Міністерство освіти і науки України
Кам’янець-Подільський національний університет імені Івана Огієнка
Фізико-математичний факультет
Кафедра комп'ютерних наук

Кваліфікаційна робота магістра

з теми: **«Розробка та дослідження моделей машинного навчання для виявлення кібератак»**

Виконав: здобувач вищої освіти групи Kn1-M24
спеціальності 122 Комп’ютерні науки
Вельма Максим Русланович

Керівник: **Моцик Ростислав Васильович,**
кандидат пед. наук, доцент

Рецензент: **Сморжевський Юрій Людвігович,**
кандидат пед. наук, доцент

Кам’янець-Подільський – 2025 р.

Зміст

АНОТАЦІЯ

Мета і завдання дослідження

Мета дослідження — розглянути використання методів машинного навчання для захисту від соціальної інженерії та для виявлення DoS/DDoS атак у мережах Інтернету речей (IoT).

Для досягнення мети необхідно виконати такі завдання:

- дослідити, як технології машинного навчання змінюють сучасні кіберзагрози;
- проаналізувати нові загрози, що виникають через використання ML і ШІ;
- вивчити, як зловмисники застосовують ML для удосконалення своїх атак;
- запропонувати засоби захисту від виявлених загроз;
- проаналізувати існуючі методи виявлення DoS/DDoS атак, включно з мережами IoT;
- розробити програмне забезпечення для виявлення таких атак;
- провести тестування на реальних або зібраних наборах даних;
- оцінити різні алгоритми ML та визначити найефективніший.

Об’єкт, предмет і методи дослідження

Об’єкт дослідження — кібербезпека, зокрема методи захисту від соціальної інженерії та DoS/DDoS атак у мережах IoT.

Предмет дослідження — застосування машинного навчання для виявлення атак і підвищення рівня захисту.

Методи дослідження — аналіз, порівняння, методи виявлення аномалій у мережевому трафіку, моделі машинного навчання.

Актуальність дослідження

Кількість пристроїв Інтернету речей постійно зростає, і разом із цим збільшується кількість потенційних загроз. Особливо складним є виявлення DoS/DDoS атак, адже у мережах IoT такий трафік часто нагадує нормальну

поведінку системи. Через це застосування машинного навчання стає одним з найперспективніших підходів для підвищення ефективності кіберзахисту.

Практичне значення

Результати дослідження можуть використовуватись для створення систем виявлення атак у мережах IoT та підвищення рівня інформаційної безпеки організацій.

Наукова новизна

Новизна роботи полягає у застосуванні алгоритмів машинного навчання для підвищення точності виявлення DoS/DDoS атак у середовищі IoT та у використанні багатокритеріальних методів оцінювання якості моделей.

Вступ

Стрімка цифровізація суспільства, розвиток інформаційних систем та активне впровадження мережевих технологій значно підвищили рівень залежності сучасних організацій від безпечного функціонування комп'ютерної інфраструктури. У таких умовах кіберзагрози стають дедалі складнішими, а масштаби та наслідки кібератак — більш руйнівними. Традиційні методи захисту, засновані на фіксованих правилах та сигнатурному аналізі, все частіше виявляються недостатньо ефективними перед новими, динамічними й адаптивними видами атак.

У відповідь на ці виклики машинне навчання поступово стає ключовим інструментом у сфері кібербезпеки. Алгоритми класифікації, навчання без учителя, глибинні нейронні мережі та інші сучасні підходи дозволяють автоматизувати процес виявлення аномальної активності, аналізувати великі масиви мережевих даних та виявляти складні, раніше невідомі кіберзагрози. Використання моделей машинного навчання значно підвищує точність і швидкість реагування, що є критичним чинником у сучасних умовах.

Актуальність даної теми для України особливо висока, зважаючи на зростання кількості кібератак як на державні, так і на приватні інфраструктури, а також на необхідність створення власних високотехнологічних рішень для забезпечення кіберзахисту. Ефективне застосування систем виявлення вторгнень, заснованих на машинному навчанні, може суттєво підвищити кіберстійкість організацій та державних установ.

Метою даної роботи є розробка та дослідження моделей машинного навчання для виявлення кібератак, оцінка їх ефективності та визначення оптимальних підходів для аналізу мережевого трафіку й ідентифікації шкідливої активності.

Для досягнення поставленої мети передбачається розв'язати такі **завдання**:

- проаналізувати сучасні методи виявлення кібератак та їх недоліки.

- дослідити популярні датасети мережевого трафіку, застосовувані у сфері кібербезпеки.

- розробити та реалізувати моделі машинного навчання для класифікації та виявлення аномалій.

- провести порівняльний аналіз точності, повноти та швидкодії створених моделей.

- сформулювати рекомендації щодо використання отриманих моделей у реальних системах виявлення вторгнень.

Практична цінність роботи полягає у можливості застосування отриманих результатів для розбудови ефективних систем кіберзахисту та інтеграції технологій машинного навчання у наявні рішення моніторингу безпеки.

РОЗДІЛ 1. АНАЛІЗ ПРОБЛЕМИ ДОСЛІДЖЕННЯ

1.1 Види соціальної інженерії та їх вплив

У сучасному цифровому середовищі соціальна інженерія стала одним із найпоширеніших та найнебезпечніших інструментів здійснення кібератак. На відміну від традиційних технічних методів злому, соціальна інженерія спрямована на маніпулювання людською поведінкою та використання психологічних вразливостей користувачів. Зловмисники відмовляються від складних технічних експлойтів на користь простіших, але значно дієвіших підходів — створення довіри, тиску на емоції, використання авторитету або відчуття терміновості. Це робить соціальну інженерію критичною загрозою, адже навіть найкращі технічні засоби безпеки не можуть повністю захистити від людського фактору. Соціальна інженерія є одним із найпоширеніших та найефективніших методів здійснення кібератак, оскільки вона спрямована не на подолання технічних бар'єрів, а на маніпулювання людською поведінкою. Людина, як елемент інформаційної системи, часто виявляється більш вразливою, ніж програмне забезпечення чи апаратні засоби захисту. Зловмисники використовують психологічні прийоми, довіру, емоції та природну схильність людей до помилок, щоб отримати конфіденційні дані, доступ до систем або можливість запустити шкідливий код. З огляду на значні втрати даних, фінансові збитки та репутаційні ризики, що виникають унаслідок кібератак соціальної інженерії, потреба в глибокому аналізі їхніх видів та механізмів дії є надзвичайно актуальною. Особлива важливість — навчати здобувачів освіти розпізнавати такі атаки, оскільки майбутні фахівці працюватимуть з великими обсягами інформації і повинні розуміти, як психологічні прийоми можуть замінити традиційні технічні атаки.

Зловмисники постійно удосконалюють соціотехнічні методи, поєднуючи маніпуляції з цифровою аналітикою, нейромережевими інструментами, генеративним штучним інтелектом та автоматизованим

збором даних. Тому сучасна підготовка кадрів має включати не лише теорію, а й практичні симуляції, що дозволяють побачити реальні механізми фішингу, вішингу, смішингу, бейтінгу та інших методів у дії.

Соціальна інженерія — це методи, за допомогою яких зловмисники обманюють людей і змушують їх самих розкрити конфіденційні дані. Найпоширеніші види таких атак:

Фішинг електронною поштою

Зловмисники надсилають підроблені листи, які зовні схожі на офіційні повідомлення від банку чи компанії.

Приклад: лист з проханням «оновити пароль».

Мета: отримати доступ до акаунтів або особистих даних.

Атаки через соціальні мережі

Шахраї збирають інформацію зі сторінок користувачів і використовують її, щоб викликати довіру.

Приклад: повідомлення, яке містить особисті дані жертви й змушує її підтвердити транзакцію.

Мета: отримати дані, доступ до акаунтів або грошей.

Психологічний тиск

Зловмисники змушують людину діяти терміново або без роздумів.

Приклад: дзвінок «від банку», де повідомляють, що «рахунок заблоковано».

Мета: виманити конфіденційну інформацію.

Зловживання довірою

Атакувальник може видавати себе за колегу або родича.

Приклад: прохання «перекинути пароль», бо «треба терміново зайти в систему».

Мета: отримати доступ до внутрішніх систем компанії.

Хто найчастіше стає жертвою?

- нові користувачі, які мало знають про кібербезпеку;
- люди, які довіряють інформації з соцмереж;

- співробітники без регулярного навчання.

Статистика

У 2019 році майже половина фінансових збитків від кіберзлочинів була пов’язана з підробленими корпоративними листами.

У 2020 році зареєстровано понад 240 тис. випадків фішингу.

Соціальна інженерія небезпечна тим, що часто достатньо зламати один корпоративний акаунт, щоб отримати доступ до всієї системи.

Як допомагають штучний інтелект і машинне навчання:

- аналізують поведінку користувачів і помічають підозрілу активність;

- фільтрують фішингові та спам-листи;

- визначають підозрілі повідомлення за допомогою аналізу тексту;

- контролюють соцмережі й помічають шахрайські схеми;

- допомагають створювати системи, які автоматично реагують на загрози.

Основні висновки:

- соціальна інженерія — серйозна й швидко зростаюча загроза;

- звичайних засобів захисту вже недостатньо;

- штучний інтелект значно покращує виявлення атак;

- найкращий підхід — це поєднання технологій та навчання користувачів;

- системи безпеки повинні постійно оновлюватись.

Соціальна інженерія включає цілий спектр технік, які постійно вдосконалюються та адаптуються до нових цифрових середовищ. Серед найбільш поширених та небезпечних видів можна виділити такі:

Фішинг є найбільш масовою формою соціальної інженерії, коли зловмисники надсилають повідомлення, що імітують офіційні листи від банків, державних установ чи популярних сервісів. У таких листах містяться посилання на підроблені сайти або вкладення з шкідливим кодом. Головною метою є примусити користувача розкрити паролі, дані банківських карток чи іншу важливу інформацію. Фішингові атаки еволюціонують: сучасні кампанії

використовують високоякісний дизайн, персоналізацію та контекст, що значно підвищує їхню ефективність. Сучасні форми фішингу включають: spear-phishing — таргетовані атаки на конкретних осіб або компанії; whaling — атаки на керівництво або топменеджмент; clone phishing — дублювання реальних листів з підміною елементів. Спіар-фішинг є більш цілеспрямованим різновидом фішингу, що орієнтований на конкретних осіб або групи. Для такої атаки зловмисники попередньо збирають детальну інформацію про жертву: її посаду, контакти, професійні інтереси тощо. Завдяки цьому повідомлення виглядає максимально правдоподібним, а отже ймовірність успішного обману суттєво зростає. Спіар-фішинг часто використовується для проникнення в корпоративні мережі або державні установи.

Вішинг — це використання телефонних дзвінків для маніпулювання жертвою. Наприклад, шахрай може представитися співробітником банку чи технічної підтримки й повідомити про «проблему», яку слід терміново вирішити. Під тиском ургентності люди часто надають конфіденційну інформацію або виконують дії, що відкривають доступ до систем. Використання сучасних технологій синтезу голосу та підміни номерів значно збільшило ефективність цього виду атак.

Смішинг поєднує методи фішингу з SMS-повідомленнями. Жертві надходить коротке повідомлення з посиланням на нібито офіційний ресурс або проханням негайно виконати певну дію. Такі атаки особливо небезпечні на мобільних пристроях, де користувачі частіше натискають на посилання, не перевіряючи їх. Маніпулятивні SMS-повідомлення з посиланням на підробні сайти або вимогою надати особисті дані підсилюються використанням підміни номера та автоматизованих сервісів розсилки.

Претекстинг полягає у створенні фальшивої історії або ситуації, яка змушує людину повірити в необхідність передачі певної інформації. Атакувальник може видавати себе за журналіста, співробітника служби безпеки, нового колегу, аудитора, постачальника, правоохоронця, сервісного

інженера або партнера по бізнесу. Цей метод покладається на логічність сценарію та психологічну впевненість того, хто його застосовує.

Клонування сайтів та веб-спуфінг – підробні вебсторінки, що повністю копіюють дизайн та функціональність реальних ресурсів. Застосовуються в комбінації з фішингом, MITM-атаками або DNS-підміною.

Бейтінг використовує прагнення людини до вигоди або цікавості. У цифровому середовищі це може бути нібито «безкоштовний» файл, музика, програма або знижка, завантаження яких призводить до зараження пристрою. Офлайн варіант — підкинуті USB-накопичувачі чи інші носії, які користувач може підключити до комп'ютера з цікавості.

Тейлгейтінг та фізичні атаки відносяться до фізичної соціальної інженерії. Зловмисник намагається потрапити в будівлю або приміщення, скориставшись ввічливістю або неуважністю працівників — наприклад, проходячи за ними в двері з електронним контролем доступу. Це дозволяє отримати доступ до фізичних пристроїв або внутрішньої мережі.

Маніпуляції в соціальних мережах стали надзвичайно поширеним методом через величезну кількість особистої інформації, яку люди добровільно публікують онлайн. Зловмисники створюють фальшиві профілі, використовують емоційні прийоми, пропонують «термінову допомогу» або інші способи втертися в довіру. Це може бути частиною ширших кампаній зі збору даних або підготовки до складніших атак.

Усі описані методи ґрунтуються на використанні людських слабкостей — довіри, страху, жадібності, порядності, бажання допомогти чи уникнути проблем. У цьому й полягає головна небезпека соціальної інженерії: навіть найдосконаліші технічні системи захисту можуть бути приреченими на провал, якщо працівники або користувачі не усвідомлюють ризиків та не дотримуються правил інформаційної гігієни.

Сучасні тенденції показують, що соціальна інженерія дедалі більше поєднується з автоматизованими технологіями, штучним інтелектом та глибокофейковими матеріалами. Це створює новий рівень складності атак і

потребує відповідного посилення систем кіберзахисту, зокрема шляхом використання інтелектуальних моделей аналізу поведінки, виявлення аномалій та машинного навчання. Соціальна інженерія залишається одним з найпотужніших інструментів здійснення кібератак, оскільки використовує не технічні вразливості, а психологічні. Фішинг, вішинг, претекстинг, смішинг, бейтинг та інші методи ефективні саме тому, що експлуатують базові людські реакції — довіру, необережність, страх або бажання швидко вирішити проблему.

1.2. Сучасний стан систем виявлення атак

В останній час велика увага приділяється методам прогнозування роботи комп'ютерних мереж в різних умовах, а особливо дослідженням оцінки функціонування інформаційної інфраструктури при проведенні націлених на неї атак.

Ідентифікація та розпізнавання впливу на мережі зв'язку на основі аналізу трафіку, що циркулює в них, відбувається на основі використання методів виявлення аномалій.

Розпізнавання порушень безпеки проводиться зазвичай за допомогою евристичних правил і аналізу сигнатур уже відомих атак.

Системи виявлення вторгнень (IDS) призначені для того, щоб помічати небезпечні дії в комп’ютерних мережах. Вони працюють або на окремому комп’ютері, або на всій мережі.

Основні види IDS:

Хостові IDS (HIDS) — встановлюються на конкретний комп’ютер.

Мережеві IDS (NIDS) — аналізують трафік у мережі загалом.

Розлянемо основні методи виявлення атак.

1. Сигнатурний метод

Працює за принципом «довідника»: система шукає збіги з відомими шаблонами атак.

Плюси: точний, мало помиляється.

Мінуси: не бачить нових атак, потребує постійного оновлення.

2. Поведінковий метод (аналітика аномалій)

Система знає, як має працювати мережа «в нормі», і помічає будь-які відхилення.

Плюси: виявляє нові та раніше невідомі атаки.

Мінуси: можливі хибні спрацьовування.

3. Комбінований метод

Поєднує два попередні і вважається найефективнішим.

Сучасний стан систем виявлення атак характеризується переходом від класичних підходів до інтелектуальних мережевих рішень. Традиційні IDS, що ґрунтувалися на сигнатурному аналізі, вже не можуть ефективно протидіяти динамічним, швидкозмінним або раніше невідомим кібератакам. Сигнатури дозволяють визначати лише ті загрози, які вже описані в базах даних, що робить такі системи менш ефективними проти zero-day атак, складних багатоступеневих (advanced persistent threats – APT) або атак соціальної інженерії, які маскуються під легітимну поведінку користувачів.

У відповідь на ці виклики особливого поширення набувають поведінкові (anomaly-based) системи виявлення атак. Вони використовують машинне навчання, статистичний аналіз і моделювання профілів нормальної діяльності, що дозволяє ідентифікувати відхилення навіть у випадках, коли точні сигнатури відсутні. Такі системи здатні аналізувати контекст подій, логіку дій користувачів, розподіл навантаження на мережу та взаємодію між вузлами, що забезпечує вищий рівень адаптивності. Їхнім важливим компонентом є технологія User and Entity Behavior Analytics (UEBA), яка дає змогу виявляти атаки всередині периметра, а також нетипові дії співробітників — як випадкові помилки, так і свідомі шкідливі дії.

Ще однією визначальною тенденцією є розвиток інтегрованих платформ виявлення й реагування, зокрема систем типу XDR (Extended Detection and Response) та SIEM (Security Information and Event Management). Такі рішення

виконують не лише моніторинг, а й кореляцію подій із різних джерел — серверів, мережевого обладнання, робочих станцій, хмарних сервісів, мобільних пристроїв. Завдяки цьому забезпечується можливість комплексного аналізу кіберінцидентів, виявлення прихованих зв'язків між подіями та формування автоматизованих алгоритмів реагування.

Важливою характеристикою сучасного етапу розвитку IDS є активне впровадження штучного інтелекту. Алгоритми глибинного навчання здатні самостійно виявляти складні кіберзагрози, прогнозувати шляхи їх поширення та пропонувати оптимальні варіанти протидії. Використання нейронних мереж у поєднанні з великими наборами даних про відомі атаки забезпечує системам виявлення здатність до самонавчання та підвищення точності. AI-орієнтовані IDS сьогодні застосовують методи кластеризації, аналізу часових рядів, обробки природної мови (наприклад, для аналізу логів) та автоматичної побудови графів взаємодії в мережі.

Поряд із цим, важливим напрямом розвитку є забезпечення високої масштабованості та гнучкості систем виявлення атак. Оскільки сучасні компанії дедалі більше переходять на хмарні інфраструктури, IDS повинні функціонувати в умовах розподілених середовищ, різнорідних платформ та віртуалізованих систем. Системи Cloud-Native IDS адаптуються до контейнеризованих середовищ, таких як Kubernetes, та інтегруються з DevSecOps-процесами, що дозволяє виявляти атаки вже на етапі розробки та тестування програмного забезпечення.

Однією з ключових тенденцій є також орієнтація на виявлення атак соціальної інженерії — фішингових кампаній, спроб компрометації облікових даних, аномальної активності користувачів. Сучасні IDS використовують кореляцію сигналів із поведінковими моделями та системами автентифікації для визначення ознак маніпулятивних впливів. Це особливо важливо з огляду на те, що більшість успішних кібератак сьогодні починаються саме з соціотехнічних методів.

Окремо варто відзначити розвиток систем раннього попередження, що використовують глобальні потоки даних про кіберзагрози (Threat Intelligence). Обмін інформацією між організаціями дозволяє IDS швидше реагувати на нові види атак та автоматично оновлювати моделі аналізу. Такі системи формують єдину екосистему, у якій дані про шкідливу активність миттєво поширюються між учасниками, підвищуючи загальний рівень кіберстійкості.

Таким чином, сучасні системи виявлення атак проходять етап глибокої трансформації: від пасивних моніторингових інструментів вони стають адаптивними, інтелектуальними платформами, здатними не лише виявляти загрози, а й прогнозувати їх, попереджувати та мінімізувати наслідки. Подальший розвиток IDS визначатимуть технології штучного інтелекту, автоматизація аналізу інцидентів, інтеграція з хмарними екосистемами та орієнтація на поведінкову модель користувачів, що дозволить ефективно протидіяти дедалі складнішим кіберзагрозам.

1.3. Методи аналізу трафіку

Аналіз трафіку допомагає визначити, що саме відбувається в мережі. Системи IDS використовують різні рівні аналізу — від перегляду заголовків пакетів до повного аналізу їхнього вмісту.

Основні рівні:

Базовий аналіз пакетів — перегляд адрес і типів трафіку.

Аналіз поведінки з’єднань — виявлення незвичних моделей.

Глибока інспекція пакетів (DPI) — аналіз вмісту на рівні додатків.

Ці методи допомагають своєчасно виявити аномалії та потенційні загрози.

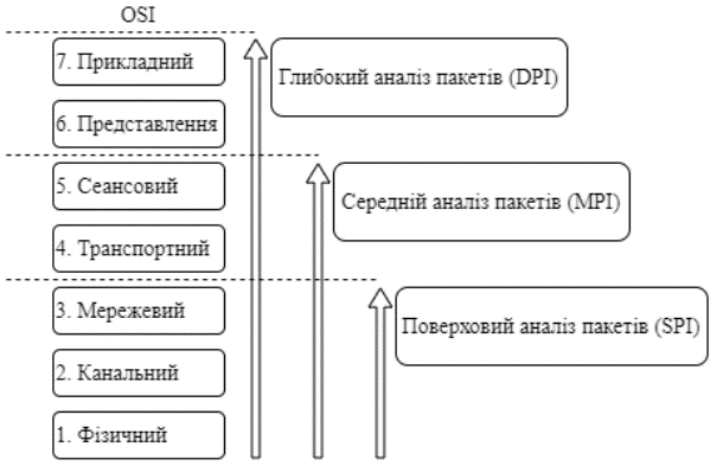


Рисунок 1.1 – Рівні розвитку технології аналізу мережевого трафіку за «глибиною».

Поверхневий аналіз пакетів (Shallow Packet Inspection, SPI).

Поверхневий аналіз пакетів — це простий спосіб перевірки мережевого трафіку. Він аналізує лише заголовки пакетів (1–3 рівні моделі OSI) — тобто перевіряє:

- IP-адресу відправника і отримувача;

- номери портів;

- тип протоколу.

Такий аналіз майже не навантажує систему і дозволяє швидко обробляти великі потоки даних.

SPI працює так:

- програма порівнює назву протоколу, який фактично використовується, з номером порту, призначеним цьому протоколу за класифікацією IANA.

Якщо все збігається — трафік вважається нормальним. Якщо ні — він підозрілий.

SPI використовується у більшості міжмережевих екранів, маршрутизаторів та інших мережевих пристроїв. На його основі створені правила доступу за IP-адресами та портами.

Середній аналіз пакетів (Medium Packet Inspection, MPI)

MPI — це більш складна технологія, яка перевіряє не лише заголовки, але й окремі частини даних, що передаються в пакеті.

Вона аналізує сеанси зв’язку, які створюються спеціальними проміжними вузлами — своєрідними шлюзами.

Принцип роботи MPI:

- пакети проходять через спеціальні пристрої-посередники;
- ці пристрої можуть читати заголовки і частково корисне навантаження
- на основі правил система вирішує, що дозволити, а що заблокувати.

MPI може використовуватися для:

- блокування небажаних файлів (відео, аудіо тощо);
- аналізу стислого або зашифрованого трафіку;
- обмеження окремих команд певних протоколів;
- кешування даних.

Недоліки MPI:

- погана масштабованість (кожний протокол потребує свого окремого шлюзу);
- робота у режимі «проксі» зменшує швидкість та навантажує систему.

Глибокий аналіз пакетів (Deep Packet Inspection, DPI). DPI — найсучасніша технологія аналізу трафіку.

Вона перевіряє як заголовки, так і весь вміст пакету.

DPI може працювати на всіх рівнях OSI, тому дуже точно визначає:

- тип даних;
- додаток, що їх передає;
- реальний зміст пакета;
- можливі загрози.

Ця технологія широко використовується:

- у системах моніторингу;
- у засобах кібербезпеки;
- для виконання вимог законодавства щодо контролю трафіку.

DPI фактично є стандартом сучасного аналізу мережевого трафіку.

Сучасний аналіз трафіку також базується на концепціях кореляційного та багатошарового аналізу, що передбачають поєднання даних із різних джерел — мережевих логів, журналів доступу, поведінкових профілів, даних систем автентифікації. Такий підхід, який реалізується у SIEM- і XDR-платформах, дає змогу визначати комплексні, багатоступеневі атаки, які на окремих етапах виглядають як звичайні або несуттєві події. Кореляція дозволяє розглядати мережевий трафік у контексті ширших кіберподій, що підвищує точність і швидкість виявлення.

Не менш важливим методом є аналіз з використанням графових моделей, у яких мережа розглядається як система взаємопов'язаних вузлів. За допомогою таких технологій можливо відстежувати шляхи поширення загроз, виявляти нетипові маршрути даних, аналізувати структуру взаємодій та визначати ключові точки ризику. Графові методи особливо ефективні у боротьбі з APT-атаками, які мають складну багаторівневу архітектуру та можуть залишатися непоміченими протягом тривалого часу.

Поряд із цим, значного поширення набувають методи аналізу трафіку на основі штучного інтелекту. Використання нейронних мереж, алгоритмів кластеризації, гібридних моделей ML та технологій обробки великих даних (Big Data) дозволяє виявляти аномальні патерни, що недоступні для традиційних алгоритмів. Особливо перспективним напрямом вважається використання глибинних рекурентних моделей (LSTM, GRU) для прогнозування поведінки мережевого трафіку та аналізу часових послідовностей, що дозволяє визначати навіть слабко виражені ознаки атак.

Сучасний аналіз трафіку базується на поєднанні класичних і інтелектуальних методів, кожен із яких виконує свою роль у процесі виявлення кібератак. У таблиці 1.1 наведено узагальнену характеристику основних підходів до аналізу мережевого трафіку, їхні переваги та обмеження.

Таблиця 1.1

Порівняльна характеристика методів аналізу мережевого трафіку

Метод аналізу	Основний принцип	Переваги	Обмеження
Сигнатурний аналіз	Пошук збігів із базою відомих шаблонів атак	Висока точність для відомих загроз; низький рівень хибних спрацювань	Не здатний виявляти нові атаки; залежність від оновлення баз
Глибокий аналіз пакетів (DPI)	Дослідження структури й вмісту кожного пакета	Виявлення шкідливого контенту; ефективність проти складних атак	Ускладнення через шифрування; потребує високих ресурсів
Аналіз зашифрованого трафіку (ETA)	Оцінка метаданих без доступу до вмісту пакета	Ефективний у сучасних умовах шифрування; корисний для хмарних систем	Нижча точність порівняно з DPI; складність побудови моделей
Кореляційний аналіз / SIEM / XDR	Об’єднання подій із різних джерел та пошук взаємозв’язків	Виявлення багатоступеневих атак; зменшення хибних тривог	Складність налаштування; вимоги до якості логів

Графові методи	Представлення мережі як множини вузлів і зв’язків	Висока ефективність проти APT; розуміння маршруту атаки	Висока складність формування графів у великих мережах
Методи машинного навчання	Автоматичне виявлення патернів у даних	Висока адаптивність; здатність працювати з великими масивами даних	Потреба в якісному датасеті; складність інтерпретації
Аналіз аномалій	Виявлення відхилень від нормальної поведінки мережі	Можливість виявляти невідомі та складні атаки	Висока кількість хибних спрацювань; потребує навчання моделей

Джерело: складено автором

Як видно з таблиці 1.1, жоден окремий метод не гарантує повного захисту, тому сучасні системи виявлення атак ґрунтуються на комбінуванні різних технологій та адаптивному аналізі трафіку. На мою думку, подальший розвиток систем аналізу трафіку буде пов’язаний із глибокою інтеграцією машинного навчання та поведінкових моделей, здатних самостійно адаптуватися до нових типів загроз. У довгостроковій перспективі саме інтелектуальні алгоритми стануть основою мережевого захисту, тоді як класичні методи виконуватимуть допоміжну роль.

Таким чином, сучасні методи аналізу трафіку є поєднанням класичних підходів і інтелектуальних технологій нового покоління. Їхній розвиток зумовлений зростанням кількості шифрованих комунікацій, еволюцією

кіберзагроз та появою високопродуктивних інструментів обробки даних. Комплексне застосування сигнатурного аналізу, моделей аномалій, глибинного аналізу пакетів і методів машинного навчання забезпечує високий рівень точності та дозволяє ефективно протидіяти сучасним кібератакам. У майбутньому роль інтелектуальних алгоритмів лише зростатиме, сприяючи створенню адаптивних, самонавчальних систем кіберзахисту нового покоління.

1.4. Постановка задачі інформаційного синтезу системи виявлення атак.

Ефективне виявлення кібератак у сучасних інформаційних системах вимагає комплексного підходу, який передбачає не лише збір та аналіз мережевого трафіку, але й узгодження різних джерел даних, методів обробки та моделей прийняття рішень. У цьому контексті постає задача інформаційного синтезу — процесу інтеграції різнорідної інформації для побудови єдиної системи, здатної своєчасно реагувати на загрози та забезпечувати високий рівень точності виявлення.

Сутність інформаційного синтезу полягає в об’єднанні даних, методів та алгоритмів у цілісну структуру, що дозволяє системі враховувати широкий спектр ознак: від низькорівневих характеристик трафіку (частота пакетів, розмір, часові інтервали) до поведінкових моделей взаємодії користувачів та пристроїв. Такий підхід забезпечує можливість комплексного аналізу ситуації, де окремі події, що самі по собі виглядають нешкідливими, у сукупності формують чітку картину шкідливої активності.

Сучасні IDS (системи виявлення вторгнень) мають низку недоліків:

- сильно навантажують систему, особливо під час роботи «в реальному часі»
- складно адаптуються до різних програмних середовищ
- не мають єдиних правил порівняння продуктивності;
- часто помиляються: багато хибних спрацювань та пропусків атак;

- погано виявляють нові або незнайомі атаки;

- не дозволяють визначити джерело атаки;

- не дають оцінки точності результаті

Мета роботи.

Створити автоматизовану систему виявлення атак, яка:

- швидко реагує на загрози;

- блокує небажані дії

- накопичує досвід виявлення атак;

- підвищує захищеність інформаційно-комунікаційної системи (ІКС).

Суть задачі інформаційного синтезу

Припустимо, що є набір класів, які описують різні типи мережевого трафіку. Для їх розпізнавання формується навчальна матриця, де вказано, які ознаки найбільше характерні для кожного класу.

Мета технології — перетворити навчальну матрицю у робочу бінарну матрицю, яка буде найкраще підходити для класифікації трафіку.

Для цього вводиться вектор параметрів:

$gm=\langle xm,d,\delta\rangle$gm=⟨xm,dm,δ⟩

де:

xm — середній (типовий) вектор класу;

dm — його «радіус» (розмір області, в якій знаходяться всі нормальні реалізації);

δ — допуск, що дозволяє враховувати невеликі відхилення.

На параметри накладаються обмеження:

- дані класу повинні мати розподіл, наближений до нормального;

- радіус класу має бути меншим за відстань до сусіднього класу;

- має бути меншим за подвійну кількість градацій контрольного поля.

Що потрібно зробити:

На етапі навчання знайти такі параметри

$\langle xm,dm,\delta\langle$xm,dm,δ⟩, які дають найбільше значення інформаційного критерію ефективності.

Після цього побудувати правила, які будуть точно розпізнавати атаки за навчальною матрицею.

У рамках постановки задачі важливим є формування критерію оптимізації — міри, за якою оцінюється якість роботи системи. Такий критерій може включати мінімізацію хибних спрацювань, максимізацію точності виявлення певних типів атак, швидкість реагування або збалансованість між точністю й продуктивністю. Залежно від обраного підходу, система може акцентувати увагу на оперативності реагування, здатності до глибокої аналітики або стійкості до масштабування у великих мережах.

З урахуванням зазначених аспектів задача інформаційного синтезу набуває вигляду багатокритеріальної проблеми, у якій необхідно узгодити гетерогенні дані, методи аналізу, алгоритми машинного навчання та механізми ухвалення рішень. Результатом успішного синтезу є створення системи, здатної функціонувати не лише як набір окремих модулів, а як інтегрований інтелектуальний інструмент, що забезпечує повноцінну ситуаційну обізнаність та ефективний захист від кібератак.

У підсумку можна зазначити, що інформаційний синтез є фундаментальною складовою побудови систем виявлення атак, оскільки саме він визначає здатність системи адаптуватися до постійно змінюваних умов та забезпечувати високу результативність навіть за умов появи нових, невідомих загроз. Саме завдяки синтезу даних та інтелектуальних моделей формується сучасний підхід до кіберзахисту, що поєднує комплексність, гнучкість і глибоку аналітичність.

Сутність інформаційного синтезу полягає в об’єднанні даних, методів та алгоритмів у цілісну структуру, що дозволяє системі враховувати широкий

спектр ознак: від низькорівневих характеристик трафіку (частота пакетів, розмір, часові інтервали) до поведінкових моделей взаємодії користувачів та пристроїв. Такий підхід забезпечує можливість комплексного аналізу ситуації, де окремі події, що самі по собі виглядають нешкідливими, у сукупності формують чітку картину шкідливої активності

2. ОПИС МЕТОДУ ДОСЛІДЖЕННЯ

2.1 Основні принципи та визначення інформаційно-екстремальної технології аналізу даних

Інформаційно-екстремальна інтелектуальна технологія (ІЕІ-технологія) є сучасним підходом до машинного навчання, що спрямований на формування чіткої структури класів розпізнавання шляхом оптимізації параметрів функціонування системи. Її сутність полягає у перетворенні апріорно нечіткого розбиття простору ознак на чітке розбиття класів за допомогою пошуку глобального максимуму статистичного інформаційного критерію у допустимій області.

На відміну від традиційних методів машинного навчання, що переважно базуються на нейроподібних структурах, ІЕІ-технологія використовує функціональний підхід до моделювання когнітивних процесів, властивих людині. Це дозволяє наблизити процес прийняття класифікаційних рішень до принципів природного інтелекту.

Принципи інформаційно-екстремального машинного навчання

Методи ІЕІ-технології ґрунтуються на низці специфічних принципів, що доповнюють традиційні положення системного аналізу:

Принцип максимізації інформації.

Реалізується шляхом введення додаткових інформаційних обмежень, які підвищують різноманітність об’єктів та забезпечують більш чітке розмежування між класами.

Принцип дуальності.

На етапі первинного моделювання застосовуються прості алгоритми, які в подальшому цілеспрямовано уточнюються в процесі глибшого машинного навчання. Це дає змогу поступово наблизити вирішальні правила до безпомилкових за навчальною матрицею.

Принцип апріорної недостатності обґрунтування гіпотез (принцип Бернуллі–Лапласа).

У разі невизначеності початкових даних всі гіпотези вважаються рівноймовірними. Система приймає рішення, виходячи з найгірших можливих статистичних умов.

рандомізації.

Випадкові перетворення вхідної інформації дозволяють дослідити статистичні властивості потоку даних та підвищити стійкість алгоритмів.

Принцип редукції даних.

Передбачає вилучення неінформативних або завадних ознак із словника ознак розпізнавання. Це зменшує розмірність простору та підвищує ефективність алгоритму.

Принцип зовнішнього доповнення.

Для оцінки функціональної ефективності машинного навчання застосовуються навчальні та контрольні вибірки, що забезпечують об’єктивну перевірку результатів.

Глибина машинного навчання в IEI-технології.

Глибина інформаційно-екстремального навчання визначається кількістю параметрів системи, що оптимізуються відповідно до інформаційного критерію.

У внутрішньому циклі навчання реалізується базовий алгоритм IEI-технології, який поступово вдосконалює геометричні параметри контейнерів класів розпізнавання у бінарному просторі ознак Хеммінга.

Мета інформаційно-екстремального машинного навчання.

Основною метою є досягнення максимальної ймовірності правильних класифікаційних рішень. На відміну від нейронних мереж, які часто мають непрозору структуру та складність інтерпретації, методи IEI-технології дозволяють безпосередньо керувати параметрами системи розпізнавання та оцінювати їх вплив на ефективність роботи.

У межах ІЕІ-технології як критерій оптимізації може використовуватися будь-яка статистична інформаційна міра різноманітності об’єктів. Достатня глибина навчання визначається досягненням максимального середнього значення інформаційного критерію за всіма класами розпізнавання відповідно до принципів відкладених рішень О. Г. Івахненка.

Побудова вирішальних правил.

Після завершення процедури оптимізації на основі отриманих параметрів формуються вирішальні правила. Їх побудова здійснюється у радіальному базисі простору Хеммінга, що забезпечує:

- майже повну інваріантність щодо високої розмірності простору ознак;
- можливість обробки великих двійкових векторів (до 285 ознак і більше);
- високу швидкість прийняття рішень у режимі реального часу.

Це є суттєвою перевагою для систем виявлення атак, які потребують оперативності та точності під час моніторингу мережевої активності.

2. 2. Машинне навчання як інструмент виявлення загроз у системах автентифікації.

Сучасні інформаційні системи постійно зазнають значного спектра кібератак, серед яких особливу небезпеку створюють атаки на механізми автентифікації. Традиційні методи забезпечення безпеки — логування, фільтрація трафіку, контроль доступу — не завжди здатні ефективно реагувати на аномальні дії, оскільки зловмисники постійно вдосконалюють техніки обходу засобів захисту. У зв’язку з цим особливо актуальним стає застосування інтелектуальних методів аналізу, здатних автоматично виявляти відхилення в поведінці користувачів.

Одним із таких підходів є машинне навчання, яке дає змогу аналізувати великі обсяги даних і виявляти закономірності, що недоступні для класичних алгоритмів або ручного аналізу. Моделі машинного навчання можуть

фіксувати нетипові патерни поведінки, наприклад, підвищену частоту авторизацій у короткий проміжок часу, використання незвичних IP-адрес чи атипову географію входів. Усе це може свідчити про спроби підбору пароля, компрометацію облікового запису або інші зловмисні дії.

Пошук аномалій у логах автентифікації дозволяє своєчасно ідентифікувати загрози, пов’язані з несанкціонованим використанням доступу, порушенням політик безпеки та маніпуляцією обліковими даними. Застосування машинного навчання у цій сфері створює можливість для автоматизованої оцінки ризиків і підвищення загального рівня кіберзахисту.

Машинне навчання є галуззю штучного інтелекту, що використовує статистичні методи для надання комп’ютерним системам здатності навчатися на основі даних без прямого програмування кожної дії. Метою машинного навчання є побудова моделей, здатних робити прогнози або приймати рішення, спираючись на закономірності, знайдені в історичних даних.

У залежності від типу задачі, можуть застосовуватися такі класи моделей:

Нейронні мережі — ефективні для виявлення складних нелінійних залежностей.

Дерева рішень та ансамблі (Random Forest, Gradient Boosting) — придатні для класифікації аномальних подій у логах.

Метод опорних векторів (SVM) — використовується для задач класифікації з чіткими межами між класами.

Моделі без учителя (DBSCAN, LOF, Isolation Forest) — ефективні для пошуку аномалій без попередньо маркованих вибірок.

Такі моделі здатні виявляти приховані патерни у великих масивах даних і застосовуються у різних сферах, зокрема для прогнозування, класифікації, рекомендаційних систем, розпізнавання облич та голосу тощо. У контексті даної роботи машинне навчання розглядається як інструмент аналізу логів автентифікації та виявлення нетипових дій, що можуть свідчити про кібератаки.

2.3. Аналіз існуючих досліджень у сфері виявлення аномалій

Під час аналізу наукових публікацій було встановлено, що більшість досліджень у сфері виявлення аномалій у логах автентифікації базуються на методах навчання з учителем. Ці роботи використовують попередньо марковані дані, у яких окремі записи чітко визначені як нормальні або аномальні. На основі таких даних моделі машинного навчання навчаються розрізняти нові вхідні події та класифікувати їх.

Серед найпоширеніших алгоритмів у цій категорії — Random Forest, Decision Tree, SVM, Logistic Regression. Водночас у дослідженнях активно використовуються і методи без учителя, наприклад:

Isolation Forest — ефективний у виявленні рідкісних подій;

DBSCAN — групує записи за щільністю, виявляючи відхилення;

LOF (Local Outlier Factor) — визначає локальні аномалії.

Аналіз літератури показує, що універсального методу для всіх типів атак не існує, проте використання машинного навчання значно підвищує точність виявлення нетипових дій у системах автентифікації.

У сучасній кіберзлочинності активно застосовуються інструменти, побудовані на основі машинного навчання. Їх аналіз дає змогу зрозуміти потенційні вектори атак та визначити, які саме методології варто використовувати для побудови захисту.

1. Соціальна інженерія на основі EvilProxy

EvilProxy є інструментом, що застосовує алгоритми штучного інтелекту для автоматизації соціальної інженерії. Він може збирати персональні дані користувачів із соціальних мереж та формувати високоточні фішингові повідомлення.

Особливістю таких атак є адаптивність: система аналізує поведінку потенційної жертви і коригує зміст повідомлень відповідно до її інтересів та активності в мережі.

Для виявлення цих атак дослідники використовують моделі машинного навчання (Random Forest, SVM, Logistic Regression), які аналізують:

структуру URL-адрес,

валідність SSL-сертифікатів,

підозрілу кількість субдоменів тощо.

2. Атаки грубої сили з використанням PassGAN

PassGAN використовує генеративні нейронні мережі для передбачення ймовірних паролів на основі масивів скомпрометованих даних. На відміну від класичного перебору паролів, PassGAN прогнозує найбільш ймовірні комбінації, ґрунтуючись на статистиці реальних витоків.

Це робить атаки грубої сили значно ефективнішими та швидшими.

3. DeepLocker: шкідливе ПЗ, кероване ШІ

DeepLocker — експериментальний зразок IBM, який демонструє нове покоління шкідливих програм. На основі алгоритмів машинного навчання він може:

- перебувати в прихованому режимі,
- активуватися лише після розпізнавання конкретної цілі (обличчя, голосу, геолокації),
- змінювати поведінку залежно від середовища.

Такий підхід суттєво ускладнює виявлення шкідливого ПЗ традиційними антивірусами.

4. Deepfake-технології: FaceSwap та Respeecher

Deepfake-інструменти дозволяють створювати правдоподібні відео та аудіо, що повністю імітують зовнішність та голос людини.

Зловмисники використовують їх для:

шахрайства, шантажу, видавання себе за керівників компаній, родичів або службовців.

Загроза полягає в підриві довіри до цифрової комунікації, включно з відеодзвінками та голосовою автентифікацією.

5. Чат-боти для автоматизації атак

Застосування великих мовних моделей (LLM) створило новий клас загроз:

SNAP_R — автоматизує персоналізовані фішингові кампанії;

WormGPT — дозволяє створювати шкідливий код без обмежень;

AutoGPT — може автономно виконувати складні зловмисні дії;

FraudGPT — генерує тексти для шахрайських схем;

PoisonGPT — поширює неправдиву інформацію та створює шкідливі сценарії.

Такі інструменти значно підвищують масштабованість і складність кібератак.

2.4. Дослідження впливу інструментів з використанням машинного навчання на методи кібератаки

Стрімкий розвиток машинного навчання та технологій штучного інтелекту істотно впливає як на методи захисту інформаційних систем, так і на інструментарій зловмисників. З одного боку, алгоритми машинного навчання ефективно допомагають аналітикам безпеки автоматизувати процеси моніторингу, сортування загроз і виявлення вразливостей. Згідно з дослідженням Sapio Research для компанії Vanta, 62 % організацій планують збільшити інвестиції в засоби безпеки на основі ШІ протягом найближчих 12 місяців.

З іншого боку, штучний інтелект дедалі активніше використовується у кібератаках. Багаторічне опитування дослідників уразливостей від Bugcrowd (2024) показало, що:

77 % зловмисників вже використовують ШІ для злому;

86 % заявляють, що ШІ повністю змінив їхній підхід до атак;

кількість тих, хто вважає ШІ «корисним для хакінгу», зросла з 21 % (2023 р.) до 71 % (2024 р.).

ШІ дає змогу зловмисникам автоматизувати підготовку атак, оптимізувати пошук вразливостей, створювати шкідливе ПЗ та обходити системи захисту з небаченою раніше ефективністю.

1. Використання ШІ у фішингових атаках

Злочинці застосовують генерацію природної мови (NLP) для створення персоналізованих фішингових листів, максимально схожих на справжню бізнес-комунікацію. Автоматизовані системи здатні адаптувати стиль, доменну лексику та структуру повідомлення під конкретну галузь або компанію.

ШІ-підсилені фішингові атаки здатні:

аналізувати поведінку користувачів у соцмережах;

імітувати стиль менеджерів, рекрутерів або партнерів;

генерувати динамічні фішингові шаблони;

обходити прості фільтри пошти.

2. Шкідливе ПЗ на основі ШІ

Штучний інтелект використовується для створення нового покоління шкідливого ПЗ, яке:

- автоматично генерує шкідливий код, навіть без участі досвідчених програмістів;
- самостійно шукає вразливі системи, аналізуючи мережевий трафік та порти;
- адаптується під умови виконання, змінюючи свій код під час роботи («мутує»);
- вміє обходити антивіруси та IDS, оскільки є непередбачуваним.

Антивірусне ПЗ, побудоване на статичних сигнатурах, стає неефективним проти динамічних моделей, здатних змінювати свій код у реальному часі.

3. DDoS-атаки нового покоління

ШІ дозволяє:

- автоматизувати управління ботнетами;

- аналізувати телеметрію в реальному часі для оптимізації навантаження;
- маскувати трафік під легітимний;
- збільшувати масштаб атак без додаткової інфраструктури.

Завдяки застосуванню моделей ШІ навіть некваліфіковані користувачі можуть запускати ефективні та добре замасковані DDoS-атаки.

4. Узагальнений вплив ШІ на етапи кібератаки (за MITRE ATT&CK)

Для систематизації впливу технологій штучного інтелекту на різні стадії атаки використано структуру MITRE ATT&CK, яка описує повний життєвий цикл атаки — від розвідки до ексфільтрації даних.

Важливою перевагою машинного навчання є можливість обробки великих масивів даних, що надходять від мережевого обладнання, серверів, додатків та кінцевих пристроїв. На основі цих даних формуються поведінкові моделі, які автоматично адаптуються під специфіку конкретної мережі. Такий підхід дозволяє виявляти аномалії, характерні для внутрішніх загроз, атаки з боку компрометованих облікових записів, складні багатоетапні атаки (APT) та нові варіанти зловмисного ПЗ.

Щоб продемонструвати різницю між основними видами ML-моделей, у таблиці 2.1 наведено порівняльний аналіз найбільш поширених алгоритмів, що застосовуються в системах виявлення кібератак.

Таблиця 2.1

Порівняльна характеристика ML-моделей у виявленні кібератак

Модель	Сильні сторони	Слабкі сторони	Типові завдання
Random Forest	Висока точність; стійкість до шуму; добре працює з різнорідними даними	Відносно повільна робота на дуже великих наборах даних	Класифікація трафіку, визначення типів атак

SVM (машина опорних векторів)	Висока якість класифікації; ефективна на малих вибірках	Погано масштабується; складність вибору ядра	Детекція аномалій, класифікація поведінки
Нейронні мережі (DNN)	Глибокий аналіз складних патернів; ефективні на великих даних	Аналіз мережевих сесій та логів	Поведінковий аналіз, робота з великими потоками
LSTM-мережі	Глибокий аналіз складних патернів; ефективні на великих даних	Довге навчання; ресурсомісткість	Аналіз мережевих сесій та логів
K-Means	Простота; висока швидкість	Чутливість до вибору кількості кластерів	Початковий аналіз аномалій, сегментація трафіку
Autoencoders	Виявлення складних аномалій; ефективні на зашифрованому трафіку	Висока складність налаштування; потреба у великій кількості навчальних даних	Аномалії, zero-day атаки

З таблиці 2.1 бачимо, що кожна ML-модель має власні переваги та обмеження, що визначають сфери її оптимального застосування в системах виявлення кібератак.

Як показано в таблиці 2.2, різні моделі демонструють значну варіативність за метриками точності, швидкістю роботи та здатністю інтерпретувати результати, що суттєво впливає на їх практичне застосування у процесі виявлення кібератак.

Таблиця 2.2

Порівняльні показники ефективності ML-моделей у системах виявлення кібератак

Модель	Точність (Accuracy)	Recall (виявлення атак)	Precision (точність класифікації	Швидкість роботи	Інтерпретованість
Random Forest	Висока	Висока	Висока	Середня	Середня
SVM	Висока	Середня-висока	Висока	Низька на великих вибірках	Низька
Logistic Regression	Середня	Середня	Середня	Висока	Висока
DNN (глибокі нейронні мережі)	Дуже висока	Дуже висока	Середня–висока	Низька (ресурсоємна)	Дуже низька
LSTM	Висока	Дуже висока	Середня	Низька	Низька

K-Means	Низька (неконтро-льоване навчання)	Середня	Низька	Дуже висока	Висока
Autoencoder	Висока для аномалій	Висока для аномалій	Низька–середня	Середня	Низька

На основі аналізу можна стверджувати, що моделі глибинного навчання демонструють найвищу ефективність у виявленні складних аномалій та багатоступеневих атак, однак їхня низька інтерпретованість і висока ресурсна вартість обмежують застосування у реальному часі. Натомість класичні моделі, такі як Random Forest або Logistic Regression, забезпечують кращий баланс між швидкістю, точністю та можливістю пояснити результат.

РОЗДІЛ 3. ЗАСТОСУВАННЯ МЕТОДІВ ЗАХИСТУ ВІД КІБЕРАТАК НА ОСНОВІ ШТУЧНОГО ІНТЕЛЕКТУ

3.1 Реалізація захисних моделей на базі ШІ

Реалізація захисних моделей на базі штучного інтелекту (ШІ) є одним із найперспективніших напрямів у сфері кібербезпеки. Завдяки здатності до навчання, аналізу великих обсягів даних та швидкому виявленню аномалій, ШІ дозволяє створювати системи захисту, які значно перевершують традиційні підходи. Моделі машинного та глибокого навчання забезпечують адаптивні рішення, здатні ефективно протистояти новим, складним атакам.

Основні принципи роботи захисних моделей на базі ШІ

Самооновлення через аналіз загроз

Алгоритми ШІ постійно аналізують дані про мережеву активність, поведінку користувачів та системні логи, що дозволяє ідентифікувати потенційні загрози на ранніх етапах

Виявлення аномалій

Моделі ШІ здатні ідентифікувати фішингові атаки, спроби злому систем аутентифікації, складні DDoS-атаки та інші загрози у режимі реального часу.

Автоматизація процесів захисту

Інтеграція моделей ШІ дозволяє автоматично фільтрувати мережевий трафік, визначати підозрілі дії та адаптивно реагувати на нові загрози

Типи алгоритмів ШІ у кібербезпеці

1. Алгоритми навчання з учителем

Приклади: дерева рішень, метод опорних векторів (SVM), нейронні мережі.

Принцип роботи: навчання на великій кількості мічених даних для виявлення відомих атак.

Переваги: висока точність при наявності структурованих даних і чітких шаблонів.

Недоліки: зниження ефективності при нових, невідомих загрозах або нестачі даних.

2. Алгоритми навчання без учителя

Приклади: кластеризація, методи пошуку аномалій.

Принцип роботи: виявлення аномальних патернів у поведінці користувачів та систем без попередньо мічених даних.

Переваги: здатність знаходити нові, невідомі загрози

Недоліки: менша точність порівняно з алгоритмами з учителем, оскільки аномалія фіксується після її прояву.

3. Глибоке навчання

Використовується для аналізу складних патернів, наприклад, у поведінці користувачів або мережевому трафіку.

Дозволяє виявляти складні та замасковані атаки, які традиційні алгоритми можуть пропустити.

4. Гібридні системи

Поєднують методи навчання з учителем, без учителя та глибоке навчання.

Забезпечують максимальний захист, одночасно знижуючи ризики від відомих і нових загроз.

Впровадження захисних моделей у реальні системи

Впровадження ШІ у системи кібербезпеки є складним, але критично важливим процесом, що включає

Аналіз існуючої інфраструктури

Оцінка наявних систем і виявлення вразливостей.

Формування бази даних для навчання моделей.

Інтеграція алгоритмів ШІ

Вибір моделей машинного або глибокого навчання залежно від типів загроз.

Приклад: алгоритми для захисту від фішингових атак аналізують контент та метадані електронних листів.

Комбінування з традиційними методами захисту

Інтеграція ШІ з антивірусними програмами, IDS/IPS, системами контролю доступу. ШІ підвищує ефективність традиційних методів і дозволяє виявляти нові аномалії.

Організаційна підготовка

Підготовка персоналу до роботи з новими технологіями.

Визначення чітких процесів реагування на кіберінциденти.

Співпраця між IT-відділами та аналітиками кібербезпеки.

Переваги захисних моделей на базі ШІ

Адаптивність до різних середовищ і сценаріїв (від малих мереж до великих корпоративних систем).

Масштабованість у разі збільшення обсягу даних чи кількості користувачів.

Можливість виявляти нові, невідомі загрози завдяки алгоритмам без учителя та глибокого навчання

Підвищення ефективності комплексних систем кібербезпеки при інтеграції з традиційними засобами захисту.

Швидка і точна реакція на сучасні складні атаки, включно з атаками, замаскованими під звичайну діяльність користувачів.

Важливим елементом реалізації захисних моделей на базі ШІ є інтеграція їх у існуючі інфраструктурні рішення, такі як системи виявлення вторгнень (IDS), системи запобігання вторгненням (IPS) та центри моніторингу безпеки (SOC). У таких середовищах моделі ШІ виконують роль автоматизованих аналітичних модулів, здатних фільтрувати шумові дані, генерувати рекомендації щодо реагування та навіть ініціювати блокування підозрілих активностей. Високий рівень автоматизації дозволяє суттєво зменшити навантаження на аналітиків SOC і підвищити швидкість реагування.

Одним із ключових етапів реалізації захисних моделей є їх тестування та калібрування. Системи ШІ мають тенденцію до помилкових спрацювань (false positive) або пропусків атак (false negative), тому моделі повинні

регулярно перевірятися на нових наборах даних, оптимізуватися та доповнюватися з урахуванням реальних інцидентів. Важливо також забезпечити баланс між чутливістю моделей до аномалій і їхньою стійкістю до шуму. У цьому контексті велике значення має використання адаптивних методів навчання, що дозволяє системі постійно оновлювати свій досвід.

Важливо відзначити, що успішна реалізація моделей штучного інтелекту вимагає не лише математично правильних алгоритмів, але й комплексного підходу до організації інформаційної безпеки. Ефективність захисної системи ШІ залежить від якості вихідних даних, наявності актуальних загрозових індикаторів, узгодженості політик безпеки та готовності організації реагувати на рекомендації, що генеруються моделями. Таким чином, штучний інтелект не замінює фахівців, а підсилює їхню роботу, надаючи потужні інструменти для аналізу й прийняття рішень.

Узагальнюючи викладене, можна стверджувати, що реалізація захисних моделей на базі штучного інтелекту є ключовим елементом сучасних систем протидії кібератакам. Поєднання алгоритмів контролюваного та неконтрольованого навчання, глибинних нейронних мереж, адаптивних методів аналізу та модулів інтеграції забезпечує високий рівень захисту навіть у складних і динамічних кіберпросторах. У наступному підрозділі буде розглянуто питання моделювання атак і тестування ефективності побудованих захисних систем, що є важливим завершальним етапом практичної реалізації підходів на основі ШІ.

Однією з фундаментальних особливостей моделей ШІ є можливість багаторівневої обробки даних. На первинному етапі відбувається збір і фільтрація мережевого трафіку, системних логів, даних про активність користувачів і службових процесів. Після нормалізації інформації застосовуються алгоритми виділення ознак, які трансформують великі необроблені набори даних у структуровані вектори, придатні для навчання моделей. Важливим етапом є вибір релевантних ознак — швидкість

передавання пакетів, кількість невдалих спроб автентифікації, нестандартні порти, нетипові часові патерни та інші індикатори потенційної атаки.

Реалізація захисних моделей на базі ШІ є невід'ємною складовою сучасних систем кібербезпеки. Впровадження алгоритмів машинного та глибокого навчання дозволяє створювати адаптивні, гнучкі та ефективні рішення, здатні протистояти як відомим, так і новим кібератакам. У поєднанні з традиційними методами захисту, ШІ значно підвищує рівень безпеки організацій та забезпечує надійний захист від сучасних кіберзагроз.

3.2 Моделювання атак та тестування захисту

Моделювання кібератак і тестування захисних механізмів є ключовими етапами створення та впровадження надійних систем кібербезпеки. Сучасні загрози постійно ускладнюються, активно використовуючи такі технології, як машинне навчання та штучний інтелект (ШІ), тому ефективний захист потребує не лише пасивного реагування, а й прогнозування потенційних атак, оцінки їх наслідків і тестування стійкості систем у реальних умовах.

Моделювання атак на основі штучного інтелекту

Використання ШІ для моделювання кібератак дозволяє створювати реалістичні сценарії, які імітують поведінку зловмисників. Алгоритми машинного навчання можуть автоматично:

- виявляти вразливості у складних системах;
- аналізувати реакцію захисних механізмів;
- адаптуватися до змін у конфігураціях системи;
- прогнозувати нові можливі вектори атак.

Основою таких симуляцій є створення моделей, що відтворюють дії кіберзловмисників, зокрема:

- генерацію шкідливих програм;
- створення фішингових повідомлень;
- обхід систем аутентифікації;

- запуск DDoS-атак;
- злом CAPTCHA за допомогою комп’ютерного зору.

Для цього застосовуються фреймворки на кшталт TensorFlow, Keras, PyTorch, а також симуляційні середовища, наприклад OpenAI Gym, що дають змогу моделювати поведінку атакувальних агентів у контрольованих умовах.

Такі підходи дозволяють проводити безперервні автоматизовані атаки, створюючи умови, максимально наближені до реальних. Це дає змогу глибше оцінити стійкість систем захисту та вдосконалити їх.

Приклади практичного моделювання атак:

1. Злом CAPTCHA за допомогою глибокого навчання

Нейронні мережі навчаються розпізнавати символи на CAPTCHA, використовуючи великі набори зображень. Після навчання алгоритми можуть автоматично проходити перевірку, обходячи захисні механізми.

2. Атаки на системи аутентифікації

Алгоритми ШІ здатні:

- аналізувати слабкі паролі;
- визначати поведінкові шаблони користувачів;
- знаходити типові вразливості в політиках доступу.

Фреймворки, такі як Metasploit, дозволяють інтегрувати машинне навчання для автоматизованого підбору паролів або моделювання складних атак на системи ідентифікації.

Тестування захисних механізмі

Тестування на проникнення та моделювання атак дають змогу оцінити, наскільки ефективно система здатна виявляти та нейтралізувати загрози, зокрема ті, які використовують ШІ.

Основні методи:

1. Penetration Testing (Пентестинг)

Проводиться з використанням інструментів:

Metasploit — моделювання експлойтів,

Nmap — аналіз мережевої інфраструктури,

Burp Suite — тестування веб-додатків.

ШІ застосовується для автоматизації складних або повторюваних атак, швидкого пошуку вразливостей та аналізу великої кількості параметрів мережі.

2. Тестування засобів виявлення вторгнень (IDS)

Системи IDS, такі як Snort або Suricata, можуть бути посилені ШІ, що аналізує мережевий трафік у реальному часі

Під час тестування фахівці запускають:

фішингові атаки

DDoS

SQL-ін’єкції,

атаки на веб-додатки.

Мета — оцінити швидкість реакції IDS, рівень хибних спрацьовувань і здатність адаптуватися до нових патернів.

3. Автоматизоване сканування вразливостей

Інструменти:

Qualys,

OpenVAS

використовують методи ШІ для:

виявлення застарілого ПЗ,

перевірки конфігурацій,

визначення слабких місць у мережі.

Це дозволяє оперативно виправляти проблеми ще до того, як вони стануть критичними.

4. Симуляційні середовища для ШІ-атак

Платформи, такі як Cylance або Darktrace, надають можливість моделювати складні багаторівневі атаки, що використовують ШІ. Це дає змогу

оцінювати поведінку системи під час реальних інцидентів,

тестувати стійкість мережі, додатків та облікових даних,

формувати рекомендації для покращення кіберзахисту

5. Red Team / Blue Team Exercises

Red Team — атакує, використовуючи інструменти ШІ, автоматизує пошук вразливостей;

Blue Team — захищається, застосовуючи ШІ для аналізу аномалій і швидкого реагування.

Такий підхід дозволяє оцінити кібербезпеку з обох сторін — атакуючої та оборонної

6. Тестування соціальної інженерії

Системи ШІ здатні створювати:

- персоналізовані фішингові листи,

- повідомлення через месенджери,

- контент у соціальних мережах.

Це дозволяє оцінювати рівень підготовки користувачів та ефективність політик безпеки.

Оцінювання ефективності методів захисту

Ефективність систем, що використовують ШІ для кіберзахисту, визначається за низкою критеріїв:

1. Швидкість виявлення загроз

Чим швидше система реагує — тим нижчі втрати.

2. Точність виявлення

Необхідно мінімізувати:

- хибнопозитивні спрацьовування (false positives),

- хибнонегативні спрацьовування (false negatives).

3. Стійкість до нових типів атак

Система повинна адаптуватися до еволюції загроз.

4. Масштабованість

Здатність працювати стабільно при збільшенні навантаження.

5. Оптимальне використання ресурсів

Захист не повинен перевантажувати систему.

6. Інтеграція з іншими засобами кіберзахисту

Комплексний підхід підвищує загальний рівень безпеки.

7. Зручність для користувача

Безпека не повинна ускладнювати роботу легітимних користувачів.

Не менш важливою частиною тестування є стійкість моделі до протидії з боку зловмисників. У сучасних умовах активно поширюється підхід adversarial attacks, коли спеціально модифіковані вхідні дані вводять модель в оману, змушуючи її неправильно класифікувати шкідливу активність. Тому тестування повинно включати перевірку моделі на таких ворожих прикладах, що підвищує її надійність і дозволяє розробити механізми захисту від маніпуляцій.

Для підвищення рівня об’єктивності оцінювання результати тестування порівнюються між різними моделями ШІ. Наприклад, традиційні алгоритми виявлення аномалій можуть забезпечувати низький рівень хибнопозитивних спрацювань, але бути менш ефективними у випадку складних атак. Натомість глибинні моделі, такі як LSTM, демонструють високу чутливість до поведінкових патернів, проте потребують значних обчислювальних ресурсів. Такий порівняльний аналіз дозволяє визначити найкращий баланс між продуктивністю та точністю системи.

Для наочності пропонуємо таблицю з підсумком тестування різних категорій моделей ШІ:

Таблиця 3.1

Порівняння результатів тестування моделей ШІ при моделюванні атак

Типи моделей	Переваги під час тестування	Слабкі сторони	Ефективність проти складних атак	Рівень хибних спрацювань
Random Forest	Висока інтерпретованість, швидке навчання	Обмежена здатність	Середня	Низький

		виявляти нові атаки		
SVM	Добра класифікація відомих загроз	Обмежена здатність виявляти нові атаки	Середня	Середній
SVM	Добра класифікація відомих загроз	Потребує ретельної нормалізації даних	Висока	Високий
LSTM	Враховує часові залежності, висока точність	Висока ресурсоємність	Дуже висока	Низький
Isolation Forest	Добре працює зі слабко структурованими даними	Часті false positives у шумному трафіку	Середня–висока	Високий

Отже, результати моделювання атак і тестування ефективності захисних моделей свідчать, що використання штучного інтелекту дозволяє суттєво підвищити здатність систем до своєчасного виявлення складних та нових кібератак. Водночас дослідження підкреслює необхідність комбінування різних підходів, оскільки окремі моделі мають специфічні слабкі місця. У наступному розділі буде наведено узагальнення отриманих результатів і сформульовано ключові висновки щодо застосування ШІ у системах кіберзахисту.

Використання ШІ для моделювання атак і тестування захисту відкриває можливість створення гнучких, адаптивних і високоефективних систем кібербезпеки. Комплексний підхід, що включає пентестинг, симуляції ШІ-атак, автоматизоване сканування вразливостей та аналіз поведінки систем, дозволяє значно підвищити рівень захищеності сучасної інфраструктури.

3.3 Практичне значення одержаних результатів

Практичне значення одержаних результатів полягає у наступному:

1. Розробка методології для виявлення вразливостей критичної інфраструктури: Впровадження цієї методології дозволить підвищити ефективність виявлення потенційних загроз.

2. Створення бази даних вразливостей: Зібрані та проаналізовані дані про вразливості можуть бути використані для подальшого вдосконалення систем захисту.

3. Практичні рекомендації для забезпечення безпеки: Рекомендації, що базуються на результатах дослідження, можуть бути використані для покращення захисту критичної інфраструктури від кіберзагроз.

4. Підвищення обізнаності: Результати дослідження можуть бути використані для навчання та підвищення обізнаності фахівців у галузі кібербезпеки.

Для кожної з вразливостей були розроблені конкретні рекомендації щодо їх усунення, включаючи оновлення програмного забезпечення, налаштування файрволів, використання двофакторної аутентифікації та впровадження систем моніторингу.

Заключні рекомендації

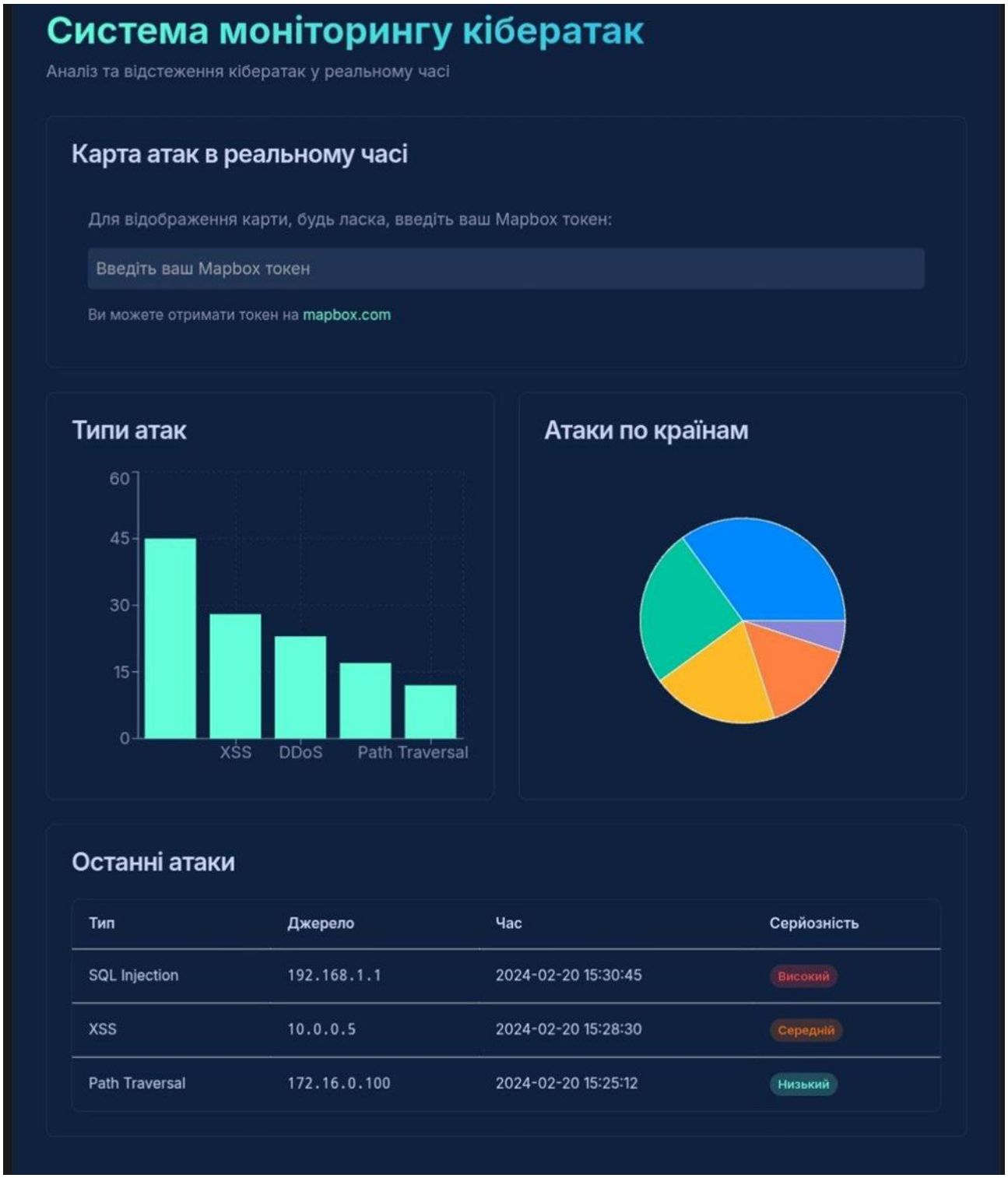
На основі проведеного дослідження можна зробити декілька ключових висновків та рекомендацій:

Регулярне оновлення програмного забезпечення: Необхідно забезпечити своєчасне оновлення всіх компонентів системи для запобігання експлуатації відомих вразливостей.

Обмеження доступу: Використання файрволів та інших засобів контролю доступу для обмеження підключень до критичних сервісів тільки довіреним IP-адресам.

Моніторинг безпеки: Впровадження систем моніторингу, що дозволяють своєчасно виявляти та реагувати на підозрілу активність.

Резервне копіювання даних: Регулярне створення резервних копій для забезпечення цілісності та доступності даних у випадку інцидентів.



Висновок

Швидкий розвиток Інтернету спричинив появу концепції Інтернету речей (англ. Internet of Things, IoT), яка охоплює такі приклади застосування, як розумні будинки та міста, системи охорони здоров'я, промислові та кіберфізичні системи. IoT являє собою сукупність взаємопов'язаних малопотужних пристроїв, якими можна керувати за допомогою веб-сервісів або інших типів інтерфейсів. Очікувано, що будь-яка нова популярна технологія привертає увагу кіберзловмисників, які намагаються зловживати її можливостями з використанням різноманітних технік злому. Ситуацію ускладнює відсутність єдиної стандартизації IoT-платформ та протоколів.

Кібератаки на пристрої IoT можуть завдавати значних збитків компаніям і державним установам, оскільки такі атаки часто є відносно простими в реалізації, а інформація про можливі вектори злому широко доступна. Основна мета подібних атак полягає у порушенні нормального функціонування системи або унеможливленні її коректного обслуговування. Водночас традиційні системи забезпечення безпеки, розроблені для класичних IT-середовищ, часто не підходять для повноцінного захисту IoT-інфраструктури. Тому для безпечного розвитку галузі необхідно впроваджувати практичні, адаптивні та оптимізовані рішення для протидії кіберзагрозам саме в мережах Інтернету речей.

Будь-який пристрій IoT є потенційно вразливим, а дані, що ним збираються чи обробляються, мають високу цінність. Злам або компрометація таких пристроїв можуть завдати шкоди об'єктам критичної інфраструктури, фізичним системам, а також вплинути на безпеку користувачів. Тому надзвичайно важливо своєчасно виявляти вразливості та фіксувати спроби атак на пристрої IoT.

Пристрої Інтернету речей є високорівнево автономними та, як правило, не потребують постійного втручання користувача. Це створює необхідність розробки інтелектуальних рішень кібербезпеки, зокрема моделей машинного навчання (ML), здатних автоматично аналізувати мережевий трафік, виявляти

аномальну поведінку та ефективно реагувати на загрози. Застосування ML-технологій дає змогу створювати адаптивні системи виявлення вторгнень та значно підвищувати рівень захищеності IoT-інфраструктури.

Попри зростання кількості досліджень, сфері виявлення атак у мережах Інтернету речей все ще бракує комплексних рішень, орієнтованих саме на специфіку IoT-систем. У зв'язку з масштабністю ризиків та стрімким розвитком технологій штучного інтелекту (ШІ) питання протидії сучасним загрозам набуває критичного значення для безпеки компаній, державних структур та суспільства загалом. Окрім того, використання ШІ зловмисниками становить окремий вектор загрози, оскільки такі технології можуть автоматизувати атаки, прискорювати пошук вразливостей, обходити системи захисту та створювати складні адаптивні моделі поведінки.

У рамках проведеного дослідження було здійснено глибокий аналіз та порівняння сучасних підходів до кіберзахисту на основі штучного інтелекту. Розглянуто моделі атаки, підходи до тестування систем безпеки та методи оцінки їх ефективності. Зокрема, Generative Adversarial Networks (GANs), ML-підходи для моделювання DDoS-атак та нейронні мережі для обходу CAPTCHA продемонстрували високу ефективність у створенні реалістичних симуляцій кіберзагроз. Генеративні моделі надають змогу формувати складні сценарії атак, але потребують значних обчислювальних ресурсів. Класичні ML-методи, такі як Random Forest чи Support Vector Machines (SVM), ефективно виявляють мережеві аномалії, проте їх продуктивність значною мірою залежить від якості навчальних вибірок.

Також було проаналізовано актуальні інструменти автоматизованого пентестингу, такі як Metasploit та Nmap, які дозволяють ідентифікувати вразливості, однак їхня результативність залежить від налаштувань і досвіду фахівців. Дослідження підтвердило, що штучний інтелект і машинне навчання створюють значні можливості для посилення кіберзахисту, забезпечуючи адаптацію до нових загроз та формування надійніших механізмів оборони.

Перспективними напрямами подальших досліджень є створення адаптивних та самонавчальних систем кіберзахисту, здатних оперативно реагувати на нові види атак та передбачати потенційні загрози. Великий потенціал мають алгоритми аналізу поведінкових патернів, методи прогнозування атак, системи глибокого навчання для детекції складних аномалій, а також інтеграція блокчейн-технологій для забезпечення цілісності даних.

У ширшому контексті машинне навчання може застосовуватися для вирішення широкого спектра задач — від прогнозування цін на нерухомість до аналізу емоцій у текстах та розпізнавання облич. Проте в межах даної дипломної роботи особливу увагу приділено питанню забезпечення безпеки в системах автентифікації та використанню методів пошуку аномалій у логах таких систем. Це дозволяє ефективно виявляти потенційні загрози та підвищувати рівень захищеності інформаційних систем на основі інтелектуальних методів аналізу даних.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Аленічев, І.В. Штучний інтелект у кібербезпеці: сучасні підходи // Наукові праці НТУУ "КПІ". 2022. №4. С. 12–23.

2. Бабенко, О.В. Методи захисту інформаційних систем на основі машинного навчання. К.: Наукова думка, 2021. 320 с.

3. Tymoshchuk, D., Yasniy, O., Mytnyk, M., Zagorodna, N., Tymoshchuk, V., (2024). Detection and classification of DDoS flooding attacks by machine learning methods. CEUR Workshop Proceedings, 3842, pp. 184 - 195.

4. Борисенко, М.А., та ін. Штучний інтелект для протидії фішинговим атакам. – К.: Вид-во «Фенікс», 2021. 288 с.

5. Василенко, Г. В. Моделювання атак за допомогою Generative Adversarial Networks (GANs) // Інформаційні системи та технології. 2022. №2. С. 71–80.

6. Семенюк Андрій Васильович - Інститут докторантури та аспірантури, .Вінницький національний технічний університет, м. Вінниця, e-mail: andrew.semeniuk.university@gmail.com

7. Lypa, B., Horyn, I., Zagorodna, N., Tymoshchuk, D., Lechachenko T., (2024). Comparison of feature extraction tools for network traffic data. CEUR Workshop Proceedings, 3896, pp. 1-11

8. Wikipedia (2023). "Соціальна інженерія". URL: https://uk.wikipedia.org/wiki/Соціальна_інженерія

9. Secureframe (2023). "60+ Social Engineering Statistics for 2023". URL: https://secureframe.com/blog/social-engineering-statistics

10. Crowdstrike (2023). "10 Types of social engineering attacks". URL: https://www.crowdstrike.com/cybersecurity-101/types-of-social-engineering-attacks/

11. Ciaramella, P. DArco, A. De Santis, C. Galdi, R. Tagliaferri. (2006). Neural Network Techniques for Proactive Password Checking. IEEE Transactions on Dependable and Secure Computing, 3(4), 327-339.

12. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. Courville. (2017). Improved training of wasserstein GANs. In Proc. of the 31st International Conference on Neural Information Processing Systems, (pp. 5769-5779).

13. Shodan search engine. (b. d.). Shodan Search Engine. https://www.shodan.io/

14. Le Roux, N., Bengio, Y. (2008). Representational power of restricted Boltzmann machines and deep belief networks. Neural computation, 20(6), 1631-1649.

15. Sharma, B., Mangrulkar, R. (2019). Deep learning applications in cyber security: a comprehensive review, challenges and prospects. International Journal of Engineering Applied Sciences and Technology, 4(8), 148-159

16. Kazennov, A. M. (2010). Basic concepts of CUDA technology. Computer Research and Modeling, 2(3), 295–308. https://doi.org/10.20537/2076-7633-2010-2-3-295-308

17. «Кібербезпека: освіта, наука, техніка»: Том 1 № 5 (2019): Кібербезпека: освіта, наука, техніка

18. Yuliia Zhdanova, Svitlana Spasiteleva , Svitlana Shevchenko, Kateryna Kravchuk, ПРИКЛАДНІ ТА МЕТОДИЧНІ АСПЕКТИ ЗАСТОСУВАННЯ ХЕШ-ФУНКЦІЙ В ІНФОРМАЦІЙНІЙ БЕЗПЕЦІ , Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»: Том 4 № 8 (2020): Кібербезпека: освіта, наука, техніка