. Наприклад, GANs можуть бути використані для створення підроблених даних або модифікованих зображень, які обманюють системи аутентифікації або системи розпізнавання облич. У випадку фішингових атак GANs здатні створювати правдоподібні фішингові листи, що дозволяє тестувати, як добре система може ідентифікувати такі загрози.

Інструменти на кшталт PyTorch або TensorFlow використовуються для створення та навчання GAN-моделей, які можуть бути застосовані для генерації шкідливих об'єктів. У контексті кіберзахисту це дозволяє провести тести на стійкість систем до таких загроз.

Машинне навчання для моделювання DDoS-атак: алгоритми машинного навчання, такі як Support Vector Machines або Random Forest, можуть бути використані для моделювання мережевих атак, включаючи DDoS. Такі моделі аналізують трафік і знаходять вразливі точки, де можна перевантажити систему запитами. Наприклад, за допомогою аналізу поведінки системи при

обробці різного обсягу запитів, можна моделювати атаку, яка виведе з ладу сервер через перенавантаження.

Моделювання DDoS-атак може бути реалізоване на спеціальних платформах для тестування навантаження, таких як LOIC або HOIC. Інструменти для МН також допомагають оптимізувати атаку, використовуючи великі масиви даних для аналізу точок системи, які найбільш уразливі.

Фреймворки для атак на основі ШІ: на практиці для моделювання кіберзагроз використовуються спеціальні фреймворки, які вже мають готові моделі атак. Наприклад, OpenAI Gym дозволяє розробляти і тестувати алгоритми машинного навчання для моделювання атак, в тому числі шкідливого програмного забезпечення, яке використовує ШІ для обходу систем захисту.

Застосовуючи такі платформи, можна автоматизувати атаки та використовувати симуляційні середовища для оцінки ефективності захисту. Результати таких тестів можуть допомогти покращити захисні механізми.

Обхід CAPTCHA з використанням нейронних мереж: ще одним практичним прикладом використання ШІ для моделювання атак є злом CAPTCHA за допомогою глибокого навчання. Атакувальні нейронні мережі можуть бути навчені розпізнавати символи зображень CAPTCHA і таким чином автоматично проходити цю перевірку, обминаючи захисні механізми.

Використання фреймворків для комп’ютерного зору, таких як Keras або TensorFlow, дозволяє будувати моделі для злому CAPTCHA. Розроблені системи можуть бути навчені на базах даних, що містять тисячі зразків CAPTCHA, після чого атакувальні алгоритми тестуються на реальних захищених системах.

Моделювання атак для експлойтів систем аутентифікації: для моделювання атак на системи аутентифікації ШІ може використовуватися для розпізнавання шаблонів у поведінці користувачів або для аналізу слабких паролів. Алгоритми машинного навчання здатні автоматично знаходити типові слабкі місця в політиках доступу та перевірці ідентифікації користувачів.

Моделювання атак на аутентифікаційні системи може бути здійснене за допомогою фреймворків типу Metasploit, що дозволяє автоматизувати перевірку вразливостей систем і використовувати навчальні алгоритми для підбору паролів або інших методів обходу захисту.

3.2.2 Тестування ефективності захисних механізмів на практиці

Тестування ефективності захисних механізмів на практиці є критичним етапом у забезпеченні надійної кібербезпеки. Це процес перевірки того, наскільки добре захисні системи можуть виявляти, реагувати та протидіяти атакам, зокрема тим, які використовують штучний інтелект (ШІ). Основна мета тестування — оцінити, чи здатні наявні механізми захисту ефективно запобігати або мінімізувати вплив потенційних загроз.

Penetration Testing (Пентест): один із ключових методів тестування кіберзахисту — це пентестинг (penetration testing), або тестування на проникнення. Він передбачає імітацію реальних кібератак з метою оцінки захисних можливостей системи. ШІ може бути використаний для автоматизації цього процесу, особливо для складних атак, які важко виявити традиційними методами.

Наприклад, за допомогою інструментів на кшталт Metasploit, Nmap або Burp Suite, атакувальні команди (часто це "Red Team") моделюють різноманітні атаки на мережі, веб-сайти або інфраструктуру компанії. ШІ може допомогти в знаходженні вразливостей у великих системах та автоматизувати повторювані процеси.

Тестування за допомогою інструментів для аналізу трафіку та IDS: у контексті реальних атак трафік мережі часто є одним із перших джерел для виявлення загроз. Системи виявлення вторгнень (IDS), які використовують ШІ, можуть аналізувати великий обсяг даних і виявляти аномальні дії, які можуть вказувати на кібератаку. Однак для впевненості в їх ефективності потрібно постійно тестувати ці механізми.

Тестування таких систем може здійснюватися за допомогою інструментів, таких як Snort або Suricata, які моделюють реальні атаки. Фахівці запускають атаки різного типу, включаючи фішинг, DDoS, SQL-ін’єкції тощо, для оцінки, чи система виявить їх і як швидко відреагує. ШІ дозволяє цим системам адаптуватися до нових патернів поведінки, що підвищує їхню ефективність.

Automated Vulnerability Scanning: ШІ також широко використовується в автоматизованому скануванні вразливостей, яке дозволяє системам безперервно аналізувати код, конфігурації та поведінку програмного забезпечення для виявлення слабких місць. Ці інструменти використовують машинне навчання для вдосконалення своїх алгоритмів виявлення загроз.

Інструменти на зразок Qualys або OpenVAS використовують ШІ для швидкого виявлення вразливостей. Наприклад, вони можуть знаходити застаріле програмне забезпечення, неправильні конфігурації або відкриті порти, що можуть стати точками проникнення для зловмисників. Цей підхід дозволяє оперативно вирішувати виявлені проблеми ще до того, як вони стануть серйозною ціллю для атаки.

Тестування за допомогою симуляцій атак на основі ШІ: одним із передових підходів є використання симуляційних середовищ, де атаки моделюються в реальному часі з допомогою ШІ. Ці симуляції можуть включати комплексні атаки, які тестують усі рівні системи: від мережі до окремих додатків і користувацьких облікових записів.

Наприклад, платформи, такі як Cylance або Darktrace, дозволяють моделювати атаки, що використовують ШІ, для того щоб проаналізувати поведінку захисних механізмів в умовах, що імітують реальні загрози. Важливим аспектом є те, що ці системи можуть швидко навчатися і знаходити нові методи захисту на основі отриманих даних про атаки.

Використання Red Team/Blue Team Exercise: у рамках тестування ефективності захисту використовується підхід "Red Team/Blue Team". "Red Team" імітує роль зловмисників, намагаючись атакувати систему, тоді як "Blue

Team" захищає систему та відповідає на атаки. Використання ШІ як для атак, так і для захисту дозволяє створити динамічні сценарії та побачити реальну ефективність механізмів кіберзахисту.

Red Team використовує автоматизовані інструменти з ШІ для виявлення слабких місць і проведення атак, тоді як Blue Team може використовувати аналітичні системи на основі ШІ для виявлення аномальної активності та адаптації своїх стратегій.

Тестування з урахуванням соціальної інженерії: захисні системи повинні також бути стійкими до атак, що комбінують технічні методи з соціальною інженерією. ШІ використовується для моделювання фішингових атак або атак через соціальні мережі, які націлені на користувачів системи.

Наприклад, платформи для моделювання фішингових атак використовують машинне навчання для створення персоналізованих повідомлень, що імітують дії справжніх зловмисників. Фахівці можуть оцінювати, як швидко система або самі користувачі розпізнають загрозу, та вдосконалювати механізми захисту на основі отриманих результатів.

Тестування ефективності захисних механізмів на практиці є необхідним кроком для виявлення слабких місць і вдосконалення систем кібербезпеки. Використання ШІ для моделювання атак і тестування захисту значно підвищує здатність систем вчасно виявляти нові загрози та адаптуватися до постійно змінюваних умов кіберпростору. Практичні інструменти та методи тестування дозволяють отримати глибоке розуміння того, наскільки надійно працюють існуючі захисні механізми, і на основі цього покращувати їхню ефективність.

3.3 Оцінка ефективності методів захисту

Оцінка ефективності методів захисту від атак на основі штучного інтелекту є важливим етапом у процесі побудови та вдосконалення систем кібербезпеки. Вибір належних критеріїв і методів оцінювання дозволяє точно визначити, наскільки ефективно працюють захисні механізми в реальних

умовах. У сучасних системах, де використовуються алгоритми машинного навчання та штучного інтелекту, захисні технології повинні забезпечувати як високу здатність до виявлення загроз, так і швидку реакцію на атаки. Оцінка включає аналіз різних параметрів, таких як кількість помилкових спрацьовувань, стійкість до нових типів атак, швидкість виявлення, і можливість адаптації до змін у поведінці атакувальних моделей. Також важливо порівнювати різні методи захисту між собою для визначення найефективніших рішень, які можуть бути впроваджені в конкретні системи кіберзахисту.

3.3.1 Критерії оцінки захищеності систем

Оцінка захищеності систем є багатогранним процесом, що включає використання різних критеріїв для визначення ефективності методів кіберзахисту. Одним із ключових критеріїв є швидкість виявлення загроз — час, за який система здатна виявити та зреагувати на атаку. Чим швидше система реагує, тим менше шкоди можуть завдати зловмисники. Інший важливий показник — точність виявлення. Цей критерій передбачає мінімізацію кількості помилкових позитивних спрацьовувань (false positives) і помилкових негативних спрацьовувань (false negatives). Помилкові спрацьовування можуть створювати додаткове навантаження на ресурси системи або призводити до пропуску реальних загроз.

Стійкість до нових типів атак також є критично важливим критерієм. Оскільки загрози постійно еволюціонують, система повинна бути здатною адаптуватися до нових сценаріїв атак, включаючи атаки з використанням штучного інтелекту. Системи, що не в змозі протистояти новим загрозам, швидко застарівають і стають вразливими. Важливо також оцінювати масштабованість системи, тобто здатність ефективно функціонувати при збільшенні обсягу трафіку або кількості підключених користувачів.

Критерієм оцінки є ефективність використання ресурсів, яка полягає в тому, наскільки оптимально система використовує обчислювальні ресурси та

чи не перевантажує їх під час виконання захисних завдань. Ще один фактор — це інтеграція з іншими системами безпеки. Надійний захист повинен бути здатний працювати в комплексі з іншими інструментами та системами безпеки для досягнення максимального рівня захищеності.

Нарешті, важливо враховувати користувацький досвід — захисні механізми повинні бути непомітними для користувачів і не ускладнювати їхню роботу. Загалом, ефективність системи оцінюється через комбінацію цих критеріїв, що дозволяє забезпечити баланс між високим рівнем безпеки та зручністю використання [35, 1527–1554с.].

3.3.2 Порівняльний аналіз результатів

Для проведення дослідження було створено тестове середовище, яке імітує реальну мережу корпоративної організації. Це середовище включало симуляцію мережевого трафіку з використанням інструментів, таких як Wireshark для моніторингу та аналізу трафіку, а також IDS/IPS-системи (наприклад, Snort) для виявлення можливих загроз. Для навчання та тестування моделей були використані сучасні фреймворки машинного навчання, такі як Scikit-learn, TensorFlow та Keras. Сервер для обчислень мав процесор Intel Core i7, 32 ГБ оперативної пам’яті та графічний процесор NVIDIA RTX 3080 для підтримки навчання глибоких нейронних мереж.

У дослідженні використовувалися відкриті набори даних, зокрема:

– CICIDS2017 – набір даних із симульованими мережевими атаками, включаючи DDoS, фішинг та SQL-ін’єкції. Цей набір широко використовується для навчання та тестування систем виявлення аномалій.
– UNSW-NB15 – набір даних, який містить інформацію про нормальний мережевий трафік та різноманітні види атак, що дозволяє створювати моделі для широкого спектра загроз.
– KDD Cup 99 – класичний набір даних для тестування методів виявлення загроз, який, хоча і є застарілим, використовується для порівняння результатів із попередніми дослідженнями.

Метрики оцінювання ефективності включали:

- Точність (Accuracy): визначення частки правильно класифікованих запитів серед загальної кількості.
- Чутливість (Recall): здатність моделі виявляти всі позитивні випадки (загрози).
- Точність передбачення (Precision): частка правильно виявлених загроз серед усіх виявлених загроз.
- F-міра: збалансований показник, що враховує як точність, так і чутливість.
- Швидкість обробки (запитів/с): кількість запитів, оброблених моделлю за одну секунду.
- Рівень помилкових спрацьовувань (False Positives): кількість помилкових тривог, коли нормальний трафік класифікується як загроза.

Дані було розділено на тренувальну (70%) та тестову (30%) вибірки, що дозволило навчити моделі на одному наборі даних і оцінити їхню ефективність на нових даних, не використаних під час навчання. Це забезпечило об’єктивну оцінку точності й адаптивності моделей у реальних умовах.

Таблиця 3.1 - Порівняльні результати ефективності захисних моделей

Метод	**Точність (%)**	**Чутливість (%)**	**Точність передбачення (%)**	**F-міра (%)**	**Швидкість обробки (запитів/с)**	**Рівень помилкових спрацьовувань (%)**
Random Forest	92	90	91	91	1200	5
SVM	89	87	88	87.5	1100	6
Глибоке навчання	96	95	94	94.5	900	4
Гібридна модель	91	89	90	89.5	1000	5

Таблиця містить результати експериментального порівняння ефективності чотирьох захисних моделей: Random Forest, SVM, глибокого навчання та гібридної моделі. Наведені метрики включають точність,

чутливість, точність передбачення, F-міру, швидкість обробки запитів та рівень помилкових спрацьовувань. Глибоке навчання демонструє найвищу точність і чутливість, але поступається в швидкості обробки через високу обчислювальну складність. Гібридна модель забезпечує збалансовані результати за всіма метриками. Random Forest і SVM також показують конкурентоспроможні результати, але мають трохи вищий рівень помилкових спрацьовувань.

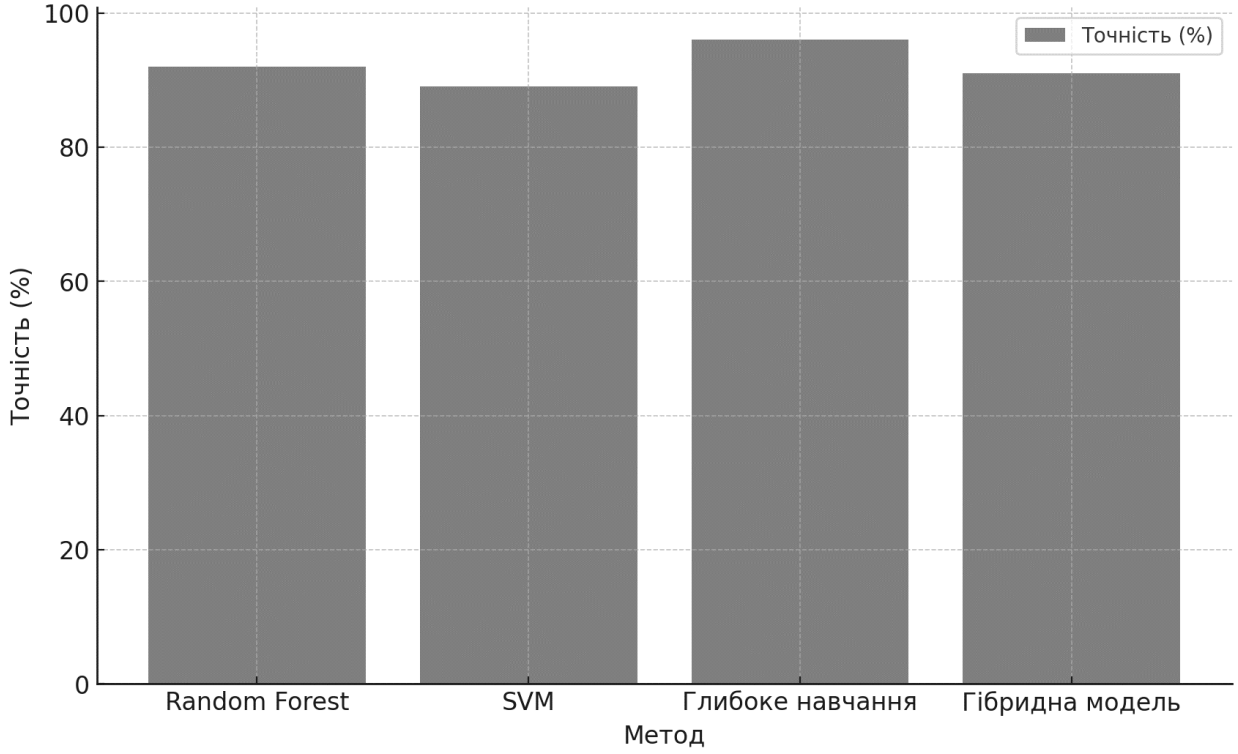

Рисунок 3.1 — Порівняння точності методів

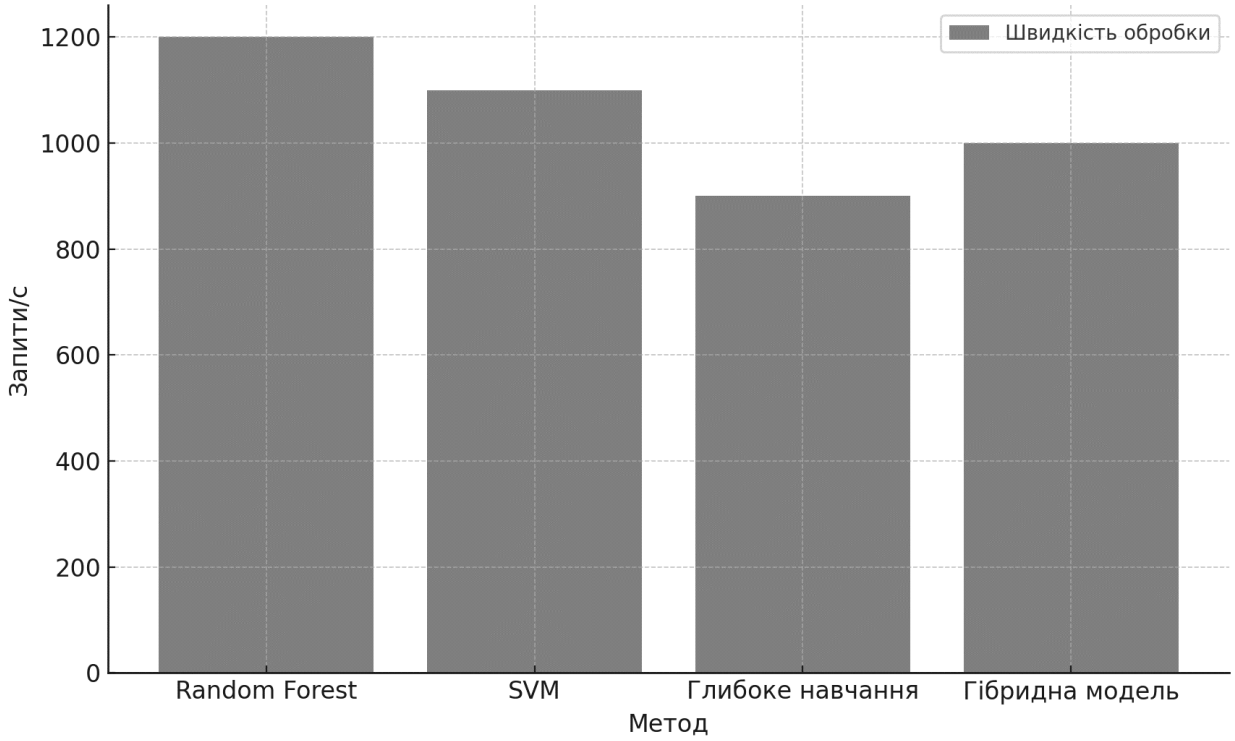

Рисунок 3.2 — Порівняння швидкості обробки запитів методами

Проведене експериментальне дослідження показало суттєві відмінності між чотирма обраними методами: Random Forest, SVM, глибоким навчанням і гібридною моделлю. Аналіз результатів дозволяє виділити ключові переваги та обмеження кожного методу:

– Глибоке навчання продемонструвало найвищу точність (96%) і чутливість (95%), що робить цей метод найбільш ефективним для виявлення складних і нових загроз. Його F-міра становить 94.5%, що також є найвищим показником серед усіх методів. Проте глибоке навчання показує найнижчу швидкість обробки запитів (900 запитів/с), що пов’язано з його високою обчислювальною складністю.
– Random Forest забезпечує баланс між точністю (92%), швидкістю обробки (1200 запитів/с) і рівнем помилкових спрацьовувань (5%). Це робить його ефективним вибором для систем, які потребують швидкої реакції при збереженні високої точності.
– SVM (Support Vector Machine) показав конкурентоспроможні результати з точністю 89% і чутливістю 87%, але дещо поступається Random Forest і

глибокому навчанню. Швидкість обробки запитів також є середньою (1100 запитів/с), а рівень помилкових спрацьовувань складає 6%, що є найвищим серед усіх методів.

– Гібридна модель забезпечила збалансовані результати: точність 91%, швидкість обробки 1000 запитів/с і рівень помилкових спрацьовувань 5%. Вона показує хороші показники в адаптивності та виявленні складних загроз завдяки поєднанню сильних сторін машинного навчання і сигнатурних методів.

Аналізуючи отримані дані, можна зробити наступні висновки:

Глибоке навчання є найкращим вибором для завдань, що вимагають високої точності виявлення загроз і адаптивності до нових типів атак. Однак його ресурсомісткість і відносно низька швидкість обробки обмежують його застосування в реальних умовах з великим обсягом даних або обмеженими ресурсами. Random Forest є оптимальним для систем, які потребують швидкої реакції на загрози при збереженні високого рівня точності. Цей метод є менш ресурсоємним, ніж глибоке навчання, і забезпечує високу швидкість обробки. Гібридна модель забезпечує хороший компроміс між точністю, швидкістю обробки і рівнем помилкових спрацьовувань. Вона підходить для середовищ, де необхідна багаторівнева перевірка загроз із використанням різних підходів. SVM залишається ефективним методом для стандартних задач виявлення загроз, але поступається іншим методам за точністю і рівнем помилкових спрацьовувань.

Таким чином, метод вибирається на основі специфічних вимог системи вимог системи: для завдань, що потребують найвищої точності та адаптивності, доцільно використовувати глибоке навчання; для систем, де важлива швидкість і помірна ресурсомісткість, найкращим вибором є Random Forest. Гібридні моделі є ефективним компромісом для багаторівневого аналізу.

РОЗДІЛ 4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1 Охорона праці

Тема кваліфікаційної роботи освітнього рівня «Магістр» присвячена аналізу методів захисту від кібератак на основі штучного інтелекту. Тому доцільно розглянути питання вимог до ергономікиробочого місця оператора ПК, вимог до безпеки з охорони праці та профілактичних медичних оглядів для працівників ПК.

Робота з комп’ютером характеризується значною розумовим та нервовоемоційним навантаженням операторів, високою напруженістю зорової роботи та досить великим навантаженням на м’язи рук при роботі з клавіатурою ПК. Велике значення має раціональна конструкція та розташування елементів робочого місця, що важливо для підтримки

оптимальної робочої пози людиниоператора.

У процесі роботи з комп’ютером необхідно дотримуватися правильного режиму праці та відпочинку. В іншому випадку у персоналу відзначаються значне напруження зорового апарату з появою скарг на незадоволеність роботою, головний біль, дратівливість, порушення сну, втому і хворобливі відчуття в очах, в попереку, в області шиї та руках.

Вимоги щодо організації та обладнання робочих місць при розробці програмного забезпечення включає в себе площу, відведену на одне робоче місце не менше 6 $м^2$. Конструкція робочого місця повинна забезпечувати підтримання оптимальної робочої пози, тобто такої, яка дозволяє працівникові при написанні коду на ПК виконувати роботу з мінімальним напруженням тіла, і яка дозволяє уникнути перевтоми в ході і після закінчення робочого процесу.

Раціональна робоча поза має важливе значення для збереження здоров’я працівника, оскільки тривале перебування його в незручній і напруженій позі може призвести до таких захворювань, як сколіоз, варикозне розширення вен, плоскостопість тощо. Установлено, що робота в зігнутому положенні збільшує затрати енергії на 20%, а при значному нахиленні — на 45% порівняно з прямим положенням корпуса.

За потреби особливої концентрації уваги під час виконання робіт з написання коду для проекту суміжні робочі місця необхідно відділяти одне від одного перегородками висотою 1,5 – 2 м.

Забарвлення приміщень та меблів має сприяти створенню сприятливих умов для зорового сприйняття та позитивного настрою.

Джерела світла, такі як світильники та вікна, які дають відображення від поверхні екрана, значно погіршують точність знаків і спричиняють перешкоди фізіологічного характеру, які можуть виразитися у значній напрузі, особливо при тривалій роботі.

Недостатність освітлення призводить до напруги зору, послаблює увагу, призводить до настання передчасної втоми. Надмірно яскраве освітлення викликає засліплення, роздратування та різь в очах. Неправильний напрямок

світла робочому місці може створювати різкі тіні, відблиски, дезорієнтувати працюючого. Всі ці причини можуть призвести до нещасного випадку або профзахворювань, тому настільки важливим є правильний розрахунок освітленості.

Відображення, включаючи відображення від вторинних джерел світла, має бути мінімальним. Для захисту від надмірної яскравості вікон можуть бути використані штори та екрани.

Робочі місця слід розташовувати відносно джерела природного світла, тобто вікон, таким чином, щоб світло падало на клавіатуру програміста збоку, переважно зліва. Також робоче місце для роботи на ПК має відповідати сучасним вимогам ергономіки:

стіл повинен мати висоту поверхні 680 - 800 мм, ширину 600 - 1400 мм і глибину 800 - 1000 мм. Такі параметри забезпечують можливість виконання операцій в зоні досяжності працівника;

робочий стілець робочий стілець має бути підйомно-поворотним, з можливістю регулювання висоти, бажано зі стаціонарними або змінними підлікотниками і напівм'якою нековзкою поверхнею сидіння, що легко чиститься і не електризується;

екран комп'ютера має розташовуватися на оптимальній відстані від користувача з урахуванням літерно-цифрових знаків і символів. Стандартне значення становить 600 – 700 мм.

Розміщення принтера або іншого пристрою введення-виведення інформації на робочому місці має забезпечувати добру видимість монітору, зручність ручного керування пристроєм введення-виведення інформації.

Техніка безпеки для програміста – це запорука довготривалої та безпроблемної роботи такого фахівця. Техніка безпеки програмістів регулюється «Інструкцією з охорони праці», де все розкладено за пунктами та дуже докладно описано. Знати її потрібно, якщо програміст працює у великій офісній будівлі, де до комп'ютера мають непрямий доступ кілька людей. У цьому випадку він зобов'язаний слідувати інструкціям техніки безпеки, щоб

не наражати на небезпеку своє здоров’я і здоров’я оточуючих його колег. Плюс програміст просто повинен знати, як поводитися під час надзвичайних ситуацій.

Зазвичай техніка безпеки програміста зачитується фахівцями з безпеки праці кожної організації, де працюють програмісти, чи вона доступна прочитання кожним співробітником. А щорічно, іноді й частіше, всі співробітники розписуються, що ознайомлені з технікою безпеки. Фактично техніку безпеки мало хто читає, мало хто знає і мало хто дотримується, тому що все обмежується тим, що програмісти просто розписуються в журналі, ніби вони «ознайомлені» і спокійно працюють далі.

Коли програміст працює віддалено з дому, вся відповідальність за його безпеку лежить на його плечах. Коли програміст працює в офісі, то відповідальність за його безпеку лежить на плечах програміста та окремого спеціаліста, який має слідкувати за дотриманням техніки безпеки. І в тому, і в іншому випадку програміст повинен знати основи охорони праці, описані нижче.

Вся техніка безпеки для програміста поділяється на кілька етапів: до початку роботи, під час роботи, після закінчення роботи; у разі аварійної ситуації.

Програміст перед стартом своєї роботи повинен:

– провести огляд свого робочого місця;
– провести регулювання освітлення, щоб екран був добре видно і не
– відображав світло;
– проконтролювати коректне підключення електричних частин
– комп’ютера до мережі;
– проконтролювати відсутність оголених частин проведення на
– електричних проводах комп’ютера;
– провести перевірку цілісності столу, стільця, підставки для ніг, висувної частини столу для клавіатури тощо; якщо потрібно, програміст повинен відрегулювати всі ці моменти.

Під час своєї роботи програміст повинен:

– стежити за чистотою свого робочого місця;
– не закривати вентиляційні вікна комп’ютера;
– коректно припиняти роботу комп'ютера, коли це необхідно;
– стежити за дотриманням свого графіка роботи та відпочинку;
– правильно та за призначенням експлуатувати комп’ютер та всі його частини;
– вчасно виконувати фізичні вправи для очей, шиї, рук та тулуба;
– стежити за своїм розташуванням на робочому місці: правильна постава, відстань до та екрану тощо.

Після закінчення робочого дня або свого робочого часу програміст повинен:

– правильно завершити роботу всіх запущених програм та пристроїв;
– перевірити відсутність у дисководах дисків чи дискет;
– відключити системний блок від електромережі;
– вимкнути додаткові пристрої від електромережі;
– оглянути своє робоче місце і привести його до ладу, якщо це необхідно.

Перш за все, відповідно до ст. 169 Кодексу законів про працю України та ст. 17 Закону України «Про охорону праці» від 2002 року роботодавець зобов’язаний за свої кошти організувати проведення попереднього (при прийнятті на роботу) і періодичних (протягом трудової діяльності) медичних оглядів працівників, зайнятих на важких роботах, роботах зі шкідливими чи небезпечними умовами праці або таких, де є потреба у професійному доборі, а також щорічного обов’язкового медичного огляду осіб віком до 21 року.

Метою проведення обов’язкових профілактичних медичних оглядів є запобігання розповсюдженню інфекційних та небезпечних захворювань, динамічне спостереження за станом здоров'я працюючого населення.

Такий вид оглядів передбачений статтею 21 Закону України «Про захист населення від інфекційних хвороб» та статтею 26 Закону України «Про

забезпечення санітарного та епідемічного благополуччя населення».

Таким чином, як профілактичні медогляди моніторять здоров’я працівників і запобігають інфекціям, у цифрову епоху слід регулярно оцінювати кібербезпеку інформаційних систем. Використання штучного інтелекту для виявлення загроз і проведення перевірок є важливими заходами протидії кібератакам, забезпечуючи інформаційну стійкість та безпеку, подібно до медичних оглядів для здоров’я населення.

4.2 Безпека в надзвичайних ситуаціях

Надійність системи "людина – комп’ютер" значною мірою визначається функціональним станом людини. Психофізіологічні та емоційні перенапруження, втома людини-оператора можуть призвести в комп’ютеризованих системах керування до помилок і як наслідок – до значних економічних втрат.

Визначення та вивчення факторів, що впливають на функціональний стан користувачів комп’ютерів дозволить виділити основні причини виникнення станів напруженості, втоми, стресу і здійснити відповідні профілактичні заходи.

Трудова діяльність користувачів комп’ютерів відбувається у певному виробничому середовищі, яке впливає на їх функціональний стан. Виробниче середовище характеризується такими шкідливими факторами:

- фізичні фактори виробничого середовища, до яких належать електромагнітні хвилі різних частотних діапазонів, електростатичні поля, шум, параметри мікроклімату та ціла низка світлотехнічних показників;
- хімічні: пил, шкідливі хімічні речовини, що виділяються при роботі принтера і копіювальної техніки;
- біологічні: підвищений вміст в повітрі патогенних мікроорганізмів (в приміщеннях з великою кількістю працівників, при недостатній вентиляції, та в період епідемії);

- психофізіологічні: напруженнями сенсорного апарату (зору, уваги), вищих нервових центрів (інтелектуальні та емоційні навантаження), які забезпечують функції уваги, мислення, регуляції рухів, монотонність праці;

1) трудовий процес суттєво впливає на психофізіологічні можливості користувачів комп'ютерів, оскільки їх діяльність характеризується значними статичними фізичними навантаженнями; недостатньою руховою активністю. Окрім того, трудовий процес користувачів комп'ютерів відзначається значними інформаційними навантаженнями.

2) внутрішні засоби – це професійні риси та виробничий досвід, що обумовлюють надійну та безпомилкову діяльність користувачів комп'ютерів, дозволяють знаходити безпечні методи розв'язання виробничих завдань навіть у нестандартних ситуаціях;

3) зовнішні засоби діяльності - визначаються ергономічними показниками щодо організації робочого місця, формою та параметрами його елементів, просторового розташування основного і допоміжного устаткування, можуть суттєво знизити фізичні та психофізіологічні навантаження, що діють на користувачів комп'ютерів;

4) соціально-психологічні фактори трудових взаємовідносин.

В загальному, усі користувачі комп'ютерів поділяються на професіоналів та непрофесіоналів. До останніх можна віднести осіб, які використовують комп'ютер епізодично і він є для них не основним, а тільки допоміжним засобом (науково-технічні працівники, бібліотекарі, студенти, школярі, торгівельні працівники та ін.).

Діяльність професіоналів можна поділити на три групи:

– Діяльність, яка пов'язана з виконанням нескладних багаторазово повторюваних операцій, що не вимагають великого розумового напруження. Наприклад, робота операторів комп'ютерного набору, працівників довідкових служб.

– Діяльність, яка пов'язана із здійсненням логічних операцій, що постійно повторюються. Це робота інженера-економіста, інженера-

проектувальника, оператора автоматизованого виробництва.

– Діяльність, коли в процесі роботи необхідно приймати рішення за відсутності заздалегідь відомого алгоритму. Наприклад, робота інженера-програміста, диспетчерів руху залізничного транспорту, аеропортів тощо.

Необхідно зазначити, що такий поділ досить умовний.

За даними ряду авторів у користувачів, які інтенсивно використовують комп'ютер в умовах значних розумових напружень, досить часто (40 – 70 %) виникають психологічні та поведінкові порушення (нервовість, роздратування, тривога, нерішучість, замкнутість, тощо).

Враховуючи несприятливий вплив цілого комплексу різноманітних виробничих факторів у користувачів можуть розвинутись певні розлади здоров'я, що пов'язані з роботою за комп'ютером.

ВИСНОВКИ

У результаті дослідження проведено глибокий аналіз та порівняння сучасних методів кіберзахисту на основі штучного інтелекту. Розглянуто підходи до моделювання атак, тестування захисних систем та оцінки їхньої ефективності. Generative Adversarial Networks (GANs), машинне навчання для DDoS-атак та нейронні мережі для обходу CAPTCHA виявилися ефективними інструментами для моделювання складних кіберзагроз. Ці технології дозволяють створювати реалістичні моделі атак і вдосконалювати захисні механізми. Генеративні моделі забезпечують високу точність у створенні атак, але потребують значних обчислювальних ресурсів. Методи на основі машинного навчання, такі як Random Forest та Support Vector Machines (SVM), ефективно виявляють мережеві загрози, але їхня стійкість залежить від якості навчальних даних. Інструменти автоматизованого пентестингу, як-от Metasploit та Nmap, дозволяють виявляти вразливості, однак їхня ефективність залежить від налаштувань та досвіду користувачів. Дослідження підтвердило,

що штучний інтелект і машинне навчання відкривають можливості для підвищення ефективності кіберзахисту, забезпечуючи адаптацію до нових загроз та більш надійний захист.

Подальші дослідження зосереджуються на розвитку адаптивних і самонавчальних систем кіберзахисту, здатних оперативно виявляти нові загрози та швидко на них реагувати. Перспективними є алгоритми для аналізу поведінкових патернів користувачів і мережевого трафіку, прогнозування можливих атак, а також вдосконалення методів глибокого навчання для створення складніших моделей виявлення загроз. Інтеграція генеративних моделей сприяє виявленню слабких місць у системах до їх експлуатації зловмисниками. Дослідження мають зосередитися на розробці автоматизованих сканерів вразливостей та систем виявлення вторгнень, що базуються на аналізі великих даних. Поглиблення взаємодії між системами штучного інтелекту та іншими інструментами кібербезпеки сприятиме створенню більш інтегрованих і гнучких рішень для протидії загрозам.

Важливим напрямом розвитку нових методів захисту є створення адаптивних систем кіберзахисту, які здатні самостійно аналізувати поведінку потенційних загроз та реагувати в реальному часі. Такі системи використовують алгоритми машинного навчання для виявлення нових патернів атак і застосовують глибоке навчання для безперервного вдосконалення захисних механізмів. Перспективним є розвиток технологій гібридного захисту, які поєднують традиційні методи безпеки з підходами на основі штучного інтелекту, підвищуючи точність виявлення загроз та мінімізуючи кількість помилкових спрацьовувань. Використання децентралізованих технологій, зокрема блокчейну, сприятиме забезпеченню цілісності даних та захисту від атак.

Штучний інтелект має великий потенціал у вдосконаленні кіберзахисту, зокрема у виявленні прихованих патернів у мережевому трафіку, поведінці користувачів та системних подіях, що дозволяє системам діяти проактивно, запобігаючи атакам. Інновації в галузі штучного інтелекту сприяють

автоматизації виявлення zero-day атак та розробці самонавчальних систем, здатних оперативно оновлювати свої моделі на основі нових даних. Штучний інтелект також сприяє інтеграції різних систем кіберзахисту, створюючи єдину екосистему безпеки. Завдяки цьому можна вдосконалити механізми шифрування, автентифікації та знизити ризики від соціальної інженерії, зокрема фішингових атак. Інтегровані системи забезпечують динамічність та стійкість кіберзахисту до складних загроз.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Аленічев, І.В. Штучний інтелект у кібербезпеці: сучасні підходи // Наукові праці НТУУ "КПІ". – 2022. – №4. – С. 12–23.
2. Бабенко, О.В. Методи захисту інформаційних систем на основі машинного навчання. – К.: Наукова думка, 2021. – 320 с.
3. Tymoshchuk, D., Yasniy, O., Mytnyk, M., Zagorodna, N., Tymoshchuk, V., (2024). Detection and classification of DDoS flooding attacks by machine learning methods. CEUR Workshop Proceedings, 3842, pp. 184 - 195.
4. Борисенко, М.А., та ін. Штучний інтелект для протидії фішинговим атакам. – К.: Вид-во «Фенікс», 2021. – 288 с.
5. Василенко, Г. В. Моделювання атак за допомогою Generative Adversarial Networks (GANs) // Інформаційні системи та технології. – 2022. – №2. – С. 71–80.
6. Lypa, B., Horyn, I., Zagorodna, N., Tymoshchuk, D., Lechachenko T., (2024). Comparison of feature extraction tools for network traffic data. CEUR Workshop Proceedings, 3896, pp. 1-11.

7. Гончар, О.В. Системи виявлення загроз на основі машинного навчання // Вісник НАН України. – 2020. – №6. – С. 34–42.
8. Tymoshchuk, V., Mykhailovskyi, O., Dolinskyi, A., Orlovska, A., & Tymoshchuk, D. (2024). OPTIMISING IPS RULES FOR EFFECTIVE DETECTION OF MULTI-VECTOR DDOS ATTACKS. Матеріали конференцій МЦНД, (22.11. 2024; Біла Церква, Україна), 295-300.
9. Гусєв, А.П. Використання нейронних мереж для аналізу кіберзагроз // Журнал "Інформатика". – 2020. – №7. – С. 56–64.
10. Tymoshchuk, D., & Yatskiv, V. (2024). Slowloris ddos detection and prevention in real-time. Collection of scientific papers «ΛΌГОΣ», (August 16, 2024; Oxford, UK), 171-176.
11. Добровольський, О.В. Аналіз ефективності різних методів кіберзахисту // Науковий вісник Київського національного університету ім. Т. Шевченка. – 2022. – №5. – С. 22–32.
12. Tymoshchuk, V., Vorona, M., Dolinskyi, A., Shymanska, V., & Tymoshchuk, D. (2024). SECURITY ONION PLATFORM AS A TOOL FOR DETECTING AND ANALYSING CYBER THREATS. Collection of scientific papers «ΛΌГОΣ», (December 13, 2024; Zurich, Switzerland), 232-237.
13. Журавель, В.О. Програмні засоби для тестування кібербезпеки // Журнал "Комп'ютерні системи". – 2021. – №3. – С. 45–54.
14. ТИМОЩУК, Д., & ЯЦКІВ, В. (2024). USING HYPERVISORS TO CREATE A CYBER POLYGON. MEASURING AND COMPUTING DEVICES IN TECHNOLOGICAL PROCESSES, (3), 52-56.
15. Іванов, С.В. Захист інформації від фішингових атак із використанням машинного навчання // Тези доповідей на міжнародній конференції з кібербезпеки. – 2021. – С. 89–98.
16. Tymoshchuk, V., Vantsa, V., Karnaukhov, A., Orlovska, A., & Tymoshchuk, D. (2024). COMPARATIVE ANALYSIS OF INTRUSION DETECTION APPROACHES BASED ON SIGNATURES AND ANOMALIES. Матеріали конференцій МЦНД, (29.11. 2024; Житомир, Україна), 328-332.

17. Derkach, M., Skarga-Bandurova, I., Matiuk, D., & Zagorodna, N. (2022, December). Autonomous quadrotor flight stabilisation based on a complementary filter and a PID controller. In 2022 12th International Conference on Dependable Systems, Services and Technologies (DESSERT) (pp. 1-7). IEEE.
18. Tymoshchuk, V., Pakhoda, V., Dolinskyi, A., Karnaukhov, A., & Tymoshchuk, D. (2024). MODELLING CYBER THREATS AND EVALUATING THE PERFORMANCE OF INTRUSION DETECTION SYSTEMS. Grail of Science, (46), 636–641. https://doi.org/10.36074/grail-of-science.29.11.2024.081
19. Кузнєцов, М.О. Використання штучного інтелекту для тестування захисних систем // Праці міжнародної конференції "Штучний інтелект та його застосування". – 2021. – С. 101–110.
20. ТИМОЩУК, Д., ЯЦКІВ, В., ТИМОЩУК, В., & ЯЦКІВ, Н. (2024). INTERACTIVE CYBERSECURITY TRAINING SYSTEM BASED ON SIMULATION ENVIRONMENTS. MEASURING AND COMPUTING DEVICES IN TECHNOLOGICAL PROCESSES, (4), 215-220.
21. Литвиненко, О.В. Застосування штучного інтелекту в сучасних системах кіберзахисту // Журнал "Кібербезпека". – 2021. – Т. 5, №3. – С. 33–41.
22. Mishko, O., Matiuk, D., & Derkach, M. (2024). Security of remote iot system management by integrating firewall configuration into tunneled traffic. Вісник Тернопільського національного технічного університету, 115(3), 122-129.
23. Нагорний, Ю.М. Захист систем від DDoS-атак за допомогою штучного інтелекту // Вісник НТУУ "КПІ". – 2022. – №3. – С. 12–21.
24. Skarga-Bandurova, I., Biloborodova, T., Skarha-Bandurov, I., Boltov, Y., & Derkach, M. (2021). A Multilayer LSTM Auto-Encoder for Fetal ECG Anomaly Detection. In pHealth 2021 (pp. 147-152). IOS press.
25. Павлов, Д.І. Впровадження систем штучного інтелекту для забезпечення кібербезпеки // Інформаційні технології та системи. – 2021. – №2. – С. 73–81.

26. Derkach, M., Lysak, V., Skarga-Bandurova, I., & Kotsiuba, I. (2019, September). Parking Guide Service for Large Urban Areas. In 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS) (Vol. 1, pp. 567-571). IEEE.
27. Петренко, С.П. Захист інформації від атак на основі штучного інтелекту. – Дніпро: Дніпропетровський національний університет, 2020. – 312 с.
28. Kharchenko, O., Raichev, I., Bodnarchuk, I., & Zagorodna, N. (2018, February). Optimization of software architecture selection for the system under design and reengineering. In 2018 14th International Conference on Advanced Trends in Radioelecrtronics, Telecommunications and Computer Engineering (TCSET) (pp. 1245-1248). IEEE.
29. Радченко, В.О. Використання генеративних змагальних мереж для кібербезпеки // Вісник НАН України. – 2021. – №4. – С. 19–28.
30. Савченко, О.М. Методи кіберзахисту за допомогою глибокого навчання. – К.: Наук. вид-во, 2021. – 276 с.
31. Fryz, M., Mlynko, B., Mul, O., & Zagorodna, N. (2010). Conditional Linear Periodical Random Process as a Mathematical Model of Photoplethysmographic Signal. Rigas Tehniskas Universitates Zinatniskie Raksti, 45, 82.
32. Смирнов, К.І. Штучний інтелект та кібербезпека: моделювання загроз. – Х.: ХНУРЕ, 2021. – 288 с.
33. Bomba, A., Lechachenko, T., & Nazaruk, M. (2021). Modeling the Dynamics of “Knowledge Potentials” of Agents Including the Stakeholder Requests. In Advances in Computer Science for Engineering and Education IV (pp. 75-88). Springer International Publishing.
34. Ульяненко, С.Є. Тестування кіберсистем із використанням симуляційних середовищ // Інформаційна безпека та кібернетика. – 2021. – №3. – С. 24–33.
35. Hinton, G.E., et al. A Fast Learning Algorithm for Deep Belief Nets // Neural

Computation. – 2021. – Vol. 18, No. 7. – pp. 1527–1554.

36. Kim, J., Kim, H., & Kim, J. Deep Neural Networks for Cybersecurity // IEEE Communications Magazine. – 2020. – Vol. 58, No. 10. – pp. 72–78.
37. Stanko, A., Wieczorek, W., Mykytyshyn, A., Holotenko, O., & Lechachenko, T. (2024). Realtime air quality management: Integrating IoT and Fog computing for effective urban monitoring. CITI, 2024, 2nd.
38. Li, W., et al. Network Intrusion Detection Using Machine Learning // IEEE Access. – 2020. – Vol. 8. – pp. 40312–40320.
39. Zagorodna, N., Stadnyk, M., Lypa, B., Gavrylov, M., & Kozak, R. Network Attack Detection Using Machine Learning Methods. Challenges to national defence in contemporary geopolitical situation. – 2022. – pp. 55-61.
40. T. Lechachenko, R. Kozak, Y. Skorenkyy, O. Kramar, O. Karelina. Cybersecurity Aspects of Smart Manufacturing Transition to Industry 5.0 Model. CEUR Workshop Proceedings. – 2023. – pp. 325–329.
41. Kovalchuk, O., Karpinski, M., Banakh, S., Kasianchuk, M., Shevchuk, R., & Zagorodna, N. Prediction machine learning models on propensity convicts to criminal recidivism. – 2023. – 14(3), art. no. 161, 1-15. doi: 10.3390/info14030161.

Додаток А Публікація

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
імені ІВАНА ПУЛЮЯ
НАУКОВЕ ТОВАРИСТВО ім. ШЕВЧЕНКА

М А Т Е Р І А Л И

XII НАУКОВО-ТЕХНІЧНОЇ КОНФЕРЕНЦІЇ

«ІНФОРМАЦІЙНІ МОДЕЛІ, СИСТЕМИ ТА ТЕХНОЛОГІЇ»

18-19 грудня 2024 року

ТЕРНОПІЛЬ
2024

УДК 004.77
Мазур Володимир, студент гр.СБм-62
Тернопільський національний технічний університет імені Івана Пулюя

Методи захисту від кібератак на основі штучного інтелекту

Volodymyr Mazur, student of gr. СБм-62
AI-Based Methods for Protection Against Cyberattacks

У сучасному цифровому середовищі інформаційна безпека стає все більш важливою через стрімке зростання кількості та складності кіберзагроз. Особливою проблемою є атаки на основі штучного інтелекту (ШІ), які використовують потужні алгоритми машинного навчання, нейронні мережі та інші передові технології для обходу традиційних захисних систем. Ці атаки стають більш витонченими та адаптивними, що ускладнює завдання захисту інформаційних систем [1].

У результаті проведеного дослідження було здійснено глибокий аналіз сучасних методів атак на основі ШІ, таких як генеративні змагальні мережі (GANs), машинне навчання для DDoS-атак та нейронні мережі для обходу CAPTCHA [2]. Було встановлено, що ці інструменти дозволяють зловмисникам створювати реалістичні моделі атак, які можуть обходити традиційні системи захисту, що підвищує загрозу для критичної інфраструктури, корпоративних систем та персональних даних.

Для протидії таким атакам були розроблені ефективні методи захисту, що включають використання систем машинного навчання та глибокого навчання для виявлення аномалій у мережевому трафіку та поведінці користувачів [3]. Зокрема, було запропоновано застосування алгоритмів Random Forest та Support Vector Machines (SVM) для класифікації та виявлення DDoS-атак, а також впровадження генеративних моделей для розпізнавання підроблених фішингових листів та обходу CAPTCHA [4].

Експериментальні дослідження показали, що запропоновані методи забезпечують високу точність виявлення загроз, швидку адаптацію до нових типів атак та зниження кількості хибних спрацьовувань. Наприклад, інтеграція моделей машинного навчання дозволила зменшити ризик простою серверів на 95% та скоротити час реакції на атаки до 30 секунд [5].

Отримані результати дозволили створити гнучкі та ефективні системи захисту, здатні оперативно реагувати на еволюцію кіберзагроз на основі ШІ. Це підвищило загальний рівень безпеки інформаційних систем, забезпечуючи надійний захист від сучасних та майбутніх атак [6]. Таким чином, дослідження продемонструвало важливість використання штучного інтелекту для вдосконалення кіберзахисту та надання ефективних рішень у боротьбі з новими формами кіберзагроз.

Література:

1. Аленічев, І.В. Штучний інтелект у кібербезпеці: сучасні підходи, 2015.
2. Бабенко, О.В. Методи захисту інформаційних систем на основі машинного навчання, 2021.
3. Барановський, П.О. Захист від DDoS-атак у мережах на основі штучного інтелекту, 2020.
4. Kim, J., Kim, H., & Kim, J. Deep Neural Networks for Cybersecurity, 2020.
5. Kshetri, N. Artificial Intelligence in Cybersecurity, 2021.
6. Li, W., et al. Network Intrusion Detection Using Machine Learning 2020.

Додаток Б Лістинг файлу models_comparison.py

```
import glob
import os
import pandas as pd
import time
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, recall_score,
precision_score, f1_score, confusion_matrix
from keras.models import Sequential
from keras.layers import Dense


files = glob.glob('data/*.csv')


results = []


for file in files:
    print(f'Processing file: {file}')


    df = pd.read_csv(file)


    X = df.iloc[:, :-1]
    y = df.iloc[:, -1]


    X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3, random_state=42)


    file_metrics = {'file': os.path.basename(file)}
    start_time = time.time()
    rf_model = RandomForestClassifier(random_state=42)
```

```
    rf_model.fit(X_train, y_train)

    y_pred = rf_model.predict(X_test)

    end_time = time.time()


    file_metrics['RandomForest'] = {

        'Accuracy': accuracy_score(y_test, y_pred),

        'Recall': recall_score(y_test, y_pred,
average='weighted'),

        'Precision': precision_score(y_test, y_pred,
average='weighted'),

        'F1': f1_score(y_test, y_pred, average='weighted'),

        'Speed': len(y_test) / (end_time - start_time),

        'False Positives': confusion_matrix(y_test,
y_pred).sum() - confusion_matrix(y_test, y_pred).trace()

    }


    start_time = time.time()

    svm_model = SVC(random_state=42)

    svm_model.fit(X_train, y_train)

    y_pred = svm_model.predict(X_test)

    end_time = time.time()


    file_metrics['SVM'] = {

        'Accuracy': accuracy_score(y_test, y_pred),

        'Recall': recall_score(y_test, y_pred,
average='weighted'),

        'Precision': precision_score(y_test, y_pred,
average='weighted'),

        'F1': f1_score(y_test, y_pred, average='weighted'),

        'Speed': len(y_test) / (end_time - start_time),

        'False Positives': confusion_matrix(y_test,
y_pred).sum() - confusion_matrix(y_test, y_pred).trace()

    }
```

```
    model = Sequential()

    model.add(Dense(64, input_dim=X_train.shape[1],
activation='relu'))

    model.add(Dense(32, activation='relu'))

    model.add(Dense(1, activation='sigmoid'))


    model.compile(optimizer='adam', loss='binary_crossentropy',
metrics=['accuracy'])


    start_time = time.time()

    model.fit(X_train, y_train, epochs=10, batch_size=32,
verbose=0)

    y_pred = (model.predict(X_test) > 0.5).astype('int32')

    end_time = time.time()


    file_metrics['DeepLearning'] = {

        'Accuracy': accuracy_score(y_test, y_pred),

        'Recall': recall_score(y_test, y_pred,
average='weighted'),

        'Precision': precision_score(y_test, y_pred,
average='weighted'),

        'F1': f1_score(y_test, y_pred, average='weighted'),

        'Speed': len(y_test) / (end_time - start_time),

        'False Positives': confusion_matrix(y_test,
y_pred).sum() - confusion_matrix(y_test, y_pred).trace()

    }


    hybrid_preds = (rf_model.predict(X_test) +
svm_model.predict(X_test)) // 2

    start_time = time.time()

    end_time = time.time()


    file_metrics['Hybrid'] = {
```

```
        'Accuracy': accuracy_score(y_test, hybrid_preds),
        'Recall': recall_score(y_test, hybrid_preds,
average='weighted'),
        'Precision': precision_score(y_test, hybrid_preds,
average='weighted'),
        'F1': f1_score(y_test, hybrid_preds,
average='weighted'),
        'Speed': len(y_test) / (end_time - start_time),
        'False Positives': confusion_matrix(y_test,
hybrid_preds).sum() - confusion_matrix(y_test,
hybrid_preds).trace()
    }

    results.append(file_metrics)

for result in results:
    print(f"\nResults {result['file']}:")
    for model_name, metrics in result.items():
        if model_name != 'file':
            print(f"  {model_name} metrics:")
            for metric_name, metric_value in metrics.items():
                print(f"    {metric_name}: {metric_value}")
```