Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)
Кафедра кібербезпеки
(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

магістр
(назва освітнього ступеня)

на тему: Використання штучного інтелекту в IPS та IDS системах

Виконав: студент VI курсу, групи СБм-61
спеціальності 125 Кібербезпека
(шифр і назва спеціальності)

(підпис) Гуменюк В.Р. (прізвище та ініціали)

Керівник (підпис) Муж.В.В. (прізвище та ініціали)

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

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

Рецензент (підпис) (прізвище та ініціали)

Тернопіль
2023

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра кібербезпеки
(повна назва кафедри)

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ Загородна Н.В.
(підпис) (прізвище та ініціали)
«__» _____ 2023 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

на здобуття освітнього ступеня Магістр
(назва освітнього ступеня)

за спеціальністю 125 Кібербезпека
(шифр і назва спеціальності)

Студенту Гуменюку Вадиму Руслановичу
(прізвище, ім'я, по батькові)

1. Тема роботи Використання штучного інтелекту в IPS та IDS системах

Керівник роботи Муж Валерій Вікторович., к.ю.н., доц.
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від «_» _ 2023 року № _

2. Термін подання студентом завершеної роботи 26.12.2023р.

3. Вихідні дані до роботи наукові літературні джерела

4
систем: сучасний стан і розвиток, 1.1 Системи виявлення вторгнень (IPS/IDS), 1.2 Ключові Компоненти Системи IPS/IDS, 1.3 Сучасні Виклики для IPS/IDS, 1.4 Висновок до першого розділу, 2 Теоретичні основи ШІ в IPS/IDS, 2.1 Машинне навчання, 2.1.1 Контрольоване машинне навчання, 2.1.2 Машинне навчання без нагляду, 2.1.3. Машинне навчання з підкріпленням, 2.1.4 Процес машинного навчання, 2.2 Технології глибокого навчання, 2.3 Оцінка ефективності роботи IDS та ML, 2.4 Аналіз наборів даних, 2.5 Аналіз досліджень з МН в IDS, 2.6 Висновок до другого розділу, 3 Тестування архітектур гібридної нейронної мережі для аналізу мережевого трафіку, 3.1 Вибір архітектури моделі та планування експерименту, 3.2 Вихідні дані до тестування, 3.3 Основні етапи проектування моделі, 3.3.1 Попередня обробка даних, 3.3.2 Підготовка нейронної мережі, 3.3.2 Тренування нейронної мережі, 3.4 Результати тестування моделей, 3.5 Висновки до третього розділу, 4. Охорона праці та безпека в надзвичайних ситуаціях, 4.1,Охорона праці, 4.2,Безпека в надзвичайних ситуаціях, 4.3,Висновок до четвертого розділу, Висновок, Перелік використаних джерел. 5.Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

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 1 Листопада 2023 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Ознайомлення з завданням до кваліфікаційної роботи	01.11 – 02.11	*Виконано*
2.	Підбір джерел про принципи роботи штучного інтелекту		
	в IPS/IDS системах	02.11 – 05.11	*Виконано*
3.	Опрацювання джерел про принципи роботи штучного		
	інтелекту в IPS/IDS системах	05.11 – 10.11	*Виконано*
4.	Виконання дослідження принципипів роботи штучного		
	інтелекту в IPS/IDS системах	11.11 – 18.11	*Виконано*
5.	Підготовка матеріалу	19.11 – 24.11	*Виконано*
6.	Оформлення розділу «Аналіз предметної області»	25.11 – 29.11	*Виконано*
7.	Оформлення розділу «Теоретична частина»	30.11 – 02.12	*Виконано*
8.	Оформлення розділу «Практична частина»	03.12 – 8.12	*Виконано*
9.	Виконання завдання до підрозділу «Охорона праці»	9.12 – 10.12	*Виконано*
10.	Виконання завдання до підрозділу «Безпека в надзви-	10.12 – 11.12	*Виконано*
	чай них ситуаціях»		
11.	Оформлення кваліфікаційної роботи	14.12 – 15.12	*Виконано*
12.	Нормоконтроль	20.12 – 21.12	*Виконано*
13.	Перевірка на плагіат	12.12	*Виконано*
14.	Попередній захист кваліфікаційної роботи		
15.	Захист кваліфікаційної роботи	26.12	

Студент _____ Гуменюк В.Р.
(підпис) (прізвище та ініціали)

Керівник роботи _____ Муж В,В
(підпис) (прізвище та ініціали)

АНОТАЦІЯ

Застосування штучного інтелекту в IPS та IDS системах // Кваліфікаційна робота на отримання освітнього рівня «Магістр» // Гуменюк Вадим Русланович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем і програмної інженерії, кафедра кібербезпеки, група СБм-61 // Тернопіль, С
,

Ключові слова: ШТУЧНИЙ ІНТЕЛЕКТ, МАШИННЕ НАВЧАННЯ, р
и
с Кваліфікаційна робота складається з пояснювальної записки та графічної частини (ілюстративний матеріал – слайди). Об'єм графічної частини кваліфікаційної роботи становить 17 слайдів. Об'єм нояснювальної записки складає 58 друкованих сторінок формату А4
,

Кваліфікаційна робота складається з чотирьох розділів в яких нараховується 15 рисунків та 10 таблиць. Використано 18 літературних джерел.

т У даній роботі детально розглядається застосування штучного інтелекту в системах запобігання та виявлення вторгнень IPS/IDS. Розглядається поточний стан і розвиток цих систем, висвітлюються їхні ключові компоненти та проблеми. Основна увага приділяється інтеграції ШІ, зокрема методам машинного навчання та глибокого навчання, для підвищення ефективності та результативності IPS/IDS.

– Були розглянуті технічні аспекти використання моделі машинного навчання BiLSTM. Практичні застосування ШІ в тестових сценаріях

б
і
б

ABSTRACT

ЗМІСТ

ВСТУП

Інтрузійні системи виявлення та запобігання (IPS/IDS) вже давно є необхідними складовими інфраструктури кібербезпеки. Їхнє завдання полягає в виявленні та реагуванні на потенційно аномальну активність у комп'ютерних мережах. Протягом довгого часу ці системи опиралася на встановлені правила та підписи, які класикували конкретні мережеві події як кібератаки. Однак з поступовим зростанням складності та різноманітності кіберзагроз, стало зрозуміло, що для ефективного захисту потрібен більш автоматизований та адаптивний підхід до виявлення та запобігання цим загрозам.

В цьому контексті одним із ключових способів покращення ефективності систем IPS/IDS є використання штучного інтелекту (ШІ). Штучний інтелект розкриває перед нами безмежні можливості для аналізу, виявлення та реагування на кібератаки, надаючи системам здатність самостійно навчатися та адаптуватися до нових загроз. У даній роботі буде розглянуто основні концепції та методи штучного інтелекту, які можуть бути успішно застосовані в контексті IPS/IDS систем.

Увага буде приділена аналізу методів машинного навчання та глибокого навчання, а також алгоритмів навчання з підкріпленням. Ці підходи дозволяють системам IPS/IDS автоматично виявляти складні та змінні кібератаки. Крім теоретичного аналізу, робота також охоплює практичну сторону застосування ШІ в IPS/IDS системах, включаючи експерименти на основі реальних даних.

Аналізуються переваги та виклики, що виникають при виборі інтеграції штучного інтелекту в традиційні системи безпеки. Обговорюються потенційні переваги, такі як підвищена точність виявлення та здатність адаптуватися до нових загроз, а також обмеження, пов'язані зі складністю налаштування та вимогами до обчислювальних ресурсів.

Сучасні кіберзагрози стали більш складнішими і небезпечним, і традиційні системи IPS/IDS, які розраховані на виявлення конкретних шаблонів атак, стають менш ефективними. Завдяки штучному інтелекту, системи можуть вивчати та розуміти відмінності між нормальною та підозрілою активністю в реальному часі, а також приймати швидкі та обґрунтовані рішення з метою запобігання загрозам.

Аналіз методів ШІ, зокрема машинного навчання та глибокого навчання, розкриває перед нами потужність алгоритмів, які можуть навчатися на великих обсягах даних та виявляти навіть найскладніші атаки, які раніше залишалися непоміченими. Глибоке навчання, включаючи нейронні мережі, стає ключовим інструментом у боротьбі зі складними загрозами.

Важливим аспектом є також використання алгоритмів навчання з підкріпленням, які дозволяють системі IPS/IDS навчатися на власних помилках і покращувати свою реакцію до майбутніх кібератак.

Практична оцінка та експерименти в цій галузі допомагають визначити ефективність підходів на основі ШІ в реальних умовах. Експерименти на основі реальних даних дозволяють оцінити, наскільки швидко та точно системи на основі ШІ можуть реагувати на нові та невідомі загрози.

Основною метою даної роботи є дослідження можливостей методів машинного навчання МН для вдосконалення ефективності у виявленні інцидентів у IPS/IDS системах.

Для досягнення зазначеної цілі, у нашій роботі потрібно виконати наступні завдання:

- Здійснити аналіз поточних труднощів та проблем в системах IPS/IDS і визначити оптимальні способи їх вирішення за допомогою ШІ.

- Провести оцінку ефективності методів машинного навчання у використанні для IPS/IDS систем шляхом визначення їх точності та виявлення можливих недоліків.
- Порівняти існуючі готові моделі машинного навчання і вибрати оптимальну, враховуючи їхню здатність до виявлення мережевих інтрузій у конкретному контексті.
- Проаналізувати набори даних, які будуть використовуватися для навчання та тестування обраної моделі, з врахуванням їхньої репрезентативності та обсягу.
- Провести тестування обраної моделі на ефективність та вдосконалити її роботу шляхом виявлення і виправлення можливих недоліків, а також знаходження оптимальних налаштувань.

Об'єкт дослідження: Системи виявлення інтрузій (IDS), включаючи їхню структуру та функціонал для забезпечення безпеки інформації..

Предмет дослідження: Застосування штучного інтелекту і методів машинного навчання для покращення ефективності роботи систем виявлення інтрузій (IDS/IPS).

Наукова новизна:

Наукова новизна полягає в детальному аналізі та оптимізації використання методів машинного навчання для підвищення ефективності IPS/IDS систем. Робота включає аналіз різних моделей машинного навчання, їхню оцінку та вибір оптимальної моделі для конкретних задач IPS/IDS. Важливим є також детальний аналіз наборів даних для навчання та тестування моделей, що дозволяє глибше зрозуміти та вдосконалити методи виявлення кіберзагроз.

ОГЛЯД ТА АНАЛІЗ IPS/IDS СИСТЕМ: СУЧАСНИЙ СТАН І РОЗВИТОК

Концепція IDS з'явилася у 1980 році в статті Джеймса Андерсона "Моніторинг та спостереження за загрозами комп'ютерній безпеці" [1]. Джеймс був членом робочої групи з комп'ютерної безпеки Наукової ради з питань оборони при Військово-повітряних силах США. У своїй статті Джеймс пояснив, як аналіз журналів аудиту може бути корисним для виявлення несанкціонованих або зловмисних дій. Наприклад, аналіз журналу файлів надає важливу інформацію, щоб визначити, чи має місце несанкціоноване використання. Ця робота лягла в основу того, що згодом стане відомим як система виявлення вторгнень на хост-комп'ютери (Host Intrusion Detection System, HIDS). Невдовзі було створено першу IDS для виявлення загроз, які були у списку відомих атак.

Наприкінці 1980-х років багато системних адміністраторів почали використовувати системи виявлення вторгнень. Однак вони були вразливі і не могли виявляти атаки нульового дня, а також

У 1990-х роках для протистояння зростаючій кількості атак було досліджено новий метод виявлення. Це був метод виявлення аномалій, який шукав незвичну поведінку/активність в системі, щоб здійняти тривогу. Тим не менш, непослідовний характер мереж у період між 1990-ми та 2000-ми роками спричинив багато хибних тривог. Таким чином, багато адміністраторів перестали використовувати IDS через їх ненадійність [2].

На щастя, завдяки, з одного боку, вдосконаленню мережі та обчислювальної потужності її компонентів, а з іншого боку, появі машинного навчання, IDS зараз є ключовим компонентом безпеки багатьох систем. Більшість методів машинного навчання були оцінені при розробці IDS. Однак використання алгоритмів глибокого навчання не було достатньо досліджено. За допомогою глибокого навчання можна досягти набагато

більше можливостей для вирішення проблеми помилкових тривог і недостатньо високої точності.

1.1 Системи виявлення вторгнень (IPS/IDS)

IDS - це інструмент, апаратний пристрій або програмне забезпечення, який здіймає тривогу, коли виявляє зловмисну активність у мережі (рис.1.1). IDS - це "сторожове око" мережі. Це важливий компонент архітектури безпеки сучасних мереж. Вони дозволяють виявляти атаки на ранніх стадіях і, таким чином, дають можливість протистояти їм. Крім того, вони дозволяють виявляти різноманітні атаки, наприклад DoS, MitM тощо. IDS можуть відстежувати та реєструвати будь-який мережевий трафік, а також весь мережевий трафік, як зазначено.

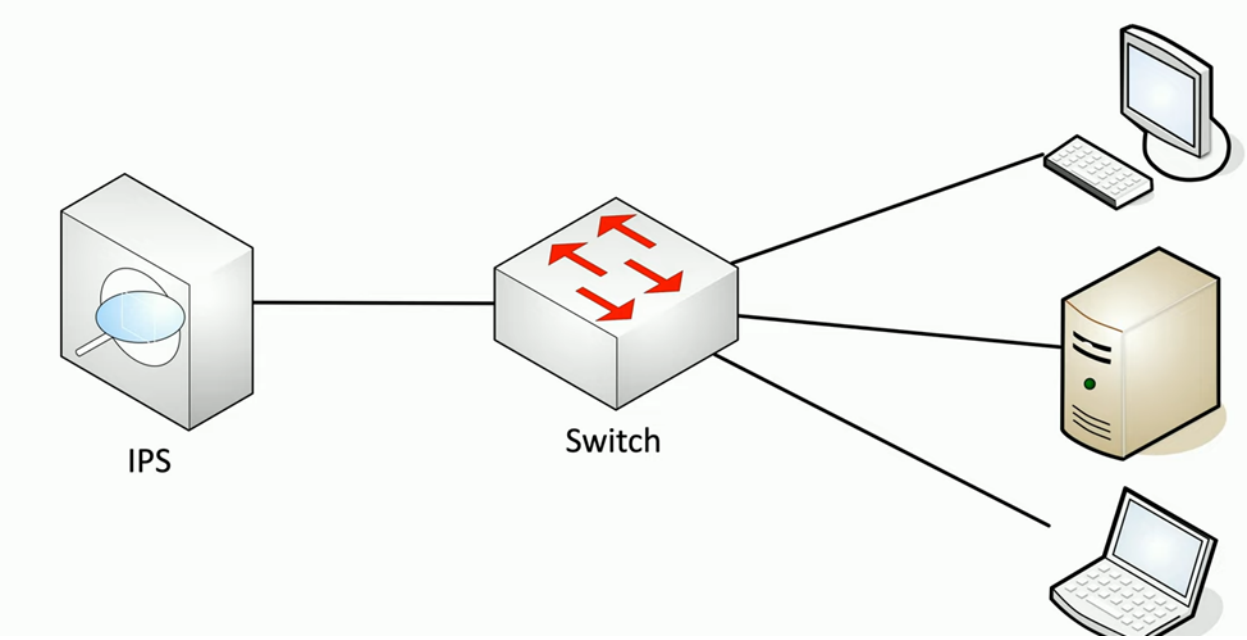

Рисунок 1.1 – Схема з’єднань системи виявлення вторгнень (Intrusion Detection Systems, IDS).

Оскільки IDS можуть надавати детальну інформацію про атаки, коли вони відбуваються, це допомагає адміністраторам безпеки зрозуміти, що

сталося. Таким чином, у випадку майбутніх подібних атак можна налаштувати безпеку системи так, щоб виявляти і запобігати атакам такого типу. Існує два типи IDS; NIDS та HIDS [3]:

- Системи виявлення мережевих вторгнень - це IDS, розміщені в мережі в стратегічних точках. NIDS аналізує загальний трафік мережі для виявлення зловмисних дій в мережі. NIDS допомагає виявляти атаки з боку ваших власних хостів і є ключовим компонентом безпеки більшості мереж організацій.
- Системи виявлення вторгнень на хост - це IDS, які встановлюються на всіх клієнтських комп'ютерах (хостах) мережі. На відміну від NIDS, HIDS аналізує трафік одного хоста, а також його активність, і якщо виявляє аномальну поведінку, то піднімає тривогу.

Існує три різних типи виявлення для IDS: Виявлення зловживань, виявлення аномалій та гібридне виявлення [3]:

- Виявлення зловживань, також відоме як виявлення сигнатур, шукає відомі шаблони вторгнення в мережі або на хості. Кожна атака має специфічну сигнатуру, наприклад, це може бути корисне навантаження пакета, IP-адреса джерела або специфічний заголовок. IDS може здійняти тривогу, якщо виявить атаку, яка має одну з сигнатур, перелічених у списку відомих сигнатур IDS. Перевагою такого підходу є висока точність виявлення відомих атак. Однак його слабкість полягає в тому, що він неефективний проти невідомих або "нульового дня" (ніколи раніше не зустрічалися атаки) шаблонів.
- Виявлення аномалій визначає нормальний стан мережі або хоста, який називається базовим, і будь-яке відхилення від нього повідомляється як потенційна атака. Наприклад, IDS на основі аномалій може створити базовий запис на основі загального мережевого трафіку, такого як послуги, що надаються кожним хостом, послуги, що використовуються кожним хостом, і обсяг активності протягом дня. Таким

чином, якщо зловмисник отримує доступ до внутрішнього ресурсу опівночі, і якщо в базовому записі опівночі не повинно бути майже ніякої активності, то IDS підніме тривогу. Перевагою виявлення аномалій є її гнучкість у виявленні невідомих атак вторгнення.

Однак у більшості випадків важко точно визначити, що є базовим записом мережі, тому рівень помилкового виявлення цих методів може бути високим.

- Гібридне виявлення поєднує в собі обидва вищезгадані методи. Як правило, вони мають нижчий рівень помилкових виявлень, ніж методи виявлення аномалій, і можуть виявляти нові атаки. На сьогоднішній день самих лише IDS недостатньо для компаній, які хочуть захиститися від атак.

На зміну IDS все частіше приходять системи запобігання вторгненням (Intrusion Prevention Systems, IPS). IPS схожа на IDS, але з активними компонентами, які зупиняють атаки до того, як вони стануть успішними. Зазвичай IPS складається з IDS та брандмауера з правилами. На відміну від IDS (рис. 1.1), IPS розміщуються в лінію (рис. 1.2), це означає, що IPS буде безперервно сканувати трафік, коли він проходить через неї. Таким чином, IPS повинна бути швидкою і мати високу обчислювальну потужність, щоб не викликати затримок в мережі, які можуть вплинути на продуктивність мережі для її користувачів.

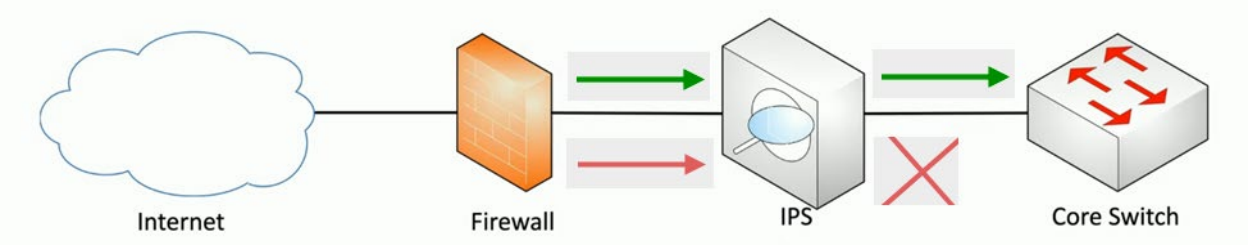

Рисунок 1.2 – Схема з’єднань системи запобігання вторгненням (Intrusion Prevention Systems, IPS).

Одним з основних недоліків багатьох IPS є хибнопозитивне виявлення атак. У випадку IDS хибне спрацьовування може бути незручним, але для IPS воно може спричинити DoS, оскільки легітимний трафік буде заблоковано. Крім того, оскільки IPS, а особливо системи запобігання мережевим вторгненням (NIPS), утворюють єдину точку відмови в мережі, вони повинні бути дуже стабільними і стійкими до атак.

1.2 Ключові Компоненти Системи IPS/IDS

Система IDS/IPS складається з наступних компонентів [4]:

- Препроцесор даних: Попередній процесор даних відповідає за збір і форматування даних для аналізу алгоритмом виявлення вторгнень.
- Алгоритм виявлення виявляє різницю між "нормальним" або "законним" і зловмисним мережевим трафіком на основі моделі виявлення.
- Фільтр попереджень оцінює серйозність вторгнення на основі критеріїв прийняття рішень і виявлених зловмисних дій. Потім фільтр сповіщень сповіщає мережевого або системного адміністратора і вживає відповідних заходів (зазвичай блокування).

На рис. 1.3 представлена блок-схема IPS/IDS системи яка використовується для моніторингу та реагування на безпекові події.

Робота починається з моніторингу системних активностей, де вхідний трафік або логи (журнали подій) перехоплюються для аналізу. Далі ці дані надходять у компонент попередньої обробки даних, де вони перетворюються на формат, придатний для аналізу. Після цього активність, вже у форматі оброблених даних, подається на вхід алгоритму виявлення, який спирається на модель виявлення.

Цей алгоритм виконує функції системи виявлення вторгнень (Intrusion Detection System, IDS), аналізуючи активність на предмет виявлення підозрілих або аномальних патернів.

Рисунок 1.3 – Загальні компоненти IPS/IDS системи

Коли алгоритм виявляє потенційну загрозу, він генерує сповіщення. Ці сповіщення потім проходять через фільтр тривог, який використовує критерії прийняття рішень для визначення, чи повинні ці сповіщення розглядатися як дійсні тривоги. Валідовані сповіщення можуть спричинити за собою вжиття заходів або створення звіту в IPS системі.

Ця блок-схема ілюструє процес, за допомогою якого системи безпеки обробляють вхідні дані, виявляють потенційні загрози та вирішують, як реагувати для забезпечення захисту системи.

Серед ключових функцій інструментів виявлення аномалій поведінки мережі, які допомагають цього досягти:

- Постійний моніторинг мережі[5]

Виявлення аномалій поведінки мережі є постійною частиною інформаційної безпеки. Це «завжди ввімкнена» діяльність, безперервний моніторинг поведінки мережі на пошук потенційних загроз. Ключові параметри поведінки порівнюються з прийнятними стандартами та історичними моделями, щоб інструмент виділяв будь-яку подію, яка виділяється.

- Аналіз зашифрованого трафіку

Інструмент має аналізувати зашифрований трафік і звичайні пакети, що передаються між загальнодоступними мережами. Це пояснюється тим, що інструменти мережевої безпеки зазвичай розгортаються в приватних корпоративних середовищах, де більшість обсягів трафіку зашифровано. Інструмент повинен відстежувати та аналізувати зашифровані потоки на наявність загроз і ризиків невідповідності, щоб забезпечити повну видимість.

- Детальне усвідомлення поведінки мережі

Недостатньо просто висвітлити незвичайну мережеву подію без будь-якої довідкової інформації. Наприклад, інструмент може повідомити менеджера мережі про незвично високе використання смуги пропускання програмою. Але чи така поведінка є законною, оскільки виникає внаслідок нового бізнес-процесу, що потребує високих вимог, чи це результат зловмисного програмного забезпечення? Інструмент має забезпечувати детальну обізнаність і контекстуальну інформацію, щоб уможливити відповідні дії.

- Сповіщення в режимі реального часу

Це основна функція інструменту виявлення аномалій поведінки мережі. Сповіщення в режимі реального часу дозволяють групі керування

мережею отримувати інформацію про потенційну загрозу одразу після її виявлення, не чекаючи запланованого звіту чи перевіряючи інформаційну панель. Інструмент в ідеалі повинен інтегруватися з системою керування інформацією про безпеку та подіями (SIEM), щоб надсилати ці сповіщення.

- Вбудовані або підключені системи реагування

Виявлення аномалій поведінки мережі є частиною більшої функції виявлення та реагування мережі (NDR). Подібно до інших систем безпеки, таких як брандмауери чи програмне забезпечення для запобігання вторгненням, які надсилають сповіщення до NDR, дані про аномалії мережі також оброблятимуться для швидкої реакції. Деякі інструменти мають вбудовану функцію NDR, щоб користувачі могли досліджувати аномалії та реагувати на них в одному робочому процесі. Інші можуть бути підключені до стороннього програмного забезпечення NDR або системи SIEM із функціями NDR.

Табл. 1.1 містить характеристики основних технологій та інструментів використовуваних в IPS/IDS системах.

Таблиця 1.1 – Огляд інструментів виявлення аномалій поведінки мережі

Інструмент виявлення аномалій	Постійний моніторинг мережі	Аналіз зашифрованого трафіку	Обізнаність про поведінку мережі	Сповіщення в реальному часі	Вбудовані / інтегровані системи реагування
	Набір власних датчик	Контекстна інформація (тип трафіку, наявність віддаленого доступу та ін.), не вимагає розшифровки	Власна технологія EntityIQ™	Інформаційна панель Awake та інтег	відкритий API для інтеграції із зовнішніми системами реагування.

Продовження таблиці 1.1

Інструмент виявлення аномалій	Постійний моніторинг мережі	Аналіз зашифрованого трафіку	Обізнаність про поведінку мережі	Сповіщення в реальному часі	Вбудовані / інтегровані системи реагування
	Телеметрія	Зашифрований трафік для отримання інформації про загрози та відповідність, не вимагає розшифровки	інформація про поведінку мережі та загрози для кінцевих точок і хмари	персоналізовані сповіщення про підозрілий доступ і ризики недотримання вимог	універсальне рішення Cisco з розширеним в и я в
	Моніторить поведінку мережі та події	Пасивні зонди та метадані трафіку з інформацією про багаторівневий протокол	інформація про ботнети, потенційне зловмисне програмне забезпечення, внутрішні загрози, витік даних тощо	Сповіщення через інтегровану	триадний підхід центру безпеки (SOC) для підключення SIEM, NDR і захисту кінцевих точок
	Моніторить мережеве середовище та оновлює свої алгоритми машинного навчання без нагляду	обробляє зашифрований трафік так само, як і звичайний мережевий трафік, створюючи групи рівноправних користувачів і виявляючи відхилення	детальні уявлення про мережу на основі показників поведінки та моніторингу на основі підписів і правил	інформаційну панель у реальному часі, яка надсилає сповіщення та інтегрується з усіма популярними каналами	власна платформа S I E M і

Продовження таблиці 1.1

Інструмент виявлення аномалій	Постійний моніторинг мережі	Аналіз зашифрованого трафіку	Обізнаність про поведінку мережі	Сповіщення в реальному часі	Вбудовані / інтегровані системи реагування
	збирає 100% інформації про мережу	Аналізатор на основі ML для аналізу всього трафіку, включаючи зашифровані потоки	попередньо запаковані правила, які виявляють сигнатури загроз, користувачі можуть створювати нові правила	Так, і генерує автоматичні часові шкали інцидентів, які допомагають під час аналізу подій	Кілька готових до використання інтеграцій, які не залежать від постачальника та можуть підключатися до будь-якої системи реагування
	Корпоративна мережа будь-якого типу та масштабу	Функція Inspector витягує методи шифрування та дані рівня з мережевого трафіку, аналіз поведінки без дешифрування	Статистичні дані та звіти про боковий рух трафіку між пристроями, викраданння даних і мережеві загрози	Так, користувачі можуть налаштовувати сповіщення та робочі процеси за допомогою готових до використання п	Може інтегруватися з будь-якою SIEM

Як видно з табл.1.1, використання власних датчиків Awake Sensors і технології EntityIQ™ в Awake Security Adversarial Modeling свідчать про високий рівень інтелектуальності та аналітики. Інтеграція з SIEM та відкритий API дозволяють впроваджувати різноманітні системи реагування. Застосування телеметрії та аналіз зашифрованого трафіку в Cisco Stealthwatch свідчать про ефективне використання штучного інтелекту в інструменті. Універсальне рішення Cisco з розширеним виявленням і реагуванням (XDR) підкреслює комплексність підходу. Щодо Flowmon NBAD, то використання пасивних зондів та аналіз метаданих

трафіку показує високий ступінь моніторингу мережі. Триадний підхід центру безпеки (SOC) дозволяє ефективно взаємодіяти з іншими системами реагування. NetWitness Detect AI є найбільш позиціонованою на ШІ системою. Активне оновлення алгоритмів машинного навчання без нагляду свідчить про постійне вдосконалення системи. Інтеграція з власною платформою SIEM і XDR підкреслює повноту та комплексність відстеження та реагування. В системі Gurucul ML XDR, збір 100% інформації про мережу та аналіз трафіку на основі машинного навчання підтримують повноцінний цикл виявлення загроз. Готові до використання інтеграції дозволяють ефективно підключатися до різних систем реагування. Функція Inspector у IBM QRadar Network Insights витягує методи шифрування та дані рівня з мережевого трафіку, підвищуючи рівень обізнаності про зашифрований рух. Можливість налаштування сповіщень та інтеграція з будь-якою SIEM дозволяють адаптувати систему до конкретних потреб користувачів.

1.3 Сучасні Виклики для IPS/IDS

Однією з основних проблем традиційних IPS/IDS є їхня схильність до генерації хибно позитивних і хибно негативних результатів[6]. Хибно позитивні вказують на вторгнення там, де їх насправді немає, що може призвести до надмірної кількості сповіщень про загрози та втрати ресурсів на перевірку фальшивих сигналів. Хибно негативні, навпаки, означають невиявлення реальних атак, що загрожує безпеці мережі.

Підтримка протоколів та аплікацій. Системи IPS/IDS повинні бути здатні розуміти і аналізувати різні мережеві протоколи та рівні аплікацій. Виникають проблеми при аналізі зашифрованого трафіку, який може маскувати потенційно шкідливі дії.

Відсутність контексту. Традиційні IPS/IDS зазвичай операційно не здатні аналізувати трафік з контекстом. Це означає, що вони можуть пропустити складні атаки, які розподіляються в часі або вимагають додаткового контексту для виявлення.

Потужність обчислювальних ресурсів. Висока обчислювальна потужність часто необхідна для обробки великих обсягів мережевого трафіку в реальному часі, особливо при застосуванні складних алгоритмів виявлення і запобігання вторгнень.

Порядок аналізу трафіку важливий. Відсутність правильного порядку обробки може призвести до недоліків у виявленні атак або збоїв в роботі системи.

Узгодженість між моделями IPS та. Традиційні методи часто не справляються зі складними кіберзагрозами, що постійно розвиваються. Щоб заповнити цю прогалину, передові технології, такі як машинне навчання (ML) і глибоке навчання (DL), все частіше інтегруються в ці системи. Алгоритми ML і DL підвищують здатність IPS та IDS навчатися на основі даних, адаптуватися до нових загроз і приймати обґрунтовані рішення на основі шаблонів і аномалій, виявлених в мережевому трафіку.

Розробка правил для невідомих загроз. Традиційні системи, засновані на правилах, часто обмежені в своїх можливостях протистояти новим типам атак, які не відповідають існуючим шаблонам або сигнатурам. Інтеграція інтелектуальних методів, зокрема моделей ML, може значно підвищити здатність цих систем виявляти та пом'якшувати невідомі загрози.

Отже, виклики, з якими стикаються IPS та. Підвищення узгодженості моделей виявлення та запобігання за допомогою передових технологій, таких як ML та DL, розробка адаптивних правил для протидії невідомим загрозам, а також зменшення кількості хибних спрацьовувань та негативних результатів завдяки обізнаності про навколишнє середовище є ключовими напрямками роботи. Ці зусилля мають вирішальне значення для

забезпечення того, щоб IPS та IDS залишалися ефективними інструментами в постійній боротьбі з кіберзагрозами.

1.4 Висновок до першого розділу

У цьому розділі було проведено детальний аналіз ролі та значення штучного інтелекту (ШІ) у системах виявлення та запобігання вторгнень (IPS/IDS). Розглядаючи сучасні виклики кібербезпеки, зокрема атаки нульового дня та складність обробки великих обсягів даних, ми визначили, що інтеграція ШІ в системи IPS/IDS здатна значно покращити реагування системи на зовнішні загрози, особливо такі які геренуються автоматично.

Автоматизація процесів та використання алгоритмів машинного навчання та глибинного аналізу даних дозволяють цим системам оперативно ідентифікувати аномальну поведінку та незвичайні активності в мережі, що можуть вказувати на потенційні безпекові інциденти. Завдяки ШІ, системи IPS/IDS здатні ефективніше виявляти як внутрішні, так і зовнішні загрози, а також знижувати час реагування на інциденти завдяки швидкому аналізу та автоматизованому вирішенню.

Інтеграція штучного інтелекту в системи IPS/IDS стає критично важливим елементом у сучасних стратегіях кібербезпеки.

ТЕОРЕТИЧНІ ОСНОВИ ШІ В IPS/IDS

У розділі розглянуто основні види машинного навчання, та варіанти їх використання для моделей IDS. Виявлено їх переваги та недоліки. Оглянуто різні набори даних які можна використати для навчання МН. Виділено методи та метрики оцінки різних моделей. Проведено аналіз готових рішень.

Машинне навчання

Машинне навчання є фундаментальним компонентом сучасних систем виявлення вторгнень, що грає ключову роль у підвищенні ефективності і точності виявлення кібератак. Використання алгоритмів машинного навчання в IDS дозволяє проводити глибокий автоматизований аналіз даних, що включають історичні дані про мережевий трафік і відомі випадки вторгнень. Ці дані стають основою для навчання моделей, які в подальшому можуть ідентифікувати незвичайну або підозрілу поведінку в мережі.

Завдяки різноманіттю алгоритмів, таких як нейронні мережі, дерева рішень, та ансамблеві моделі, машинне навчання у IDS адаптується до специфіки мережі та різних видів загроз. Це дозволяє системам не лише виявляти відомі типи атак, але й ефективно реагувати на нові, невідомі вторгнення. Алгоритми машинного навчання аналізують мережевий трафік у реальному часі, відстежуючи відхилення від нормального патерну поведінки, що є ключовим для виявлення зловмисної активності.

Одним з найважливіших аспектів використання машинного навчання в IDS є його здатність до адаптації та самонавчання. Системи можуть постійно оновлюватися на основі нових даних, що забезпечує високий рівень гнучкості у відповіді на змінюючіся кіберзагрози. Це також сприяє

зменшенню кількості помилкових спрацьовувань, підвищуючи точність виявлення реальних атак.

Інтеграція алгоритмів машинного навчання з іншими системами кібербезпеки, такими як фаєрволи та антивірусні програми, створює більш комплексну та ефективну оборону проти кібератак. Крім того, можливість прогнозування та запобігання атакам, заснована на аналізі поточних тенденцій, стає важливою перевагою, що дозволяє попереджати потенційні вторгнення, навіть перш ніж вони відбудуться.

Як показано на рис. 2.1, існує три основні типи машинного навчання: контрольоване, неконтрольоване та з навчання з підкріпленням [7]. Ці методи розглядаються далі в цьому розділі.

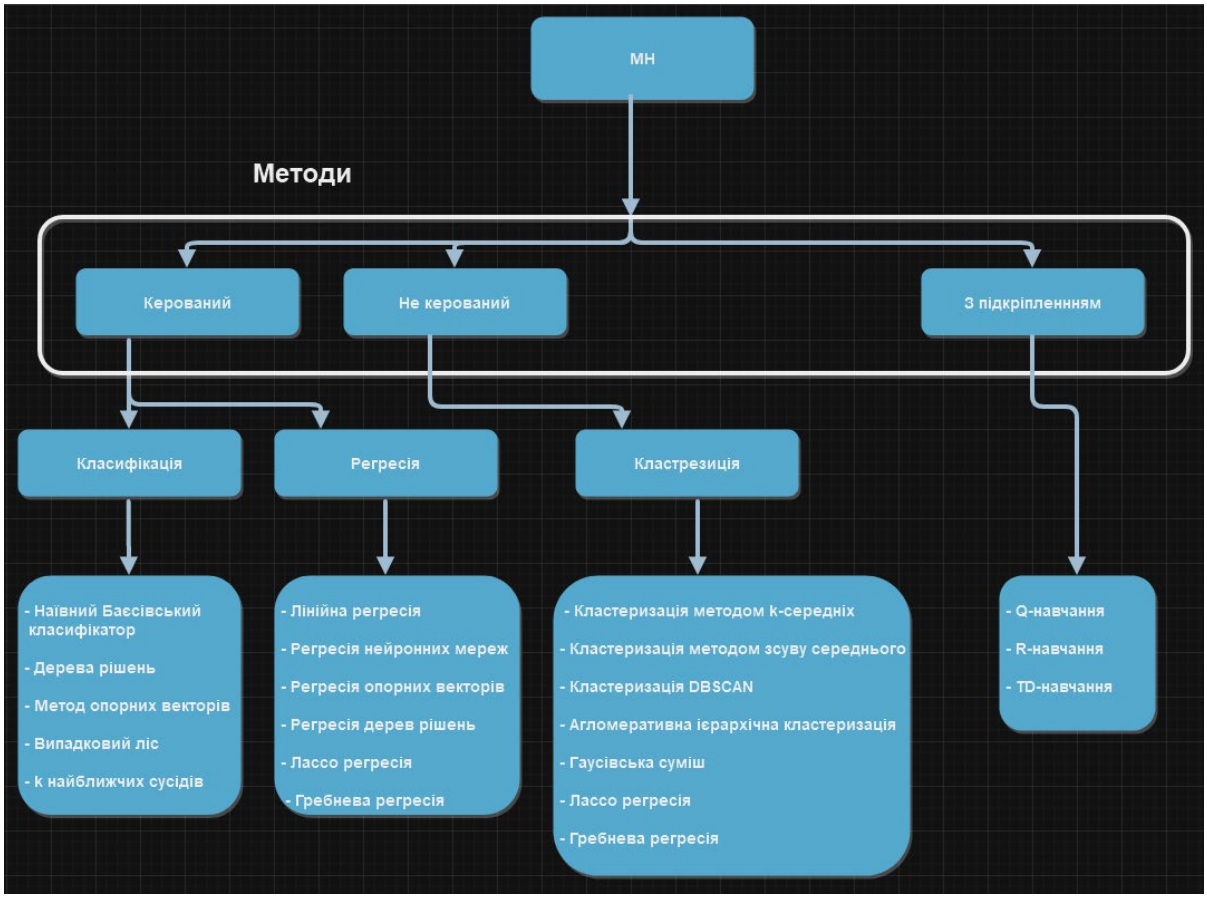

Рисунок 2.1 – Основні типи методів машинного навчання

Контрольоване машинне навчання

У сфері кібербезпеки, зокрема у системах виявлення вторгнень (IDS), контрольоване машинне навчання[8] відіграє вирішальну роль, надаючи потужні інструменти для аналізу та класифікації мережевих даних. У контрольованому навчанні алгоритми ML використовують заздалегідь підготовлений набір даних, який складається з вхідних даних (наприклад, мережевих пакетів) та відповідних міток, що визначають, чи є ці дані нормальними або містять сліди вторгнення.

Цей підхід вимагає ретельної підготовки датасету, де кожен елемент має бути коректно анотований. Така анотація може включати мітки, які ідентифікують певні типи атак або підтверджують безпечну поведінку. Використання добре підготовлених даних дозволяє алгоритму точно "вчитися" на випадках зловмисної та безпечної поведінки, що згодом покращує точність класифікації в реальних умовах.

Алгоритми контрольованого навчання можуть включати різноманітні техніки, такі як логістична регресія, нейронні мережі, рішаючі дерева та ансамблеві методи, як-от випадкові ліси. Кожен з цих методів має свої особливості та переваги у різних сценаріях виявлення вторгнень.

Однією з ключових переваг контрольованого навчання є його здатність ефективно виявляти відомі типи атак, для яких існують попередньо підготовлені дані. Це робить цей підхід особливо цінним у боротьбі з поширеними та добре документованими загрозами. Однак, однією з обмежень є залежність від якості та охоплення тренувального датасету, що може обмежувати здатність алгоритму виявляти нові, раніше невідомі види атак.

Машинне навчання без нагляду

Неконтрольоване машинне навчання використовується з немаркованими даними. Як випливає з назви, неконтрольоване машинне навчання не контролюється користувачем для покращення моделі. Модель буде вдосконалюватися сама по собі, оскільки вона буде виявляти закономірності та інформацію з набору даних, який їй надали. Зазвичай алгоритм групує різні дані в категорії, які мають схожість або відмінності. Навчання без вчителя корисне для аналізу великих даних. Як показано на рис. 2.2, неконтрольоване навчання можна розділити на три типи проблем: кластеризація, асоціація та зменшення розмірності:

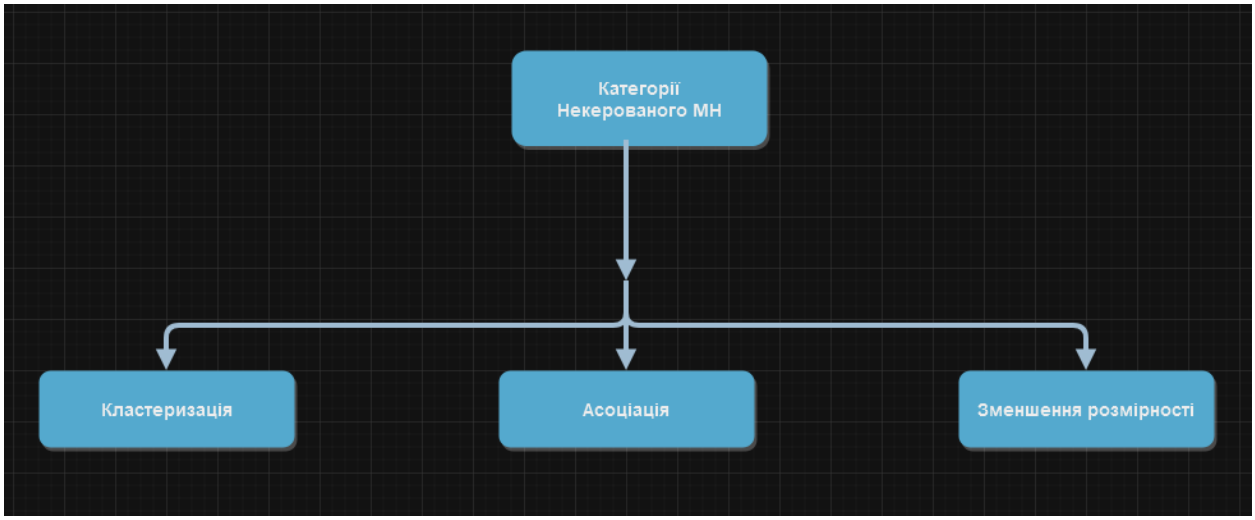

Рисунок 2.2 – Категорії неконтрольованого машинного навчання

Машинне навчання без нагляду, яке не вимагає попередньо маркованих даних для тренування, включає в себе кілька ключових методів: кластеризацію, асоціацію та зменшення розмірності. Ці методи знаходять широке застосування у різних сферах, включаючи системи виявлення вторгнень (IDS), де вони використовуються для ідентифікації прихованих шаблонів та відносин у даних.

Кластеризація полягає у групуванні набору об'єктів таким чином, що об'єкти в одному кластері (групі) між собою схожі, а об'єкти з різних кластерів – відрізняються. У контексті IDS, кластеризація може бути використана для виявлення аномальних патернів поведінки в мережі, групуючи подібні види мережевого трафіку та виділяючи нетипові активності, які можуть вказувати на зловмисні дії.

Методи асоціації зосереджені на виявленні правил або зв'язків між різними елементами в наборах даних. Це може бути особливо корисним для виявлення складних загроз та атак, які мають кілька фаз або використовують різні вектори атаки. Наприклад, виявлення взаємозв'язку між певними видами мережевих запитів та послідовністю зловмисних дій.

Зменшення розмірності є ще одним ключовим елементом машинного навчання без нагляду, яке дозволяє спрощувати дані, знижуючи кількість випадкових змінних, які їх описують. Це може бути здійснено за допомогою методів як головні компоненти аналізу (PCA) чи t-SNE. У контексті IDS, це дозволяє виділити основні фактори або атрибути, які характеризують нормальний або аномальний мережевий трафік, полегшуючи процес аналізу та виявлення вторгнень.

Загалом, машинне навчання без нагляду забезпечує важливі інструменти для аналізу та розуміння складних мережевих даних. Воно дозволяє виявляти приховані взаємозв'язки та аномалії, які можуть не бути очевидними при традиційному підході, і є незамінним у виявленні та реагуванні на складні та високотехнологічні кібератаки. Однак через велику кількість даних, необхідних для навчання, воно вимагає великих обчислювальних потужностей і багато часу. Крім того, неконтрольоване навчання має вищий ризик отримання неточних результатів порівняно з контрольованим навчанням. Таким чином, наприкінці навчання часто потрібне втручання людини, щоб перевірити, чи правильні вихідні змінні. Ця перевірка також займає багато часу.

Машинне навчання з підкріпленням

Навчання з підкріпленням є одним із ключових підходів у сфері машинного навчання, яке знаходить застосування у багатьох областях, включаючи кібербезпеку та системи виявлення вторгнень (IDS). Цей тип навчання відрізняється від контрольованого та неконтрольованого навчання тим, що зосереджений на взаємодії агента з динамічним середовищем, де він намагається максимізувати деяку міру винагороди через свої дії.

У контексті IDS, навчання з підкріпленням може бути використане для розробки систем, які адаптуються до змінюваних патернів атак та ворожого середовища. Наприклад, алгоритм з підкріпленням може вчитися визначати найбільш ефективні стратегії для виявлення нових видів вторгнень або навіть адаптуватися до змін в мережевій інфраструктурі.

Процес навчання включає в себе визначення стратегій (політик), які вказують агенту, які дії вибирати в певних станах середовища, щоб максимізувати очікувану винагороду. В контексті IDS, ця "винагорода" може бути пов'язана з успішним виявленням вторгнення або ефективним запобіганням атакам.

Одним з важливих аспектів навчання з підкріпленням є його здатність вчитися у багатоетапних середовищах, де ефекти певних дій можуть бути не відразу очевидні. Це дозволяє системі розвивати складніші стратегії, які враховують як негайні, так і довготривалі наслідки рішень.

Однак, навчання з підкріпленням також має свої виклики, особливо у контексті IDS. По-перше, воно вимагає чітко визначеного механізму винагороди, який іноді складно розробити у динамічному середовищі кіберзагроз. По-друге, моделі з підкріпленням можуть вимагати значного обсягу взаємодії з середовищем для ефективного навчання, що може бути складно забезпечити в реальних умовах.

Завдяки цим особливостям, навчання з підкріпленням відкриває нові перспективи для розвитку IDS, здатних адаптуватися та еволюціонувати у відповідь на змінюючі сценарії кіберзагроз, хоча й вимагає ретельного підходу до розробки та валідації. У табл. 2.1 наведені методи ML.

Таблиця 2.1 – Огляд методів машинного навчання

Методи ML	**Частовикористовувані алгоритми**	**Переваги**	**Недоліки**
Контрольоване навчання	Логістична регресія, Нейронні мережі, Рішаючі дерева, SVM	Висока точність для відомих шаблонів, добре підходить для класифікації та регресії	Вимагає великих попередньо маркованих датасетів, менш ефективний для невідомих шаблонів
Неконтрольоване навчання	K-середніх, DBSCAN, Асоціативні правила	Виявлення прихованих шаблонів та аномалій без попереднього маркування даних	Може бути складним у інтерпретації, висока залежність від якості вхідних даних
Навчання з підкріпленням	Q-навчання, Глибоке Q-навчання, Політики градієнтного спуску	Адаптація до змінних середовищ, підходить для складних задач з множинними кроками	Вимагає значної кількості взаємодій для ефективного навчання, ризик перенавчання

Процес машинного навчання

Процес ML складається в основному з двох етапів: навчання та прогнозування (рис. 2.3). На етапі навчання в алгоритм ML подаються певні набори даних як вхідні дані. Алгоритм ML створює навчену модель, яка потім дозволяє прогнозувати вхідні дані. Важливо підкреслити, що дані, які використовуються для навчання, не можуть використовуватися на етапі прогнозування, щоб перевірити, чи навчена модель працює правильно, оскільки модель не дасть надійних результатів на даних, які вже використовувалися для навчання моделі.

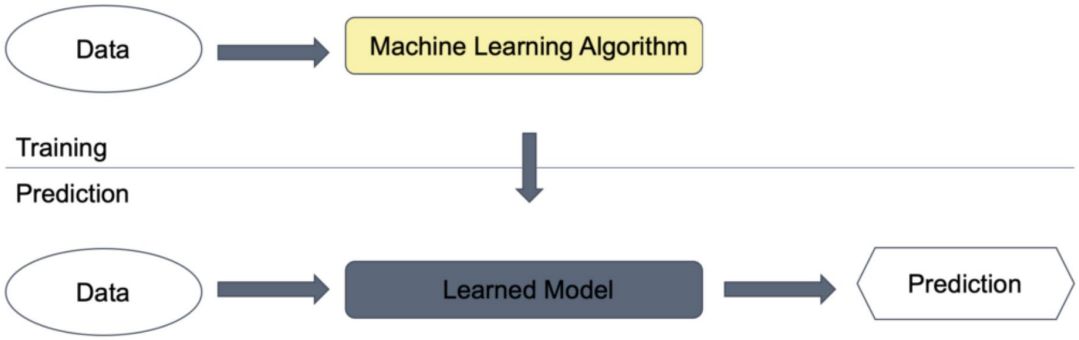

Рисунок 2.3 – Узагальнена схема процесу машинного навчання

2 Технології глибокого навчання

Методи DL з’явилися в 2006 році і стали важливою темою досліджень. Слово «deep» означає багато прихованих шарів нейронної мережі. Це підкатегорія штучних нейронних мереж (ШНМ) і має більше прихованих рівнів, ніж традиційні нейронні мережі, які досягають 150. Незважаючи на те, що це гілка машинного навчання, складність у структурі та представлення даних для навчання робить його ширшою версією ML. DL має справу з алгоритмами, які навчаються на прикладах, як і в ML. Продуктивність алгоритмів ML і DL змінюється зі збільшенням масштабу даних. Щоб знайти мережеві шаблони, алгоритми DL вимагають масивних даних, тоді як алгоритми ML вимагають менше даних. Структуру можна зробити глибокою, додавши один або кілька прихованих шарів у ШНМ, і оскільки дані обробляються на кожному рівні, таким чином завдання навчання стає глибшим.

У прямому порівнянні між DL і ML, DL показує значні переваги щодо продуктивності та точності, і він може обробляти набагато більший обсяг даних. Але в чому його причина? DL має дві важливі переваги, оскільки не вимагає ручного проектування функцій і використовує кілька рівнів.

Вилучення функцій людьми та вручну неможливо з огляду на величезну кількість неструктурованих даних, а також у випадках, коли аналітики та інженери безпеки не мають знань про те, які характеристики мають значення для виявлення загроз через величезну потенційну кількість поведінки зловмисників. Як ви можете бачити на зображенні нижче, DL складається з одного входу, одного виходу та кількох повністю пов’язаних прихованих шарів між ними. Кожен рівень представлений у вигляді серії нейронів і витягує характеристики вищого рівня даних, доки останній рівень не вирішить, що показує вхід. Чим більше рівнів мають мережі, тим функції вищого рівня вивчатимуться

D

L – Розмір даних - алгоритми DL працюють набагато краще з великими обсягами даних (мільйони), тоді як алгоритми ML найкраще працюють з невеликими наборами даних.

р – Час - хоча алгоритми DL вимагають більше часу для навчання, цей додатковий час компенсується на етапі виробництва та експлуатації в реальному часі.

о – Вибір характеристик для аналізу - алгоритми DL вибирають функції (вхідні дані) самостійно, а експерти з безпеки інтерпретують результати на основі їх підходу. Навпаки, алгоритми ML вимагають визначення своїх функцій і міток (виходів).

В останні роки методи DL використовувалися для ідентифікації атак типу «Відмова в обслуговуванні» (DoS), і вдалося їх правильно класифікувати.

г

а

т

о

Оцінка ефективності роботи IDS та ML

п

е

р

Алгоритми машинного навчання (ML) відіграють ключову роль в оцінці систем виявлення вторгнень (IDS), пропонуючи складні засоби для виявлення та пом'якшення загроз безпеці. Ефективність цих алгоритмів зазвичай оцінюється за допомогою ряду показників продуктивності

Для опису матриці плутанини використовуються наступні терміни:

- Істинно позитивний (TP): Зразок атаки було правильно ідентифіковано як атаку.
- Істинно негативний (TN): Нормальний зразок було правильно ідентифіковано як нормальний трафік.
- Хибнопозитивний результат (FP): нормальний зразок було помилково ідентифіковано як атаку.
- Хибнонегативне спрацьовування (FN): Зразок атаки було помилково ідентифіковано як звичайний трафік

IDS повинен мати низький рівень помилкових спрацьовувань, щоб уникнути помилкових тривог у системі, оскільки це створює збої в роботі мережі. Крім того, необхідно мати низький рівень хибних спрацьовувань, щоб уникнути невиявлених атак, які потрапляють у вашу мережу.

Використовуючи вищезгадані терміни та матрицю з Таблиці 2.2, ми можемо оцінити різні метрики, які часто використовуються для оцінки ефективності IDS.

Таблиця 2.2 – Метрики для оцінки IDS

		Прогнозований Клас	
		Нормальний	Атака
Фактичний клас	Нормальний	TN	FP
	Атака	FN	TP

Основні метрики в оцінці IDS[10]:

Точність: Це найпростіша метрика. Вона вимірює частку загальних прогнозів (як позитивних, так і негативних), які модель визначила правильно.

Формула виглядає так:

=

де Accuracy – точність;

TP-кількість істинно позитивних прогнозів;

TN- кількість істинно негативних прогнозів;

- загальна кількість прогнозів.

Чіткість: Чіткість оцінює правильність, досягнуту в позитивному класі. Вона відповідає на питання: "Скільки з усіх випадків, класифікованих як позитивні, є насправді позитивними?"

«Формула виглядає так:

c

де Precision – чіткість; =

TP-кількість істинно позитивних прогнозів;

FP – кількість хибнопозитивних результатів;»

Відтворення (чутливість): Відтворення обчислює здатність моделі знаходити всі релевантні випадки (позитивні). Це відношення правильно передбачених позитивних спостережень до всіх фактичних позитивних спостережень.

Формула має вигляд:

=

де Recall - Відтворення (чутливість);

T

P F

N кількість істинно позитивних прогнозів;

кількість хибно негативних результатів;

Оцінка F1. Показник F1 – це середнє гармонійне значення точності та відгуку, що забезпечує баланс між ними. Це особливо корисно, коли розподіл класів нерівномірний.

Формула виглядає наступним чином:

=

де F1- середнє гармонійне значення точності та відгуку; ×

Recall - Відтворення (чутливість);

Precision – чіткість.»

Час виконання: Це міра обчислювальної ефективності алгоритму, яка показує, скільки часу потрібно моделі для прогнозування.

Важливість цих показників:

Точність є життєво важливою для розуміння загальної ефективності моделі, але може вводити в оману, якщо класи незбалансовані.

Точність має вирішальне значення в сценаріях, де помилкові спрацьовування є дорогими або небезпечними.

Повторюваність стає критичною в ситуаціях, коли пропуск істинно позитивного результату може мати серйозні наслідки, наприклад, при виявленні порушень безпеки.

F1-Score забезпечує більш реалістичну оцінку у випадках, коли існує нерівномірний розподіл класів або коли і точність, і відгук є однаково важливими.

Час виконання є важливим для IDS в реальному часі, де час відгуку є критично важливим.

Таким чином, оцінка алгоритмів ML в IDS за допомогою цих метрик дає повне уявлення про їхню продуктивність, балансуючи між точністю та обчислювальною ефективністю. Цей баланс особливо важливий в динамічних і складних середовищах.

Аналіз наборів даних

Огляд найбільш відомих наборів даних[11], які використовуються для навчання та тестування систем виявлення вторгнень:

- KDDcup99: Цей набір даних, похідний від DARPA'98, широко використовується для оцінки IDS. Він містить близько 4.9 мільйона зразків, кожен з яких має 41 характеристику, класифікуючись як нормальний або атакуючий. Атаки поділяються на 4 типи: DoS, U2R, R2L, і зондування. Є три версії цього набору: повний, 10% вибірка, і тестовий набір. Головна проблема KDDcup99 - незбалансованість, особливо в класах, як DoS та Probe.
- Кіото 2006: Створений за допомогою збору даних через різні мережеві заходи безпеки в Кіотському університеті. На базі 41 ознаки KDDcup99 визначено 14 статистичних та 10 додаткових ознак, утворюючи 24-ознакові вибірки. Включає трафік з 2006 по 2015 рік.
- NSL-KDD: Розроблений для вирішення проблем KDDcup99, запропонований у 2009 році. Зберігає чотири категорії атак з KDDcup99 і включає навчальний та тестовий набори з 21 і 37 різними типами атак відповідно.

- UNSW-NB15: Створений Австралійським центром кібербезпеки, цей набір даних є гібридом звичайної діяльності та атак. Містить дев'ять типів атак і включає два файлы: навчальний і тестовий з різними типами трафіку.

  [12]: Створений Канадським інститутом кібербезпеки у 2017 році, цей набір включає реальний трафік з звичайними та атакуючими зразками. Атаки аналізуються за допомогою CICFlowMeter з врахуванням часових міток, IP-адрес і протоколів.Узагальнення різних наборів даних наведено нижче в Таблиці 2.3.

Таблиця 2.3 – Короткий опис наборів даних

Набір даних	Рік	Типи Атак	Атаки
KDDcup99	1999	4	Probe, DoS, U2R, R2L
Kioto-2006	2006	2	Відомі атаки, Невідомі атаки
NSL-KDD	2009	4	Probe, DoS, U2R, R2L
UNSW-NB15	2015	7	Fuzzers, Analysis, Backdoors, DoS, Exploits, Generic, Reconnaissance, Shellcode, Worms
CICIDS2017	2017	7	Brute Force, DoS, HeartBleed, Web attack, Infiltration, Botnet, DDoS

2.5 Аналіз досліджень з МН в IDS

Опираючись на дослідження [12], було проведено аналіз десяти наукових робіт, що зосереджуються саме на застосуванні методів машинного навчання в IPS та IDS системах. Цей аналіз дозволив виявити актуальні тренди у виборі моделей машинного навчання та найпопулярніші датасети, які використовуються у сучасних дослідженнях.

В роботі [13] Bamhdi та ін. було розроблено ансамбль класифікаторів (SVM, RF, KNN, PSO, ETC, MLP) з системою голосування більшістю голосів для підвищення точності виявлення вторгнень. Датасет: NSL-KDD. Точність: 99% для DoS, 97,2% для Probe, 93,2% для U2R-атак.

Sun та ін. [14] створили глибоку навчальну систему виявлення вторгнень (DL-IDS), використовуючи гібридну архітектуру CNN та LSTM для аналізу мережевого трафіку. Датасет: CICIDS2017. Загальна точність: 98,67%.

Vinayakumar та ін. (2019) [15] розробили вдосконалену IDS, використовуючи глибокі нейронні мережі (DNN) з моделлю MLP. Датасети: KDDCup 99, NSL-KDD, UNSW-NB15, Kyoto, WSN-DS, CICIDS 2017.

В роботі [16], Wisawanichthan і Thammawichai (2021): Представили двошаровий гібридний підхід (DLHA) для ефективного виявлення різноманітних типів мережевих атак. Використали наївний класифікатор Байєса і SVM. Датасет: NSL-KDD. Точність: 96,67% для R2L та 100% для U2R.

Фокус Mari, Zinca, and Dobrota [17] був на створенні ворожих екземплярів мережевого трафіку з використанням GAN для підвищення стійкості IDS. Датасет: NSL-KDD.

Lirim та ін. [18] використали CNN у поєднанні з багатошаровим персептроном для підвищення точності виявлення вторгнень. Датасет: UNSW-NB15. Точність: 94,4% і 95,6%.

Lin Chen та ін. [19] розробили IDS на основі CNN, інтегровану з онлайн-фазою виявлення. Датасет: CICIDS2017. Точність: 96,55% на наборі ознак і 99,56% на необробленому трафіку.

Yu та ін. [20] впровадили IDS, засновану на методі Few-Shot Learning, використовуючи CNN та DNN для вилучення ознак. Датасети: UNSW-NB15 та NSL-KDD. Точність: 92,34% і 92%.

Andresini та ін. [21] поєднали неконтрольований підхід з автокодерами та контрольованим етапом з CNN для побудови наборів даних. Датасети: KDDcup99, UNSW-NB15, CICIDS2017. Точність: 92,49%, 93,40% і 97,90%.

Elhefnawy та ін. [22] розробили гібридний вкладений генетично-нечіткий алгоритм (HNGFA) для виявлення атак. Датасети: KDDcup99 і UNSW-NB15. Точність: 98,19% і 80,54%.

Таблиця 2.4 – Узагальнені результати аналізу

Автори	Рік	Метод виділення ознак	Використаний класифікатор	Виявлені Атаки
Bamhdi, Abrar, and Masoodi	2021	PSO	Ансамблевий(SVM)(RF)(KNN)(PSO)(ETC)(MLP)	DoS Probe R2L U2R
Sun et al	2022	CNN	LSTM model	DoS .DDoS Port Scanning Botnet Attacks Web Attacks: Attacks targeted towards web applications, including SQL injection, XSS (Cross-Site Scripting), Infiltration
Vinayakumar et al.	2019	N/A	DNN з можливістю масштабування прихованих шарів	Normal, DoS, Probe, R2L, U2R
Wisanwanichthan et al.	2021	ICFS та PCA	Наївний Байєс та SVM	Employed Double Layered Hybrid Approach (DLHA) for detecting DoS, Probe, R2L and U2R
Mari, Zinca, and Dobrota	2023	ANN, Random Forest and kNN	GAN	Normal, DoS, Probe, R2L, U2R
Lin et al	2020	NA	CNN	FTP Brute Force, SSH Brute Force, DoS, Web attacks, penetration attacks
Yu et al	2020	Вбудована функція з використанням CNN та DNN	Навчання з кількох пострілів	DoS, Probe, U2R, R2L. Other attack types tested: normal, generic, fuzzers, reconnaissance, shellcode, worms, backdoor, exploits
Andresini et al	2020	Подвійний автокодер	Soft-max	NA
Elhefnawy et al]	2020	Наївний Байєс	Гібридний вкладений генетичний нечіткий алгоритм	Probe, DoS, U2R, and R2L. Other attack types tested: normal, generic, fuzzers, reconnaissance, shellcode, worms, backdoor, exploits
Lirim et al	2021	NA	CNN з багатошаровим персептроном	DoS, DDoS, PortScan, Web Attack, Heartbleed, Benign, Infiltration, Brute Force, SSH, FTP

При аналізі було виявлено, що використання машинного навчання значно підвищує ефективність IDS (систем виявлення вторгнень). Якість наборів даних є критичною для ефективності IDS, особливо в умовах зростання обсягів даних.

Глибоке навчання, включаючи CNN (конволюційні нейронні мережі), стає популярнішим, адже воно краще справляється з великими обсягами даних та ефективне проти "нульових" атак. Для обробки цих методів часто

використовуються графічні процесори та хмарні платформи. Проте, багато існуючих рішень використовують застарілі набори даних, такі як KDDcup99 та NSL-KDD, що може знижувати точність систем при тестуванні на сучасному мережевому трафіку.

Проблема дисбалансу класів у наборах даних також залишається критичною, оскільки це може ускладнити виявлення рідкісних, але небезпечних типів атак. Розробники стикаються з необхідністю балансування між складністю моделі та потребами в обчислювальних ресурсах. Ефективність моделі може покращуватися через ретельний відбір важливих ознак для навчання, що знижує потребу в ресурсах. Загалом, це аналіз готових моделей підкреслює важливість використання актуальних наборів даних для підвищення ефективності IDS в сучасних мережевих умовах.

2.6 Висновок до другого розділу

У другому розділі було розглянуто, та проаналізовано які є моделі машинного навчання, їх переваги та недоліки в контексті IPS/IDS. Розглянуто метрики та формули оцінювання ефективності моделі МЛ в IDS. Проаналізовані набори даних, які використовують для навчання та тестування IDS систем, та було чітко зрозуміло, що найкращими датасетами є ті що були опубліковані найпізніше, та ті що мають широкий спект атак.

Також був проведений порівняльний аналіз досліджень готових МЛ моделей. Згідно з цим аналізом досліджень, останні тенденції показують, що методи глибокого навчання все частіше використовуються для виявлення атак. Однак це збільшує складність моделей, що вимагає більше обчислювальних ресурсів.

Аналіз та порівняння усіх цих даних, дав нам зрозуміти, як і які моделі можна застосовувати для покращення роботи IDS системи. Які недостатки та переваги має та чи інша модель, на яких датасетах краще навчати модель.

Тому ґрунтуючись на цих даних у третьому розділи ми оберемо модель, датасет, оцінимо її точність.

ТЕСТУВАННЯ АРХІТЕКТУР ГІБРИДНОЇ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ АНАЛІЗУ МЕРЕЖЕВОГО ТРАФІКУ

Мета даного дослідження полягає у виборі та оптимізації моделі машинного навчання для системи виявлення вторгнень (IDS) з метою ефективнішого виявлення аномалій у мережі, що допоможе у попередженні потенційних кібератак. Важливо забезпечити високу точність моделі, щоб мінімізувати кількість хибних позитивів та хибних негативів. Висока точність дозволить системі IDS ефективно виявляти реальні загрози, зменшуючи ризик пропуску атак. Модель повинна бути не лише точною, але й ефективною з точки зору часу навчання. Це зменшить обчислювальне навантаження на систему та спростить процес її впровадження та оновлення. Автоматизація процесу виявлення аномалій зменшує потребу у безперервному моніторингу спеціалістами, що дозволяє їм зосередитися на більш складних задачах та стратегічному рішенні проблем.

3.1 Вибір архітектури моделі та планування експерименту

У цьому контексті, дослідження зосереджується на використанні гібридної нейронної мережі CNN-BiLSTM [23], яка комбінує переваги конволюційних нейронних мереж (CNN) для ефективного виявлення шаблонів у вхідних даних та двонаправлених довгострокових короткочасних мереж (BiLSTM) для аналізу послідовностей даних. Це дає можливість глибше аналізувати мережевий трафік та виявляти складні веб-атаки з більшою точністю та ефективністю. В оригінальному формулюванні, CNN-BiLSTM застосовується до розпізнавання іменованих об’єктів, і вивчає функції як на рівні символів, так і на рівні слів. Компонент CNN використовується для створення функцій рівня символа. Для кожного слова модель використовує згортку та максимальний рівень об’єднання,

щоб витягти новий вектор ознак із векторів ознак для кожного символу, таких як вбудовування символів і (опціонально) тип символу.

Двонаправлений LSTM або BiLSTM – це термін, який використовується для моделі послідовності, яка містить два рівні LSTM, один для обробки вхідних даних у прямому напрямку, а інший для обробки у зворотному напрямку (рис. 3.1). Це призводить до кращого вивчення функцій за крок часу. Зазвичай використовується в завданнях, пов'язаних з НЛП. Інтуїція, яка лежить в основі цього підходу, полягає в тому, що, обробляючи дані в обох напрямках, модель може краще зрозуміти зв’язок між послідовностями (наприклад, знаючи наступні та попередні слова в реченні).

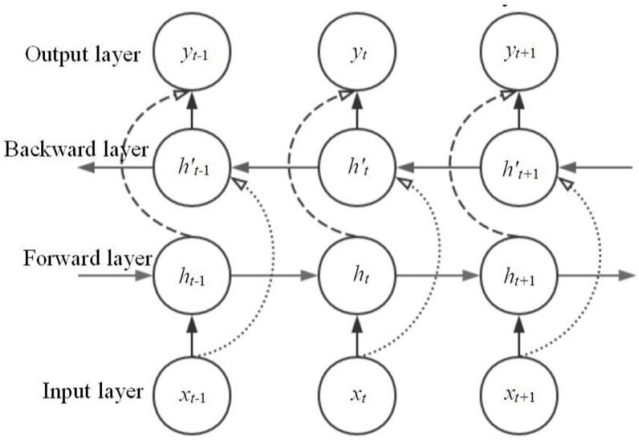

Рисунок 3.1 – Структура BiLSTM мережі [24]

Загальна модель почитається з 1-D CNN шару який дозволяє швидко розвивати просторове навчання для даного часового ряду даних. За рівнем 1-DCNN слідує шар Max Pooling, який дає змогу дискретизувати параметри на основі вибірки, щоб розпізнавати відповідні функції, що призводить до скорочення часу навчання та запобігання переобладнанню. За Max Pooling, іде рівень пакетної нормалізації (Batch Normalisation Layer), який дає змогу нормалізувати параметри між проміжними шарами, щоб досягти меншого часу навчання. Двошарові LSTM використовуються для вивчення даних як

прямого, так і зворотного часових рядів, при цьому приховані шари використовують дві одиниці, які мають однаковий вхід і підключені до однакового виходу. Один із блоків обробляє прямий часовий ряд, а інший – зворотний часовий ряд.

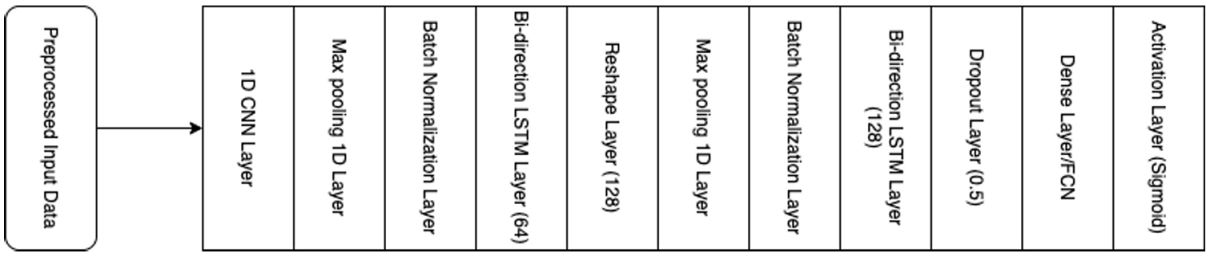

Рисунок 3.2 – Блок-схема використаної моделі

Два шари Bi-LSTM у моделі розташовані таким чином, що вони подвоюють розмір ядра в кожній ітерації. Відповідно до блок-схеми моделі, перший рівень Bi-LSTM починається з 64 одиниць, а наступний — останній рівень Bi-LSTM — 128 одиниць. Між кожним рівнем Bi-LSTM є рівень Max Pooling, щоб відхилити найменш відповідні функції, і рівень пакетної нормалізації, щоб нормалізувати вихідні дані попереднього проміжного рівня, щоб підвищити продуктивність і зменшити час тренувань. Dropout встановлюється для виключення надмірної підгонки, хоча моделі використовують максимальне об’єднання між кожним шаром. Далі йде шар

Виходячи з вибраної архітектури, план експерименту містив наступні ідеї щодо її оптимізації.

- Тестування базової моделі з повним набором шарів
- Модифікація моделі згідно одного з варіантів:

  ервірити як впливає на навчання моделі кількість епох навчання.

  м

  ідключити шар Dropout для перевірки впливу на перенавчання.

  н

  а

Якщо модель перенавчається без цього шару, це може бути індикатором його важливості.

мінити кількість одиниць Bidirectional(LSTM) або заміна їх на звичайні (не бідирекційні) LSTM шари.

3.2 Вихідні дані до тестування

Згідно аналізу у розділі 2 було обрано набір даних CICIDS2017 [25], який є повністю публічним набором що включає безпечні та актуальні типові атаки, які нагадують реальні дані (PCAP). Також він містить результати аналізу мережевого трафіку за допомогою CICFlowMeter із позначеними потоками на основі міток часу, IP-адрес джерела та призначення, портів джерела та призначення, протоколів і атак (файли CSV). Для цього набору даних було абстраговано поведінку 25 користувачів на основі протоколів HTTP, HTTPS, FTP, SSH і електронної пошти. Таким чином, набір даних містить змішані типи трафіку, включаючи звичайний застосунки, інфільтрація та інші. Це дозволяє використовувати датасет для різноманітних досліджень в області безпеки. Потрібно зауважити що він не є повністю збалансованим, оскільки кількість звичайного трафіку значно переважає порівняно з атаками. Це відображає реальний розподіл мережевого трафіку, але може створювати виклики при навчанні моделей машинного навчання. Для створення цього набору було використано реальний мережевий трафік, який був зібраний і анотований вручну з метою створення точного та релевантного набору даних. Набір даних зазвичай доступний у формі CSV файлів, які містять різні атрибути трафіку, включаючи IP-адреси, порти, протоколи, тривалість з'єднання, кількість пакетів, обсяг даних тощо.

У наборі представлено понад 50 ГБ "сирих" даних у форматі PCAP і включає 8 попередньо оброблених файлів, що містять розмічені сесії з виділеними ознаками в різні дні спостереження. У табл. 3.1 описано які типи даних представлені у кожному CSV файлі.

Таблиця 3.1 – Характеристики файлів вхідних даних

№	Назва файлу	Атаки
1	Monday-WorkingHours.pcap_ISCX.csv	Benign (Звичайний трафік)
2	Tuesday-WorkingHours.pcap_ISCX.csv	Benign, FTP-Patator, SSH-Patator
3	Wednesday-workingHours.pcap_ISCX.csv	Benign, DoS GoldenEye, DoS Hulk, DoS Slowhttptest, DoS slowloris, Heartbleed
4	Thursday-WorkingHours-Morning-WebAttacks.pcap_ISCX.csv	Benign, Web Attack – Brute Force, Web Attack – Sql Injection, Web Attack – XSS
5	Thursday-WorkingHours-Afternoon-Infiltration.pcap_ISCX.csv	Benign, Infiltration
6	Friday-WorkingHours-Morning.pcap_ISCX.csv	Benign, Bot
7	Friday-WorkingHours-Afternoon-PortScan.pcap_ISCX.csv	Benign, PortScan

У нашому дослідженні використаємо Thursday-WorkingHours-Morning-WebAttacks.pcap_ISCX.csv. У табл. 3.2 представлені типи та кількість записів наявних у цьому наборі даних.

Таблиця 3.2 – Типи та кількість записів

№	Тип запису	Кількість записів
1	BENIGN	168187
2	Brute Force	1580
3	SQl injection	22
4	XSS	653

Щоб у подальшому використовувати цей набір даних потрібно збалансувати його. Для усунення дисбалансу класів підійде метод випадкового семплювання (undersampling), що полягає у видаленні випадково обраних екземплярів класу "BENIGN". Цільовим співвідношення кількості екземплярів " BENIGN " та екземплярів які відповідають типам атак встановлено 70%/30%.

У наборі даних представлено класи BENIGN, Brut Force, SQl injection,

- BENIGN – звичайний, нешкідливий мережевий трафік. Він використовується для порівняння та визначення аномалій або атак у даних
- Brute Force – Спроба вгадати паролі або ключі шляхом автоматичного перебору великої кількості комбінацій
- SQl injection– Атака, що включає вставку шкідливого SQL коду через веб-форми або URL-запити, що може призвести до несанкціонованого доступу до баз даних.
- XSS– Атака, яка включає вставку шкідливих скриптів у веб-сторінки, які переглядаються іншими користувачами.

Основні етапи проектування моделі

Основні етапи створення та тестування моделі містять наступне:

Попередня обробка даних

- Оцінювання важливості функції
- Аналіз обраних ознак
- Вибір гіперпараметрів
- Тренування базової моделі й оцінювання результатів
- Заміна параметрів моделі, тренування та оцінювання результатів
- Порівняння моделей.

Попередня обробка даних

На цьому етапі, завантажений набір даних було перевірено на цілісність, збалансованість та наявність пропущених значень.

Функція для перевірки унікальних міток представлена на рис. 3.3.

```
df['Label'].unique()
```

```
array(['BENIGN', 'Web Attack - Brute Force', 'Web Attack - XSS',
       'Web Attack - Sql Injection', nan], dtype=object)
```

```
df['Label'].value_counts()
```

```
BENIGN                        168186
Web Attack - Brute Force        1507
Web Attack - XSS                 652
Web Attack - Sql Injection        21
Name: Label, dtype: int64
```

Рисунок 3.3 – Перевірка міток

На наступному кроці було перевірено тип значень у наборі даних.

Оскільки стовпці «Flow Bytes/s» і «Flow Packets/s» мали нечислові значення, вони були замінені (рис. 3.4).

```
[ ]  df = df.drop(df[pd.isnull(df['Flow ID'])].index)
     df.shape
```

```
     (170366, 84)
```

```
[ ]  df.replace('Infinity', -1, inplace=True)
     df[["Flow Bytes/s", "Flow Packets/s"]] = df[["Flow Bytes/s", "Flow Packets/s"]].apply(pd.to_numeric)
```

Замінемо значення NaN і значення нескінченності на -1

```
[ ]  df.replace([np.inf, -np.inf, np.nan], -1, inplace=True)
```

Перетворюємо рядкові символи на числа, використовуючи LabelEncoder

```
[ ]  string_features = list(df.select_dtypes(include=['object']).columns)
     string_features.remove('Label')
     string_features
```

```
     ['Flow ID', 'Source IP', 'Destination IP', 'Timestamp']
```

Рисунок 3.4 – Код і результат перетворення відсутніх та нечислових значень

При оцінці розподілу міток виявилося, що з 458968 записів було багато порожніх записів ("BENIGN" - доброякісний фоновий трафік). Такі записи було видалено оскільки контрольоване середовище начання потребує данні з мітками.

Наступним кроком було перевірка на збалансованість (рис. 3.5)

```
benign_total = len(df[df['Label'] == "BENIGN"])
benign_total
```

```
168186
```

```
attack_total = len(df[df['Label'] != "BENIGN"])
attack_total
```

```
2180
```

```
df.to_csv("web_attacks_unbalanced.csv", index=False)
df['Label'].value_counts()
```

```
BENIGN                         168186
Web Attack - Brute Force         1507
Web Attack - XSS                  652
Web Attack - Sql Injection         21
Name: Label, dtype: int64
```

Рисунок 3.5 – Перевірка кількості парамертів в наборах

Як видно з рис. 3.5., набір даних незбалансований - загальна кількість записів = 170366, доброякісний фоновий трафік = 168186, записів з атаками набагато менше: 1507 + 652 + 21 = 2180. Для того щоб зробити набір більш збалансованим, був використаний підхід недостатньої вибірки, - було видалено більшість "BENIGN" записів.

Таким чином, сформовано збалансований набір даних безпечних даних (2180 / 30 * 70 ~ = 5087 записів.

В отриманому наборі даних було встановлено 2 мітки: 0 ("BENIGN") і 1 (attack) (рис. 3.6).

```
[ ]  df = pd.read_csv('web_attacks_balanced.csv')
```

```
Label колонка закодована як: "BENIGN" = 0, attack = 1.
```

```
[ ]  df['Label'] = df['Label'].apply(lambda x: 0 if x == 'BENIGN' else 1)
```

Рисунок 3.6 – Встановлення значень міток до збалансованого набору

Підготовка нейронної мережі

Далі проводилося визначення та компіляція нейронної мережі. Як зазначено вище, в якості базової моделі було використано послідовну модель із шарами CNN і BiLSTM (рис. 3.7).

```
batch_size = 32
model = Sequential()
model.add(Convolution1D(64, kernel_size=32, padding="same",
          activation="relu", input_shape=(76, 1)))
model.add(MaxPooling1D(pool_size=(5)))
model.add(BatchNormalization())
model.add(Bidirectional(LSTM(64, return_sequences=False)))
model.add(Reshape((128, 1), input_shape=(128, )))

model.add(MaxPooling1D(pool_size=(5)))
model.add(BatchNormalization())
model.add(Bidirectional(LSTM(128, return_sequences=False)))

model.add(Dropout(0.5))
model.add(Dense(4))
model.add(Activation('softmax'))
model.compile(loss='categorical_crossentropy',
              optimizer='adam', metrics=['accuracy'])
```

```
model.summary(line_length=100)
```

Рисунок 3.7 – Код структури базової моделі

Перед шарами BiLSTM застосовано пакетну нормалізацію. Перед фінальним шаром наноситься шар відсіву, щоб запобігти переобладнанню. Категоріальна крос-ентропія обрана як базова функція втрат через багатокласову класифікацію. Щоб отримати результат класифікації, функція активації softmax застосовується до виходу останнього, повністю підключеного шару. Результат представлено на рис. 3.8.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv1d (Conv1D)              (None, 76, 64)            2112
_________________________________________________________________
max_pooling1d (MaxPooling1D) (None, 15, 64)            0
_________________________________________________________________
batch_normalization (BatchNormalization) (None, 15, 64)  256
_________________________________________________________________
bidirectional (Bidirectional) (None, 128)              66048
_________________________________________________________________
reshape (Reshape)            (None, 128, 1)            0
_________________________________________________________________
max_pooling1d_1 (MaxPooling1D) (None, 25, 1)           0
_________________________________________________________________
batch_normalization_1 (BatchNormalization) (None, 25, 1)  4
_________________________________________________________________
bidirectional_1 (Bidirectional) (None, 256)            133120
_________________________________________________________________
dropout (Dropout)            (None, 256)               0
_________________________________________________________________
dense (Dense)                (None, 4)                 1028
_________________________________________________________________
activation (Activation)      (None, 4)                 0
=================================================================
Total params: 202,568
Trainable params: 202,438
Non-trainable params: 130
```

Рисунок 3.8 – Архітектура базової моделі

Тренування нейронної мережі

Етап навчання призначений для використання з різними версіями Tensorflow. Також в базовій моделі можна вибрати, чи має проводитись навчання за допомогою CPU чи GPU. Цей крок може зайняти деякий час, 13 хвилин або більше залежно від продуктивності процесора (швидше звичайно з графічним процесором).

У першому експерименті було використано перехресну перевірку показники оцінки (рис. 3.9).

```
def print_metrics(y_eval: np.ndarray, y_pred: np.ndarray, average: str = 'binary') -> List[float]:
    accuracy = metrics.accuracy_score(y_eval, y_pred)
    precision = metrics.precision_score(y_eval, y_pred, average=average)
    recall = metrics.recall_score(y_eval, y_pred, average=average)
    f1 = metrics.f1_score(y_eval, y_pred, average=average)

    print('Accuracy =', accuracy)
    print('Precision =', precision)
    print('Recall =', recall)
    print('F1 =', f1)

    return [accuracy, precision, recall, f1]
```

Рисунок 3.9 – Функція для отримання показників оцінки якості моделі

Результати тестування моделей

Тест 1. Зміна параметру Bidirectional(LSTM)

В даному тесті було вирішено дослідити вплив зміни розміру нейронів у шарі Bidirectional LSTM на ефективність моделі. Базова модель, яка має Bidirectional LSTM шар з 64 нейронами, уже показала певні результати. Для глибшого аналізу, ми вирішили варіювати кількість нейронів у цьому шарі, встановлюючи їх кількість відповідно до 32, 128, 256 та 512. Це дозволить нам оцінити, як зміна кількості нейронів впливає на загальну продуктивність моделі, зокрема на її точність та здатність до узагальнення. Результати цих тестів будуть зібрані та проаналізовані у Табл

Таблиця 3.3 – Результати першого тесту

Параметри		Базова			
Точність класифікації					
Невірно класифікувано					
Чіткість					
Повторюваність					

Результати тесту зі зміною кількості нейронів у шарі Bidirectional LSTM виявились досить інформативними. Спостереження показують, що збільшення кількості нейронів до 256 сприяло значному підвищенню точності класифікації. Це можна пояснити тим, що збільшення кількості нейронів до цього рівня дозволило моделі ефективніше вловлювати та аналізувати складні шаблони в даних, що призвело до зростання її точності.

Водночас, подальше збільшення кількості нейронів до 512 спричинило зниження точності, що може бути ознакою перенавчання моделі. Перенавчання відбувається, коли модель занадто детально адаптується до тренувального набору даних, включаючи нерелевантні шуми та особливості, що не характерні для загального розподілу даних. Це призводить до зниження здатності моделі до узагальнення та її ефективності на нових даних.

Щодо покращення показників при зменшенні кількості нейронів до 32, це може бути пов'язано з тим, що компактніша модель з меншою кількістю параметрів ефективніше уникає перенавчання та краще узагальнює дані. Це підкреслює важливість знаходження оптимальної кількості нейронів для даної архітектури та набору даних.

Тест 2. Заміна кількості епох навчання

Наступний тест зміна кількості епох навчання.

Таблиця 3.4 – результати 2-го тесту

Параметри	5 епох	епох Базова	20 епох	40 епох
Точність класифікації				
Невірно класифіковано				
Чіткість				
Повторюваність			0.8101	

Згідно із результатів наведених у табл 3.4 модель показує найкращі результати при 20 епохах, що може свідчити про те, що це оптимальна кількість епох для навчання даної моделі з урахуванням розміру та складності набору даних. Зі збільшенням кількості епох після цього показника якість моделі починає знижуватися, що може бути ознакою перенавчання. Також із зменшенням кількості епох замітні кращі результати, але при тестуванні даної моделі були виявлені помилки та замітний процес недонавчання. Модель не достатньо добре класифікує дані на тестових даних.

Тест 3. Заміна кількості епох навчання

Зміна оптимізатора у функції втрат, для знаходження кращого оптимізатора для знаходження найоптимальнішого варіанту.

(SGD) відрізняється від класичного градієнтного спуску тим, що оновлює параметри моделі, використовуючи лише один або кілька тренувальних прикладів за кожен крок, що робить процес оновлення швидшим та з вищою ймовірністю виходу з локальних мінімумів, дозволяючи потенційно знайти більш оптимальне рішення..

: Це поєднання RMSprop та моментуму Нестерова, яке може прискорити збіжність в деяких випадках.

Adam автоматично коригує швидкість навчання для кожного параметра. Це означає, що параметри з великими градієнтами матимуть

меншу швидкість навчання, а параметри з малими градієнтами матимуть більшу швидкість навчання.

Таблиця 3.5 – результати 3-го тесту

Параметри		adam Базова	
Точність класифікації			
Невірно класифіковано			
Чіткість			
Повторюваність			0.8101
Час, с			

Оптимізатор SGD показав кращі результаті ніж у базовій моделі із adam проте збільшився і час навчання. Найкращі ж результати продемонстрував Nadam із найвищою точністю класифікації даних, проте із найдовшим часом виконання.

В результаті тестів у табл 3.5 було визначено що для конкретної архітектури найкращим оптимізатором функції втрат є SGD. Показавши найоптимальніші результати при тестуванні моделі.

Тест 4. Тестування моделі із увімкненим та вимкненим шаром

Таблиця 3.6 – результати 4-го тесту

Параметри	Без шару	З шаром Dropout Базова	
Точність класифікації			
Невірно класифіковано			
Чіткість			
Повторюваність			

Було вирішено ще протестувати модуль при 256 та вимкненому параметрі, результат показав зменшення точності. Це говорить що цей параметр може бути корисним при більшому значені LSTM, так як на менших значеннях модель не так часто перенавчається і цей параметр може знижувати результативність, а при збільшені важливо його вмикати та змінювати його значення, щоб модель не так часто перенавчалась.

Після проведення попередніх тестів було виявленно як саме обрані параметри впливають на дану модель, згідно із цих даних, можна налаштувати модель так щоб вона видавала більш точний результат, та знайти оптимальне рішення для даної моделі

Тест 5. Останній тест підбір накращих параметрів.

Табл 3.7 – результати 5-го тестування

	Найкраща точність	Базова	Найоптимальніша
Точність класифікації			
Невірно класифіковано			
Чіткість			
Повторюваність			
Час, с			

У Моделі із найкращою точністю змінені такі параметри:

- Bidirectional(LSTM) параметр змінено на 256
- Оптимізатор використано NADAM
- Dropout параметр змінено на 0.9 від базового 0.5, так як у цій моделі велика кільсть нейронів та збільшена у двічі кількість епох, щоб запобігти перенавчанню.
- Кількість епох навчання моделі 20.

Оптимізована модель була сконфігурована так щоб покращити результати та щоб навчання виконувалось за оптимальний час. Це дозволяє швидше реагувати ціною невилокої втрати точності. У ній змінено наступні параметри:

- Bidirectional(LSTM) параметр змінено на 128
- Оптимізатор SGD
- Dropout параметр вимкнено так як не спорстерігалось перенавчяння при даних параметрах моделі.
- Кількість епох навчання моделі 15.

Висновки до третього розділу

Завершуючи даний розділ, можна відзначити, що використання гібридної нейронної мережі CNN-BiLSTM в системі виявлення вторгнень (IDS) є перспективним напрямком. Ця модель поєднує переваги здатності CNN виявляти шаблони у мережевому трафіку та аналізу послідовностей, який виконує BiLSTM. Важливими аспектами є вибір правильного набору даних для тестування та збалансованість цього набору для точного аналізу. Тому було обрано останній доступний набір даних CICIDS2017, та в ньому обрано набір даних в кому мітстився трафік із веб атаками. Також довелося нормалізувати дані перед використаннями

Проведено ряд тестів моделі CNN-BiLSTM на наборі даних ключову роль у її продуктивності. Наприклад, кількість нейронів у BiLSTM, тривалість навчання та вибір оптимізатора можуть суттєво впливати на точність класифікації аномалій в мережевому трафіку. Додавання шару Dropout виявилося ефективним для запобігання перенавчанню при великій кількості нейронів.

В цілому, вдалося знайти набір параметрів, який показав найкращі результати з точністю 92.81%, порівняно з базовою моделлю, яка при стандартних налаштуваннях мала точність лише 77.60%. Це є значним покращенням, хоча час навчання мережі збільшився. Збільшення часу обумовлено вибором параметрів, які ускладнили мережу, дозволивши їй краще класифікувати загрози серед чистого трафіку. Одним із рішень цієї проблеми може бути використання більш потужних апаратних компонентів або використання ресурсів хмарного обчислення.

Крім того, було сформовано 87.32% набір налаштувань моделі, який виявився найоптимальнішим з точки зору балансу точності та часу навчання. Звісно, у складних мережевих середовищах компаній буде доцільно використовувати модель з найвищою точністю, аби мінімізувати ризик хибнопозитивних та хибнонегативних спрацьовувань. Однак, потрібно враховувати, що чим складніша модель машинного навчання, тим більше ресурсів потрібно для її обчислень. Тому підбір параметрів моделі, як описано в даному розділі, є дуже важливим етапом при інтеграції машинного навчання в IPS та IDS системи.

ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1 Охорона праці

Інтеграція штучного інтелекту для ефективного виявлення аномалій в системах IPS та IDS вимагає значного обсягу робочого часу від фахівця за комп'ютером. Тому важливо дотримуватись норм робочого часу для забезпечення ефективної праці працівників.

Згідно з нормативними актами про охорону праці (НПАОП 0.00-7.15-18) є такі вимоги безпеки до робочих місць працівників з електронними пристроями:

– Площа, відведена на одне робоче місце має становити не менше 6 кв.м., а об’єм – не менше 20 куб.м.[27]

– Організація робочого місця повинна гарантувати підтримку оптимальної робочої позиції, що дозволяє працівникові виконувати завдання з мінімальним навантаженням на тіло і запобігає виникненню перевтоми під час та після завершення робочого процесу.

– Для забезпечення безпеки та захисту здоров’я працівників усе випромінювання від екранних пристроїв має бути зведене до гранично допустимого рівня (вплив на людину факторів довкілля - шуму, вібрації, забруднювачів, температури тощо, який не спричиняє соматичних або психічних розладів, а також змін стану здоров’я, працездатності, поведінки,що виходять за межі пристосувальних реакцій) з погляду безпеки та охорони здоров’я працівників.[28]

– Організація робочого місця працівника з екранними пристроями має забезпечувати відповідність усіх елементів робочого місця та їх розташування ергономічним, антропологічним, психофізіологічним вимогам, а також характеру виконуваних робіт.[29]

– Освітлення робочого місця працівника з екранними пристроями має створювати відповідний контраст між екраном і навколишнім

середовищем (з урахуванням виду роботи) та відповідати вимогам ДСанПІН 3.3.2.007-98.

– Мікроклімат приміщень з робочими місцями працівників з екранними пристроями має підтримуватись на постійному рівні та відповідати вимогам Санітарних норм мікроклімату виробничих приміщень ДСН 3.3.6.042-99, затверджених постановою Головного державного санітарного лікаря України від 01 грудня 1999 року No 42.[30]

Вимоги щодо розміщення ІТС

Приміщення, в яких планується установка та подальша робота з комп'ютером, повинні відповідати проектній документації будинку, погодженій з уповноваженими державними органами. Крім того, роботодавець повинен враховувати санітарні нормативи освітлення, вимоги до параметрів мікроклімату (температура, відносна вологість), ступеня і сили вібрації, звукового шуму і вогнестійкості приміщення, а також характеристики електромагнітного, ультрафіолетового та інфрачервоного полів. Робочі місця, обладнані персональними комп'ютерами, заборонено облаштовувати у підвальних або цокольних приміщеннях будівель.[31] При обладнанні приміщень забороняється використання полімерних матеріалів, що виділяють шкідливі хімічні речовини.

Природне і штучне освітлення

Згідно документу ДБН В.2.5-28:2018 "Природне і штучне освітлення" приміщення з постійним перебуванням людей повинні мати природне освітлення. Природне освітлення поділяється на бокове, верхнє і комбіноване. Що до штучного освітлення воно поділяється на робоче, аварійне, охоронне і чергове.[32]

Для загального штучного освітлення доцільно використовувати розрядні та світлодіодні джерела світла, які за однакової потужності з тепловими джерелами мають більшу світлову віддачу та більший термін експлуатації.

Види інструктажів з охорони праці

Працівники, під час прийняття на роботу та періодично, повинні проходити на підприємстві інструктажі з питань охорони праці, надання першої медичної допомоги потерпілим від нещасних випадків, а також з правил поведінки та дій при виникненні аварійних ситуацій, пожеж і стихійних лих.

За характером і часом проведення інструктажі з питань охорони праці (далі – інструктажі) поділяються на вступний, первинний, повторний, позаплановий та цільовий.)

Вступний інструктаж Проводиться:

– з усіма працівниками, які приймаються на постійну або тимчасову роботу, незалежно від їх освіти, стажу роботи та посади;

– з працівниками інших організацій, які прибули до організації і беруть безпосередню участь у робочому процесі або виконують інші роботи для підприємства;

Вступний інструктаж проводиться спеціалістом служби охорони праці або іншим фахівцем відповідно до наказу (розпорядження) по організації, який в установленому типовим положенням порядку пройшов навчання і перевірку знань з питань охорони праці.[33] Первинний інструктаж.

Первинний інструктаж проводиться до початку роботи безпосередньо на робочому місці з працівником:

– новоприйнятим (постійно чи тимчасово) до організації або до фізичної особи, яка використовує найману працю;

– який переводиться з одного структурного підрозділу організації до іншого;

Повторний інструктаж

Повторний інструктаж на робочому місці індивідуально з окремим працівником або групою працівників, які виконують однотипні роботи, за обсягом і змістом переліку питань первинного інструктажу.

Повторний інструктаж проводиться в терміни, визначені нормативно-правовими актами з охорони праці, які діють у галузі, або роботодавцем (фізичною особою, яка використовує найману працю) з урахуванням конкретних умов праці, але не рідше:

– на роботах з підвищеною небезпекою - 1 раз на 3 місяці;

– для решти робіт - 1 раз на 6 місяців. Позаплановий інструктаж.

Позаплановий інструктаж проводиться з працівниками на робочому місці або в кабінеті охорони праці:

– при введенні в дію нових або переглянутих нормативно-правових актів з охорони праці, а також при внесенні змін та доповнень до них;

– при порушеннях працівниками вимог нормативно-правових актів з охорони;

– праці, що призвели до травм, аварій, пожеж тощо;

– при перерві в роботі виконавця робіт більш ніж на 30 календарних днів для робіт з підвищеною небезпекою, а для решти робіт - понад 60 днів.

Цільовий інструктаж.

Цільовий інструктаж проводиться з працівниками:

– при ліквідації аварії або стихійного лиха;

– при проведенні робіт, на які відповідно до законодавства оформлюються наряд-допуск, наказ або розпорядження.

Цільовий інструктаж проводиться індивідуально з окремим працівником або з групою працівників. Обсяг і зміст цільового інструктажу визначаються залежно від виду робіт, що виконуватимуться.

4.2 Безпека в надзвичайних ситуаціях

Здоров'я людини грунтується на основі генетичних факторів, способу життя та екологічних умов. Однак певною мірою воно залежить також від свідомого ставлення людини до себе та оточуючого середовища. Здоров'я людини — стан повного соціально-біологічного комфорту коли функція всіх органів і систем організму виважені з природним і соціальним середовищем, відсутні будь-які хвилювання, хворобливі стани та фізичні дефекти. Критерій здоров'я визначається комплексом показників. Однак за найзагальнішими рисами здоров'я індивідуума можна визначити як природний стан організму, що хактеризується повною зрівноваженістю будь-яких виражених хворобливих змін. Слід пам'ятати, що здоров'я залежить від багатьох факторів які об'єднуються в одне інтегральне поняття —здоровий спосіб життя. Його метою є навчити людину розумно ставитися до свого здоров'я, фізичної та психічної культури, загартовувати свій організм, вміло організовувати працю і відпочинок.

До основних складових здорового способу життя належать декілька основних чинників.

Спосіб життя має велике значення для здоров'я людини і складається з чотирьох категорій:

- Економічної (рівень життя).
- Соціологічної (якість життя).
- Соціально-психологічної.
- Соціально-економічної.

Отже, до способу життя людини належать: активна участь людини в процесі формування умов життя, її адекватна реакція на зміну умов навколишнього середовища, а також праця, побут, задоволення матеріальних і духовних потреб у суспільному житті, норми і правила поведінки.

Слід пам'ятати, що людина — суб'єкт і одночасно — головний результат своєї діяльності. Культура з цієї точки зору — це самосвідоме ставлення до самого себе. Однак люди дуже часто нехтують своїм здоров'ям, ведуть неправильний спосіб життя, не дотримуються режиму переїдають, курять. Тому для здоров'я потрібні знання, які увійшли б у повсякденну звичку людини.

Не завжди в житті людини здоров'я займає перше місце порівняно з речами та іншими матеріальними благами. У результаті це призводить до шкоди не лише своєму здоров'ю, а й здоров'ю майбутніх поколінь. Отже, здоров'я повинно займати перше місце в ієрархії потреб людини.

На превеликий жаль, ціну здоров'я більшість людей усвідомлює лише тоді, коли воно значно похитнулось. Лише тоді виникає прагнення вилікувати захворювання, стати здоровим.

Нерозумне і довге випробовування стійкості свого організму нездоровим способом життя (алкоголь, нікотин). Тільки через певний час спрацьовують зворотні зв'язки у людини, коли вона полишає шкідливі звички, проте, часто запізно.

Джерелом навичок з цього питання є, передусім, приклад батьків, допомагає також і санітарна освіта. Важливим фактором, що визначає реакцію людини на екстремальну ситуацію, є її психофізичні якості та загальний стан. Вони проявляються через чутливість людини до виявлення сигналів небезпеки перед реакцією на них. Показники, які зумовлюють можливості людини виявити небезпечну ситуацію та адекватно відреагувати на неї, залежать від її індивідуальних особливостей, зокрема від її нервової системи. На поведінку людини у небезпечній ситуації впливає й її психічний та фізичний стан.

Відомо, що 80 % більшості хвороб мають психосоматичний характер, тобто значною мірою залежить від стану душі людини, який визначає її безпечну поведінку.

Сучасна людина зустрічається з багатьма факторами ризику, що негативно впливають на стан й нервової та серцево-судинної систем, знижує опірність організму. При цьому виникає стресова реакція організму. Так, наприклад, психічна травма, отримана внаслідок конфлікту, виводить людину з нормального психічного стану, що може призвести до суттєвих змін у виконанні професійних функцій і загального функціонального стану. У перекладі «стрес» означає «напруження», тобто відповідь організму на поставлену перед ним проблему.

Велике значення для розвитку стресового стану має поведінка в екстремальних умовах (аварія, кримінальна ситуація, стихійне лихо). Неправильна поведінка у таких ситуаціях найчастіше є причиною шкідливих наслідків стресу. Вона зумовлює результат стресу більше, ніж фактори зовнішнього середовища. У цих випадках стрес може виявитись у вигляді паніки, суєти, істерики.

Це захисна реакція організму на зовнішні надзвичайні подразники і ситуації, тривалі негативні емоції. Він супроводжується підвищеним серцебиттям, виснаженням і зривом адаптаційних і імунних систем організму та іншими змінами. До певної межі стрес сприяє вирішенню людиною певних завищених завдань і навантажень. Однак, у разі перевищення цієї межі в організмі людини виникають порушення механізмів саморегуляції, відбувається погіршення трудової діяльності і стануться зриви, які призводять до виникнення небезпечних ситуацій. При стресових ситуаціях різко підвищується вміст адреналіну у крові, посилюється робота серця, звужуються кровоносні судини, підвищується температура тіла і рівень глюкози у крові. У результаті в організмі виникають фізіологічні порушення, розлади нервової, серцево-судинної систем та ін. До цих розладів належать нервовість, роздратованість, тривога, агресивність, втома, загострення хворобливих станів.

Тривала стресова ситуація призводить до багатьох психосоматичних захворювань: психозів, неврозів, захворювань мозку, серцево-судинних захворювань, інфаркту, гіпертонічної хвороби, шлунково-кишкових захворювань, зниження імунітету, онкологічних захворювань.

4.3 Висновок до четвертого розділу

Таким чином, у результаті аналізу вимог щодо охорони праці користувачів комп’ютерів, визначено особливості організації робочих місць, вимог з електробезпеки, природного та штучного освітлення для ефективної і безпечної роботи.

Також розглянуто питання здорового способу життя та його вплив на професійну діяльність, структури системи БЖД, елементів теорії, що відповідають моделі безпеки життєдіяльності.

ВИСНОВОК

В результаті роботи всі поставлені завдання було виконано. А саме: Здійснено аналіз поточних труднощів та проблем в системах IPS/IDS і визначено, що інтеграція штучного інтелекту в системи IPS/IDS стає критично важливим елементом у сучасних стратегіях кібербезпеки. Однією з основних проблем традиційних IPS/IDS є їхня схильність до генерації хибнопозитивних і хибнонегативних результатів. Це створює багато зайвих сповіщень, що змушує витрачати зайві людино-години на розгляд цих інцидентів. У випадку IPS систем, можуть бути задіяні дії та заблоковані хости або зв'язки, які не несуть зловмисний характер та є важливими для коректної роботи мережі. Цю проблему може вирішити штучний інтелект, оскільки він може краще виявляти шкідливий трафік як шкідливий, а нормальний - як нормальний.

Було виконано порівняння існуючих моделей машинного навчання як за аналізом літератури так і, частково, експериментально. За результатами експерименту було вибрано оптимальну конфігурацію моделі, враховуючи її здатність до виявлення мережевих інтрузій у конкретному контексті.

Для оцінки ефективності МН було використано метрики оцінювання якості роботи ids системи. Точність вимірює частку загальних прогнозів (як позитивних, так і негативних), які модель визначила правильно. Чіткість оцінює правильність, досягнуту в позитивному класі. Відтворення це відношення правильно передбачених позитивних спостережень до всіх фактичних позитивних спостережень. Показник F1 - це середнє гармонійне значення точності та відгуку, що забезпечує баланс між ними.

Згідно аналізу існуючих моделей навчання було обрано CNN-використовуваних моделей в цій темі, що дає нам можливість краще класифікувати типи подій у мережі. Саме така багатокласова модель дає дає

можливість налаштувати автоматичне реагування на конкретні типи атак, наприклад атаки класифіковані як метод грубої сили, можуть заблокувати обліковий запис, замість повного блокування з’єднання.

Обрано набір даних CICIDS2017. А саме Thursday-WorkingHours-Morning-WebAttacks.pcap_ISCX.csv. У цьому наборі дані класифіковані наступним чином: BENIG (нормальний трафік), Web Attack - Brute Force, Web Attack - Sql Injection, Web Attack – XSS. Для збалансування набору даних було використано метод випадкового семплювання, для досягнення співвідношення 70% "BENIGN" до 30% атак.

Проведено ряд тестів моделі CNN-BiLST. Виявлено які шари моделі впливають на покращення її навчання та точності класифікації атак.

Сконфігуровано модель яка видала найкращі результати за оптимальніший час має наступні змінені параметри:

- Bidirectional(LSTM) параметр змінено на 128
- Оптимізатор SGD
- Dropout параметр вимкнено так як не спорстерігалось перенавчяння при даних параметрах моделі.
- Кількість епох навчання моделі 15.

Такі параметри ускладнили модель, що дало їй змогу точніше класифікувати атаки на 14%, ніж модель із базовими параметрами але не ускладнили настільки щоб модель перенавчалась.

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Anderson P. Computer Security Threat Monitoring and Surveillance. [Електронний ресурс]. – Режим доступу: https://csrc.nist.gov/csrc/media/publications/conference-paper/1998/10/08/proceedings-of-the-21st-nissc-1998/documents/early-cs-papers/ande80.pdf, 18.12.2023.
2. ThreatStack. The History of Intrusion Detection Systems (IDS)—Part 1. [Електронний ресурс]. – Режим доступу: https://www.threatstack.com/blog/the-history-of-intrusion-detection-systems-ids-part-1,18.12.2023.
3. IBM Cloud Education. Supervised Learning. [Електронний ресурс]. – Режим доступу: www: https://www.ibm.com/cloud/learn/supervised-learning, 18.12.2023.
4. Thapa S., Mailewa A. The role of intrusion detection/prevention systems in modern computer networks. In Conference: Midwest Instruction and Computing Symposiu (MICS). 2020. Vol. 53. pp. 1-14. URL: https://www.researchgate.net/publication/340581541_THE_ROLE_OF_INTRUSION_DETECTIONPREVENTION_SYSTEMS_IN_MODERN_COMPUTER_NETWORKS_A_REVIEW
5. Vijay S., Gharakheili H.H., Vishwanath A., Boreli R., Mehani O. Network-level security and privacy control for smart-home IoT devices. In 2015 IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), IEEE. 2015. pp. 163-167. URL: https://www.researchgate.net/publication/281275810_Network-Level_Security_and_Privacy_Control_for_Smart-Home_IoT_Devices
6. Adeleke O. (2020). Intrusion detection: issues, problems and solutions. In 3rd International Conference on Infдormation and Computer Technologies (ICICT). IEEE. 2020 pp. 397-402. URL: https://www.researchgate.net/publication/341400652_Intrusion_Detection_Issues_Problems_and_Solutions
7. Seldon Machine Learning Regression Explained. [Електронний ресурс]. – Режим доступу: https://www.seldon.io/machine-learning-regression-explained, 18.12.2023.
8. Terence S. All Machine Learning Models Explained in 6 Minutes. [Електронний ресурс]. – Режим доступу: https://www.ibm.com/cloud/learn/unsupervised-learning, 18.12.2023.
9. IBM Cloud Education. Unsupervised Learning. [Електронний ресурс]. – Режим доступу: https://www.ibm.com/cloud/learn/unsupervised-learning, 18.12.2023.

10. Sabahi F., Movaghar A. Intrusion Detection: A Survey. In Proceedings of the Third International Conference on Systems and Networks Communications, Sliema, Malta. 2008. pp. 23–26. URL: https://www.researchgate.net/publication/232623012_Intrusion_Detection_A_Survey
11. Mahbod Tavallaee, Ebrahim Bagheri, Wei Lu, Ali A. Ghorbani. A detailed analysis of the KDD CUP 99 data set. IEEE. 2009. pp 1-6. URL: http://dx.doi.org/10.1109/CISDA.2009.5356528
12. Vanin P., Newe T., Dhirani L.L., O'Connell E., O'Shea D., Lee B. and Rao M. A Study of Network Intrusion Detection Systems Using Artificial Intelligence/Machine Learning. Applied Sciences, 2022, p.11752. URL:https://doi.org/10.3390/app122211752..
13. Bamhdi A.M., Abrar I., Masoodi F. An ensemble based approach for effective intrusion detection using majority voting. Telkomnika (Telecommunication Computing Electronics and Control), 2021. pp 664-671. URL: https://www.researchgate.net/publication/348620765_An_ensemble_based_approach_for_effective_intrusion_detection_using_majority_voting
14. Sun P., Liu P., Li Q., Liu C., Lu X., Hao R., Chen J. DL-IDS: Extracting features using CNN-LSTM hybrid network for intrusion detection system. Security and communication networks, 2022. pp 1-11. URL: https://www.researchgate.net/publication/343965128_DL-IDS_Extracting_features_using_CNN-LSTM_hybrid_network_for_intrusion_detection_system
15. Vinayakumar R., Alazab M., Soman K.P., Poornachandran P., Venkatraman S. Robust intelligent malware detection using deep learning. IEEE. 2019. pp 46717-46738. URL: https://www.researchgate.net/publication/332200619_Robust_Intelligent_Malware_Detection_Using_Deep_Learning
16. Wisanwanichthan, T., & Thammawichai, M. A double-layered hybrid approach for network intrusion detection system using combined naive bayes and SVM. IEEE. 2021, 138432-138450. URL: https://www.researchgate.net/publication/355125907_A_Double-Layered_Hybrid_Approach_for_Network_Intrusion_Detection_System_Using_Combined_Naive_Bayes_and_SVM
17. Mari, A.-G., Zinca, D. and Dobrota, V. Development of a Machine-Learning Intrusion Detection System and Testing of Its Performance Using a Generative Adversarial Network. Sensors. 2023. p.1315. URL: https://doi.org/10.3390/s23031315.
18. Lirim Sharafaldin, I. Lashkari, A.H.; Ghorbani, A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization. Canadian Institute for Cybersecurity (CIC). 2018. pp. 108–116. URL. https://www.scitepress.org/papers/2018/66398/66398.pdf

19. Lin Chen, Kuang X., Xu A., Suo S., Yang Y. A Novel Network Intrusion Detection System Based on CNN. In Proceedings of the Eighth International Conference on Advanced Cloud and Big Data (CBD). 2020; pp. 243–247. URL: https://www.researchgate.net/publication/352398091_A_Novel_Network_Intrusion_Detection_System_Based_on_CNN
20. Yu Y., Bian N. An Intrusion Detection Method Using Few-Shot Learning. IEEE. 2020. pp 49730–49740. URL: https://www.researchgate.net/publication/339858151_An_Intrusion_Detection_Method_Using_Few-Shot_Learning
21. Andresini G., Appice A., Mauro N.D., Loglisci C., Malerba D. Multi-Channel Deep Feature Learning for Intrusion Detection. IEEE. 2020. pp 53346–53359. URL: https://www.researchgate.net/publication/339977574_Multi-Channel_Deep_Feature_Learning_for_Intrusion_Detection
22. Elhefnawy R., Abounaser H., Badr A. A Hybrid Nested Genetic-Fuzzy Algorithm Framework for Intrusion Detection and Attacks. IEEE. 2020. pp 98218–98233. URL: https://www.researchgate.net/publication/341558679_A_Hybrid_Nested_Genetic-Fuzzy_Algorithm_Framework_for_Intrusion_Detection_and_Attacks
23. Protic D. Review of KDD Cup '99, NSL-KDD and Kyoto 2006+ Datasets. Vojnoteh. Glas. 2018. 580–596. URL: https://www.researchgate.net/publication/326000849_Review_of_KDD_Cup_%2799_NSL-KDD_and_Kyoto_2006_datasets
24. Sharafaldin I., Lashkari A.H., Ghorbani A. Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization; Canadian Institute for Cybersecurity (CIC). 2018. pp. 108–116. URL: https://www.scitepress.org/papers/2018/66398/66398.pdf
25. www.unb.ca.com . IDS 2017 Datasets. Canadian Institute for Cybersecurity. 2017. URL: Available at: https://www.unb.ca/cic/datasets/ids-2017.html.
26. RFID-технології та магнітні мітки. [Електронний ресурс] – Режим доступу до ресурсу: http://allta.com.ua/what-is-rfid.
27. The Basics of Bluetooth Low Energy (BLE) [Електронний ресурс] – Режим доступу до ресурсу: https://www.novelbits.io/basics-bluetooth-low-energy/.
28. Wi-Fi HaLow (IEEE 802.11ah) [Електронний ресурс] – Режим доступу до ресурсу: https://www.ixbt.com/news/2016/01/05/wi-fi-halow-ieee-802- 11ah.html.
29. Наказ Міністерства соціальної політики України «Про затвердження Вимог щодо безпеки та захисту здоров'я працівників під час роботи з

екранними пристроями [Електронний ресурс] – Режим доступу до ресурсу: https://zakon.rada.gov.ua/laws/show/z0508-18.
30. Закон України «Про охорону праці» [Електронний ресурс] – Режим доступу до ресурсу: https://zakon.rada.gov.ua/laws/show/2694-12.
31. Наказ Міністерства внутрішніх справ України «Про затвердження Правил пожежної безпеки в Україні» [Електронний ресурс] – Режим доступу до ресурсу: https://zakon.rada.gov.ua/laws/show/z0252-15.
32. Державний комітет ядерного регулювання України. Проект від 01.03.2008 р. Консультації щодо підвищення безпеки джерел іонізуючого випромінювання в Україні. Київ, 2008.
33. Білявський Г.О, Бутченко Л.І., Навроцький В.М. Основи екології: Теорія і практикум: Навч. Посібник. Київ, 2002. 352

Додаток А

УДК 004.056
Гуменюк В. Р., Муж Валерій Вікторович, к.ю.н., доцент Кафедри кібербезпеки.
Тернопільський національний технічний університет імені Івана Пулюя, Україна

ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ В IPS ТА IDS СИСТЕМАХ

Humeniuk V. R., V. V. Muzh, Cand. Sc. (Law), Associate Professor of the Department of Cybersecurity.

USING ARTIFICIAL INTELLIGENCE IN IPS AND IDS SYSTEMS

В сучасному світі кібербезпеки, системи виявлення та запобігання вторгненням (IPS/IDS) відіграють ключову роль у захисті організацій від широкого спектру кіберзагроз. Однак, зі зростанням складності та хитромудрості кібератак, традиційні IPS/IDS системи стикаються з рядом викликів, що обмежують їхню ефективність. Серед цих викликів:

1. Невідповідність до нових загроз: Традиційні IPS/IDS системи часто базуються на відомих сигнатурах атак, що робить їх менш ефективними проти нових чи модифікованих загроз, які не відповідають існуючим сигнатурам.
2. Високий рівень помилкових позитивних сповіщень: Системи, що базуються на сигнатурах, схильні до великої кількості помилкових сповіщень, що може призводити до "втоми" аналітиків безпеки та ігнорування справжніх загроз.
3. Обмежена масштабованість та гнучкість: Традиційні системи вимагають постійного оновлення сигнатур та правил, що може бути трудомістким та не завжди встигає за швидкістю розвитку загроз.

Впровадження AI і ML у IPS/IDS може допомогти подолати ці виклики. Штучний інтелект може вчитися з даних про вже відомі атаки, підлаштовуючись під нові методи ведення кібервійни. Це забезпечує глибший аналіз поведінкових патернів, дозволяючи виявити загрози, які не можуть бути визначені традиційними методами.Застосування ML може значно зменшити кількість помилкових позитивних сповіщень, підвищуючи точність системи і зменшуючи навантаження на аналітиків. Також, AI та ML можуть забезпечити більшу гнучкість та масштабованість систем безпеки, адаптуючись до змін у кіберзагрозах в реальному часі. Основною мотивацією дослідження є покращення ефективності IPS/IDS систем шляхом інтеграції AI та ML, що дозволить організаціям ефективніше виявляти та реагувати на кіберзагрози, забезпечуючи більш високий рівень безпеки.

Основна мета цього дослідження покликана вивчити, як AI і ML можуть бути інтегровані в існуючі IPS/IDS системи, з метою підвищення їхньої ефективності у виявленні, запобіганні, реагуванні та відновленні після інцидентів безпеки. Також вивчаються методи адаптації цих систем до постійно змінюваних загроз. У роботі розглядаються наступні дослідницькі запитання:

1. Які AI/ML технології та алгоритми можна використовувати в IPS/IDS системах, та які їхні переваги та недоліки?
2. Як можна інтегрувати AI/ML у існуючі системи безпеки, зокрема в IPS/IDS?
3. Як впровадження AI/ML впливає на ефективність систем IPS/IDS у боротьбі з кіберзагрозами?

Література

1. Intrusion Detection System using AI and Machine Learning Algorithm (2017). [online] irjet.net. Available at: https://www.irjet.net/archives/V4/i12/IRJET-V4I12314.pdf.