МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ПОЛІСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет інформаційних технологій,
обліку та фінансів
Кафедра комп'ютерних технологій
і моделювання систем

Кваліфікаційна робота
на правах рукопису

Сус Василь Юрійович

УДК 004.056:004.94

КВАЛІФІКАЦІЙНА РОБОТА

Розробка моделі прогнозування кіберзагроз на основі великих даних

125 «Кібербезпека та захист інформації»

Подається на здобуття освітнього ступеня магістр

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____
(підпис, ініціали та прізвище здобувача вищої освіти)

Керівник роботи:
Веретюк Сергій Михайлович,
кандидат технічних наук, доцент

Житомир – 2024

Висновок кафедри_____

за результатами попереднього захисту:_____

Протокол засідання кафедри_____

№ ____ від «_____» _____ 20____ р.

Завідувач кафедри_____

_____ _____ _____

(науковій ступінь, вчене звання) (підпис) (прізвище, ім'я, по батькові)

«_____» _____ 20____ р.

Результати захисту кваліфікаційної роботи

Здобувач вищої освіти _____ захистив (ла)

(прізвище ,ім'я, по батькові)

кваліфікаційну роботу з оцінкою:

сума балів за 100-бальною шкалою _____

за шкалою ECT8 _____

за національною шкалою _____

Секретар ЕК

_____ _____ _____

(науковій ступінь, вчене звання) (підпис) (прізвище, ім'я, по батькові)

АНОТАЦІЯ

Дослідження присвячено актуальній проблемі прогнозування кіберзагроз у динамічному середовищі інформаційної безпеки. Зростаюча складність і частота атак, а також критичність вразливостей інформаційних систем вимагають розробки ефективних проактивних підходів для випередження потенційних загроз. У роботі запропоновано метод прогнозування кіберзагроз на основі моделі ARIMA (AutoRegressive Integrated Moving Average), що дозволяє аналізувати часові ряди даних і визначати закономірності в динаміці вразливостей.

Розроблений підхід інтегрується з великими даними, зокрема базами CVE (Common Vulnerabilities and Exposures), для аналізу частотності появи нових загроз та їхньої критичності. Проведено аналіз існуючих підходів до прогнозування, таких як методи OSINT, Cyber Kill Chain та ARIMA, і обґрунтовано доцільність використання часових рядів для ідентифікації трендів і аномалій у вразливостях.

У процесі апробації моделі було виконано оцінку її ефективності, включаючи точність прогнозування короткострокових змін у динаміці загроз. Аналіз результатів підтвердив, що ARIMA демонструє високу точність у прогнозуванні частотності та критичності вразливостей. Крім того, інтеграція моделі з платформами моніторингу та обробки великих даних забезпечує можливість автоматизації аналізу в реальному часі.

Практична цінність дослідження полягає в можливості застосування методу для побудови проактивних стратегій кіберзахисту, що включають передбачення нових загроз, оптимізацію розподілу ресурсів для реагування на інциденти та зниження загального впливу атак на інформаційні системи. Запропонований підхід забезпечує адаптивність і масштабованість, що важливо для великих організацій, зокрема в державному секторі та критичній інфраструктурі.

Таким чином, результати роботи сприяють підвищенню стійкості інформаційних систем до сучасних викликів у сфері кібербезпеки та створюють наукове підґрунтя для подальших досліджень у галузі прогнозування загроз на основі великих даних.

Робота містить 42 сторінки, 14 рисунків, 2 таблиці, 32 літературних джерела.

SUMMARY

The study addresses a pressing issue of forecasting cyber threats in the dynamic environment of information security. The growing complexity and frequency of attacks, coupled with the critical nature of information system vulnerabilities, necessitate the development of effective proactive approaches to anticipate potential threats. This paper proposes a method for cyber threat forecasting based on the ARIMA (AutoRegressive Integrated Moving Average) model, which enables time series analysis and pattern identification in the dynamics of vulnerabilities.

The developed approach integrates with big data, particularly CVE (Common Vulnerabilities and Exposures) databases, to analyze the frequency and criticality of emerging threats. The research includes an analysis of existing forecasting methods, such as OSINT, Cyber Kill Chain, and ARIMA, substantiating the applicability of time series analysis for identifying trends and anomalies in vulnerabilities.

During model validation, the effectiveness of the proposed approach was assessed, including the accuracy of short-term predictions of threat dynamics. The results confirmed that ARIMA demonstrates high accuracy in forecasting the frequency and criticality of vulnerabilities.

The practical value of the study lies in the applicability of the method for developing proactive cybersecurity strategies. These strategies include predicting emerging threats, optimizing resource allocation for incident response, and mitigating the overall impact of attacks on information systems. The proposed approach ensures adaptability and scalability, which are essential for large organizations, particularly in the public sector and critical infrastructure.

Thus, the findings contribute to enhancing the resilience of information systems to modern cybersecurity challenges and provide a scientific foundation for further research in threat forecasting using big data.

The work comprises 42 pages, 14 figures, 2 tables, and 32 references.

ЗМІСТ

ВСТУП

Сучасний етап розвитку цифрових технологій супроводжується значним зростанням кількості та складності кіберзагроз, які ставлять під загрозу безпеку інформаційних систем у різних галузях. Це включає атаки на державні установи, бізнес-структури та приватних користувачів, що супроводжується фінансовими збитками, порушенням конфіденційності даних та втратами репутації. Ефективне прогнозування кіберзагроз стає необхідною складовою для забезпечення проактивного захисту, зменшення ризиків і мінімізації наслідків атак.

Актуальність теми дослідження зумовлена необхідністю впровадження інноваційних підходів до аналізу та прогнозування кіберзагроз. Використання моделі ARIMA, у поєднанні з технологіями аналізу великих даних, відкриває нові можливості для виявлення тенденцій, закономірностей і потенційних ризиків у сфері кібербезпеки. Це дозволяє розробляти стратегії захисту, адаптовані до динамічних змін у цифровому середовищі.

Кваліфікаційна робота присвячена дослідженню процесів прогнозування кіберзагроз з використанням методів аналізу часових рядів та обробки великих даних. У роботі систематизовано теоретичні основи прогнозування кіберзагроз, проведено аналіз існуючих підходів та моделей, а також визначено роль великих даних у побудові ефективних моделей захисту інформаційних систем. У практичній частині роботи розроблено метод прогнозування кіберзагроз, що базується на моделі ARIMA, адаптованій до обробки великих обсягів даних. Цей метод дозволяє передбачати динаміку загроз, зокрема частотність нових вразливостей, їх критичність та потенційний вплив на інформаційні системи. Апробація розробленого методу продемонструвала його ефективність у задачах управління кіберризиками та формуванні проактивних заходів кіберзахисту.

Мета: удосконалення та розробка методу прогнозування кіберзагроз на основі аналізу відкритих великих даних із застосуванням моделі ARIMA.

Об’єкт: процеси прогнозування кіберзагроз.

Предмет: Методи аналізу та прогнозування кіберзагроз із застосуванням моделей обробки великих даних і часових рядів.

Завдання кваліфікаційної роботи:

1. Провести аналіз відкритих даних щодо інформації про підтверджені вразливості.
2. Дослідити математичні моделі аналізу часових рядів для прогнозування появи нових вразливостей.
3. Розробити метод прогнозування кіберзагроз на основі відкритих даних.
4. Розробити практичні рекомендації щодо застосування розробленого методу

Наукова новизна роботи:

Удосконалено метод прогнозування кіберзагроз на основі інтеграції моделі ARIMA із великими даними, що дозволяє враховувати залежності в часових рядах для точного аналізу та передбачення динаміки загроз. Розроблений підхід дозволяє динамічно оновлювати модель ARIMA на основі нових даних, що забезпечує адаптацію до змін у сфері кіберзагроз і підвищення актуальності прогнозів. Виконано перевірку моделі на даних CVE, що підтвердило її високу точність і ефективність у короткостроковому прогнозуванні (з точністю понад 85% для горизонту в 3 місяці).

Розроблено метод оцінки критичності вразливостей на основі аналізу історичних даних із застосуванням відкритих баз даних, таких як CVE, що забезпечує підвищення точності оцінки кіберризиків.

Запропоновано алгоритм прогнозування, який забезпечує ефективне розподілення ресурсів для протидії загрозам за рахунок пріоритизації найбільш критичних вразливостей.

Практична цінність отриманих результатів:

Розроблений метод прогнозування дозволяє організаціям передбачати появу нових вразливостей і оптимізувати управління кіберризиками, а отже оптимізувати задіяння ресурсів (кошти, обладнання, програмне забезпечення, персонал, час)

Інтеграція з платформами моніторингу та системами підтримки прийнятт рішень забезпечує автоматизацію процесів аналізу та прогнозування загроз у реальному часі.

Застосування методу ARIMA сприяє побудові проактивних стратегій кіберзахисту, що надає можливість економити час і ресурси на виявлення й усунення критичних вразливостей.

За темою кваліфікаційної роботи опубліковано наукові публікації, а саме:

- Сус В.Ю. Аналіз моделі прогнозування кіберзагроз «Cyber Kill Chain». Моделювання, керування та інформаційні технології (МСІТ–2024) : збірник праць учасників Міжнародної науково-практичної конференції, 2024. 367 с.

- Сус В.Ю. Огляд відкритих даних для прогнозування та оцінки кіберзагроз. Litteris et Artibus: Нові горизонти : збірник матеріалів Всеукраїнської науково-практичної конференції. Випуск IX / за заг. ред. О. В. Тригуби. Кременець : ВЦ КОГПА ім. Тараса Шевченка, 2024. 358 с.

- Сус В.Ю. Підхід до прогнозування частності та критичності нових вразливостей на основі відкритих даних. Безпека, технології, інновації: нові горизонти : збірник праць учасників міжфакультетської науково-практичної інтернет-конференції здобувачів вищої освіти і молодих вчених, 12 листопада 2024 р. Житомир : Поліський національний університет, 2024. 102 с.

Робота виконана на 42 сторінках, містить 14 рисунків, 2 таблиці, 32 літературних джерела.

РОЗДІЛ 1. ТЕОРЕТИЧНІ ОСНОВИ ПРОГНОЗУВАННЯ КІБЕРЗАГРОЗ

1.1 Основні поняття прогнозування кіберзагроз

Прогнозування кіберзагроз є важливим елементом кібербезпеки, спрямованим на випередження зловмисників та запобігання потенційним атакам на інформаційні системи. Цей процес охоплює використання сучасних технологій і методів аналізу даних для передбачення майбутніх кібератак, що дозволяє забезпечити проактивний захист.

Вразливість – це потенційна слабкість або недолік у системі, програмі або процесі, який може бути використаний зловмисниками для виконання атак або неправомірного доступу до інформації. Це може включати помилки в програмному забезпеченні, недостатнє керування доступом, або недоліки в фізичній безпеці обладнання. Вразливість є точкою входу для потенційних загроз.

Загроза – це будь-яка дія, подія або обставина, яка може спричинити шкоду інформаційним ресурсам, порушення конфіденційності, цілісності або доступності даних. Загрози можуть включати як внутрішні, так і зовнішні фактори, такі як кібератаки, зловмисне програмне забезпечення, природні катастрофи або помилки людського фактору [1]. Важливо відзначити, що загроза без вразливості може бути неефективною, як і вразливість без загрози може залишатися неексплуатованою.

Ризик є результатом взаємодії між вразливістю та загрозою. Це ймовірність того, що вразливість буде використана загрозою для здійснення атаки. Ризик включає як ймовірність виникнення загрози, так і масштаб можливих наслідків. Ризик може варіюватися від низького до високого залежно від того, наскільки критична вразливість та наскільки ймовірна її експлуатація.

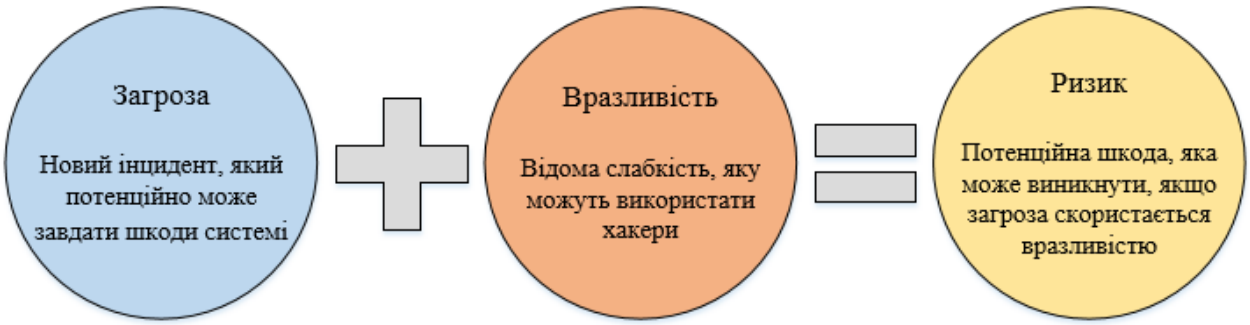

Рис. 1.1 - Взаємозв'язок між загрозою, вразливістю та ризиком

Розуміння взаємозв'язку між загрозою, вразливістю та ризиком дозволяє краще оцінювати рівень небезпеки, з яким може зіткнутися система. Однак для ефективного управління ризиками недостатньо лише знати формулу їх утворення — необхідно мати доступ до актуальних і точних даних, що відображають стан системи в реальному часі. Саме тут ключову роль відіграє моніторинг, який забезпечує збирання та аналіз інформації про потенційні загрози, наявні вразливості та інші критично важливі параметри безпеки [3].

Моніторинг кіберзагроз базується на збиранні та аналізі ключових даних, що дозволяють відстежувати потенційні загрози та вчасно реагувати на них. Ці дані охоплюють різні аспекти активності в мережі, поведінки користувачів, а також вразливостей у системах та програмах.

Ключові дані для моніторингу становлять собою основу для побудови прогнозів і оцінки ризиків, оскільки саме вони відображають динаміку досліджуваних явищ і процесів. Збір та аналіз таких даних дозволяє ідентифікувати закономірності та аномалії, що можуть вказувати на потенційні загрози або перспективні можливості.

Основними джерелами для моніторингу загроз є лог-файли систем та додатків, які фіксують події в мережі, такі як спроби входу, зміни прав доступу та аномальні активності, що дозволяють виявляти підозрілі дії. Також

важливими є дані телеметрії, що включають показники продуктивності, трафік мережі та частоту запитів, які допомагають виявити аномалії, наприклад, DDoS-атаки або спроби проникнення. Інші джерела включають інформацію про вразливості в системах та додатках, виявлені за допомогою сканерів вразливостей і патч-менеджменту, що дозволяє вчасно усувати слабкі місця. Окрім того, дані з розвідки загроз (Threat Intelligence) надають інформацію про нові типи атак, тактики зловмисників та індикатори компрометації , що допомагають оцінити ймовірність і потенційний вплив загроз [3]. Основні маркери прогнозування стану систем наведені у таблиці 1.1.

Таблиця 1.1 – Основні маркери для прогнозування стану системи

Джерело даних	Маркер	Важливість маркерів
Лог-файли систем і додатків	- Спроби входу - Зміни прав доступу - Збої в системі - Аномальні активності	Допомагають виявити підозрілі дії на ранніх етапах, що дозволяє швидко реагувати на потенційні загрози або атаки.
Телеметрія	- Піки мережевого трафіку - Раптове збільшення використання ресурсів - Нетипові запити	Виявляє аномалії, які можуть бути індикаторами атак, таких як DDoS або несанкціоновані спроби доступу, дозволяючи оперативно реагувати.
Сканери вразливостей і патч-менеджмент	- Виявлені вразливості - Патчі, що не були застосовані - Застарілі системи	Прогнозує потенційні слабкі місця в системі, дозволяє вжити заходів до того, як ці вразливості будуть використані для атаки.
Threat Intelligence	- Нові типи атак - Техніки і тактики зловмисників	Допомагають розпізнати нові загрози і оцінити їх потенційний вплив на організацію, дозволяючи адаптувати стратегію безпеки.

Маркери в даних для прогнозування стану системи – це ключові показники або сигнали, які можна відстежувати в реальному часі чи аналізувати історично для виявлення потенційних загроз, аномалій або небезпечних станів у роботі інформаційної системи.

1.2 Аналіз існуючих моделей прогнозування кіберзагроз

Моделі прогнозування кіберзагроз включають системи та методи, які дозволяють передбачати можливі кіберінциденти, зокрема, атаки, витоки даних та інші загрози в кіберпросторі. Ці моделі здебільшого базуються на аналізі великої кількості даних про попередні інциденти, індикатори зламів та інші сигнали безпеки, що дозволяє виявляти тенденції та потенційні загрози. Серед найбільш широко використовуваних моделей у сфері кібербезпеки особливе місце займають Cyber Kill Chain, OSINT (Open Source Intelligence) та ARIMA (AutoRegressive Integrated Moving Average).

Cyber Kill Chain – це модель, розроблена компанією Lockheed Martin у 2011 році. В умовах зростаючого числа складних кібератак і появи нових типів загроз компанії потребували систематизованого підходу до аналізу та протидії атакам. Lockheed Martin застосувала концепцію kill chain (ланцюг знищення), що використовується у військових операціях для опису послідовності дій, необхідних для успішного нападу на противника. Військовий ланцюг складався з етапів розвідки, націлювання, атак і завершення місії [4]. Модель Cyber Kill Chain, є частиною концепції Intelligence Driven Defense для ідентифікації та запобігання кіберзагрозам. Модель була представлена таким чином, що складається з семи етапів: розвідка, озброєння, доставка, експлуатація, встановлення, командування та контроль та дії на об’єкті [5].

У компанії Lockheed Martin кібератака проходить через серію чітко визначених етапів або процесів, кожен з яких є необхідним для успішного завершення атаки. Якщо захисник зможе заблокувати хоча б один із цих кроків, зловмисник не зможе перейти до наступного етапу, що порушить всю атаку і зробить її неефективною. Цей підхід базується на ідеї про те, що кібератаки — це складні та багатоступеневі операції, які потребують точного виконання кожного етапу для досягнення мети. Відповідно, кожен етап атаки — від збору

інформації та підготовки інструментів до встановлення контролю і виконання злочинних дій – може бути ціллю для контрзаходів [4].

OSINT (Open Source Intelligence) є моделлю, що орієнтується на збір і аналіз відкритих джерел інформації для отримання розвідувальних даних, зокрема в галузі кібербезпеки. Вона базується на виявленні, обробці та інтерпретації інформації, яка є загальнодоступною, як-от вебсайти, соціальні мережі, публічні реєстри, новинні ресурси та інші відкриті джерела. У контексті кібербезпеки OSINT дозволяє прогнозувати потенційні загрози, аналізуючи зовнішні сигнали, такі як активність у даркнеті, витоки даних чи технічні індикатори атак. Ця модель забезпечує організації можливість попередньо оцінювати ризики, виявляти уразливості та формувати стратегії захисту, використовуючи переваги доступу до великого обсягу загальнодоступної інформації [6].

ARIMA (AutoRegressive Integrated Moving Average) є популярною моделлю для аналізу часових рядів у прогнозуванні кіберзагроз. Вона використовується для моделювання та передбачення сплесків атак, таких як DDoS, або для виявлення довгострокових тенденцій і сезонності в загрозах [11]. Завдяки поєднанню авторегресії, інтеграції та ковзного середнього, ARIMA дозволяє точно аналізувати історичні дані, наприклад, журнали мережевого трафіку чи записи про інциденти зловмисної активності. Модель особливо ефективна, коли дані демонструють стаціонарні властивості після попередньої обробки. Порівняльна таблиця існуючих моделей наведена в таблиці 1.2.

Всі наведені моделі мають суттєві відмінності у своїх підходах до кібербезпеки. OSINT зосереджується на превентивному аналізі та прогнозуванні, використовуючи інформацію, отриману з відкритих джерел. Cyber Kill Chain, у свою чергу, спрямована на реагування та зупинку атак у режимі реального часу, ідентифікуючи їх структуру на основі етапів виконання, Водночас Cyber Kill Chain демонструє високу ефективність у

запобіганні складним атакам завдяки своїй деталізованій моделі дій зловмисника [8].

Таблиця 1.2 – Порівняльний аналіз OSINT та Cyber Kill Chain

Критерій	OSINT	Cyber Kill Chain	ARIMA
Мета	Збір відкритих даних для прогнозування загроз та ідентифікації ризиків.	Аналіз і зупинка кібератак ляхом переривання ключових етапів атаки.	Прогнозування кількості кібератак чи їх інтенсивності на основі історичних часових рядів.
Тип підходу	Превентивний (попередження ризиків до їх виникнення).	Респонсивний (зупинка атак на різних етапах виконання).	Аналітичний (моделювання та аналіз часових рядів для передбачення загроз).
Основні етапи	Збір, аналіз, створення звітів, інтеграція.	Розвідка, озброєння, доставка, експлуатація, встановлення, командування та контроль, дії.	Перевірка стаціонарності даних, Параметризація (AR, I, MA), Прогнозування.
Джерела даних	Відкриті джерела (публічні реєстри, соціальні мережі, форуми, даркнет, OSINT-інструменти).	Внутрішні дані системи, телеметрія, сигнали зламу, вразливості системи.	Логи подій, системний трафік, записи IDS/IPS, мережевий трафік, дані з бази кіберінцидентів.
Сфера Застосування	Виявлення нових тенденцій, моніторинг зовнішнього середовища, попередження атак. Загальний огляд ризиків та тенденцій.	Локалізація та зупинка атак на рівні інфраструктури. Деталізовані етапи дій зловмисника.	Прогнозування сплесків атак, визначення трендів загроз; аналіз сезонності.

Перевага OSINT полягає в її здатності забезпечувати гнучкість і масштабованість аналізу ризиків без необхідності доступу до внутрішньої інфраструктури організації. На відміну від OSINT та Cyber Kill Chain, ARIMA працює з глибоким аналізом структурованої інформації та передбачає майбутні загрози. Перевагою моделі ARIMA є можливість аналізу часових рядів, враховуючи сезонність, тренди та сплески атак, що дозволяє виявляти закономірність в даних про вразливості.

1.3 Роль великих даних у прогнозуванні кіберзагроз

1.3.1 Використання великих даних у прогнозуванні кіберзагроз

Використання відкритих даних у сфері прогнозування та оцінки кіберзагроз відіграє ключову роль у класифікації інформації про кібератаки, що сприяє розробці високоточних моделей для виявлення та попередження кіберзлочинів. Аналіз відкритих джерел даних дозволяє не лише детально досліджувати характеристики відомих типів атак, але й ідентифікувати нові шаблони та тенденції у поведінці кіберзлочинців. Такий підхід забезпечує більш проактивну стратегію захисту інформаційних систем, спрямовану на зниження ризиків і мінімізацію наслідків потенційних кібератак [9].

Популярним джерелом для ознайомлення із відкритими даними є сервіс Kaggle. На сервісі Kaggle зібрано близько 400 тисяч датасетів, за допомогою, яких є змога аналізувати записи для прогнозування кіберзагроз. Прикладом датасетів із сервісу Kaggle є набір даних "UNSW-NB15 Dataset". Це один із популярних наборів даних для аналізу та досліджень у сфері кібербезпеки, зокрема для виявлення мережевих атак [10]. На основі даних з цього набору можна легко класифікувати різні види атак.

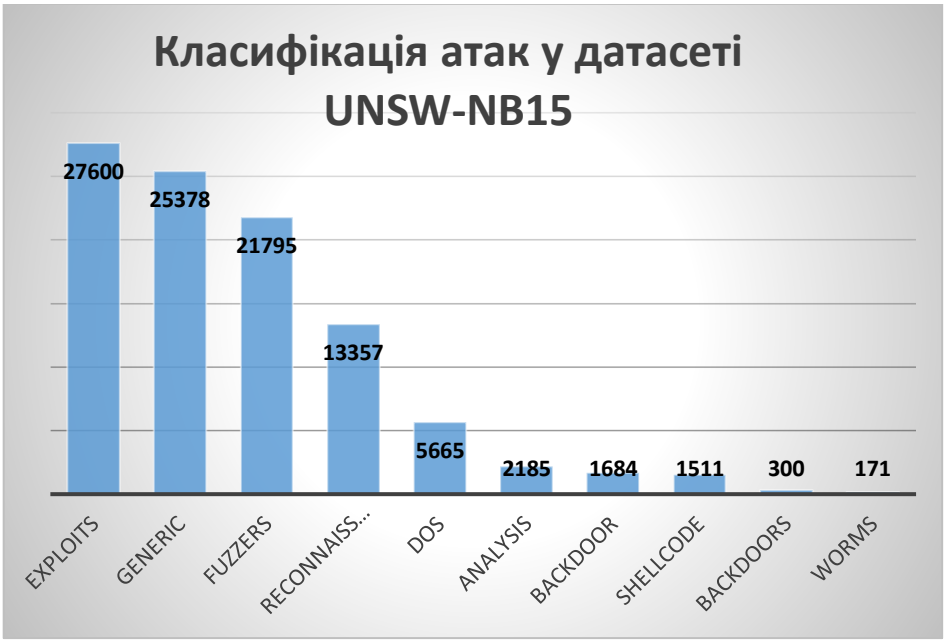

Рис. 1.2 – класифікація атак у датасеті UNSW-NB15

1.3.2 Огляд відкритих великих даних для кібербезпеки

Великі дані є критично важливим елементом у процесі розробки моделей прогнозування кіберзагроз. Сервіс CVEdetails надає доступ до значної бази даних, яка містить інформацію про вразливості програмного забезпечення, його версії та рівень ризику. Платформа регулярно оновлює інформацію про нові вразливості та проводить їх класифікацію, що робить її корисним інструментом для розробки моделей прогнозування.

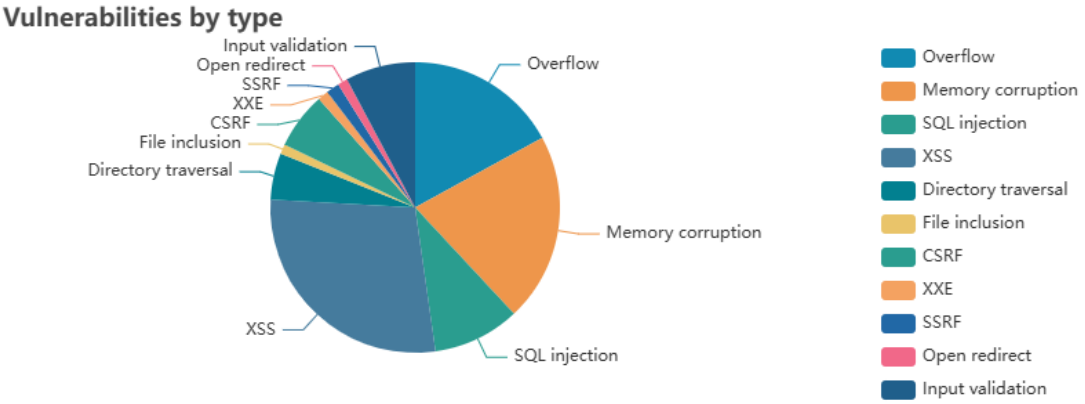

Рис 1.3 – розподіл вразливостей за типом на сервісі CVEdetails

За допомогою CVEdetails можна здійснювати аналіз нових вразливостей, що виникають на глобальному рівні. Сервіс постійно оновлює свої записи та надає їм оцінку "небезпечності", що дозволяє точно оцінити ступінь загрози. Окрім того, система сортує інформацію за організаціями, що надає змогу здійснювати детальний аналіз загроз, з якими зіткнулися різні організації.

▾ організація	≎ Домен
1E Limited	1e.com
360 Security Technology, Inc.	360.cn
42Gears Mobility Systems Pvt Ltd	42gears.com
9передній	9front.org
Acronis International GmbH	acronis.com
Adobe Systems Incorporated	adobe.com
Advanced Micro Devices Inc.	amd.com
Airbus	airbus.com
AlgoSec	algosec.com
Alias Robotics SL	aliasrobotics.com
Alibaba, Inc.	list.alibaba-inc.com
AMI	ami.com
Обчислення Ампера	amperecomputing.com

Рис. 1.4 – сортування записів про загрози з якими зіткнулися різні організації

Великі дані є основою сучасних підходів до прогнозування кіберзагроз, оскільки вони дозволяють здійснювати детальний аналіз великих обсягів інформації, що надходить з різних джерел. Завдяки таким сервісам, як CVEdetails, що надають доступ до баз даних вразливостей, можна здійснювати моніторинг загроз у реальному часі, класифікувати їх за рівнем небезпеки та аналізувати потенційні ризики для різних організацій. Це забезпечує можливість своєчасного виявлення вразливостей, мінімізації їхніх наслідків і запобігання атакам [9].

Висновки до розділу 1

Актуальність обраної теми дослідження визначається швидким зростанням складності та частоти кіберзагроз у сучасному цифровому середовищі, що супроводжується постійним ускладненням методів атак. Ефективне прогнозування кіберзагроз, засноване на аналізі великих даних, дозволяє не лише ідентифікувати потенційні ризики, а й своєчасно реагувати на можливі атаки, знижуючи їх вплив на інформаційні системи.

У розділі проаналізовано основні поняття, моделі та роль великих даних у прогнозуванні кіберзагроз. Проведене дослідження дозволило узагальнити ключові аспекти, що визначають ефективність прогнозування у сфері кібербезпеки. Встановлено, що прогнозування кіберзагроз є багатокомпонентним процесом, який включає аналіз взаємозв’язків між вразливістю, загрозами та ризиками. Важливість чіткого розуміння цих взаємозв’язків підкреслюється необхідністю інтегрованого підходу до оцінювання потенційних загроз, включаючи моніторинг. Проведено Аналіз існуючих моделей прогнозування кіберзагроз. Аналіз показав різну ефективність моделей у різних контекстах. Модель Cyber Kill Chain дозволяє систематизувати аналіз та переривання атак, OSINT прогнозує загрози за даними з відкритих джерел, а ARIMA виявляє тенденції та аномалії за

часовими рядами. Особлива увага приділена ролі великих даних у прогнозуванні кіберзагроз. Застосування відкритих даних і спеціалізованих платформ дозволяє отримувати точну та актуальну інформацію про уразливості й типи загроз.

РОЗДІЛ 2. РОЗРОБЛЕННЯ МЕТОДУ ПРОГНОЗУВАННЯ КІБЕРЗАГРОЗ

2.1 Використання методу ARIMA для прогнозування часових рядів.

Прогнозування часових рядів є важливим інструментом аналізу, який застосовується в кібербезпеці. Часовий ряд — це впорядкована послідовність значень, що змінюються в часі, наприклад, щоденні реєстрації кібератак або щотижнева кількість нових вразливостей. Основною метою прогнозування часових рядів є передбачення майбутніх значень на основі історичних даних. У випадку кібербезпеки це дозволяє виявляти закономірності в активності загроз, планувати захисні заходи та оптимізувати управління ризиками.

Модель ARIMA (Autoregressive Integrated Moving Average) є ефективним методом аналізу часових рядів і прогнозування, заснованим на виявленні статистичних залежностей між минулими та майбутніми значеннями. Модель демонструє високу точність у випадках, коли часові ряди відображають чіткі патерни залежності, наприклад, щоденну частоту кіберінцидентів.

Модель ARIMA успішно застосовується для прогнозування кількості кіберінцидентів на основі історичних даних. Наприклад, у випадку аналізу атак типу DDoS ARIMA показала здатність ідентифікувати сезонні патерни, такі як збільшення активності під час святкових періодів. У дослідженні, опублікованому в журналі Journal of Information Security [12], зазначено, що для отримання точних прогнозів модель повинна працювати з очищеними даними та враховувати автокореляції, наприклад, між кількістю атак у попередні дні та їх імовірністю в майбутньому.

У моделі ARIMA обчислення базуються на трьох ключових компонентах: авторегресії (AR), інтеграції (I), яка усуває тренд і робить ряд стаціонарним, та ковзному середньому (MA), що враховує шум у даних. Завдяки цим складовим модель може адаптуватися до різних типів залежностей у часових рядах [13].

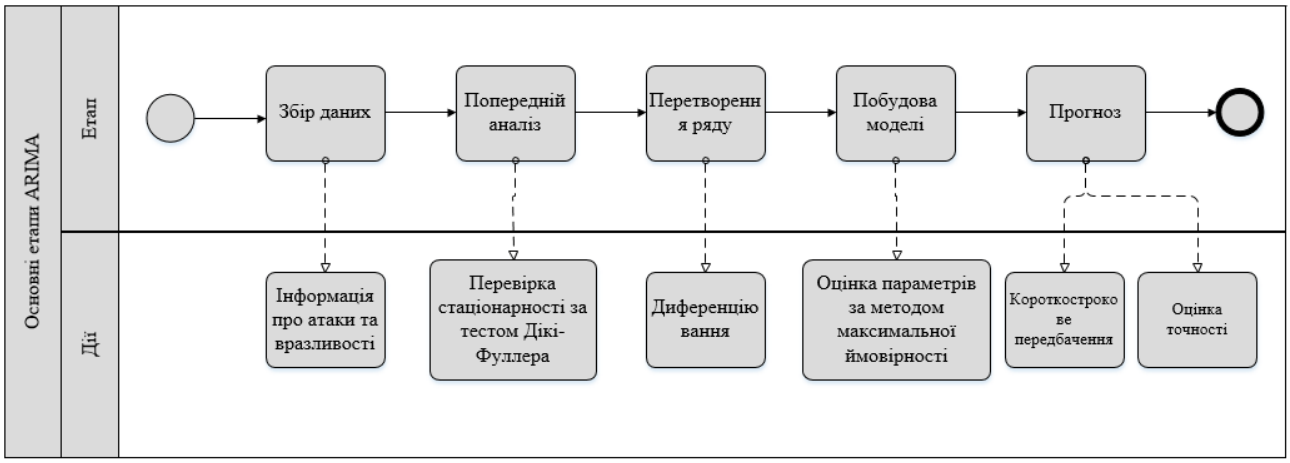

Рис 2.1 - Схема основних етапів роботи моделі ARIMA (авторська доробка)

2.2 Представлення даних щодо інцидентів, вразливостей та загроз у вигляді часового ряду.

Ефективний аналіз і прогнозування кіберзагроз значною мірою залежать від того, наскільки структуровано й точно представлені дані про інциденти, вразливості та загрози. Часовий ряд, що являє собою впорядковану послідовність спостережень за визначеними часовими інтервалами, є оптимальним форматом для аналізу такої динамічної інформації. Представлення даних у вигляді часових рядів дозволяє ідентифікувати тренди, сезонність та інші патерни, які важливі для виявлення потенційних ризиків і розробки проактивних заходів безпеки.

Дані про кіберінциденти можуть бути отримані з різних джерел, включаючи великі дані. Великі дані, зокрема ті, що зібрані з баз даних про вразливості, таких як CVE, забезпечують контекст, багатовимірність і масштабність аналізу.

Common Vulnerabilities and Exposures (CVE) — це стандартизована база даних, яка використовується для класифікації відомих вразливостей програмного забезпечення. Вона забезпечує ідентифікацію загроз за допомогою унікальних ідентифікаторів та використання системи оцінки CVSS

(Common Vulnerability Scoring System), яка визначає ступінь критичності кожної вразливості. У рамках аналітики кіберзагроз, CVE-дані є важливим джерелом для моделювання потенційних атак і планування превентивних заходів [15]. Особливістю великих даних є їх нерівномірність та висока мінливість.

Часові ряди, побудовані на основі CVE, дають змогу аналізувати динаміку появи нових вразливостей та пов'язаних із ними загроз. Наприклад, використовуючи часові мітки публікації CVE, можна виявити закономірності у темпах розкриття вразливостей для різних типів програмного забезпечення або операційних систем. Такі дані також відображають сезонні або періодичні патерни, пов’язані із запуском нових продуктів, випуском оновлень або активністю хакерських угруповань, які швидко експлуатують опубліковані уразливості. У дослідженні, представленому в Cybersecurity Big Data Analytics Review [16], зазначено, що аналіз часових рядів на основі CVE дозволяє виявляти піки активності, які співпадають із великими релізами програмного забезпечення.

У рамках великих даних аналіз CVE стає особливо цінним, оскільки репозиторій містить тисячі записів, що постійно оновлюються. Було обрано 10000 записів виявлених вразливостей з 1999 по 2011 р.

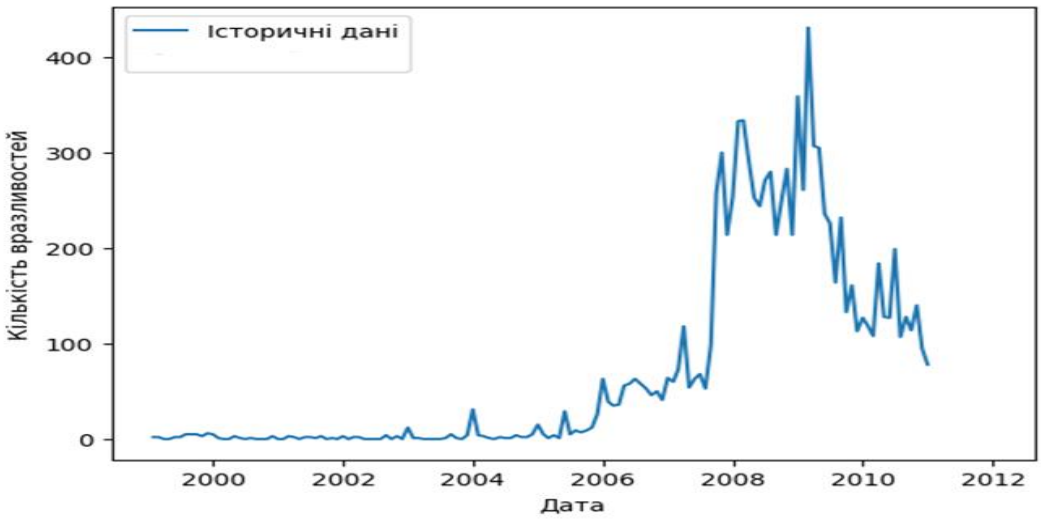

Рис 2.2 – Динаміка виявлених вразливостей (CVE [15])

2.3 Розроблення методу прогнозування динаміки загроз на основі ARIMA

Для математичного опису прогнозування частотності нових вразливостей використовується модель ARIMA, яка поєднує три складові: авторегресію, інтеграцію та ковзне середнє. Основна ідея моделі полягає у використанні взаємозв'язків між попередніми спостереженнями в часовому ряді та випадковими похибками для передбачення майбутніх значень. Це дозволяє виявляти тенденції, цикли та аномалії у даних про кількість нових вразливостей, що публікуються в датасетах класифікації типу CVE [18].

Авторегресійна частина (AR) моделі забезпечує врахування залежності поточного значення ряду від попередніх.

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \epsilon_t$$

$\phi_1, \phi_2, \ldots, \phi_p$ – параметри авторегресії , p — порядок моделі AR, а ϵ_t — випадкова похибка з нульовим середнім і сталою дисперсією. Цей компонент дозволяє відстежувати регулярні цикли у даних.

Інтеграція (I) застосовується для усунення нелінійних трендів, які можуть призводити до нестабільності в моделюванні, через перетворення даних у стаціонарний ряд. Таке перетворення досягається шляхом диференціюванням різниць між значеннями ряду, що забезпечує їх статистичну однорідність. Якщо $d = 1$, то диференційований ряд Y_t' визначається як:

$$Y_t' = Y_t - Y_{t-1}.$$

Ковзне середнє (MA): Враховує вплив попередніх помилок у прогнозах на поточне значення.

$$Y_t = \mu + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

$\theta_1, \theta_2, \ldots, \theta_q$ — параметри ковзного середнього, а q — порядок моделі MA [18].

Усі три компоненти об'єднуються в єдине рівняння ARIMA(p, d, q), що включає в себе модель прогнозування Y_t, та записується як:

$$\Delta^d Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

Δ^d. Кожна складова впливає на підсумковий прогноз, забезпечуючи як локальну, так і глобальну адаптацію до змін у даних.

Процес побудови моделі включає кілька етапів. Попередній аналіз ряду здійснюється для перевірки його стаціонарності за допомогою статистичних тестів. У разі виявлення тренду чи сезонності застосовується диференціювання, що дозволяє усунути довгострокові залежності. Далі визначаються оптимальні параметри p, d, q моделі за допомогою аналізу автокореляційної (ACF) та часткової автокореляційної (PACF) функцій для визначення значущих значень p і q. Ці функції вказують на потенційні значення порядків авторегресії та ковзного середнього. Вибір підходящої моделі здійснюється шляхом оцінки критеріїв якості, таких як інформаційний критерій Акаіке (AIC) чи критерій Байєса (BIC). Критерії Акаіке та Байєса, які дозволяють порівнювати моделі та оцінювати їхню ефективність. Обидва критерії базуються на понятті правдоподібності, однак вони мають різні підходи до балансування між якістю підгонки моделі та її складністю. Критерій Акаіке є інформаційним критерієм, що спрямований на мінімізацію інформаційної втрати при побудові статистичних моделей. Критерій Байєса, у свою чергу, включає додатковий фактор, пов'язаний із розміром вибірки, цей підхід робить BIC більш чутливим до надлишкових параметрів.

Після визначення параметрів моделі та їх оцінки прогнозування здійснюється через ітеративне обчислення прогнозних значень. Для цього використовуються рекурсивні методи, які враховують історичні дані ряду та внесок кожного з компонентів моделі. Такий підхід забезпечує високу точність у прогнозуванні, особливо при короткострокових горизонтах, і дозволяє моделі адаптуватися до динамічних змін у кількості вразливостей, сприяючи більш ефективному плануванню кіберзахисту.

Прогнозування критичності нових вразливостей можна здійснювати, використовуючи інформацію з минулих вразливостей, таких як тип вразливості, тип програмного забезпечення, вплив на конфіденційність, цілісність і доступність. Основною цільовою змінною в цьому випадку буде CVSS-оцінка критичності [19]

Для прогнозування критичності використовують такі моделі:

1. Регресія для передбачення числової оцінки CVSS

- Лінійна регресія: Припускає лінійну залежність між критичністю та ознаками.
- Ридж-регресія: Додає регуляризацію для уникнення переобучення, коли ознак багато.
- Регресійні дерева або ансамблеві методи (Random Forest, Gradient Boosting): Враховує нелінійні залежності та взаємодії між ознаками.

2. Класифікація для передбачення категорій критичності

- Random Forest, Decision Tree: Дерева ухвалення рішень добре підходять для класифікації на категорії критичності.
- Градієнтний бустинг (XGBoost): Покращена ансамблева модель, яка поєднує предиктори, що будуються послідовно, з метою підвищення точності.

Прогнозування частотності та критичності вразливостей дозволяє організаціям ефективніше керувати кіберризиками, зосереджуючи ресурси на найбільш небезпечних загрозах. Це підвищує захищеність, мінімізує

потенційні збитки та підтримує безперервність бізнес-процесів, що є критичним у сучасному цифровому середовищі [18].

На основі проведеного аналізу та використання моделі ARIMA можна розробити ефективний підхід для прогнозування кіберзагроз із застосуванням великих даних. Головна ідея методу полягає у поєднанні можливостей часових рядів для виявлення закономірностей у даних про вразливості та потужності платформ великих даних для обробки великих обсягів інформації в реальному часі. ARIMA дозволяє враховувати як лінійні залежності, так і стохастичні складові, що особливо важливо для моделювання складних явищ, таких як динаміка появи кіберзагроз. Схема моделі прогнозування кіберзагроз на основі ARIMA та великих даних відображає послідовний процес, що включає кілька ключових етапів. Вона демонструє, як необроблені дані про кіберзагрози трансформуються в прогнози, які можна використовувати для забезпечення проактивного захисту систем. Схему зображено на рисунку 7.

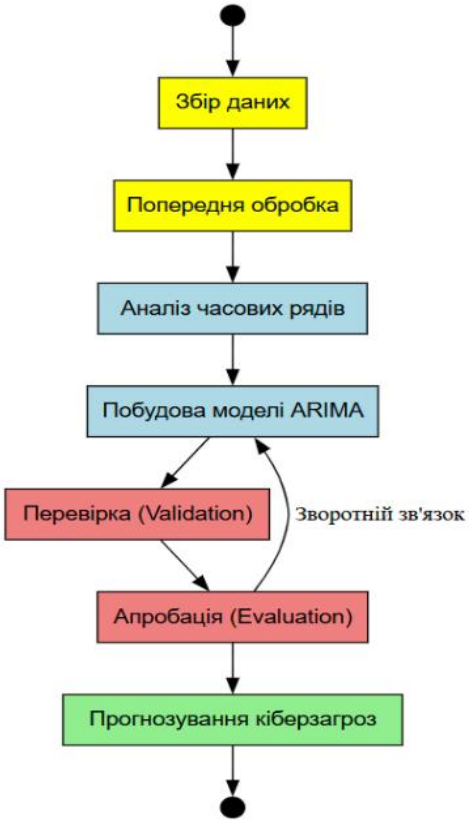

Рис 2.3 – Схема методу прогнозування кіберзагроз на основі ARIMA та великих даних

Прогнозування базується на аналізі великих обсягів даних, які включають історичну інформацію про вразливості, типи атак та пов'язані з ними часові закономірності. Зібрані дані, такі як журнали подій, мережевий трафік, записи про виявлені загрози (CVE-записи), проходять попередню обробку. Цей етап включає очищення, нормалізацію та видалення аномалій, що може сприяти зменшенню впливу шуму в даних. Наступним етапом є аналіз часових рядів, Це дозволяє отримати попереднє розуміння динаміки змін у кількості або типах кіберзагроз, що є основою для створення моделі ARIMA.

На етапі побудови моделі ARIMA визначаються оптимальні параметри моделі — порядок авторегресії (AR), ступінь інтеграції (I) та порядок ковзного середнього (MA). Вибір цих параметрів здійснюється за допомогою аналізу автокореляційної та часткової автокореляційної функцій. Крім того, використовуються статистичні критерії, такі як AIC або BIC, які допомагають обрати модель, що забезпечує найкращу рівновагу між точністю та складністю.

Перевірка адекватності моделі включає два ключові етапи: перевірку (Validation) та апробацію (Evaluation). На етапі перевірки проводиться тестування моделі на основі даних, не включених у навчання, з метою оцінки її точності та здатності прогнозувати ключові закономірності. Апробація ж спрямована на практичну перевірку моделі в умовах, наближених до реальних. Ці етапи взаємопов’язані через зворотній зв’язок, який дозволяє виявляти недоліки моделі, коригувати її параметри й підвищувати її загальну ефективність.

Завершальним етапом є прогнозування кіберзагроз, яке включає оцінку ймовірності нових атак, виявлення критичних вразливостей та автоматизацію попереджень. На основі отриманих прогнозів організації можуть проактивно реагувати на потенційні загрози, впроваджуючи захисні заходи або перенаправляючи ресурси на найбільш вразливі ділянки інфраструктури.

Метод прогнозування кіберзагроз на основі ARIMA із застосуванням великих даних забезпечує високу точність і адаптивність до змін у динаміці

загроз завдяки врахуванню взаємозв’язків у часових рядах. Його інтеграція з технологіями обробки великих обсягів інформації дозволяє масштабувати процеси аналізу та прогнозування, зберігаючи ефективність навіть у випадку потокових або розподілених даних. Використання оптимізованих алгоритмів та автоматизація етапів обробки даних мінімізують вплив людського фактора і сприяють швидкому реагуванню на нові ризики.

Висновки до розділу 2

У другому розділі представлено підхід до прогнозування кіберзагроз на основі аналізу часових рядів із застосуванням моделі ARIMA та великих даних. Модель ARIMA дозволяє аналізувати минулі тенденції та цикли, а також враховувати вплив попередніх змін у даних для передбачення майбутніх значень. Застосування великих даних, зокрема інформації з баз CVE, забезпечує глибший аналіз завдяки доступу до великої кількості даних про вразливості та атаки.

Розроблений підхід дозволяє прогнозувати динаміку загроз, визначати частоту появи нових вразливостей та їх критичність, що дає змогу проактивно реагувати на потенційні ризики. Інтеграція моделі з інструментами обробки великих обсягів даних підвищує її ефективність і адаптивність до змін, що робить її корисною для управління кібербезпекою та захисту інформаційних систем.

РОЗДІЛ 3. АПРОБАЦІЯ МЕТОДУ ПРОГНОЗУВАННЯ ВРАЗЛИВОСТЕЙ

3.1 Реалізація методу прогнозування з використанням даних CVE

Реалізація методу прогнозування вразливостей здійснювалася на основі інтеграції великих даних з бази Common Vulnerabilities and Exposures. CVE є загальновизнаним стандартом для каталогізації інформації про вразливості програмного забезпечення, забезпечуючи унікальну ідентифікацію кожної вразливості. Це джерело обрано завдяки його масштабності, актуальності та регулярному оновленню. На етапі використання даних CVE застосовувалися основні етапи схеми прогнозування, які включали попередню обробку даних, спрямовану на очищення записів від дублікатів та помилок, усунення пропущених значень для забезпечення цілісності інформації, а також аналіз часових рядів. Усі ці етапи спрямовані на підвищення точності та надійності моделі прогнозування, що враховує характерні тренди й циклічність у публікації нових вразливостей.

Для прогнозування використовувалася модель ARIMA, яка демонструє високу ефективність при роботі з часовими рядами. Першим етапом стала попередня обробка даних, що включала очищення, видалення дублікатів та усунення пропущених значень. Окрім того, для забезпечення стаціонарності рядів використовувалося диференціювання. Цей етап був критично важливим для адаптації моделі до змін у трендах публікації вразливостей. Основні етапи роботи з моделлю включали визначення параметрів моделі ARIMA (p, d, q) шляхом аналізу автокореляційних та часткових функцій, інтеграцію моделі з великими даними для навчання моделі на історичних даних, що забезпечило врахування довгострокових тенденцій,

Результати прогнозування моделі ARIMA наведено на рисунку нижче. Графік відображає кількість записів у базі CVE (синя лінія) до 2020 року, а також прогноз на 16 місяців уперед (червона пунктирна лінія).

Рис. 3.1 – Прогнозування кіберзагроз на основі ARIMA та великих даних

Аналіз графіка демонструє стабільне зростання кількості вразливостей. Ця тенденція вказує на необхідність посилення заходів кіберзахисту, адже кількість нових уразливостей, згідно з прогнозом, продовжуватиме зростати. Прогноз показує тенденцію до збільшення середньої кількості записів із можливими сезонними коливаннями. Висока точність моделі в короткостроковій перспективі дозволяє виявляти тренди та забезпечувати проактивний підхід до кібербезпеки.

3.2 Перевірка моделі прогнозування на адекватність та визначення основних параметрів моделі

Оптимальні параметри моделі визначалися на основі автокореляційної (ACF) та часткової автокореляційної (PACF) функцій, а також із застосуванням інформаційних критеріїв Акаіке (AIC) і Байєса (BIC).Параметр p (порядок авторегресії): відображає кількість минулих значень, які використовуються для прогнозу. Для моделі було встановлено значення p=2, що відображає вплив останніх двох спостережень на поточне значення. Параметр d (рівень диференціювання): забезпечує стаціонарність ряду. Для даного набору часових рядів d=1, що відповідає першому диференціюванню даних для усунення трендів. Параметр q (порядок ковзного середнього): описує вплив випадкових

шумів у моделі. Значення q=1 вказує на використання одного попереднього залишку у формулі прогнозу. Ці параметри забезпечили баланс між складністю моделі та її здатністю точно описувати структуру даних.

Модель ARIMA була протестована на адекватність і точність за допомогою часових рядів, отриманих із бази CVE. Дані до 2012 року використовувалися для навчання моделі, тоді як дані з 2012 по 2020 рік застосовувалися для перевірки її точності. Наведено результати моделювання з використанням ARIMA для прогнозів на 1, 3, 6 та 9 місяців. Графіки демонструють порівняння фактичних даних (історичних значень) із прогнозованими показниками для кожного з обраних часових горизонтів. Це дозволяє проаналізувати поведінку моделі, її адаптивність до змін у часовому ряді та тенденцію до накопичення помилок на довших періодах прогнозування.

На першому графіку представлено прогноз на 1 місяць. Червона пунктирна лінія відображає прогнозовані значення, тоді як синя лінія демонструє фактичні дані. Збіг трендів та характерна стабільність короткострокового прогнозу свідчать про високу точність моделі при малих горизонтах прогнозування.

Рис. 3.2 – Застосування методу прогнозування. Перевірка моделі із періодом прогнозування 1 місяць

Графік, що ілюструє прогноз на 3 місяці, демонструє схожу поведінку. Попри невеликі відхилення від фактичних даних, модель загалом ефективно відтворює загальну тенденцію зростання. Прогнозовані значення адекватно відображають ключові коливання тренду.

Рис. 3.3 – Застосування методу прогнозування. Перевірка моделі із періодом прогнозування 3 місяці

Прогноз на 6 місяців відображає вже більші розбіжності між реальними та прогнозованими значеннями. Це обумовлено впливом додаткових факторів, які стають значущими на середньострокових часових горизонтах.

Рис. 3.4 – Застосування методу прогнозування. Перевірка моделі із періодом прогнозування 6 місяці

Довгостроковий прогноз на 9 місяців демонструє більші відхилення, водночас модель зберігає здатність відтворювати основні тренди, хоч і з меншою точністю. Це свідчить про те, що ARIMA підходить переважно для коротко- та середньострокового прогнозування.

Рис. 3.5 – Застосування методу прогнозування. Перевірка моделі із періодом прогнозування 9 місяці

Для перевірки точності прогнозу було використано дві основні метрики: середню відносну помилку (MPE) та середньоквадратичну помилку (RMSE).

MPE є метрикою, що вимірює середнє відхилення прогнозованих значень від фактичних, виражене у відсотках. Цей показник дозволяє оцінити відносну помилку моделі, зокрема її схильність до переоцінювання чи недооцінювання даних. Формула обчислення

$$MPE = \left(\frac{1}{n}\right) \times \Sigma \left(\frac{(y_t - \hat{y}_t)}{y_t}\right) \times 100\%$$

y_t — фактичне значення в момент часу t, $\hat{y}_t$— прогнозоване значення та n — кількість спостережень.

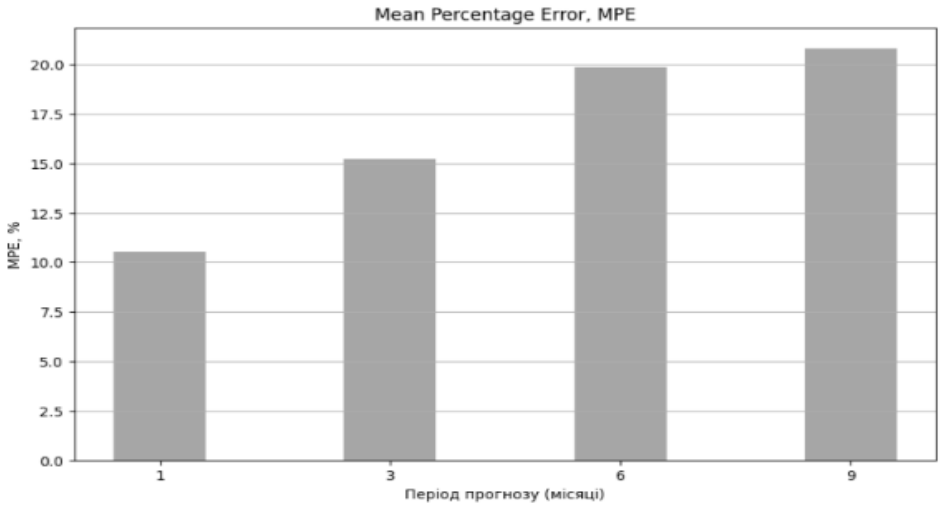

Рис. 3.6 – Графік динаміки MPE для різних періодів прогнозування.

RMSE вимірює середню величину квадратних відхилень між фактичними та прогнозованими значеннями. Вона відображає абсолютний рівень похибки моделі, що робить її особливо корисною для оцінки точності в одиницях вимірювання прогнозованих даних. Формула розрахунку RMSE виглядає так:

$$\sqrt{\left(\frac{1}{n}\right) \times \Sigma\,(y_t - \hat{y}_t)^2}$$

y_t – фактичне значення в момент часу t, $\hat{y}_t$– прогнозоване значення та n – кількість спостережень.

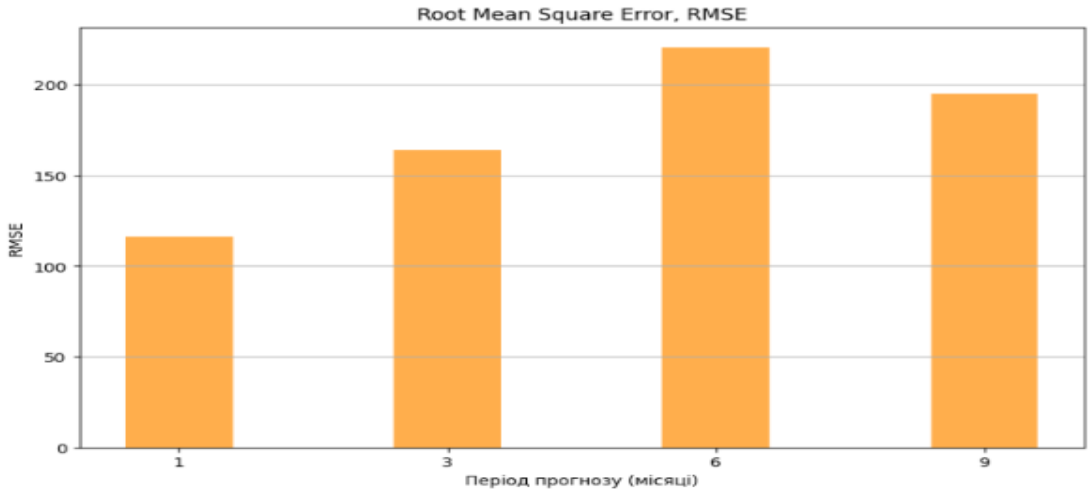

Рис. 3.7 – графік відображення зміну RMSE

Найкращий прогноз демонструється при короткостроковому горизонті прогнозування (1 місяць). Для цього періоду модель ARIMA забезпечує найнижчий рівень середньої відносної помилки (MPE ~10%) та найменшу середньоквадратичну помилку (RMSE ~150 одиниць). Зі збільшенням горизонту прогнозування (до 3, 6 і 9 місяців) точність прогнозу поступово знижується, що проявляється у зростанні обох метрик похибки. Таким чином, модель ARIMA найефективніше працює на коротких періодах прогнозу, підтверджуючи її придатність для коротко- та середньострокових завдань.

3.3 Практичні рекомендації щодо застосування методу прогнозування вразливостей

Метод прогнозування на основі моделі ARIMA є потужним інструментом для оптимізації управління кіберризиками в організаціях. Його застосування дозволяє ефективно прогнозувати розвиток вразливостей і відповідно формувати стратегії для своєчасного реагування на кіберзагрози. Для забезпечення максимальної ефективності цього методу, важливо впроваджувати низку практичних рекомендацій.

Однією з основних рекомендацій є інтеграція моделі ARIMA з платформами моніторингу, такими як системи управління інформацією та подіями безпеки, наприклад - Emergency Responce Tools Management SolutionsТака інтеграція дозволяє автоматизувати процес аналізу та прогнозування кіберзагроз, що є критично важливим для своєчасного виявлення потенційних вразливостей. Використання моделі в реальному часі з вбудованими механізмами моніторингу дозволяє організаціям оперативно реагувати на зміни у ситуації та передбачати виникнення нових загроз. Це зменшує часові витрати на виявлення вразливостей та покращує загальний рівень безпеки.

Іншою важливою рекомендацією є пріоритизація ресурсів на основі результатів прогнозу. Модель ARIMA дозволяє визначати найбільш критичні вразливості, на які необхідно зосередити зусилля. Завдяки прогнозу можна виявити найбільш ймовірні загрози і оцінити їх потенційний вплив на організацію, що дозволяє ефективно розподіляти ресурси на їх усунення. Це дозволяє організаціям зосередити увагу на найбільш небезпечних вразливостях, мінімізуючи ризики та підвищуючи ефективність витрат.

Не менш важливим є забезпечення адаптивності моделі. Для цього необхідно регулярно оновлювати модель ARIMA на основі нових даних з баз вразливостей, таких як CVE. Постійне оновлення моделі гарантує її відповідність змінним умовам у сфері кібербезпеки, дозволяючи враховувати нові загрози та вдосконалювати прогнози. Це забезпечує точність прогнозів та дозволяє своєчасно коригувати стратегії реагування на вразливості. Крім того, результати прогнозування можуть бути використані для формування проактивної стратегії кіберзахисту. Оскільки модель ARIMA враховує сезонні та трендові коливання вразливостей, її прогнози можуть бути використані для довгострокового планування заходів з кіберзахисту.

Висновки до розділу 3

У розділі було розглянуто апробацію методу прогнозування вразливостей на основі даних CVE, зокрема через використання моделі ARIMA. Описано етапи реалізації цього методу, починаючи від попередньої обробки даних до застосування моделі для прогнозування кількості вразливостей. Застосування ARIMA дозволило ефективно прогнозувати динаміку публікацій вразливостей, підтверджуючи високу точність у короткостроковій перспективі та виявлення важливих трендів. Було проведено перевірку моделі на адекватність. Окрім теоретичних аспектів, було також запропоновано практичні рекомендації щодо інтеграції цієї моделі з платформами моніторингу та забезпечення адаптивності

методу для підтримки своєчасного реагування на кіберзагрози. Результати демонструють потенціал методу ARIMA як потужного інструменту для підвищення ефективності управління кіберризиками та розробки стратегії кіберзахисту в умовах динамічно змінюваного середовища загроз.

ВИСНОВКИ

У межах виконаної роботи було розроблено метод прогнозування кіберзагроз із використанням ARIMA, що враховує часові ряди й інтеграцію великих даних.

Проведено аналіз ефективності моделі на базі даних CVE, що продемонстрував точність прогнозів для короткострокового аналізу Перевірка моделі на адекватність, підтвердила спроможність моделі з точністю не гірше ніж 85% прогнозувати появу нових вразливостей в горизонті 3 місяців.

Запропонований підхід сприяє своєчасному реагуванню на загрози, зниженню ризиків і мінімізації впливу атак на інформаційні системи.

Виявлено, що інтеграція ARIMA з платформами обробки великих даних забезпечує адаптивність і масштабованість моделі, що важливо для динамічного кіберсередовища.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. "Про основні засади забезпечення кібербезпеки України" : Закон України від 5 жовтня 2017 р. № 2163-VIII. Відомості Верховної Ради України. URL: https://zakon.rada.gov.ua/laws/show/2163-19#Text (дата звернення: 29.10.2024).

2. An introduction to the cyber threat environment. Canada: Communications Security Establishment, 2022. 18p. URL: https://www.cyber.gc.ca/sites/default/files/ncta-2022-intro-e.pdf

3. Cybersecurity and Infrastructure Security Agency (CISA). Cyber Threat Intelligence. – URL: https://www.cisa.gov/sites/default/files/2024-08/Federal_Government_Cybersecurity_Incident_and_Vulnerability_Response_Playbooks_508C.pdf (дата звернення: 16.11.2024).

4. Сус В.Ю. Аналіз моделі прогнозування кіберзагроз «Cyber Kill Chain» // *Моделювання, керування та інформаційні технології (МСІТ–2024)* : збірник праць учасників Міжнародної науково-практичної конференції. 2024. С. 201-211.

5. Lockheed Martin, The Cyber Kill Chain. URL: https://www.lockheedmartin.com/en-us/capabilities/cyber/cyber-kill-chain.html

6. Bazzell, M. Open Source Intelligence Techniques: Resources for Searching and Analyzing Online Information. IntelTechniques, 2022. 580 p.

7. Open Web Application Security Project (OWASP). OSINT Framework. URL: https://osintframework.com/ (дата звернення: 17.11.2024).

8. Hutchins, E. M., Cloppert, M. J., Amin, R. M. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. Lockheed Martin Corporation, 2011. 34 p.

9. Сус В.Ю. Огляд відкритих даних для прогнозування та оцінки кіберзагроз // Litteris et Artibus: *Нові горизонти* : збірник матеріалів

Всеукраїнської науково-практичної конференції. Випуск IX / за заг. ред. О.В. Тригуби. Кременець : ВЦ КОГПА ім. Тараса Шевченка, 2024. С. 105-108.

10. Teamincribo. UNSW-NB15 Dataset. Kaggle. URL: https://www.kaggle.com/datasets/dhoogla/unswnb15 (дата звернення: 17.11.2024).

11. MDPI. Використання ARIMA для прогнозування кіберзагроз на основі часових рядів. Rapid Forecasting of Cyber Events. URL: https://www.mdpi.com/2078-2489/15/1/36

12. Rajasooriya, S. M., Tsokos, C. P., & Kaluarachchi, P. K. (2017). Cybersecurity: Time Series Predictive Modeling of Vulnerabilities. *Journal of Information Security*, 8, 362–382.

13. Shumway, R. H., Stoffer, D. S., Shumway, R. H., & Stoffer, D. S. (2017). ARIMA models. Time series analysis and its applications: with R examples, 75-163.

14. SpringerOpen, 2018. Forecasting cyberattacks with incomplete, imbalanced, and insignificant data. *Cybersecurity Journal*. URL : https://link.springer.com/article/10.1186/s42400-018-0016-5

15. CVE Dataset – Kaggle. URL: https://www.kaggle.com/datasets/andrewkronser/cve-common-vulnerabilities-and-exposures .

16. Journal of Cybersecurity Studies, 2021. "Time Series Analysis for Cybersecurity: Challenges and Applications." URL : https://link.springer.com/article/10.1007/s10207-024-00921-0

17. Cybersecurity Big Data Analytics Review, 2022. "Time Series Analysis of Vulnerability Disclosure in the CVE Database." URL : https://www.researchgate.net/publication/318116583_Big_Data_for_Cybersecurity_Vulnerability_Disclosure_Trends_and_Dependencies

18. Сус В.Ю. Підхід до прогнозування частності та критичності нових вразливостей на основі відкритих даних // *Безпека, технології, інновації: нові горизонти* : збірник праць учасників міжфакультетської науково-практичної

інтернет-конференції здобувачів вищої освіти і молодих вчених, 12 листопада 2024 р. Житомир : Поліський національний університет, 2024. С. 17-19.

19. Scarfone, K., & Mell, P. (2009, October). An analysis of CVSS version 2 vulnerability scoring. In 2009 3rd International Symposium on Empirical Software Engineering and Measurement (pp. 516-525). IEEE. URL : https://www.researchgate.net/publication/221494966_An_analysis_of_CVSS_version_2_vulnerability_scoring

20. Hyndman, R. J., & Athanasopoulos, G. (2018): URL: https://www.scirp.org/journal/paperinformation?paperid=103718

21. ls, Paul The Unified KillChain. URL: UnifiedKillChain.com.

22. Протидія кіберзагрозам URL : https://www.kmu.gov.ua/news/yak-protydiiaty-kiberzahrozam-ta-zakhystyty-systemy-vid-vorozhykh-kiberatak-vazhlyvi-rekomendatsii-ta-dopomoha-cert-ua.

23. Shumway, R. H., Stoffer, D. S., Shumway, R. H., & Stoffer, D. S. (2017). ARIMA models. *Time series analysis and its applications: with R examples*, 75-163.

24. Маєрс Т., Браун Р. UNSW-NB15 Dataset для систем виявлення вторгнень на основі мереж. *Журнал мережевої безпеки*. 2022. Вип. 5. С. 57–79.

25. Носко В. А., Сергєєв А. В. Комплексний огляд машинного навчання у сфері кібербезпеки. *Журнал кібербезпеки*. 2021. Вип. 3. С. 23–45

26. EURASIP Journal on Information Security. Long Short-Term Memory для аналізу часових рядів у кібербезпеці. URL: https://jis-eurasipjournals.springeropen.com.

27. Mitnick K. D., Simon W. L. The Art of Deception: Controlling the Human Element of Security. – Wiley Publishing, 2011. – 368 с.

28. CERT-UA Vulnerability Database. CERT-UA: Актуальні дані про виявлені вразливості та їх вплив на інформаційні системи. URL: https://cert.gov.ua

29. Microsoft Security Intelligence Report. 2023. Глобальний звіт з аналізу кіберзагроз. URL: https://www.microsoft.com/security

30. SANS Institute Whitepapers on Cybersecurity. Огляд сучасних методів аналізу та прогнозування кіберзагроз. URL: https://www.sans.org

31. Kaufman, C., Perlman, R., & Speciner, M. (2021). Network Security: Private Communication in a Public World. 3rd Edition. Pearson. URL : https://www.nist.gov/publications/network-security-private-communication-public-world-3rd-edition

32. McAfee, A., & Brynjolfsson, E. (2017). Machine, Platform, Crowd: Harnessing Our Digital Future. W.W. Norton & Company.